

# The Termite Group I Phylum Is Highly Diverse and Widespread in the Environment<sup>∇</sup>

Daniel P. R. Herlemann, Oliver Geissinger, and Andreas Brune\*

Max Planck Institute for Terrestrial Microbiology, Karl-von-Frisch-Strasse, 35043 Marburg, Germany

Received 29 March 2007/Accepted 9 August 2007

**The bacterial candidate phylum Termite Group I (TG-1) presently consists mostly of “Endomicrobia,” which are endosymbionts of flagellate protists occurring exclusively in the hindguts of termites and wood-feeding cockroaches. Here, we show that public databases contain many, mostly undocumented 16S rRNA gene sequences from other habitats that are affiliated with the TG-1 phylum but are only distantly related to “Endomicrobia.” Phylogenetic analysis of the expanded data set revealed several diverse and deeply branching lineages comprising clones from many different habitats. In addition, we designed specific primers to explore the diversity and environmental distribution of bacteria in the TG-1 phylum.**

Termite Group I (TG-1) represents a deep branch in the tree of bacterial 16S rRNA gene sequences (18) and has been recognized as a candidate phylum (10). TG-1 comprises a large number of the bacteria in the hindguts of *Reticulitermes* species (6, 25), where they occur as intracellular symbionts of flagellate protists (22). These symbionts, for which the name “Endomicrobia” has been proposed, form a monophyletic lineage occurring exclusively in the hindguts of termites and wood-feeding cockroaches (11, 22).

However, a few sequences only distantly related to the “Endomicrobia” but clearly affiliated with the TG-1 phylum have been reported to also occur in habitats other than termite guts (3, 19, 20, 24). Moreover, Nakajima et al. (17) obtained two sequences from the gut of *Reticulitermes speratus* that fall outside the “Endomicrobia” lineage. At present, public databases contain a growing number of sequences from various habitats that are phylogenetically unassigned but seem to be affiliated with the TG-1 phylum.

In this study, we screened public databases for hitherto unrecognized TG-1 sequences and conducted a comprehensive phylogenetic analysis of the expanded data set. In addition, we designed specific PCR primers to investigate the diversity and environmental distribution of major lineages of TG-1 bacteria in soils, sediments, and intestinal tracts.

**Data mining.** Sequences affiliated with the TG-1 phylum were retrieved from GenBank (<http://www.ncbi.nlm.nih.gov/>) using various characteristic oligonucleotide signatures deduced from the originally available 16S rRNA gene sequences, and they were added to the database of the ARB program suite (15). By continuously adapting the signatures to the growing data set (data not shown), we obtained approximately 50 previously unassigned phylotypes that fell into the radiation of the TG-1 phylum. Most of these phylotypes were from large-scale diversity studies of various environments, including soils, sediments, and intestinal tracts. Sequences were aligned with the

ARB Fast Aligner tool. The alignment was manually corrected, and highly variable regions and ambiguous positions were excluded from the analysis. Rigorous chimera checking with Belerophon (9) and fractional treeing (14) identified only one sequence as a putative chimera (accession no. DQ830579), which was removed from the data set. All shorter sequences (500 to 1,300 bp) were added to the core tree using the parsimony tool implemented in ARB.

**Primer design and PCR.** Phylogenetic analysis of the nearly full-length sequences (>1,300 bp) fully supported the conclusion that the TG-1 phylum forms a separate line of descent in the bacterial tree, consisting of several diverse and deeply branching lineages (Fig. 1). Designing a single phylum-specific primer that excluded all representatives of other phyla proved to be impossible. A primer set for specific amplification of “Endomicrobia” (primer set 1) (Table 1) was designed in a previous study (22). Based on the expanded data set, we designed three additional primer sets (primer sets 2 to 4) that covered most of the other lineages in the TG-1 phylum. When applicable, information from shorter sequences was included to improve primer design.

The primers were used for PCR-based screening of various habitats for the presence of TG-1 bacteria. DNA was extracted by bead beating (16), and humic substances were removed by passing aqueous extracts over an Autoseq G-50 column (Amersham Bioscience). For PCR amplification a standard protocol optimized for the primer pairs was used (Table 1). The products of two identical reactions were combined, cleaned with a MinElute PCR purification kit (QIAGEN), and cloned with an pGEM-T Easy vector kit (Promega). Positive clones were amplified with M13 vector primers and checked for inserts on a 1% agarose gel. Clones with inserts that were the correct lengths were sorted by restriction fragment length polymorphism analysis as previously described (21). Inserts were sequenced on both strands.

