The cockroach origin of the termite gut microbiota: patterns in bacterial community structure reflect major evolutionary events

Carsten Dietrich*, Tim Köhler* & Andreas Brune§

Department of Biogeochemistry, Max Planck Institute for Terrestrial Microbiology, Karl-von-Frisch-Strasse 10, 35043 Marburg, Germany

Running Title: The cockroach origin of the termite gut microbiota

Keywords: Termites / cockroaches / gut microbiota / diversity / community structure / evolution

* These authors contributed equally to this work.

§ Corresponding author. Mailing address: Department of Biogeochemistry, Max Planck Institute for Terrestrial Microbiology, Karl-von-Frisch-Strasse 10, 35043 Marburg, Germany. Phone: +49-6421-178700. Fax: +49-6421-178709. E-mail: brune@mpi-marburg.mpg.de.
Termites digest wood and other lignocellulosic substrates with the help of their intestinal microbiota. While the functions of the symbionts in the digestive process are slowly emerging, the origin of the bacteria colonizing the hindgut bioreactor is entirely in the dark. Recently, our group discovered numerous representatives of bacterial lineages specific for termite guts in a closely related omnivorous cockroach, but it remains unclear whether they derive from the microbiota of a common ancestor or were independently selected by the gut environment. Here, we studied the bacterial gut microbiota in 34 species of termites and cockroaches using pyrotag analysis of the 16S rRNA genes. Although the community structure strongly differed between the major host groups, with dramatic changes in the relative abundance of particular bacterial taxa, we found that the majority of sequence reads belonged to bacterial lineages that were shared among most host species. When mapped onto the host tree, the changes in community structure coincided with major events in termite evolution, such as acquisition and loss of cellulolytic protists and the ensuing dietary diversification. Unifrac analysis of the core microbiota of termites and cockroaches and phylogenetic treeing of individual genus-level lineages revealed a general host signal, whereas the branching order often did not match the detailed phylogeny of the host. It remains unclear whether the lineages in question were associated already with the ancestral cockroach since the early Cretaceous (cospeciation) or are diet-specific lineages that were independently acquired from the environment (host selection).
INTRODUCTION

Termites digest wood and other lignocellulosic substrates with the help of their intestinal microbiota—a symbiosis that has fascinated biologists for more than a century (1). The ability to mineralize lignocellulose and humus lends termites an important place in carbon and nitrogen cycling in tropical soils (2) and makes them promising models for the industrial conversion of lignocellulose into microbial products and the production of biofuels (3).

The ancestors of termites were presumably detritivorous, subsocial cockroaches (4, 5). About 130 million years ago, they gained the ability to digest wood through acquisition of cellulolytic flagellates (6, 7). These eukaryotic protists, which fill up the bulk of the hindgut volume, are the major habitat of the prokaryotic community present in the digestive tract of all phylogenetically lower termites (1, 8). The complete loss of all gut flagellates in the youngest termite family, the Termitidae—another hallmark in the evolutionary history of termites—led to dietary diversification and an enormous ecological success (6, 9). While the Macrotermitinae established a unique symbiosis with a lignocellulolytic fungus (10), other lineages of higher termites started to exploit diets of increasing humification, a development accompanied by further differentiation of the hindgut (9) and its entirely prokaryotic microbiota (8, 11).

While the role of the cellulolytic flagellates in lower termites is well defined, the functions of the mostly uncultivated bacterial symbionts in the digestive process, particularly in the flagellate-free higher termites, are just emerging (1, 8). Most importantly, the origin of the bacteria colonizing the hindgut bioreactor is entirely in the dark (11).

Although the gut microbiota differs substantially between termite species, it comprises many phylogenetic clusters that are unique to termites (e.g. 12, 13, 14). The origin of these lineages remains unclear, but their detection also in the gut of several cockroaches (15, 16, 17), the closest relatives of termites, together with occasional evidence of co-cladogenesis with the termite hosts (18, 19) gave rise to the hypothesis that the bacterial microbiota of extant termites and cockroaches is derived from their common dictyopteran ancestors.

In this study, we used a cultivation-independent high-throughput approach to characterize diversity and structure of the intestinal microbiota in a broad selection of termites and...
cockroaches. 16S rRNA gene fragments (V3–V4 region) were amplified with universal, bar-
coded primers and classified using a comprehensive reference database of all homologs
previously obtained from insect guts, which had been optimized to resolve termite- and
cockroach-specific groups (20). Comparative analysis of the datasets was employed to detect the
presence and distribution of common bacterial lineages across the major host groups.

MATERIALS AND METHODS

Insect samples. Termites were taken from colonies maintained in the laboratory or were
collected in the field. Cockroaches and other insects were purchased from commercial breeders,
and hindguts were dissected immediately upon arrival (17, 20). In some cases, field-collected
termites had to be preserved in ethanol for transport. Since the entire guts of ethanol-preserved
specimens were processed within less than one week, detrimental effects of this treatment on
community structure can be excluded (21). Moreover, our dataset includes several closely related
species that differ with respect to this pre-treatment but yielded highly similar profiles, which
further disseminates a potential bias introduced by the inclusions of ethanol-stored samples.
Details on the nature and origin of each sample are shown in Table 1. Field-collected specimens
were routinely identified by sequencing their mitochondrial cytochrome oxidase II (COII) gene
(22). The COII gene sequences of all species that were not represented in public databases have
been submitted to NCBI Genbank (accession numbers KF372028–KF372033).

