

**ECOLOGICAL AND EVOLUTIONARY  
DRIVERS OF MICROBIAL  
COMMUNITY STRUCTURE  
IN TERMITE GUTS**

**Dissertation**

zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.)  
am Fachbereich Biologie der Philipps-Universität Marburg,

vorgelegt von

**Carsten Dietrich**

aus Erfurt.

Universitätsstadt Marburg, 2015





---

Die Untersuchungen zur folgenden Arbeit wurden von Oktober 2011 bis Februar 2015 am Max-Planck-Institut für terrestrische Mikrobiologie in Marburg in der Forschungsgruppe von Prof. Dr. Andreas Brune durchgeführt.

Vom Fachbereich Biologie der Philipps-Universität Marburg als Dissertation angenommen am: 11.05.2015

Tag der Disputation: 27.05.2015

Erstgutachter: Prof. Dr. Andreas Brune

Zweitgutachter: Prof. Dr. Roland Brandl

Drittgutachter: Prof. Dr. Stefan Rensing

---

---

## Folgende Publikationen sind aus dieser Dissertation entstanden:

- Brune A und **Dietrich C** (2015). The termite gut microbiota: Digesting the diversity in the light of ecology and evolution. *Ann. Rev. Microbiol.* 69.
- Dietrich C** und Brune A (2014). The complete mitogenomes of six higher termite species reconstructed from metagenomic datasets (*Cornitermes* sp., *Cubitermes ugandensis*, *Microcerotermes parvus*, *Nasutitermes corniger*, *Neocapritermes taracua*, and *Termes hospes*). *Mitochondr. DNA.* Dec 4:1-2.
- Dietrich C\***, Köhler T\* und Brune A (2014). The cockroach origin of the termite gut microbiota: patterns in bacterial community structure reflect major evolutionary events. *Appl. Environ. Microbiol.* 80:2261-2269.
- Dietrich C\***, Nonoh J\*, Lang K, Mikulski L, Meuser K, Köhler T, Boga HI, Ngugi DK, Sillam-Dussès S und Brune A (eingereicht). Habitat selection and vertical inheritance drive archaeal community structure in arthropod guts.
- Köhler T, **Dietrich C**, Scheffrahn RH und Brune A (2012). High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation of the gut in wood-feeding higher termites (*Nasutitermes* spp.). *Appl. Environ. Microbiol.* 78:4691-4701.
- Li H\*, **Dietrich C\***, Zhu N, Mikaelyan A, Ma B, Pi R, Liu Y, Yang M, Brune A und Mo J (eingereicht). Age polyethism drives community structure of the bacterial gut microbiota in the fungus-cultivating termite *Odontotermes formosanus*.
- Mikaelyan A, **Dietrich C**, Köhler T, Meuser K, Sillam-Dussès D und Brune A (eingereicht). Dietary and phylogenetic determinants of bacterial gut community structure in higher termites.
- Zheng H, **Dietrich C**, Thompson CL, Meuser K und Brune A (2015). Population structure of Endomicrobia in single host cells of termite gut flagellates (*Trichonympha* spp.). *Microb. Environ.* 30: 92-98.

## Folgende Publikationen aus dieser Dissertation sind in Vorbereitung:

- Dietrich C**, Rossmassler K und Brune A (in Vorbereitung). Genome sequences of *Fibrobacteres* and TG3 obtained by marker-assisted compositional binning of multiple metagenomes.
- Rossmassler K, **Dietrich C**, Thompson C, Mikaelyan A und Brune A (in Vorbereitung). Comparative metagenomics reveal bacterial functional profiles in the hindgut compartments of higher termites.

## Folgende Publikationen sind nicht in dieser Dissertation enthalten, aber während meiner Studien entstanden:

- Guichard P, Desfosses A, Maheshwari A, Hachet V, **Dietrich C** und Brune A, Ishikawa T, Sachse, C und Gönczy P (2012). Cartwheel architecture of *Trichonympha* basal body. *Science* 337:553-554.
- Meyer BH, Peyfoon E, **Dietrich C**, Hitchen P, Panico M, Morris HR, Dell A und Albers SV (2013). Agl16, a thermophilic glycosyltransferase mediating the last step of N-glycan biosynthesis in the thermoacidophilic crenarchaeon *Sulfolobus acidocaldarius*. *J. Bacteriol.* 195:2177-2186.
- Zheng H, **Dietrich C**, Radek R und Brune A (eingereicht). *Endomicrobium proavitum*, the first isolate of Endomicrobia class. nov. (phylum *Elusimicrobia*) – an ultramicrobacterium with an unusual cell cycle that fixes nitrogen with a Group IV nitrogenase.

## Weiterhin sind folgende Beiträge während meiner Studien entstanden:

- Lang K, Mikulski L, **Dietrich C**, Meuser K und Brune A (2014). Methanogenic community structure in the digestive tracts of tropical millipedes. In Lang K. Diversity and ultrastructure of the seventh order of methanogens. Doctoral Thesis. University of Marburg.
- Nonoh JO, **Dietrich C** und Brune A (2013). Distribution of archaeal populations and methanogenic potentials in the highly compartmented gut of soil-feeding termites (*Cubitermes ugandensis* and *Ophiotermes* sp.). In Nonoh JO. Archaeal diversity and community structure in the compartmented gut of higher termites. Doctoral Thesis. University of Marburg.
- Nonoh JO, **Dietrich C** und Brune A (2013). Microenvironment and community structure of Archaea and Bacteria in the highly compartmented gut of *Amitermes* sp. (*Isoptera: Termitinae*). In Nonoh JO. Archaeal diversity and community structure in the compartmented gut of higher termites. Doctoral Thesis. University of Marburg.

\* Diese Autoren haben zu gleichen Teilen beigetragen.



---

## Acknowledgements

### Ich danke...

- ...meinem Doktorvater und Mentor **Prof. Dr. Andreas Brune** für sein großes Engagement und das Vertrauen, das er mir entgegengebracht hat. Besonders bedanke ich mich für seine Motivation, Ratschläge und stete Diskussionsbereitschaft, die immer wieder neue Impulse für diese Arbeit gab. Ganz besonders danke ich ihm für den Freiraum, den er mir für meine eigenen Ideen ließ und ich mich dadurch immer weiter entwickeln konnte.
- ...**Prof. Dr. Roland Brandl, Prof. Dr. Stefan Rensing, Prof. Dr. Uwe Maier** und **Prof. Dr. Martin Thanbichler** für ihre Bereitschaft diese These zu begutachten und für ihre hilfreiche Kritik und anregenden Vorschläge.
- ...**Prof. Dr. Ralf Conrad, Prof. Dr. Johann Heider** und **Dr. Marc Dumont** für ihre hilfreiche Kritik in meinem Thesis Advisory Committee und auch dafür, dass sie mich abseits des TACs immer wieder unterstützt haben.
- ...**Prof. Dr. Ute Windisch** und **Prof. Dr. Harald Platen** für deren immer noch anhaltende Unterstützung in meinem Werdegang und deren wertvolle Lehrveranstaltungen, die mich bis heute prägen.
- ...der **International Max Planck Research School for Environmental, Cellular and Molecular Microbiology** für das Stipendium, das ich erhalten habe. Weiterhin bedanke ich mich bei allen Mitgliedern meines Jahrgangs für die hilfreichen Anmerkungen während der Seminare.
- ...der gesamten AG Brune inklusive der Ehemaligen für ihre Forschungsergebnisse, die die Formulierung der Hypothesen meiner Arbeit erst erlaubt haben. Weiterhin bedanke ich mich für die anregende Kritik im Termiten-Seminar. Insbesondere möchte ich mich noch bei **Dr. Claire Thompson, Dr. Tim Köhler, Dr. Jürgen Strassert, Dr. Aram Mikaelyan, Niclas Lampert** und **Hao Zheng** für das tägliche wissenschaftliche Kaffeekränzchen bedanken. Weiterhin bedanke ich mich bei **Dr. David Ngugi**, der mich neugierig auf die AG Brune gemacht hat. Ich danke zudem der unverzichtbaren **Katja Meuser** dafür, dass sie meinen überschwänglichen „Ordnungsdrang“ ausgehalten hat und stets jedem hilft.
- ...der gesamten Abteilung der Biogeochemie und deren Ehemaligen. Besonders bedanke ich mich bei **Dr. Marcela Hernández** und **Dr. Marc Dumont** für deren Interesse an meiner Arbeit. Mein besonderer Dank gilt **Dr. Jennifer Pratscher**, die mir in meiner Anfangszeit sehr geholfen hat.
- ... **Dr. Jennifer Pratscher, Miriam Runkel** und **Dr. Aram Mikaelyan** und für das Gegenlesen dieser Arbeit.
- ...meiner Familie insbesondere meinen **Eltern, meiner Schwester** und meiner **Großmutter** für deren Unterstützung und Glauben an mich. Meiner **Partnerin** danke ich sehr für die große Unterstützung besonders in der letzten Phase dieser Arbeit.