**Phylogenetic analysis.** Phylogenetic analysis of the resulting clone libraries documented that all primer pairs were highly specific and amplified only 16S rRNA genes of bacteria in the TG-1 phylum.

Primer set 1 gave a PCR product only with the termite hindgut samples (Table 2). This corroborates the specificity of

\* Corresponding author. Mailing address: Max Planck Institute for Terrestrial Microbiology, Karl-von-Frisch-Strasse, 35043 Marburg, Germany. Phone: 49-6421-178701. Fax: 49-6421-178709. E-mail: brune@staff.uni-marburg.de.

<sup>∇</sup> Published ahead of print on 17 August 2007.



FIG. 1. Phylogenetic tree of all sequences in the TG-1 phylum obtained in this and previous studies. A core tree was constructed with 1,262 unambiguously aligned sequence positions of all nearly full-length sequences retrieved from GenBank (in bold type), using maximum-likelihood analysis (fastDNAMl). Short sequences (>500 bp), positionally filtered by base frequency (50%), were added without changing the global tree topology using the ARB parsimony tool. The scale bar is only approximate because the procedure distorts branch length. Representatives of 10 other phyla were used as the outgroup. Clones obtained in this study are indicated by filled squares. The topology of the core tree and individual clusters was tested separately by neighbor-joining and parsimony analysis (DNAPARS) with bootstrapping (seqboot; 1,000 bootstraps). Only nodes supported by high bootstrap values are marked (filled circles, >95%; open circles, >75%); nodes not supported by all analyses are shown as multifurcations. Original sequence definitions in the GenBank database were replaced with a consistent nomenclature including sequence length (in base pairs), habitat, geographic origin, and accession number. Roman numerals indicate the lineages referred to in the text and in Table 1.

TABLE 1. Primer sets used for amplification of major lineages in the TG-1 phylum and annealing temperatures and MgCl<sub>2</sub> concentrations used in the PCR assay<sup>a</sup>

Primer set	Target group(s) <sup>b</sup>	Primer	<i>E. coli</i> position	Sequence (5' to 3')	Annealing temp (°C)	MgCl <sub>2</sub> concn (mM)
1	I	TG1-209f <sup>c</sup> TG1-1325r <sup>c</sup>	210–237 1325–1343	AATGCGTTTTGAGATGGTCCTG GATTCCTACTTCATGTGG	54.0	2
2	II	EluD540f EluD1300r	539–562 1286–1306	AGGTGGCAAGCGTTACTCGGAAT TCTGAACCTGGGGCCGGCTTTT	63.9	1.5
3	III	EluB22f EluB998r	22–40 981–998	GCTCAGAGTTAACGCTGGC GTCGTTTCGAGCCCAGGTAA	61.7	1.5
4	IV	EluC104f EluC1173r	104–126 1153–1173	GGCAGACGAGTGAGTARACAGTA ACGTTATCCGCGGCAGTCTCC	57.0	2
T	II-IV	27f <sup>d</sup> Elu1058r	8–27 1042–1058	AGAGTTTGATCCTGGCTCAG CCATGCAGCACCTCGGC	55.0	2

<sup>a</sup> See Fig 1 and Table 2.

<sup>b</sup> With the exception of primer set T, all primer sets had at least two or three mismatches to sequences outside the TG-1 phylum present in the greengenes database (2; <http://greengenes.lbl.gov/>).

<sup>c</sup> See reference 22.

<sup>d</sup> See reference 12.

this primer set for its target group, the “Endomicrobia” (lineage I in Fig. 1), which seem to be restricted to termites and wood-feeding cockroaches harboring gut flagellates (1, 22). “Endomicrobia” sequences from termite guts were not further investigated since they are the subject of a separate study (11).

Primer set 2 matched the sequences of three deeply branching lineages (lineages IIa, IIb, and IIc) comprising clones from soil or sediment samples. An amplification product of the ex-

pected size was obtained only with Italian rice soil. The resulting clone library contained several phylotypes falling into lineage IIa and a single phylotype falling into lineage IIb (Fig. 1). Generally, the large number of sequences from soils and sediments retrieved from public databases suggests that bacteria in lineage II are widely distributed in these habitats.