Pyrotag sequencing. DNA was extracted from the pooled gut homogenates of 3 to 10
individuals of each species (depending on gut volume) using a bead-beating protocol with
phenol–chloroform purification (23). PCR amplification of the V3–V4 region of the bacterial
16S rRNA genes with a bar-coded primer set (343Fmod–784Rmod) modified to optimize
coverage of the taxa known to prevail in termite and cockroach guts was as previously described
(20). Amplicons were mixed in equimolar amounts and commercially sequenced (454 GS FLX
Titanium technology; GATC Biotech, Konstanz, Germany). Pyrotag sequences were pre-
processed and aligned using the mothur software suite (24) (version 1.27.0) using stringent
conditions (25) (reads > 200 bp; no ambiguous bases, maximum number of homopolymers ≤ 8).
The sequences in each sample where denoised with the Acacia program (26) using default
parameters, except that standard deviation from mean read length was set to 5 to avoid a loss of
entire taxa from individual datasets due to sequence length heterogeneity between phylotypes. Denoising reduced the number of OTUs (3% sequence dissimilarity) in the samples by 0 to 5%. The pyrotag datasets were submitted to the NCBI Sequence Read Archive (accession numbers in Table 1).

**Classification.** Sequence reads were classified with the *Naïve Bayesian Classifier* implemented in *mothur*, using a bootstrap value of 60% as cutoff. Since classification success with public reference databases was limited due to lack of taxonomic resolution, particularly in the groups represented in termites and cockroaches (20; Table 2), we used a customized reference database to improve resolution (DictDb v. 2.3). The reference database was built on the basis of the *silva* database (*silva* SSU REF NR 114), to which additional sequences from bacterial microbiota of dictyopteran insects were added, including both sequences from published studies and unpublished data from our laboratory. The taxonomy of relevant lineages was refined by incorporating genus-level taxa that have been identified either in published phyllogenies of relevant groups (e.g. 13, 27) or additional hitherto unresolved monophyletic groups. The reference database is available from the authors upon request; a publication of the latest version documenting the detailed classification of termite- and cockroach-specific clusters is in preparation.

**Statistical analyses.** All samples were subsampled to the smallest number of reads per sample in the dataset (5,045 reads). Classification-dependent ordinations (genus level) were based on the Bray-Curtis dissimilarity coefficient (28). To reduce the dimensions of the dataset, the results were displayed using non-metric multidimensional scaling (NMDS). Classification-independent ordinations were carried out using the same strategy with reads grouped into operational taxonomic units (OTUs), or a phylogeny-based analysis of the reads with UniFrac (29) displayed using principle-coordinate analysis (PCoA). For all analyses, the significance of clusters was tested by analysis of variance using distance matrices (ADONIS). The significance of clusters in OTU and taxon-based analyses was tested independently with the multi-response permutation procedure (MRPP). To determine the contribution of the genera to the ordination patterns, we carried out principal-component analysis (PCA) of the entire dataset (frequency of reads at genus level) and calculated the contribution of each genus to all dimensions relative to all other genera (30). Multivariate statistics were carried out using the *R* software (version 2.15.1) with the *vegan* package.
Phylogeny-based analysis of community similarity (unweighted UniFrac) of the core microbiota of cockroaches and termites was conducted with the genus-level taxa that were present in > 70% of the species in each of the major host groups; to account for differences in read number, each taxon was randomly subsampled to ten sequences per sample. A cladogram was constructed based on the resulting dissimilarity matrix using a neighbor-joining algorithm.

**Phylogenetic analysis of the pyrotag reads.** After random subsampling to 5045 reads per sample, all sequences were classified and sorted into genus-level bins. All samples in the same bin were grouped into OTUs (3% dissimilarity), and one representative sequence per OTU was selected for each sample using the *mothur* command *get.oturep*, which also returns the number of reads in each OTU. Maximum-likelihood trees were calculated for each genus-level lineage with *FastTree 2* (34), transformed into ultrametric trees using *PATHd8* (35), and visualized using the *R* package *APE* (36).

**RESULTS AND DISCUSSION**

The bacterial 16S rRNA genes in hindgut DNA of 34 termites and cockroaches and a few other insects were amplified with universal, bar-coded primers for the V3–V4 region. For each host species, we obtained an average of 10,000 high-quality sequence reads (Table 1). This is the first such data for most of these species, and classification against the RDP database (41) yielded large fractions of unclassified sequences, particularly at lower taxonomic ranks (Table 2). Our curated reference database (20) significantly improved classification, increasing the fraction of classified sequences in the different samples at the genus level from 24–68% (RDP) to 61–93% (our database) (Table 2).

Classification yielded 200–300 genus-level taxa for the majority of samples (between 90 and 550 in extreme cases) (Table 1). The detailed classification results for all taxonomic ranks can be found in interactive Table S1 in the supplemental material. The number of operational taxonomic units (OTUs) obtained by similarity-based clustering of the sequences (3% sequence dissimilarity) was 2–10-fold higher, indicating additional diversity at the species level (Table 1). Predictions of species richness and coverage (Good's coverage and Chao1 estimators) that are
Based on the abundance of singletons in a data set, high-throughput sequencing fails to cover the entire bacterial diversity in a gut community; even the samples with the largest numbers of reads (> 100,000) still contain a large fraction of populations present only in low abundance (Table 1).