---

[...]

**„Abseits der Straße  
beginnen die Entdeckungen“**  
— Hellmut Walters

**“Nothing in biology makes sense  
except in the light of evolution”**  
— Theodosius Dobzhansky

---

---

## Zusammenfassung

Termiten sind eine Ordnung sozialer Insekten, die sich vermutlich vor 150 Millionen Jahren aus subsozialen Schaben entwickelt hat. Die Fähigkeit Holz zu verwerten erhielten (niedere) Termiten durch die Aufnahme cellulolytischer Flagellaten – eukaryotische Einzeller, die den Großteil des Darmvolumens einnehmen und das wichtigste Mikrohabitat für prokaryotische Mikroorganismen sind. Der Verlust der Flagellaten in der jüngsten Termitenfamilie, der *Termitidae* oder auch höhere Termiten genannt, führte zu einer ausschließlich prokaryotischen Darmmikrobiota sowie zu beachtlicher Substrat-Diversifizierung und enormem ökologischen Erfolg. Während die Unterfamilie *Macrotermitinae* eine Symbiose mit holzabbauenden Pilzen der Gattung *Termitomyces* einging, konnten andere höhere Termiten Substrate mit höherem Humifizierungsgrad verwerten.

Vorhergehende Studien zeigten, dass sich die Termitendarmmikrobiota nah verwandter Termiten sehr ähnelt, aber zwischen entfernt verwandten Termiten deutliche Unterschiede zu finden sind. Da in der Literatur allerdings nur äußerst wenige Termiten auf ihre Darmmikrobiota hin untersucht wurden, ist nicht klar ob diese Unterschiede bestimmten Mustern folgen. Daher enthält diese Dissertation Studien, die die archaeelle und bakterielle Diversität der Darmmikrobiota über eine große Bandbreite von Termiten mittels Hochdurchsatzsequenzierung der 16S-rRNA-Gene untersuchen. Während die archaeelle Mikrobiota hauptsächlich aus methanogenen Organismen besteht, zeigt die bakterielle Darmmikrobiota eine deutlich höhere Diversität. Die Phyla *Bacteroidetes*, *Firmicutes*, *Proteobacteria* und *Spirochaetes* kommen in den Därmen fast aller Termiten vor, allerdings in variabler Abundanz. Dagegen treten andere Phyla wie *Elusimicrobia*, *Fibrobacteres* und die *Candidatus Division* TG3 nur in bestimmten Termitengruppen auf. Interessanterweise ändert sich die Abundanz verschiedener Archaeen und Bakterien zwischen unterschiedlichen Termiten substanziell und spiegelt dabei sowohl die Wirtsphylogenie als auch Änderungen der Ernährungsstrategie des Wirts wider. Phylogenetische Analysen archaeeller und bakterieller Organismen ergaben dagegen, dass diese zwar aus wirtsspezifischen Gruppen bestehen, aber nicht mit ihren Wirten kospezieren. Erkenntnisse aus Studien in dieser Dissertation und anderer publizierter Studien wurden innerhalb des Diskussionskapitels erneut ausgewertet, um die Ursachen der Unterschiede in den mikrobiellen Darmgemeinschaften zu identifizieren und diskutieren. Als Ergebnis werden als treibende „Kräfte“ der Struktur der Darmmikrobiota von Termiten Habitat- und Nischenselektion vorgeschlagen. Allerdings wird der Einfluss des stochastischen Elements dieser Mechanismen durch proktodeale Trophallaxis stark reduziert; was einer potenziellen Koevolution stark Vorschub leistet, die ultimativ in Kospeziation münden könnte. Während Koevolution sich wahrscheinlich in Form der zahlreichen wirtsspezifischen mikrobiellen Gruppen darlegt, sind Belege für Kospeziation innerhalb der Darmmikrobiota von Termiten äußerst selten und betreffen fast ausschließlich Flagellaten und ihre Symbionten in niederen Termiten.

In höheren holzverzehrenden Termiten kommen Bakterien des Phylums *Fibrobacteres* und der *Cand. Div. TG3* besonders häufig vor. Kürzlich wurde gezeigt, dass Organismen dieser Gruppen mit Holzfasern in höheren Termiten assoziiert sind. Allerdings blieb ihr funktionelles Potenzial weiterhin nahezu unbekannt. Eine Studie innerhalb dieser Arbeit beschäftigt sich mit Metagenomen der Darmmikrobiota verschiedener höherer Termiten. In holzfressenden Vertretern konnte zwar eine hohe Abundanz von *Fibrobacteres* und *Cand. Div. TG3* festgestellt werden, aber es konnten nur wenige oder gar keine Gene über die üblichen datenbankabhängigen Klassifikationsprogramme zugeordnet werden – es wurde daher versucht diese Diskrepanz auszugleichen. In einer Anschlussstudie ist es mittels einer neuartigen Methode gelungen über 30 Populations-Genome von *Fibrobacteres* und *Cand. Div. TG3* aus den Metagenom-Datensätzen zu rekonstruieren. Eine anschließende komparative Analyse belegt, dass Organismen beider Gruppen zwar ein unterschiedlich hohes Potenzial zum Abbau von Holz besitzen, sich aber wahrscheinlich gegenseitig im Abbau ergänzen. Weitergehende Analysen deuten zudem an, dass Vertreter beider Gruppen das Potenzial zur Stickstofffixierung und Respiration unter hypoxischen Bedingungen besitzen und damit gut an das Termitendarmökosystem angepasst sind.

---

---

## Abstract

Presumably descending from subsocial cockroaches 150 million years ago, termites are an order of social insects that gained the ability to digest wood through the acquisition of cellulolytic flagellates. These eukaryotic protists fill up the bulk of the hindgut volume and are the major habitat of the prokaryotic community present in the digestive tract of lower termites. The complete loss of gut flagellates in the youngest termite family *Termitidae*, also called higher termites, led to an entirely prokaryotic gut microbiota as well as a substantial dietary diversification and enormous ecological success. While the subfamily *Macrotermitinae* established a symbiosis with wood-degrading fungi of the genus *Termitomyces*, other higher termites exploit diets with a higher degree of humification.

Previous studies on the gut communities of termites have observed that while the gut microbiota of closely related hosts is very similar, those of more distantly related hosts are characterized by considerable differences in gut communities. Since these observations are based on highly limited samplings of hosts, it is uncertain if these differences reflect important evolutionary patterns. This dissertation includes studies examining the archaeal and bacterial diversity of the gut microbiota over a wide range of termites using high-throughput sequencing of the 16S rRNA genes. In comparison to the rather simple archaeal communities, which were mainly composed of methanogens, the bacterial gut microbiota were characterized by considerably higher diversity. At the phylum-level, *Bacteroidetes*, *Firmicutes*, *Proteobacteria* and *Spirochaetes* were ubiquitously distributed among the termites, albeit with differences in relative abundance. Other phyla, however, such as *Elusimicrobia*, *Fibrobacteres* and the *candidate division* TG3, occurred only in certain host groups of termites. The distribution pattern of archaeal and bacterial lineages reflects both host phylogeny and differences in the digestive strategy of the host. Although several genus-level bacterial lineages showed a certain degree of host-specificity, phylogenetic analyses of the amplified rRNA genes showed that these bacterial lineages do not appear to be cospeciating with their hosts. The findings of studies included in this dissertation and other published studies were evaluated to identify potential drivers of community structure and other shaping mechanisms. Thus, gut community structure in termites is primarily shaped by habitat and niche selection. The stochastic element of these mechanisms, however, is strongly attenuated by proctodeal trophallaxis, which facilitates coevolution and might ultimately lead to cospeciation. While coevolution is likely true for many lineages and documented by host-specific microbial lineages, there is only little evidence of cospeciation in the gut microbiota of termites. If present, it is restricted almost exclusively to flagellates and their symbionts in lower termites.