Primer set 3 was designed to target the sequences in lineage III, which consists of sequences from intestinal habitats and also includes strain Pei191, the first isolate from the TG-1 phylum obtained from the gut of a beetle larva (O. Geissinger and A. Brune, unpublished data). Several new phylotypes were also obtained from the hindguts of *Reticulitermes santonensis* and *Zootermopsis nevadensis* (Table 2). All were distantly related to the clones previously retrieved from an *R. speratus* gut wall sample (17), underlining the finding that termite guts harbor a second lineage of TG-1 bacteria besides the “Endomicrobia.” Cow rumen amplification products yielded a diverse but monophyletic group of sequences clustering with a single TG-1 sequence (accession no. AB034017) in a bacterial clone library from this habitat and erroneously assigned to the *Proteobacteria* (23).

Primer set 2 also yielded PCR products with the termite gut samples, but cloning analysis revealed that the products consisted exclusively of “Endomicrobia” sequences, which indicated that this primer set insufficiently discriminates against this group. Also, the termite clone libraries obtained with primer set 3 contained sequences belonging to the “Endomicrobia” (Table 2). Although both primer sets had two or more mismatches with all nontarget sequences within the TG-1 phylum, they apparently lack differentiating power for the corresponding subgroups if too many “Endomicrobia” are present in a sample, a problem encountered only with lower termites.

None of the DNA samples used in this study gave a PCR product with primer set 4, designed to detect sequences from lineage IV, which comprises clones from many different habitats (Fig. 1). Lahn River sediment yielded PCR products only with primer set T (Table 1), which was designed to detect most

TABLE 2. Detection of TG-1 bacteria in various habitats using the new designed primer sets for specific amplification of major lineages in the phylum<sup>a</sup>

Habitat or strain	Detection of TG-1 bacteria with <sup>b,c</sup> :			
	Primer set 1	Primer set 2	Primer set 3	Primer set 4
Lahn River sediment <sup>d</sup>	–	–	–	–
Marburg forest soil <sup>d</sup>	–	–	–	–
Italian rice soil <sup>e</sup>	–	6/0 (14)	–	–
Cow rumen <sup>f</sup>	–	–	5/0 (16)	–
<i>Reticulitermes santonensis</i> gut <sup>g</sup>	+	3/3 (16)	5/3 (17)	–
<i>Zootermopsis nevadensis</i> gut <sup>g</sup>	+	3/3 (19)	5/2 (12)	–
<i>Pachnoda ehippiata</i> gut <sup>g</sup>	–	–	±	–
Strain Pei191 <sup>h</sup>	–	–	+	–

<sup>a</sup> See Table 1.

<sup>b</sup> +, PCR product of the expected size; –, no PCR product; ±, results with different preparations varied. If clone libraries were analyzed, the number of TG-1 phylotypes/number of “Endomicrobia” phylotypes among the total clones tested (in parentheses) is indicated.

<sup>c</sup> Each reaction mixture (25 μl) contained reaction buffer (Invitrogen), MgCl<sub>2</sub> (see Table 1), deoxynucleoside triphosphates (200 μM each), primers (0.3 μM each), dimethyl sulfoxide (1 μl), DNA extract (200 to 300 ng), and *Taq* DNA polymerase (1.25 U; Invitrogen). The thermal cycling consisted of an initial denaturation step of 5 min at 96°C, followed by 30 cycles consisting of 30 s at 94°C, 30 s at the annealing temperature (Table 1), and 45 s at 72°C.

<sup>d</sup> Topsoil under *Fagus sylvatica* and anoxic sediment of the Lahn River were freshly collected in Marburg, Germany.

<sup>e</sup> Dried rice soil (*Oryza sativa*) from wetland rice fields of the Italian Rice Research Institute in Vercelli (Italy) was regenerated for 3 days as described by Frenzel et al. (5).

<sup>f</sup> Rumen contents of a freshly slaughtered Holstein cow.

<sup>g</sup> Prepared as previously described (13, 21).

<sup>h</sup> Isolated from the hindgut of *Pachnoda ehippiata*.