The bacterial communities of each host group differed strongly already at the phylum level (Fig. 1). *Spirochaetes* were rare in cockroaches but abundant in lower termites and wood-feeding higher termites, often representing the majority of the reads, which is in agreement with the general notion that spirochaetes are the most characteristic element of the termite gut microbiota (42). *Firmicutes* and *Bacteroidetes* were generally more abundant in cockroaches than in termites, except for a large proportion of *Firmicutes* in all soil-feeding higher termites and *Bacteroidetes* in the lower termite *Coptotermes niger*, again confirming results previously obtained for selected species (17, 43, 44, 45). Members of *Elusimicrobia* were highly represented only in lower termites.

Ordination analysis revealed high similarities among the bacterial microbiota of the different host groups. The robust clustering of samples based on genus-level classification (Fig. 2) was found also with classification-independent (OTU-based) and phylogeny-based (UniFrac) approaches (Fig. S2 in the supplementary material). In all cases, cockroaches were clearly separated from termites, and lower termites from higher termites. Also the different subfamilies of *Termitidae* formed discrete clusters, with the fungus-cultivating *Macrotermiteinae* showing a strong affinity to the cockroaches. A notable exception was the wood-feeding cockroach *Cryptocercus punctulatus*, the closest relative of termites (4, 6). It did not cluster among the other cockroaches but was always more similar to lower termites, with which it shares the presence of cellulolytic flagellates. The (unrelated) wood-feeding cockroaches *Panesthia angustipennis* and *Salganea esakii* (family Blaberidae), whose gut microbiota lacks such flagellates, clustered with the omnivorous cockroaches.

When we ranked all 884 genus-level taxa in the dataset according to their contribution to the ordination results, it became apparent that already the top 100 genera were responsible for almost 70% of the pattern and represented 90% of the sequences in the dataset (Table S3). Many genus-level taxa occurred in all major host lineages, extending the previously postulated presence of...
termite-specific bacterial lineages to all cockroaches (17, 19), but with distinct differences in their relative distribution (Fig. 3, groups A and D). An obvious break in the pattern between cockroaches and termites (Fig. 3, groups B, C, and D) indicated that the transition from an omnivorous to a wood-feeding lifestyle had a strong impact on bacterial community structure. Bacterial lineages abundant in cockroaches decreased in frequency in lower termites, and rare lineages dramatically increased. The latter was most obvious in the spirochetal cluster Treponema Ia and matches with the dominance of Spirochaetes in the gut of wood-feeding termites (11, 42). Since this cluster comprises the homoacetogenic Treponema primitia (46), its upshift is also consistent with changes in the distribution of several functional marker genes (formyltetrahydrofolate synthetase, CO dehydrogenase and hydrogenase) (e.g. 47, 48, 49), which indicated that the bacteria responsible for reductive acetogenesis in omnivorous cockroaches are not the same as those in wood-feeding termites and C. punctulatus.

Several of the genus-level taxa that predominated only in lower termites (Fig. 3, group C) represent lineages that harbor specific symbionts of termite gut flagellates. Taxa comprising ectosymbiotic spirochetes (Treponema II) (50) and endosymbiotic ‘Endomicrobium’ (15, 51) and Desulfovibrio spp. (TC I) (52, 53) were abundant only in those termites that harbor the respective host flagellates. However, low numbers of Endomicrobia were consistently present also in cockroaches and higher termites, corroborating the presence of putatively free-living relatives (54) that were recruited as endosymbionts presumably long after the flagellates had established their symbiosis with lower termites (55). Also the dynamic patterns of Cluster-V Bacteroidetes among the lower termites (Fig. 3, groups A and C), which harbor several lineages of symbionts that have strictly cospeciated with their respective flagellate host (19, 56), is in agreement with their recruitment from free-living relatives that are present but of low abundance in termites lacking these flagellates (19).

The second obvious break in the community patterns was between lower and higher termites, marking a decrease in abundance of the flagellate-associated bacterial lineages and a strong increase in several other taxa (Fig. 3, groups C and E). The dominance of termite-specific clusters of Fibrobacteres, the TG3 phylum, and certain Treponema lineages (Ib and Ic) in wood- and grass-feeding termites is consistent with previous reports on the distribution of these groups (13). There is strong evidence from enzymatic (57) and metagenomic (58, 59) studies of
Nasutitermes and Amitermes spp. that bacterial members of the gut microbiota—particularly Fibrobacteres (possibly including the related TG3 phylum) and Spirochaetes—took over the function of the flagellates in fiber digestion. Our results indicate that these putative cellulose-digesting bacteria are apparently represented already among lower termites but cannot form large populations because the protists sequester all wood particles into their food vacuoles, restricting the bacteria to soluble substrates. Thus, the dramatic changes in the bacterial community between lower and higher termites are probably due to both the gain of new substrates and the loss of the flagellate niche.

Also the resurgence in the Macrotermiteinae of taxa that prevail in cockroaches may be related to the dietary diversification of higher termites following the loss of flagellates. The high similarity between the gut microbiota of omnivorous cockroaches and Macrotermiteinae, first discovered in a study of the gut microbiota of the cockroach S. lateralis (17), is rooted in the shared presence of lineages that are only of low abundance in wood- or soil-feeding termites (Fig. 3, group D). A predominance of Bacteroidetes and Firmicutes seems to be a common feature of omnivorous mammals (60, 61); their abundance in the Macrotermiteinae may be caused by the protein-rich fungal biomass included in the diet of the fungus-cultivating species (62).