The higher wood-feeding termites have long been associated with a marked abundance of the phyla *Fibrobacteres* and *cand. div.* TG3. Although these phyla have been shown to be members of a specific cellulolytic community associated with wood particles in the hindguts of higher termites, their full functional potential still remains unknown. In order to elucidate the role of these organisms, a study in this dissertation carries out metagenomic analyses of various higher termites. In wood-feeding representatives, *Fibrobacteres* and *cand. div.* TG3 were in fact highly abundant, but only a few or no genes could be assigned to both groups by the usual database-dependent classification programs due to the lack of suitable genomes in these databases. In response, a new study was conceived to compensate this discrepancy. By further development of a new reference-independent method, over 30 population genomes of *Fibrobacteres* and *cand. div.* TG3 could be reconstructed from the metagenomic data sets. Subsequent comparative analysis revealed that organisms of both groups differ in their potential of wood degradation, but likely complement each other. Further analyses indicate that representatives of both groups might be able to fix nitrogen and respire under hypoxic conditions — two favourable adaptations to the unique termite gut environment.



---

## Table of contents

	Acknowledgements	VII
	Zusammenfassung	XI
	Abstract	XIII
<b>Chapters</b>		
<b>1</b>	General Introduction	1
<b>2</b>	High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation in the gut of wood-feeding higher termites ( <i>Nasutitermes</i> spp.)	15
<b>3</b>	The cockroach origin of the termite gut microbiota: patterns in bacterial community structure reflect major evolutionary events	39
<b>4</b>	Dietary and phylogenetic determinants of bacterial gut community structure in higher termites	65
<b>5</b>	Structure, spatial distribution, and succession of the bacterial community in the hindgut and fungus comb of the termite <i>Odontotermes formosanus</i> ( <i>Macrotermitinae</i> )	77
<b>6</b>	Habitat selection and vertical inheritance drive archaeal community structure in arthropod guts	99
<b>7</b>	Comparative metagenomic analyses reveal bacterial functional profiles in the hindgut compartments of higher termites	139
<b>8</b>	Genome sequences of <i>Fibrobacteres</i> and TG3 obtained by marker-assisted compositional binning of multiple metagenomes	163
<b>9</b>	General discussion	193
<b>Appendix</b>		
<b>A</b>	The complete mitogenomes of six higher termite species reconstructed from metagenomic datasets	211
<b>B</b>	Population structure of Endomicrobia in single host cells of termite gut flagellates ( <i>Trichonympha</i> spp.)	215
	Curriculum vitae	233
	Contributions	235
	Erklärung der Eigenständigkeit	237





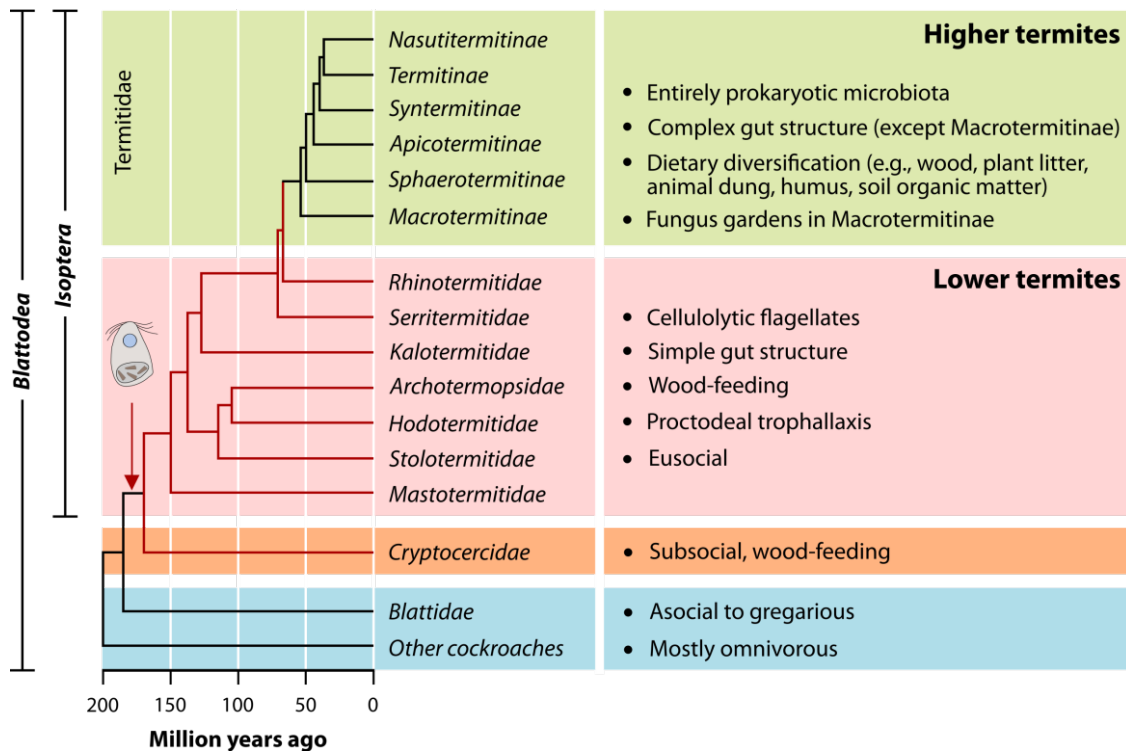
# Chapter 1

## General introduction

Carsten Dietrich and Andreas Brune

*Affiliations:* Max Planck Institute for Terrestrial Microbiology, 35043 Marburg, Germany | *This chapter is based on the manuscript:* Brune A and Dietrich C (2015). The termite gut microbiota: Digesting the diversity in the light of ecology and evolution. *Ann. Rev. Microbiol.* 69 | *Contributions:* C.D. contributed to the manuscript and designed the figures. A.B. contributed to the manuscript.

## Evolution of symbiotic digestion in termites



**Fig. 1 | Phylogeny of termites (*Isoptera*), illustrating their origin within the radiation of cockroaches (together forming the *Blattodea*), with the omnivorous *Blattidae* as sister group, and important events in the evolution of the digestive symbiosis.** Red branches indicate the presence of cellulolytic flagellates that were acquired by a common ancestor of lower termites and *Cryptocercidae*, which gave rise to their wood-feeding lifestyle. The loss of flagellates in the higher termites gave rise to an enormous dietary diversification (chronogram redrawn from Bourguignon et al., 2015).

It is now generally accepted that termites originated within the radiation of cockroaches from a presumably wood-feeding, subsocial ancestor (Lo and Eggleton, 2011; Nalepa, 2011). The most recent molecular phylogenies revealed that the split between the termite line and their sister group, the blattid cockroaches, occurred already during the Jurassic (at least 150–170 mya) (Fig. 1). A key event in the evolution of termites was the acquisition of cellulolytic flagellates by a common ancestor of termites (*Isoptera*) and their sister group, the *Cryptocercidae* (hereafter included in all statements concerning lower termites), which must have provided a strong boost of their capacity for lignocellulose digestion. The symbiosis between termites and flagellates was stabilized by the development of proctodeal trophallaxis; this trait ensures the reliable transfer of flagellates among nestmates and across generations and is part of their complex social behavior, which started with long-lasting biparental care in the subsocial *Cryptocercidae* and culminated in an elaborate caste system and the sharing of labor in termites (see Klass et al., 2008; Lo and Eggleton, 2011). Together with the cellulolytic activities of the host, which are present in the entire blattodean lineage, the flagellates form the dual cellulolytic system of lower termites (Brune and Ohkuma, 2011; Watanabe and Tokuda, 2010) – a by far more efficient means for the symbiotic digestion of lignocellulose than that of other detritivorous and xylophagous cockroaches (Klass et al., 2008).

It is therefore surprising that the youngest of all termite families, the *Termitidae*, which arose about 50 mya (Bourguignon et al., 2015), is no longer associated with cellulolytic protists (Ohkuma and Brune, 2011). The loss of flagellates in the higher termites resulted in an entirely prokaryotic gut microbiota and was accompanied by numerous symbiotic innovations and an enormous dietary diversification, which in turn brought about an enormous ecological and evolutionary success (Bignell and Eggleton, 2000).