TG-1 sequences other than those of lineage I. This primer set yielded PCR products for all habitats tested but turned out to be nonspecific. Sequence analysis revealed that only about 10% of the clones in each clone library fell into the TG-1 phylum (lineage III in the *Pachnoda ephippiata* gut; lineages IIb and IIc in Lahn River sediment). Other clones were mostly representatives of the *Bacteroidetes* and the *Acidobacteria*, which was in agreement with the lack of discrimination of this primer set with a few representatives of these phyla (data not shown).

**Abundance in the environment.** TG-1 bacteria are highly abundant only in bacterial clone libraries from the hindguts of lower termites (6, 7, 8, 18, 25). Fluorescence in situ hybridization corroborated the conclusion that “Endomicrobia” constitute a significant portion of the gut microbiota (11, 22). In contrast, bacterial clone libraries from other habitats generally contain only small numbers of sequences from members of the TG-1 phylum, if they contain any at all. In rumen fluid clone libraries (150 clones) and farm soil clone libraries (1,700 clones), only a single clone each fell into the TG-1 phylum (23, 24). Clone libraries of *P. ephippiata* larva gut homogenates (113 clones) (4) did not contain any clones affiliated with the TG-1 phylum, although a representative of the “intestinal cluster” was obtained from this species with primer set T (this study). Possible explanations for these phenomena may lie in either a low relative abundance of such bacteria in the respective communities or a mismatch in the “universal” *Bacteria* primers used in these studies.

**Conclusion.** The results of this study document that bacteria affiliated with the TG-1 phylum form a separate line of descent, as proposed previously on the basis of a much smaller data set (10, 18). This phylum consists of numerous diverse and deeply branching lineages comprising bacteria from a wide range of chemically and geographically distinct habitats, including soils, sediments, and intestinal tracts. Although TG-1 bacteria seem to be numerically abundant only in the hindguts of lower termites (lineage I; “Endomicrobia”), the great diversity and wide environmental distribution of other lineages suggest a hitherto unrecognized role in the environment.

**Nucleotide sequence accession numbers.** Sequences determined in this study have been deposited in the EMBL database under accession numbers AM491071 to AM491086, AM491098, AM491123, and AM491125 (<http://www.ebi.ac.uk/>).

This study was funded by the Max Planck Society. D.P.R.H. received a fellowship from the International Max Planck Research School for Environmental, Cellular, and Molecular Microbiology.