Most genus-level taxa were unevenly distributed across cockroaches and lower and higher termites, but many of them were consistently represented among the members of the major host groups (Fig. 4). Although the variable taxa were at least five times more numerous than these core taxa, they represented less than a quarter of the reads in the respective datasets (Table 3).

The 30 taxa common to the three groups included members of the Spirochaetes (Table S3 in the supplementary material), underlining that small populations of the most typical element of the termite gut microbiota are present also in cockroaches. Although these core taxa represented only a small fraction of the genus-level diversity, they made up almost half of the reads in the entire dataset (Fig. 5). Taxa common to termites but not regularly present in cockroaches (Table S3 in the supplementary material) represented only 8.7% of the reads. It is important to note that the core taxa are not always restricted to dictyopteran hosts. Almost half of the core taxa (14 of 30) were present also in the three other insect species included in this study, representing a large proportion of the reads in these samples (Pachnoda ephippiata, 22%; Acheta domesticus, 45%; Gryllus assimilis, 26%). The most abundant lineages in these insects that were shared with most
dictyopteran samples were Gut cluster 2 (*Lachnospiraceae*) in the scarab beetle larva (*P. ephippiata*) and *Alistipes* 2 (*Rikenellaceae*) and *Dysgonomonas* (*Porphyromonadaceae*) in the crickets (*A. domesticus* and *G. assimilis*).

Despite the abundant presence of lineages that are not restricted to dictyopteran hosts, a UniFrac analysis of the core taxa retained a clear host signal in the phylogeny of its components (Fig. 6). Cockroaches formed a sister group of the *Cryptocercus*/termite clade, and higher termites were apical to all lineages with gut flagellates. However, the internal topology of the cladogram often did not match the branching order of the host tree (Fig. 3F), particularly in the cockroaches and lower termites, indicating that the dictyopteran core microbiota is not caused by cospeciation. Rather, the lack of clustering among the gut microbiota of blattid cockroaches and the proximity of wood-feeding blaberid cockroaches (*Panesthia angustipennis* and *Salganea esakii*) to the *Cryptocercus*/termite clade suggest that factors other than host phylogeny must shape the bacterial community structure.

Closer inspection of the genus-level taxa that contribute most to the separation of the major host groups revealed that phylogenetic clustering is often restricted to sequences from particular host groups (e.g., Fig. S4O,R). Although the results of phylogenetic analyses based on short sequences must be interpreted with caution, it is noteworthy that the basal position of sequences from cockroach guts relative to those from termites and the apical position of sequences from higher termites are frequent themes. However, it remains unclear whether the lineages in question were associated already with the ancestral cockroaches (cospeciation) or are diet-specific lineages that were independently acquired from the environment (host selection). The latter would explain the frequently observed quasi-random occurrence of the same genus-level lineages among different host groups. A prominent example is *Alistipes* 2, which is highly abundant also in the cricket (*Achaeta domesticus*) and contributes to the similarity of its core microbiota to that of several cockroaches (Fig. S4C). The unexpected presence of cockroach clusters in higher termites (e.g., Fig. S4H) suggests that also the horizontal transfer of microbiota between different host lineages has to be considered. A puzzling phenomenon is the presence of a small number of sequence-identical phylotypes within sequence clusters of entirely different hosts (e.g., Fig. S4N). Here, horizontal transfer or environmental uptake are unlikely
explanations, and the possibility of methodological artifacts (e.g., mistagging of templates during the emulsion PCR; 63) has to be considered.

It remains to be investigated whether traces of host phylogeny can be found also in the archaenal microbiota in the guts of termites and cockroaches. Although archaea are much less abundant than bacteria (0.1–3% of prokaryotes in termite guts; 64), methanogens seem to be present in all dictypteran lineages (65). The diversity of the archaenal community is much smaller than that of bacteria and comprises both termite-specific clusters and lineages with representatives from many environments (e.g., 23, 66).

**Conclusions.** This study provides a new view of the complex bacterial communities in the gut microbiota of termites. Clearly, phylogeny is not the only driver of community structure in the dictyopteran microbiota. Changes in the quality of the diet (lignin and fiber content, humification state) or the provision of new niches for nitrogen-fixing or -upgrading symbionts promoted bacteria from different functional guilds that were either already present in the microbial seed bank of the gut (19) or newly acquired from the environment—and caused their decline when such services were no longer required.

The results of our study are the foundation for future studies targeting the specific roles of important bacterial populations by metagenomic and metatranscriptomic analysis or single-cell approaches. Of particular interest will be mechanisms of bacterial cellulose degradation and humus digestion in higher termites (59), microbial interactions in hydrogen metabolism and methanogenesis (67), and the emerging role of the flagellate symbionts in the nitrogen economy of the digestive symbiosis (8). Following Dobzhansky's famous dictum (68), the complex patterns in the gut microbiota of this ancient group of insects make sense only in the light of evolution.