## The termite gut habitat

The intestinal tracts of termites are small ecosystems with a wide range of microhabitats that strongly differ in their abiotic and biotic environment. Many of the environmental features are intrinsic properties of the gut, whereas others result from physiological activities of the host or the microbial residents in the respective location. In addition, the types of habitats available for microbial colonization have changed during the evolutionary history of the host.

**Gut structure.** Termites share the basic gut structure of cockroaches (Fig. 2). The foregut transports the food from the mouthparts into a spacious crop, where it is incubated with the secretions of the salivary glands. After further comminution by the gizzard, food passes into the midgut, where it is digested by enzymes secreted by the midgut. The digestion products are resorbed by the midgut epithelium and gastric caeca. The remaining material is transported into the hindgut, (proctodeum), which consists of a short ileum (P1), followed by enteric valve (P2), colon (P3 and P4), and rectum (P5) and harbors the bulk of the gut microbiota. After the removal of water and ions, the residues of digestion are released as feces (Chapman et al., 2013).

The crop of termites is much smaller than that of cockroaches, the midgut is shortened, and the caeca are reduced or entirely absent. In all lower termites, the anterior colon (P3), which may be somewhat expanded already in cockroaches, is strongly dilated into a single, voluminous paunch (Noirot, 1995). In higher termites (with the exception of *Macrotermitinae*), the hindgut is further elongated and differentiated into several proctodeal compartments (Noirot, 2001).

Cockroaches and termites harbor bacteria in all gut regions. Although foregut and midgut are loaded with digestive enzymes (proteases, lysozyme, chitinases) secreted by the salivary glands and epithelia and the passage of food through the midgut is relatively rapid (Eggleton, 2011), there seem to be sites of microbial fermentations, as indicated by the accumulation of lactate and short-chain fatty acids (e.g. Köhler et al., 2012; Schauer et al., 2012; Bauer et al., 2015). The hindgut is the major site of microbial colonization, which is reflected also in the high concentrations of short-chain fatty acids in most hindgut compartments (e.g., Köhler et al., 2012; Schauer et al., 2012). An exception is the alkaline P1 compartment of wood- and soil-feeding higher termites, which is populated only in relatively low numbers (Schmitt-Wagner et al., 2003; Köhler et al., 2012).

**Microhabitats.** The rapid passage of the digesta requires that to prevent washout, a microorganism must either swim fast enough or associate with particles that are retained in the gut longer than the liquid fraction. The gut flagellates of lower termites either are highly motile and able to actively maintain their position in the gut, or possess organelles for attachment to the cuticle of the gut wall (e.g., Radek et al., 2014). Since flagellates make up the bulk of the hindgut volume, it is not surprising that the majority of prokaryotes in the hindgut of lower termites colonize the surface, cytoplasm, and even nucleus of these protists (Brune, 2014). The transfer of gut contents among nestmates extends the life span of the flagellate habitats beyond that of an individual termite.

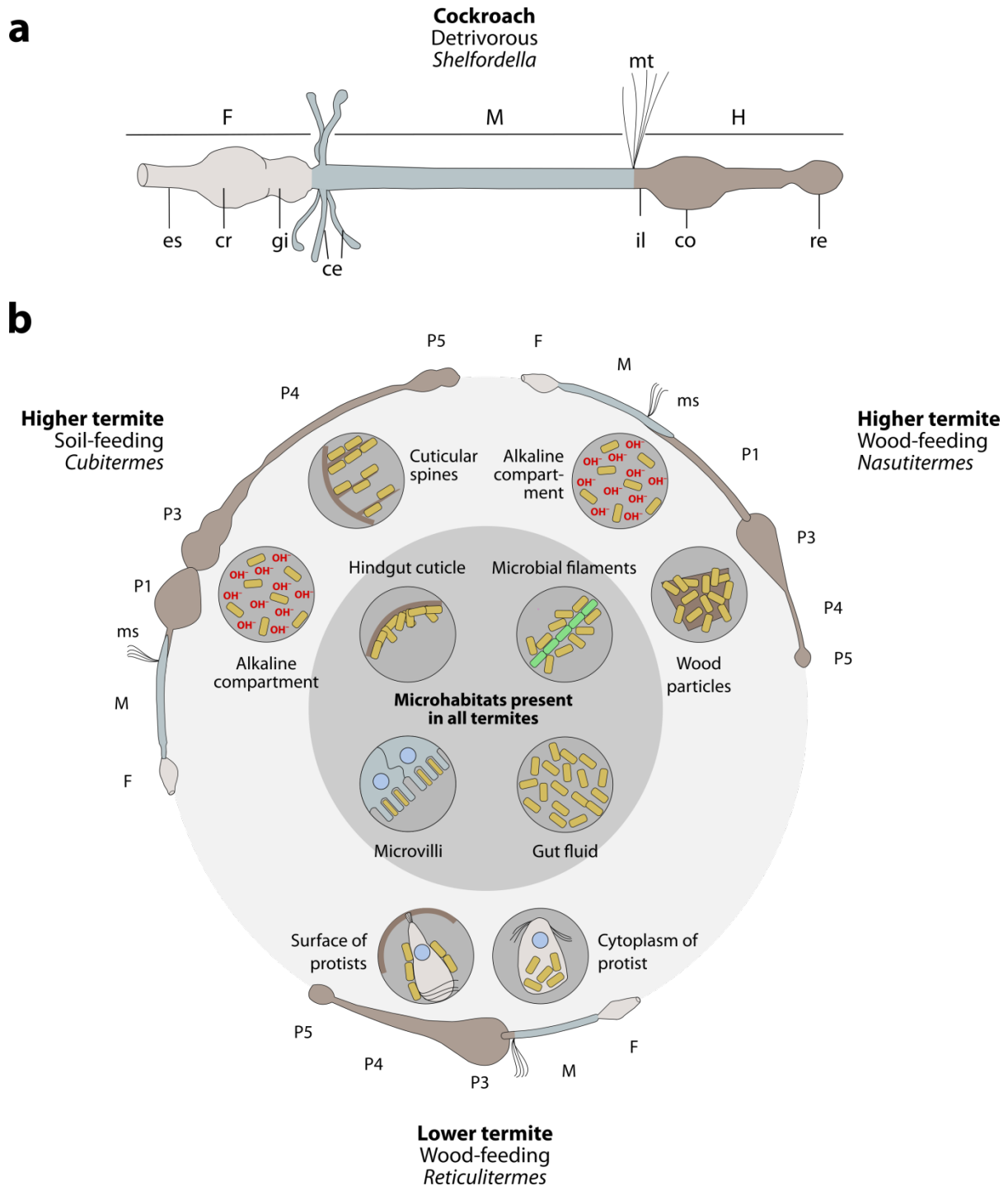
The lumen itself is not a favorable microhabitat for bacterial cells, except for the large spirochetal forms, which swim fast enough to actively maintain their position. In wood-feeding higher termites, the retention time of wood particles is longer than that of the gut fluid. Similar mechanisms seem to exist in soil feeders, where small clay particles (rich in organic matter) are retained longer than large sand grains (Eggleton, 2011). It has been estimated (based on DNA content) that almost one-third of the microbial biomass in the luminal fluid of the hindgut paunch (P3) of a wood-feeding *Nasutitermes* species is firmly associated with the fiber fraction (Mikaelyan et al., 2014).

In insect guts, only the midgut is endodermal and allows direct access of microorganisms to the epithelial surfaces, which are protected by the peritrophic membrane. Associations of bacterial cells with the microvilli of the epithelial brush border or the ectoperitrophic space of the mixed segment have been observed (Breznak and Pankratz, 1977; Tokuda et al., 2001). Foregut and hindgut are of ectodermal origin and are always lined with a cuticle. However, in the hindgut region, the cuticle has characteristic pores or pits (Breznak and Pankratz, 1977; Czolij et al., 1985), which may increase the permeability for acetate and other short-chain fatty acids. The cuticle offers plenty of surfaces and is usually covered by a dense microbial biofilm. Cuticular spines in the P4 compartment of certain higher termites provide additional attachment sites for the microbiota (Bignell et al., 1980). However, during ecdysis insects replace their entire cuticle, and the hindgut has to be recolonized after each molt.