#### REFERENCES

1. Brune, A., and U. Stingl. 2005. Prokaryotic symbionts of termite gut flagellates: phylogenetic and metabolic implications of a tripartite symbiosis, p. 39–60. *In* J. Overmann (ed.), *Molecular basis of symbiosis*. Springer, Berlin, Germany.
2. DeSantis, T. Z., P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, and G. L. Andersen. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* **72**:5069–5072.
3. Dojka, M. A., P. Hugenholtz, S. K. Haack, and N. R. Pace. 1998. Microbial diversity in a hydrocarbon- and chlorinated-solvent-contaminated aquifer undergoing intrinsic bioremediation. *Appl. Environ. Microbiol.* **64**:3869–3877.
4. Egert, M., B. Wagner, T. Lemke, A. Brune, and M. W. Friedrich. 2003. Microbial community structure in midgut and hindgut of the humus-feeding larva of *Pachnoda ephippiata* (Coleoptera: Scarabaeidae). *Appl. Environ. Microbiol.* **69**:6659–6668.
5. Frenzel, P., F. Rothfuss, and R. Conrad. 1992. Oxygen profiles and methane turnover in a flooded rice microcosm. *Biol. Fertil. Soils* **14**:84–89.
6. Hongoh, Y., M. Ohkuma, and T. Kudo. 2003. Molecular analysis of bacterial microbiota in the gut of the termite *Reticulitermes speratus* (Isoptera, Rhinotermitidae). *FEMS Microbiol. Ecol.* **44**:231–242.
7. Hongoh, Y., H. Yuzawa, M. Ohkuma, and T. Kudo. 2003. Evaluation of primers and PCR conditions for the analysis of 16S rRNA genes from a natural environment. *FEMS Microbiol. Lett.* **221**:299–304.
8. Hongoh, Y., P. Deevong, T. Inoue, S. Moriya, S. Trakulnaleamsai, M. Ohkuma, C. Vongkaluang, N. Noparatnaraporn, and T. Kudo. 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.
9. Huber, T., G. Faulkner, and P. Hugenholtz. 2004. Bellerophon: a program to detect chimeric sequences in multiple sequence alignments. *Bioinformatics* **20**:2317–2319.
10. Hugenholtz, P., B. M. Goebel, and N. R. Pace. 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. *J. Bacteriol.* **180**:4765–4774.
11. Ikeda-Ohtsubo, W., M. Desai, U. Stingl, and A. Brune. Phylogenetic diversity of “Endomicrobia” and their specific affiliation with termite gut flagellates. *Microbiology*, in press.
12. Lane, D. J. 1991. 16S/23S rRNA sequencing, p. 115–175. *In* E. Stackebrandt and M. Goodfellow (ed.), *Nucleic acid techniques in bacterial systematics*. John Wiley & Sons, New York, NY.
13. Lemke, T., U. Stingl, M. Egert, M. W. Friedrich, and A. Brune. 2003. Physicochemical conditions and microbial activities in the highly alkaline gut of the humus-feeding larva of *Pachnoda ephippiata* (Coleoptera: Scarabaeidae). *Appl. Environ. Microbiol.* **69**:6650–6658.
14. Ludwig, W., S. H. Bauer, M. Bauer, I. Held, G. Kirchhof, R. Schulze, I. Huber, S. Spring, A. Hartmann, and K. H. Schleifer. 1997. Detection and in situ identification of representatives of a widely distributed new bacterial phylum. *FEMS Microbiol. Lett.* **153**:181–190.
15. Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, et al. 2004. ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**:1363–1371.
16. Lueders, T., and M. W. Friedrich. 2003. Evaluation of PCR amplification bias by terminal restriction fragment length polymorphism analysis of small-subunit rRNA and *mcrA* genes by using defined template mixtures of methanogenic pure cultures and soil DNA extracts. *Appl. Environ. Microbiol.* **69**:320–326.
17. Nakajima, H., Y. Hongoh, R. Usamib, T. Kudo, and M. Ohkuma. 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.
18. Ohkuma, M., and T. Kudo. 1996. Phylogenetic diversity of the intestinal bacterial community in the termite *Reticulitermes speratus*. *Appl. Environ. Microbiol.* **62**:461–468.
19. Reardon, C. L., D. E. Cummings, L. M. Petzke, B. L. Kinsall, D. B. Watson, B. M. Peyton, and G. G. Geesey. 2004. Composition and diversity of microbial communities recovered from surrogate minerals incubated in an acidic uranium-contaminated aquifer. *Appl. Environ. Microbiol.* **70**:6037–6046.
20. Spear, J. R., J. J. Walker, T. M. McCollom, and N. R. Pace. 2005. Hydrogen and bioenergetics in the Yellowstone geothermal ecosystem. *Proc. Natl. Acad. Sci. USA* **102**:2555–2560.
21. Stingl, U., and A. Brune. 2003. Phylogenetic diversity and whole-cell hybridization of oxymonad flagellates from the hindgut of the wood-feeding lower termite *Reticulitermes flavipes*. *Protist* **154**:147–155.
22. Stingl, U., R. Radek, H. Yang, and A. Brune. 2005. “Endomicrobia”: cytoplasmic symbionts of termite gut protozoa form a separate phylum of prokaryotes. *Appl. Environ. Microbiol.* **71**:1473–1479.
23. Tajima, K., S. Arai, K. Ogata, T. Nagamine, H. Matsui, M. Nakamura, R. I. Aminov, and Y. Benno. 2000. Rumen bacterial community transition during adaptation to high-grain diet. *Anaerobe* **6**:273–284.
24. Tringe, S. G., C. von Mering, A. Kobayashi, A. A. Salamov, K. Chen, H. W. Chang, M. Podar, J. M. Short, E. J. Mathur, J. C. Detter, P. Bork, P. Hugenholtz, and E. M. Rubin. 2005. Comparative metagenomics of microbial communities. *Science* **308**:554–557.
25. Yang, H., D. Schmitt-Wagner, U. Stingl, and A. Brune. 2005. Niche heterogeneity determines bacterial community structure in the termite gut (*Reticulitermes santonensis*). *Environ. Microbiol.* **7**:916–932.