**Acknowledgements** We thank Christine Nalepa (North Carolina State University), Rudy Plarre (Federal Institute for Materials Research and Testing, Berlin), Rudolf Scheffrahn (University of Florida), Gaku Tokuda (University of the Ryukyus), and our coworkers David Ngugi and James Nonoh (MPI Marburg) for providing insect samples, and Katja Meuser for technical assistance.
References


**Figures**

**FIG 1** Relative abundance of the most prevalent bacterial phyla in the gut microbiota of different host groups. O, other insects; C, cockroaches; L, lower termites; M, T and N, the higher termites *Macroterminae, Termitinae*, and *Nasutitermitinae*, respectively. *Cryptocercidae* and *Apicotermitinae* were not included because each group was represented by only a single species. Bars show the range and median number of sequence reads assigned to the respective phylum. Detailed results for all bacterial phyla and individual host species are shown in Fig. S1 in the supplementary material. Colors of the major host groups are the same in all figures.

**FIG 2** Phylogenetic patterns in the community structure of the bacterial gut microbiota of the different host species. Community similarities (Bray-Curtis) were calculated based on distribution of genus-level taxa (Table S1 in the supplementary material) and visualized by non-metric multidimensional scaling (NMDS; stress 11.3%). The clusters formed by samples from the major host lineages have the same colors in all figures. Symbols indicate feeding habits: generalists (○), wood feeding (▲), grass feeding (▲), soil/humus feeding (○), and fungus cultivating (■). Two species of crickets (*Gryllidae*) and a beetle larva (*Scarabaeidae*) were included as outgroups. For the species behind each data point, see (Fig. 3). The clusters were supported by both ADONIS and MRPP analyses (p < 0.001).

**FIG 3** Relative abundance of selected bacterial lineages in the gut microbiota that contribute strongly to the separation of cockroaches, lower termites, and higher termites in the ordination analyses. The lineages were selected from the top 50 taxa (see Table S3 in the supplementary material) and subjectively sorted according to patterns (A–E) explained in the text. The color scale is logarithmic to emphasize rare taxa. Numbers indicate host species (see Table 1). Symbols indicate feeding habits (see legend of Fig. 2). The tree (F) illustrates the simplified phylogeny of major host lineages (a, other cockroaches; b, *Blattidae*; c, *Cryptocercidae*; d, *Mastotermidae*; e, *Termopsidae*; f, *Hodotermitidae*; g, *Kalotermitidae*; h, *Rhinotermitidae*; i, *Macroterminae*; j, *Apicotermitinae*; k, *Termitinae*; l, *Nasutitermitinae*). The branches connecting species that harbor gut flagellates are marked in red.
FIG 4  Ternary plot of the distribution of genus-level taxa across the major host groups. The area of the circles represents the relative abundance of the reads in the entire dataset, the position specifies their average abundance in the respective host groups, and the colors indicate the number of host groups in which core status is attained (presence in > 70% of the hosts; data from Table S3 in the supplementary material). An interactive version that allows one to identify the genus behind each data point of the figure is included as a supplementary file (Fig. S3).

FIG 5  Representation of core taxa in the dataset based on the number of genera or sequence reads in the entire dataset. Core status was assigned if a taxon was represented by > 70% of the host species in a major host group (cockroaches, lower termites, and higher termites).

FIG 6  Cladogram of community similarities based on the core taxa common to all major host groups (unweighted UniFrac of the sequences belonging to the core genera). For sample numbers, see Table 1. Symbols indicate lifestyle: generalists (○), wood feeding (▲), grass feeding (Δ), soil/humus feeding (○), and fungus cultivating (■). The branches connecting species that harbor gut flagellates are marked in red.
Tables

**TABLE 1** Characteristics of the 16S rRNA gene amplicon libraries of the bacterial hindgut microbiota of each host species. The same numbers are used to identify the samples in all tables and figures.

**TABLE 2** Improvement of classification success using our curated reference database that included all homologs previously obtained from insect guts and optimized to resolve all termite- and cockroach-specific groups (20) over that using the RDP database (Release 10, Update 30). The proportions of classified sequences in representative samples are reported for different taxonomic levels. Classification success for all samples is shown in Table S2 in the supplementary material.