**Environmental factors.** The physicochemical conditions in the different gut compartments are affected by both the biotic and abiotic environment. Although oxygen continuously enters the gut via the host epithelia, its efficient removal by the gut microbiota renders the center of all dilated hindgut compartments anoxic. It is important to realize that due to the small size of termite guts, diffusive transport of metabolites along their steep radial concentration gradients is much more important than axial transport and convective mixing by peristalsis and also not affected by the activity of the flagellates (see Brune and Ohkuma, 2011).

Compared to larger guts of most vertebrates, the surface-to-volume ratio of insect guts is enormous (Brune, 1998), which increases the relative importance of aerobic processes but also facilitates the exchange of both gaseous and dissolved products of microbial metabolism at the epithelial surfaces. The redox potential of the different microhabitats is modulated by their oxygen status, the production of redox-active compounds like hydrogen or ferrous iron (in soil-feeding termites), or differences in intestinal pH (see Brune and Ohkuma, 2011).

Also host secretions should have a strong effect on the microbiota in different gut compartments. In the anterior gut, digestive enzymes in saliva and in midgut secretions provide sugars or amino acids as substrates for the resident microbiota, but they also digest microbial biomass (Fujita, 2004, Watanabe and Tokuda, 2010). Passage through the anterior gut may represent a barrier to colonization by foreign microorganisms or pathogens, and the mechanisms by which hindgut microbiota transferred to nestmates via proctodeal trophallaxis evades digestion (and in the case of flagellates, mechanical disruption by the gizzard) remains unknown. Also the extreme alkalinity in the anterior hindgut of soil-feeding termites should affect the viability of transient microbiota and select for lineages adapted to this habitat.



**Fig. 2 | Schematic anatomy of termite guts and important microbial habitats in different host groups.** (a) The basic plan of the undifferentiated cockroach gut. Abbreviations: F, foregut; es, esophagus; cr, crop; gi, gizzard; M, midgut; ce, caeca; mt, Malpighian tubules; H, hindgut; il, ileum; co, colon; re, rectum. (b) Termite guts are derived from the same basic plan, but while crop and midgut are reduced relative to those of cockroaches, the hindgut (P, proctodeum) is increasingly elongated and may be differentiated into a mixed segment (ms) and several proctodeal compartments (P1–P5), which provide additional microhabitats for microbial colonization (nomenclature after Noirot, 2001).

## Functional niches

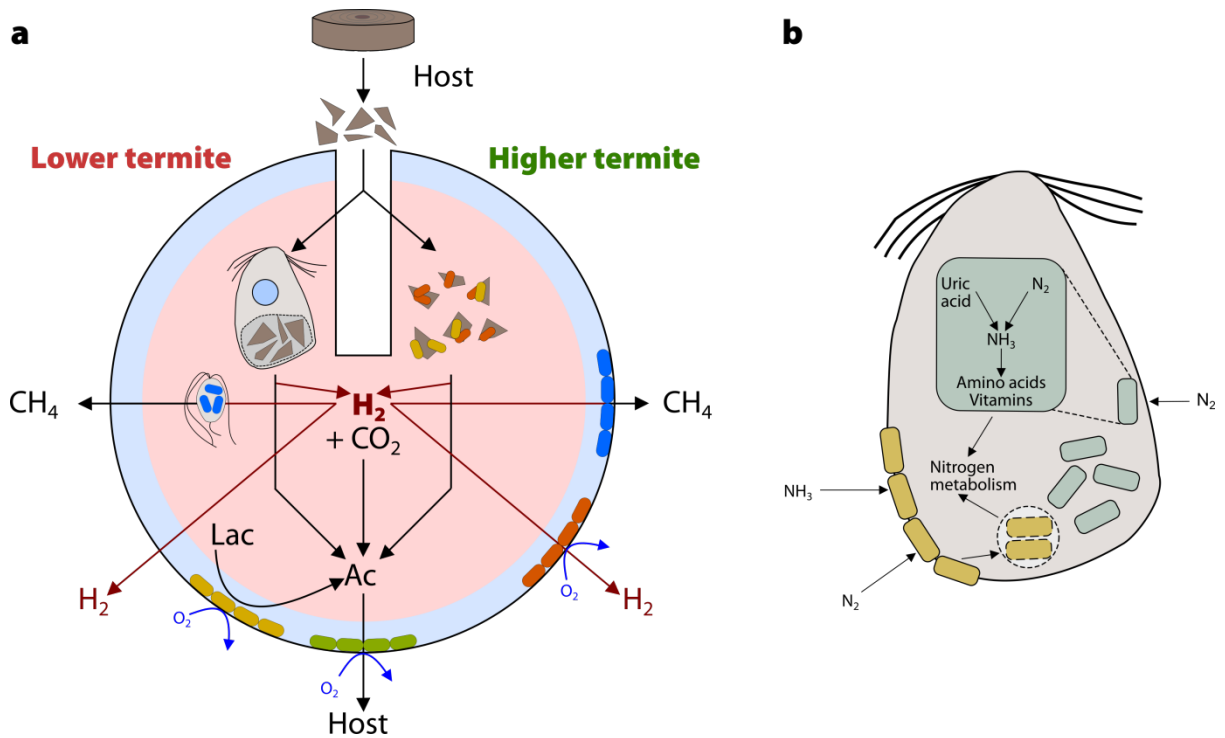
The major products of symbiotic digestion of lignocellulose in the termite hindgut are short-chain fatty acids and microbial biomass. While the fermentation products are resorbed by the hindgut epithelium and drive the energy metabolism of the termite, the microbial biomass has an important nutritive value for the host (Brune, 2014). The functional roles of individual microbial populations are not always clear, but after decades of research, it is possible to sketch major niches of the hindgut ecosystem that are relevant for termites of all feeding guilds (Fig. 3a).

**Polymer degradation.** The primary function of the hindgut microbiota is the depolymerization of recalcitrant plant fiber (Hongoh, 2011; Brune, 2014). In the hindgut of lower termites, the flagellates produce a broad suite of glycoside hydrolases for the efficient breakdown of phagocytized wood, including various cellulases (exoglucanases and endoglucanases) and diverse hemicellulases (e.g., xylanases, arabinosidases, mannosidases) (see Ni and Tokuda, 2013). Recently, the protist community has been identified also as the major source of chitinase activity in the hindgut of *Zootermopsis angusticollis* (Rosengaus et al., 2014). It is important to note that hydrolysis takes place in the digestive vacuoles of flagellates, which not only sequesters the wood particles from the hindgut fluid but should also prevent access of luminal bacteria to the sugars produced by the depolymerization process.

This scenario differs in higher termites, where the absence of flagellates requires new modes of fiber digestion. The oldest strategy is that of the *Macrotermitinae*, which digest wood or lignocellulosic plant litter with the help of a lignin-degrading basidiomycete fungus (*Termitomyces* spp.). This fungus is not part of the gut microbiota but is cultivated in fungus gardens (combs) in the nests (Nobre et al., 2011; Poulsen, 2015). The termites harvest older combs, which consist of partially digested lignocellulose and fungal biomass. The composition of the ingested material varies among the genera of fungus-cultivating termites and may be the reason for differences in the composition of their bacterial microbiota (Poulsen, 2015). Fiber digestion in the fungus comb is incomplete and continues in the gut, as indicated by metagenomic analysis of several species of fungus-cultivating termites (Liu et al., 2013; Poulsen et al., 2014). The gut bacteria, specifically members of the *Bacteroidetes*, encode large numbers of glycosyl hydrolases implicated in the breakdown of polysaccharides of plant and fungal cell walls.

The other subfamilies of higher termites evolved strategies to unlock also partially humified lignocellulose as a new dietary resource. During humification, which may start with the dung of herbivorous mammals or decaying wood and plant litter, there is a continuous decrease in cellulose content and a relative increase in the residual complex polysaccharides and nitrogenous products derived from microbial biomass. The community structure and abundance of different glycoside hydrolase families clearly differ between wood- and dung-feeding termites (He et al., 2013). Peptides derived from microbial biomass are an important component of soil organic matter and probably form a major dietary resource of true soil feeders (see Brune and Ohkuma, 2011), which may explain the high abundance of *Firmicutes* (Schmitt-Wagner et al., 2003).

Several lineages of higher termites apparently returned to a wood-feeding lifestyle (Bourguignon et al., 2015). In the absence of flagellates, the wood particles entering the hindgut are available for bacterial colonization. A recent study of *Nasutitermes* spp. linked the cellulolytic activities in the fiber fraction to fiber-associated members of *Fibrobacteres*, *Spirochaetes*, and the *candidate division* TG3 (Mikaelyan et al., 2014), which explains the abundance of cellulase genes in hindgut metagenomes that were assigned to *Fibrobacteres* and *Spirochaetes* (He et al., 2013, Warnecke et al., 2007).



**Fig. 3 | Major functional niches of the hindgut ecosystem.** (a) The fermentative breakdown of wood polysaccharides to acetate and  $\text{CO}_2$  differs between lower and higher termites (flagellates vs. fiber-associated bacteria). Hydrogen is a major intermediate, giving rise to methanogenesis and reductive acetogenesis. Microorganisms located in the microoxic gut periphery catalyze oxygen-dependent processes. Arrows indicate fluxes of carbon (black) and hydrogen (red). (b) Contribution of symbionts to the nitrogen metabolism of their flagellate host. Ectosymbionts fix nitrogen and/or assimilate ammonia; they are phagocytized and subsequently digested. Endosymbionts fix nitrogen and recycle uric acid (after Hongoh, 2011; Brune, 2014).

**Hydrogen metabolism.** Hydrogen is a central intermediate in the fermentative processes (Fig. 3). In lower termites, hydrogen is released during the oxidation of polysaccharides to acetate and  $\text{CO}_2$  by the cellulolytic flagellates (Pester and Brune, 2007). Hydrogen accumulates strongly also in higher termites (Köhler et al., 2012; Li et al., 2012), but the identity of the primary fermenters is not clear. Even in cases where hydrogen does not accumulate strongly (Pester and Brune, 2007; Desai and Brune, 2012), this seems to be due to a close coupling of  $\text{H}_2$  production and consumption rather than production of other reduced fermentation products (see Brune, 2014).

In wood-feeding termites, most of the hydrogen produced in the gut is converted to additional acetate via reductive acetogenesis from  $\text{CO}_2$  (see Brune and Ohkuma, 2011). The capacity of several isolates for reductive acetogenesis (Breznak and Leadbetter, 2006) and inventories of functional genes (*fhs*, *coo*) involved in the Wood-Ljungdahl pathway (e.g., Matson et al., 2011; Ottesen and Leadbetter, 2011; Rosenthal et al., 2013) indicate that both lower and higher termite species possess diverse populations of homoacetogenic spirochetes. However, the particular role of the diverse FeFe hydrogenases (*hyd*) present in the gut microbiota (e.g., Ballor and Leadbetter, 2012; Zheng et al., 2013) in the production and consumption of hydrogen is not entirely clear. Even the closely related isolates from *Treponema* cluster Ia may differ in their capacity for reductive acetogenesis (Breznak and

Leadbetter, 2006; Dröge et al., 2008), underscoring that not all of the diverse lineages of termite gut treponemes are necessarily homoacetogenic (Mikaelyan et al., 2015).

Also hydrogenotrophic methanogenesis is a characteristic process of the termite gut microbiota, but it is typically of minor importance in wood-feeding species, most likely because of the hydrogen limitation of methanogens owing to their locations in the hindgut periphery (see Brune, 2010). The reasons for the much stronger methane emissions of fungus-cultivating and soil-feeding termites are not entirely clear, but may be rooted either in the structure of the methanogenic communities, which includes large populations of lineages with obligately methylophilic representatives (Paul et al., 2012), or in their spatial organization, which allows the transfer of reducing equivalents between gut compartments (see Brune, 2010). In lower termites with unusually high methane emission rates, the methanogens are associated with flagellates located in the hydrogen-rich gut center (see Hongoh and Ohkuma, 2010).

**Oxygen consumption.** Due to the small size of termite guts, the influx of oxygen across the hindgut wall is enormous (Brune, 1998). Therefore, the isolation of facultative and strictly aerobic bacteria from termite guts is not surprising. Particularly the acetate-oxidizing microaerophiles colonizing the hindgut wall appear to be well adapted to the hypoxic conditions in the periphery of the hindgut (Wertz and Breznak, 2007; Wertz et al., 2012; Isanapong et al., 2013). Oxygen is used as electron sink also by fermenting bacteria, as indicated by the shift from propionate to acetate during the metabolism of lactate (see Brune and Ohkuma, 2011). Even strict anaerobes, such as methanogens colonizing the hindgut wall, can remove oxygen from their environment using H<sub>2</sub> as reductant (see Brune, 2010). The utilization of hydrogen for the removal of O<sub>2</sub> increases the yield of acetate and other useful fermentation products over methane and hydrogen, which are worthless for the termite. Oxygen is also an essential cosubstrate for the oxidative metabolism of aromatic compounds; genome sequences revealed cryptic capacities for oxygenase activities in anaerobic termite gut spirochetes (Lucey and Leadbetter, 2014).

**Nitrogen metabolism.** The low nitrogen content of lignocellulose is a serious constraint to the growth of wood-feeding termites. The hindgut microbiota plays an important role in the fixation, recycling, and upgrading of nitrogen (see Hongoh, 2011; Brune, 2014). While the microbial fermentation products formed in the hindgut are directly absorbed by the host epithelia, the biomass can be accessed only by the digestive enzymes in the midgut after proctodeal trophallaxis. Even in dung-feeding and humivorous species, where the mineralization of nitrogenous food constituents leads to a net formation of ammonia in the digestive process (Ngugi and Brune, 2012), the need of the insect for essential amino acids and vitamins can be met only the digestion of microbial biomass in the midgut.

The diversity of *nifH* genes, a functional marker for nitrogen fixation, indicates a broad diazotrophic potential in the termite gut communities (e.g., Yamada et al., 2007; Inoue et al., 2015). In lower termites, symbionts of flagellate seem to play an important role in the fixation and/or upgrading of nitrogen (Hongoh et al., 2008a; Hongoh et al., 2008b; Desai and Brune, 2012) (Fig. 3b). The identity of the microorganisms responsible for the high rates of nitrogen fixation in wood-feeding higher termites is unclear (Yamada et al., 2007).



## The prokaryotic gut microbiota of termites

Taxonomic inference of termite gut prokaryotes by morphology is very limited or even impossible. Early cultivation-based studies of termite gut microbiota resulted in several isolates. Many of these organisms represent lineages that are low in abundance in their native environment, and are likely responsible for unimportant ecological roles. Only culture-independent studies based on the 16S rRNA gene can access the true prokaryotic diversity of the termite gut microbiota, which consists of many hitherto uncultivated phylotypes. While initial molecular studies were based on a small number of sequences (e.g. Berchtold et al., 1994; Ohkuma and Kudo, 1996), analyses within the last 10 years got 1,000 long sequences and more (Hongoh et al., 2003). Recent studies employ high-throughput sequencing such as MiSeq sequencing. These new methods allow a high degree of multiplexing and provide millions of sequences per run (Caporaso et al. 2012; Loman, et al., 2012).

**Bacteria.** The bacterial gut microbiota of termites varies substantially among major termite host groups (see Brune, 2014). The microbiota of the lower termite *Coptotermes formosanus* (Shinzato et al., 2005) and the fungus-feeding higher termite *Macrotermes gilvus* (Hongoh et al., 2006b) are dominated by the phyla *Bacteroidetes* and *Firmicutes* (in the mentioned order), whereas other wood-feeding higher and lower termites are dominated by bacteria of the phylum *Spirochaetes* (e.g. *Reticulitermes speratus* from Hongoh et al. 2005; *Nasutitermes takasagoensis* from Hongoh et al. 2006a). Lower termites also harbor bacteria of the phylum *Elusimicrobia*, which comprise flagellate symbionts known to participate in the nitrogen economy of their flagellate host (Stingl et al., 2005; Hongoh et al., 2008b). In wood-feeding higher termites, *Fibrobacteres* and *candidate division TG3* are suddenly present in high abundance, whereas in other termites these lineages are absent or very low in abundance. Bacteria of these two divisions were shown to reside on the wood fibers in these termites (Mikaelyan et al., 2014). Their metabolism, however, remains unknown.