**TABLE 3** Number of genus-level taxa considered variable and core taxa, and their average contribution to the gut communities in the major host groups. Core taxa were defined as those present in > 70% of the species of the respective host groups (see Table S3 in the supplementary material).
Relative abundance of the most prevalent bacterial phyla in the gut microbiota of different host groups. O, other insects; C, cockroaches; L, lower termites; M, T and N, the higher termites Macrotermitea, Termitea, and Nasutitermitea, respectively. Cryptocercidae and Apicotermitidae were not included because each group was represented by only a single species. Bars show the range and median number of sequence reads assigned to the respective phylum. Detailed results for all bacterial phyla and individual host species are shown in Fig. S1 in the supplementary material. Colors of the major host groups are the same in all figures.
FIG 2 Phylogenetic patterns in the community structure of the bacterial gut microbiota of the different host species. Community similarities (Bray-Curtis) were calculated based on distribution of genus-level taxa (Table S1) and visualized by non-metric multidimensional scaling (NMDS; stress 11.3%). The clusters formed by samples from the major host lineages have the same colors in all figures. Symbols indicate feeding habits: generalists (□), wood feeding (▲), grass feeding (△), soil/humus feeding (○), and fungus cultivating (■). Two species of crickets (Gryllidae) and a beetle larva (Scarabaeidae) were included as outgroups. For the species behind each data point, see (Fig. S2). The clusters were supported by both ADONIS and MRPP analyses (p < 0.001).
FIG 3 Relative abundance of selected bacterial lineages in the gut microbiota that contribute strongly to the separation of cockroaches, lower termites, and higher termites in the ordination analyses. The lineages were selected from the top 50 taxa (see Table S3 in the supplementary material) and subjectively sorted according to patterns (A–E) explained in the text. The color scale is logarithmic to emphasize rare taxa. Numbers indicate host species (see Table 1). Symbols indicate feeding habits (see legend of Fig. 2). The tree (F) illustrates the simplified phylogeny of major host lineages (a, other cockroaches; b, Blattidae; c, Cryptocercidae; d, Mastotermitidae; e, Termopsidae; f, Hodotermitidae; g, Kalotermitidae; h, Rhinotermitidae; i, Macrotermitinae; j, Apicotermitinae; k, Termitinae; l, Nasutitermitinae). The branches connecting species that harbor gut flagellates are marked in red.
**FIG 4** Ternary plot of the distribution of genus-level taxa across the major host groups. The area of the circles represents the relative abundance of the reads in the entire dataset, the position specifies their average abundance in the respective host groups, and the colors indicate the number of host groups in which core status is attained (presence in > 70% of the hosts; data from Table S3 in the supplementary material). An interactive version that allows to identify the genus behind each data point of the figure can be found as supplementary file Fig_S3.svg. It has been tested with the common browsers Internet Explorer, Firefox and Chrome.
FIG 5: Representation of core taxa in the dataset based on the number of genera or sequence reads in the entire dataset. Core status was assigned if a taxon was represented by >70% of the host species in a major host group (cockroaches, lower termites, and higher termites).
FIG 6 Cladogram of community similarities based on the core taxa common to all major host groups (unweighted UniFrac of the sequences belonging to the core genera). For sample numbers, see Table 1: generalists (□), wood feeding (▲), grass feeding (△), soil/humus feeding (○), and fungus cultivating (■). The branches connecting species that harbor gut flagellates are marked in red.
<table>
<thead>
<tr>
<th>Sample</th>
<th>Host species</th>
<th>Origin</th>
<th>Total reads</th>
<th>Genus-level taxa</th>
<th>OTUs (3% dissim.)</th>
<th>Coverage (%)</th>
<th>Diversity indices</th>
<th>NCBI accession</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Richness</td>
<td>number</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Diversity</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Evenness</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cockroaches</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Ergaula capucina</td>
<td>B1</td>
<td>6,020</td>
<td>232</td>
<td>891</td>
<td>70.5</td>
<td>1,266</td>
<td>5.52</td>
</tr>
<tr>
<td>2</td>
<td>Symphoche maconterta</td>
<td>B1</td>
<td>5,045</td>
<td>135</td>
<td>499</td>
<td>80.9</td>
<td>431</td>
<td>4.70</td>
</tr>
<tr>
<td>3</td>
<td>Rhyparobia maderae</td>
<td>B1</td>
<td>12,164</td>
<td>268</td>
<td>1346</td>
<td>70.8</td>
<td>2,540</td>
<td>5.42</td>
</tr>
<tr>
<td>4</td>
<td>Elliptorhina chopardi</td>
<td>B1</td>
<td>6,794</td>
<td>200</td>
<td>663</td>
<td>79.6</td>
<td>798</td>
<td>5.25</td>
</tr>
<tr>
<td>5</td>
<td>Panchlora sp.</td>
<td>B1</td>
<td>11,889</td>
<td>212</td>
<td>2042</td>
<td>66.6</td>
<td>1,064</td>
<td>4.41</td>
</tr>
<tr>
<td>6</td>
<td>Diploptera punctata</td>
<td>B1</td>
<td>5,708</td>
<td>161</td>
<td>543</td>
<td>80.8</td>
<td>627</td>
<td>4.93</td>
</tr>
<tr>
<td>7</td>
<td>Opisthoplatia orientalis</td>
<td>B1</td>
<td>11,707</td>
<td>291</td>
<td>1515</td>
<td>70.5</td>
<td>3,153</td>
<td>5.72</td>
</tr>
<tr>
<td>8</td>
<td>Panesthia angustipennis</td>
<td>B1</td>
<td>5,394</td>
<td>202</td>
<td>1141</td>
<td>72.5</td>
<td>1,710</td>
<td>6.01</td>
</tr>
<tr>
<td>9</td>
<td>Saliganea esakii</td>
<td>B1</td>
<td>17,412</td>
<td>296</td>
<td>1916</td>
<td>80.8</td>
<td>2,955</td>
<td>6.27</td>
</tr>
<tr>
<td>10</td>
<td>Eublaberus posticus</td>
<td>B1</td>
<td>103,530</td>
<td>416</td>
<td>5743</td>
<td>79.9</td>
<td>12,034</td>
<td>5.34</td>
</tr>
<tr>
<td>11</td>
<td>Schultesi lampyridiformis</td>
<td>B1</td>
<td>5,085</td>
<td>217</td>
<td>857</td>
<td>70.3</td>
<td>1,482</td>
<td>5.42</td>
</tr>
<tr>
<td>12</td>
<td>Euryctes floridana</td>
<td>B1</td>
<td>41,336</td>
<td>354</td>
<td>3410</td>
<td>77.2</td>
<td>6,855</td>
<td>5.80</td>
</tr>
<tr>
<td>13</td>
<td>Shelfordella lateralis</td>
<td>B1</td>
<td>6,226</td>
<td>186</td>
<td>714</td>
<td>82.4</td>
<td>674</td>
<td>5.30</td>
</tr>
<tr>
<td>14</td>
<td>Blatta orientalis</td>
<td>B1</td>
<td>8,024</td>
<td>246</td>
<td>1069</td>
<td>68.6</td>
<td>2,045</td>
<td>5.14</td>
</tr>
<tr>
<td>15</td>
<td>Cryptocercus punctatus</td>
<td>F1</td>
<td>6,715</td>
<td>180</td>
<td>715</td>
<td>75.5</td>
<td>844</td>
<td>4.90</td>
</tr>
<tr>
<td>Lower termites</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>Mastotermes darwiniensis</td>
<td>L1</td>
<td>7,596</td>
<td>137</td>
<td>398</td>
<td>86.3</td>
<td>583</td>
<td>3.94</td>
</tr>
<tr>
<td>17</td>
<td>Zootermopsis nevadensis</td>
<td>L2</td>
<td>6,129</td>
<td>278</td>
<td>1617</td>
<td>72.6</td>
<td>3,451</td>
<td>5.18</td>
</tr>
<tr>
<td>18</td>
<td>Hodotermopsis sjoestedtii</td>
<td>L1</td>
<td>7,600</td>
<td>272</td>
<td>1584</td>
<td>73.9</td>
<td>3,569</td>
<td>5.25</td>
</tr>
<tr>
<td>19</td>
<td>Hodotermes mossambicus</td>
<td>F2</td>
<td>16,520</td>
<td>204</td>
<td>978</td>
<td>74.3</td>
<td>1,840</td>
<td>5.33</td>
</tr>
<tr>
<td>Sample</td>
<td>Host species</td>
<td>Origin</td>
<td>Total reads</td>
<td>Genus-level taxa</td>
<td>OTUs (3% dissim.)</td>
<td>Coverage (%)</td>
<td>Diversity indices (based on OTUs)</td>
<td>NCBI accession number</td>
</tr>
<tr>
<td>--------</td>
<td>----------------------------</td>
<td>---------</td>
<td>-------------</td>
<td>------------------</td>
<td>-------------------</td>
<td>--------------</td>
<td>-----------------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>20</td>
<td>Incisitermes marginipennis</td>
<td>L1</td>
<td>16,491</td>
<td>299</td>
<td>2807</td>
<td>79.0</td>
<td>6,354 4.27 0.56 087</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>Neotermes jouteli</td>
<td>F3</td>
<td>6,256</td>
<td>276</td>
<td>2354</td>
<td>78.4</td>
<td>4,547 4.70 0.63 088</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Reticulitermes santonensis</td>
<td>L2</td>
<td>48,066</td>
<td>112</td>
<td>427</td>
<td>85.1</td>
<td>602 3.92 0.67 089</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>Coptotermes niger</td>
<td>L1</td>
<td>53,003</td>
<td>91</td>
<td>166</td>
<td>87.2</td>
<td>202 2.26 0.45 090</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>Higher termites</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>Odontotermes sp.*</td>
<td>F4</td>
<td>12,898</td>
<td>307</td>
<td>1005</td>
<td>63.1</td>
<td>1,391 5.77 0.86 091</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>Macrotermes sp.</td>
<td>L1</td>
<td>12,073</td>
<td>260</td>
<td>1358</td>
<td>69.5</td>
<td>2,790 5.34 0.76 092</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>Macrotermes subhyalinus*</td>
<td>F5</td>
<td>27,297</td>
<td>211</td>
<td>4805</td>
<td>68.4</td>
<td>1,182 5.23 0.84 093</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>Alyscotermes testus*</td>
<td>F5</td>
<td>24,582</td>
<td>550</td>
<td>3203</td>
<td>78.6</td>
<td>5,940 6.57 0.82 094</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>Cubitermes ugandensis</td>
<td>F6</td>
<td>22,832</td>
<td>211</td>
<td>5413</td>
<td>49.7</td>
<td>2,020 6.49 0.97 095</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>Ophiotermes sp.*</td>
<td>F7</td>
<td>8,418</td>
<td>328</td>
<td>1336</td>
<td>76.0</td>
<td>2,026 6.13 0.85 096</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>Amitermes meridionalis*</td>
<td>F8</td>
<td>23,840</td>
<td>354</td>
<td>1556</td>
<td>85.5</td>
<td>2,246 5.04 0.70 097</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>Microcerotermes sp.*</td>
<td>F5</td>
<td>34,626</td>
<td>291</td>
<td>2358</td>
<td>79.1</td>
<td>4,407 4.55 0.61 098</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>Nasutitermes comiger</td>
<td>L3</td>
<td>10,363</td>
<td>175</td>
<td>1998</td>
<td>65.5</td>
<td>1,208 4.15 0.67 099</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nasutitermes</td>
<td>F9</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>Takasagoensis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>Trinervitermes sp.</td>
<td>F5</td>
<td>25,173</td>
<td>232</td>
<td>1103</td>
<td>84.2</td>
<td>1,943 4.68 0.67 101</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>Others</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>Pachnoda ephippiata</td>
<td>B2</td>
<td>10,033</td>
<td>339</td>
<td>1325</td>
<td>80.2</td>
<td>1,335 5.77 0.85 102</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>Acheta domesticus</td>
<td>B2</td>
<td>5,326</td>
<td>104</td>
<td>241</td>
<td>84.2</td>
<td>276 4.14 0.80 103</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>Gryllus assimilis</td>
<td>B1</td>
<td>26,800</td>
<td>190</td>
<td>669</td>
<td>90.3</td>
<td>712 4.13 0.70 104</td>
<td></td>
</tr>
</tbody>
</table>
Origin of samples: B, commercial breeders: B1, Jörg Bernhardt, Helbigsdorf, Germany (http://www.schaben-spinne.de); B2, b.t.b.e. Insektenzucht, Schnürpflingen, Germany. F, field collections: F1, Heywood County, NC, USA (by C. Nalepa); F2, near Pretoria, South Africa (by J. Rohland); F3, Fort Lauderdale, Florida, USA (by R.H. Scheffrahn); F4, near Kajiado, Kenya; F5, near Nairobi, Kenya (by J.O. Nonoh); F6, Lhiranda Hill, Kakamega, Kenya; F7, Kalunja Glade, Kakamega, Kenya (by D.K. Ngugi); F8, Lakefield NP, Cape York, Australia (by A. Brune); F9, near Nishihara, Japan (by G. Tokuda). L, laboratory colonies: L1, R. Pfarre, Federal Institute for Materials Research and Testing, Berlin, Germany; L2 MPI Marburg; L3, R.H. Scheffrahn, University of Florida, Fort Lauderdale, Florida, USA.