**Archaea.** Archaeal lineages detected in termite guts can be classified to either to the phylum *Thaumarchaeota*, which is only detected in higher termites so far (Friedrich et al., 2001), or to four orders of methanogens, i.e. *Methanobacteriales*, *Methanomassiliicoccales*, *Methanomicrobiales*, and *Methanosarcinales* within the *Euryarchaeota* phylum. The distribution of these lineages differs among termite hosts (Brune, 2010). The predominant archaea in lower termites are members of the *Methanobacteriales* (e.g. Ohkuma and Kudo, 1998; Shinzato et al., 1999). In contrast, the archaeal community in higher termites is more variable, and most abundant methanogenic orders differ among host subfamilies (Brune, 2010): *Methanobacteriales* (*Termitinae*), *Methanosarcinales* (*Macrotermitinae*), or *Methanomicrobiales* (*Nasutitermitinae*). Although findings of past studies indicate certain trends in the archaeal structure, no large scale comparative studies have been carried out, and the determinants of archaeal community structure in termite guts still remain unknown.

**Monophyletic lineages of termite bacteria.** Phylogenetic analysis of the small-subunit rRNA genes from the termite guts suggest that most bacterial lineages from termite guts fall into monophyletic clusters that are exclusively comprised of sequences from termite guts (Hongoh et al., 2005; Mikaelyan et al., 2015) — leading to the hypothesis that the majority of the termite gut microbiota coevolves with their termite host (Hongoh et al., 2005). The ultimate form of coevolution, cospeciation, seems to be restricted exclusively to flagellates and their symbionts in lower termites (e.g. Noda et al., 2007; Ikeda-Ohtsubo and Brune, 2009)

Coevolution implies reciprocal adaptations of partners in a symbiosis. However, detailed information on the adaptation of termite gut bacteria is often missing. It is likely that termite gut lineages represent new species, genera or even higher taxonomic ranks that might have acquired functions from distant relatives by horizontal gene transfer. Such exchanges of genetic material have been known to occur frequently in dense microbial communities (Cordero and Polz, 2014) such as the termite gut.

## Aims of this thesis

Termite guts harbor a dense and diverse microbiota that is essential for symbiotic digestion. The major players in lower termites are unique lineages of cellulolytic flagellates, whereas higher termites harbor only bacteria and archaea. Functions and diversity of these uncultivated microorganisms are poorly understood. Hence, studies included in this thesis cover the following areas:

**Diversity.** Previous studies showed that the gut microbiota of closely related termites is very similar, but highly different between distantly related species. However, these results are only based on a few termite hosts and it is not known whether it is possible to extrapolate these findings to the entire termite host range. Therefore, studies included in this thesis aim to describe the bacterial and archaeal diversity by high-throughput sequencing of 16S rRNA genes in the guts of many termites to reveal patterns in community structure between host groups, different dietary groups, gut compartments and microhabitats.

**Drivers of community structure.** Published studies and studies included in this thesis conclude that fundamental changes in the course of termite evolution seem to have caused major shifts in the gut microbiota of termites. However, the patterns in microbial community structure and the underlying ecological drivers are poorly understood. The aims of the studies in this thesis are the identification and the discussion of potential drivers of the termite gut microbiota.

**Functional potential of termite-specific groups.** In wood-feeding higher termites, members of *Fibrobacteres* and the *candidate division* TG3 are reported to be attached to wood-fibers. However, nothing is known about their metabolic potential. Chapter 7 of this thesis analyzes the metagenomes of the microbiota in guts of different higher termites. The data sets of wood-feeding representatives indicate that *Fibrobacteres* and *cand. div.* TG3 can be detected in high abundance based on 16S rRNA genes, but little or none of the other genes are assigned to these lineages by conventional reference-dependent methods. Therefore chapter 8 aims to reconstruct population genomes of the termite-specific lineages of *Fibrobacteres* and *cand. div.* TG3 from metagenome data sets by a new reference-independent method to detect potential adaptations to the termite gut environment.

## References

- Ballor NR and Leadbetter JR (2012).** Patterns of [FeFe] hydrogenase diversity in the gut communities of lignocellulose-feeding higher termites. *Appl. Environ. Microbiol.* 78:5368–5374.
- Bauer E, Lampert N, Mikaelyan A, Köhler T, Maekawa K and Brune A (2015).** Physicochemical conditions, metabolites, and community structure of the bacterial microbiota in the gut of wood-feeding cockroaches (*Blaberidae: Panesthiinae*). *FEMS Microbiol. Ecol.* (in Press).
- Berchtold M, Ludwig W and Koenig H (1994).** 16S rDNA sequence and phylogenetic position of an uncultivated spirochete from the hindgut of the termite *Mastotermes darwiniensis* Froggatt. *FEMS Microbiol. Lett.* 123:269–273.
- Bignell DE and Eggleton P (2000).** Termites in ecosystems, pp. 363–387. *In* Abe T, Bignell DE, Higashi M. (eds.), *Termites: Evolution, Sociality, Symbioses, Ecology*. Kluwer Academic Publisher, Dordrecht.
- Bignell DE, Oskarsson H and Anderson JM (1980).** Distribution and abundance of bacteria in the gut of a soil-feeding termite *Procupitermes aburiensis* (*Termitidae, Termitinae*). *J. Gen. Microbiol.* 117:393–403.
- Bourguignon T, Lo N, Cameron SL, Šobotnik J, Hayashi Y, Shigenobu S, Watanabe D, Roisin Y, Miura T and Evans TA (2015).** The evolutionary history of termites as inferred from 66 mitochondrial genomes. *Mol. Biol. Evol.* 32(2):406–421.
- Breznak JA and Leadbetter JR (2006).** Termite gut spirochetes, pp. 318–329. *In* Dworkin M, Falkow S, Rosenberg E, Schleifer K-H and Stackebrandt E (eds.), *The Prokaryotes*, 3rd ed., Volume 7: *Proteobacteria: Delta and Epsilon Subclasses*. Deeply Rooting Bacteria, vol. 7. Springer, New York.
- Breznak JA and Pankratz HS (1977).** In situ morphology of the gut microbiota of wood-eating termites [*Reticulitermes flavipes* (Kollar) and *Coptotermes formosanus* Shiraki]. *Appl. Environ. Microbiol.* 33:406–426.
- Brune A (1998).** Termite guts: the world's smallest bioreactors. *Trends Biotechnol.* 16:16–21.
- Brune A (2010).** Methanogens in the digestive tract of termites, pp. 81–100. *In* Hackstein JHP (ed.), *(Endo)symbiotic Methanogenic Archaea*. Springer, Heidelberg.
- Brune A (2014).** Symbiotic digestion of lignocellulose in termite guts. *Nat. Rev. Microbiol.* 12:168–180.
- Brune A and Ohkuma M (2011).** Role of the termite gut microbiota in symbiotic digestion, pp. 439–475. *In* Bignell DE, Roisin Y and Lo N (eds.), *Biology of Termites: A Modern Synthesis*, Springer, Dordrecht.
- Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, Gormley N, Gilbert JA, Smith G and Knight R (2012).** Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J.* 6:1621–1624.
- Chapman RF, Simpson SJ, Douglas AE (2013).** *The insects: structure and function*, vol. 5, Cambridge University Press, Cambridge.
- Cordero OX and Polz MF (2014).** Explaining microbial genomic diversity in light of evolutionary ecology. *Nat. Rev. Microbiol.* 12:263–273.
- Czolij R, Slaytor M and O'Brien RW (1985).** Bacterial flora of the mixed segment and the hindgut of the higher termite *Nasutitermes exitiosus* Hill (*Termitidae, Nasutitermitinae*). *Appl. Environ. Microbiol.* 49:1226–1236.
- Desai MS and Brune A (2012).** *Bacteroidales* ectosymbionts of gut flagellates shape the nitrogen-fixing community in dry-wood termites. *ISME J.* 6:1302–1313.
- Dröge S, Rachel R, Radek R and König H (2008).** *Treponema isoptericolens* sp. nov., a novel spirochaete from the hindgut of the termite *Incisitermes tabogae*. *Int. J. Syst. Evol. Microbiol.* 58:1079–1083.
- Eggleton P (2011).** An introduction to termites: biology, taxonomy and functional morphology, pp. 1–26. *In* Bignell DE, Roisin Y, Lo N (eds.), *Biology of Termites: A Modern Synthesis*. Springer, Dordrecht.
- Friedrich MW, Schmitt-Wagner D, Lueders T and Brune A (2001).** Axial differences in community structure of *Crenarchaeota* and *Euryarchaeota* in the highly compartmentalized gut of the soil-feeding termite *Cubitermes orthognathus*. *Appl. Environ. Microbiol.* 67:4880–4890.
- Fujita A (2004).** Lysozymes in insects: what role do they play in nitrogen metabolism? *Physiol. Entomol.* 299:305–310.