Good's coverage estimator (37)

Richness: Chao 1 estimator (38); diversity: non-parametric Shannon index (39); evenness index (40).

All datasets were submitted to the Sequence Read Archive of NCBI. The full accession number is SAMN02228nnn – the last three digits are indicated in the table.

Ethanol-preserved specimen; the entire gut was used for DNA extraction.
TABLE 2 Improvement of classification success using our curated reference database that included all homologs previously obtained from insect guts and optimized to resolve all termite- and cockroach-specific groups (20) over that using the RDP database (Release 10, Update 30). The proportion of classified sequences in representative samples are reported for different taxonomic levels. Classification success for all samples is shown in Table S2 in the supplemental material.

<table>
<thead>
<tr>
<th>Host species</th>
<th>Phylum</th>
<th>Classification success (%)</th>
<th>RDP</th>
<th>This study</th>
<th>Class</th>
<th>Classification success (%)</th>
<th>RDP</th>
<th>This study</th>
<th>Order</th>
<th>Classification success (%)</th>
<th>RDP</th>
<th>This study</th>
<th>Family</th>
<th>Classification success (%)</th>
<th>RDP</th>
<th>This study</th>
<th>Genus</th>
<th>Classification success (%)</th>
<th>RDP</th>
<th>This study</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cockroaches</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ergaula capucina</td>
<td>94.9</td>
<td>99.3</td>
<td>79.3</td>
<td>96.1</td>
<td>73.4</td>
<td>93.1</td>
<td>63.3</td>
<td>88.0</td>
<td>33.2</td>
<td>61.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Panesthia angustipennis</td>
<td>95.9</td>
<td>99.3</td>
<td>79.3</td>
<td>96.1</td>
<td>77.2</td>
<td>94.2</td>
<td>59.0</td>
<td>86.0</td>
<td>24.8</td>
<td>61.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Salganea esakii</td>
<td>93.8</td>
<td>99.4</td>
<td>79.7</td>
<td>95.3</td>
<td>78.8</td>
<td>93.5</td>
<td>71.1</td>
<td>90.1</td>
<td>40.6</td>
<td>74.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blatta orientalis</td>
<td>96.2</td>
<td>99.3</td>
<td>79.8</td>
<td>97.7</td>
<td>77.9</td>
<td>96.3</td>
<td>70.5</td>
<td>93.4</td>
<td>48.0</td>
<td>66.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cryptocercus punctulatus</td>
<td>93.8</td>
<td>99.0</td>
<td>74.8</td>
<td>93.4</td>
<td>72.5</td>
<td>90.2</td>
<td>59.0</td>
<td>82.8</td>
<td>33.1</td>
<td>67.7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Termites</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reticulitermes santonensis</td>
<td>85.5</td>
<td>99.2</td>
<td>80.9</td>
<td>96.2</td>
<td>79.2</td>
<td>95.6</td>
<td>76.1</td>
<td>94.5</td>
<td>68.0</td>
<td>93.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cubitermes ugandensis</td>
<td>92.0</td>
<td>98.7</td>
<td>84.4</td>
<td>97.4</td>
<td>78.4</td>
<td>94.9</td>
<td>64.7</td>
<td>83.3</td>
<td>24.3</td>
<td>69.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Nasutitermes corniger</td>
<td>87.8</td>
<td>98.9</td>
<td>83.0</td>
<td>97.4</td>
<td>81.6</td>
<td>96.5</td>
<td>77.9</td>
<td>95.3</td>
<td>40.9</td>
<td>91.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TABLE 3 Number of genus-level taxa considered variable and core taxa, and their average contribution to the gut communities in the major host groups. Core taxa were defined as those present in > 70% of the species of the respective host groups (Table S3 in the supplemental material).

<table>
<thead>
<tr>
<th></th>
<th>Cockroaches</th>
<th>Lower termites</th>
<th>Higher termites</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Variable</td>
<td>Core</td>
<td>Variable</td>
</tr>
<tr>
<td>Number of taxa</td>
<td>363</td>
<td>67</td>
<td>270</td>
</tr>
<tr>
<td>Percent of reads</td>
<td>21.2</td>
<td>78.8</td>
<td>22.1</td>
</tr>
</tbody>
</table>