- Hara K, Shinzato N, Seo M, Oshima T and Yamagishi A (2002). Phylogenetic analysis of symbiotic archaea living in the gut of xylophagous cockroaches. *Microb. Environ.* 17:185–190.
- He S, Ivanova N, Kirton E, Allgaier M, Bergin C, Scheffrahn RH, Kyripides NC, Warnecke F, Tringe SG, and Hugenholtz P (2013). Comparative metagenomic and metatranscriptomic analysis of hindgut paunch microbiota in wood- and dung-feeding higher termites. *PLoS ONE* 8:e61126.
- Hongoh Y and Ohkuma M (2010). Termite gut flagellates and their methanogenic and eubacterial symbionts, pp. 55–79. *In* Hackstein JHP (ed.), (Endo)symbiotic Methanogenic Archaea, Springer, Heidelberg.
- Hongoh Y and Ohkuma M and Kudo T (2003). Molecular analysis of bacterial microbiota in the gut of the termite *Reticulitermes speratus* (Isoptera: Rhinotermitidae). *FEMS Microbiol. Ecol.* 44:231–242.
- Hongoh Y, Deevong P, Hattori S, Inoue T, Noda S, Noparatnaraporn N, Kudo T and Ohkuma M (2006a). Phylogenetic diversity, localization and cell morphologies of the *candidate phylum* TG3 and a subphylum in the phylum *Fibrobacteres*, recently found bacterial groups dominant in termite guts. *Appl. Environ. Microbiol.* 72:6780–6788.
- Hongoh Y, Deevong P, Inoue T, Moriya S, Trakulnaleamsai S, Ohkuma M, Vongkaluang C, Noparatnaraporn N and Kudo T (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.
- Hongoh Y, Ekpornprasit L, Inoue T, Moriya S, Trakulnaleamsai S, Ohkuma M, Noparatnaraporn N and Kudo T (2006b). Intracolony variation of bacterial gut microbiota among castes and ages in the fungus-growing termite *Macrotermes gilvus*. *Mol. Ecol.* 15:505–516.
- Hongoh Y, Sharma VK, Prakash T, Noda S, Toh H, Taylor TD, Kudo T, Sakaki Y, Toyoda A, Hattori M and Ohkuma M (2008a). Genome of an endosymbiont coupling N<sub>2</sub> fixation to cellulolysis within protist cells in termite gut. *Science* 322:1108–1109.
- Hongoh Y, Sharma VK, Prakash T, Noda S, Taylor TD, Kudo T, Sakaki Y, Toyoda A, Hattori M and Ohkuma M (2008b). Complete genome of the uncultured Termite Group 1 bacteria in a single host protist cell. *Proc. Natl. Acad. Sci. USA* 105:5555–5560.
- Hongoh, Y (2011). Toward the functional analysis of uncultivable, symbiotic microorganisms in the termite gut. *Cell. Mol. Life Sci.* 68:1311–1325.
- Ikeda-Ohtsubo W und Brune A (2009). Cospeciation of termite gut flagellates and their bacterial endosymbionts: *Trichonympha* species and 'Candidatus Endomicrobium trichonymphae'. *Mol. Ecol.* 18:332–342.
- Inoue JI, Oshim K, Suda W, Sakamoto M, Iino T, Noda S, Hongoh Y, Hattori M and Ohkuma M (2015). Distribution and evolution of nitrogen fixation genes in the phylum *Bacteroidetes*. *Microb. Environ.* (in press) <http://dx.doi.org/10.1264/jsme2.ME14142>.
- Isanapong J, Sealy Hambright W, Willis AG, Boonmee A, Callister SJ, Burnum KE, Pasa-Tolic L, Nicora CD, Wertz JT, Schmidt TM and Rodrigues JL (2013). Development of an ecophysiological model for *Diplosphaera colotermitum* TAV2, a termite hindgut Verrucomicrobium. *ISME J.* 7:1803–1813.
- Klass K-D, Nalepa CA and Lo N (2008). Wood-feeding cockroaches as models for termite evolution (*Insecta: Dictyoptera*): *Cryptocercus* vs. *Parasphaeria boleiriana*. *Mol. Phylogenet. Evol.* 46:809–817.
- Köhler T, Dietrich C, Scheffrahn RH and Brune A (2012). High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation of the gut in wood-feeding higher termites (*Nasutitermes* spp.). *Appl. Environ. Microbiol.* 78:4691–4701.
- Li H, Sun J, Zhao J, Deng T, Lu J, Dong Y, Deng W and Mo J (2012). Physiochemical conditions and metal ion profiles in the gut of the fungus-growing termite *Odontotermes formosanus*. *J. Insect Physiol.* 58:1368–1375.
- Liu N, Zhang L, Zhou H, Zhang M, Yan X, Wang Q, Long Y, Xie L, Wang S, Huang Y and Zhou Z (2013). Metagenomic insights into metabolic capacities of the gut microbiota in a fungus-cultivating termite (*Odontotermes yunnanensis*). *PLoS ONE* 8:e69184.
- Lo N and Eggleton P (2011). Termite phylogenetics and co-cladogenesis with symbionts, pp. 27–50. *In*: Bignell DE, Roisin Y and Lo N (eds.), *Biology of Termites: A Modern Synthesis*. Springer, Dordrecht.

- Loman NJ, Constantinidou C, Chan JZM, Halachev M, Sergeant M, Penn CW, Robinson ER and Pallen MJ (2012). High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nat. Rev. Microbiol.* 10:599–606.
- Lucey KS and Leadbetter JR (2014). Catechol 2,3-dioxygenase and other meta-cleavage catabolic pathway genes in the 'anaerobic' termite gut spirochete *Treponema primitia*. *Mol. Ecol.* 23:1531–1543.
- Matson EG, Gora KG and Leadbetter JR (2011). Anaerobic carbon monoxide dehydrogenase diversity in the homoacetogenic hindgut microbial communities of lower termites and the Wood Roach. *PLoS ONE* 6:e19316.
- Mikaelyan A, Lampert N, Köhler T, Rohland J, Boga H, Meuser K and Brune A (2015). DictDb: an expanded reference database for the highly resolved classification of the bacterial gut microbiota of termites and cockroaches. Submitted manuscript.
- Mikaelyan A, Strassert JFH, Tokuda G and Brune A (2014). The fiber-associated cellulolytic bacterial community in the hindgut of wood-feeding higher termites (*Nasutitermes* spp.). *Environ. Microbiol.* 16:2711–2722.
- Nalepa CA (2011). Altricial development in wood-feeding cockroaches: the key antecedent of termite eusociality, pp. 69–95. *In* Bignell DE, Roisin Y, Lo N (eds.), *Biology of Termites: A Modern Synthesis*. Springer, Dordrecht.
- Ngugi DK and Brune A (2012). Nitrate reduction, nitrous oxide formation, and anaerobic ammonia oxidation to nitrite in the gut of soil-feeding termites (*Cubitermes* and *Ophiotermes* spp.). *Environ. Microbiol.* 14:860–871.
- Ni J and Tokuda G (2013). Lignocellulose-degrading enzymes from termites and their symbiotic microbiota. *Biotechnol. Adv.* 31:838–850.
- Nobre T, Rouland-Lefèvre C and Aanen DK (2011). Comparative biology of fungus cultivation in termites and ants, pp. 193–210. *In* Bignell DE, Roisin Y, Lo N (eds.), *Biology of Termites: A Modern Synthesis*. Springer, Dordrecht.
- Noda S, Kitade O, Inoue T, Kawai M, Kanuka M, Hiroshima K, Hongoh Y, Constantino R, Uys V, Zhong J, Kudo T and Ohkuma M (2007). Cospeciation in the triplex symbiosis of termite gut protists (*Pseudotrichonympha* spp.), their hosts, and their bacterial endosymbionts. *Mol. Ecol.* 16:1257–1266.
- Noirot C (1995). The gut of termites (isoptera). Comparative anatomy, systematics, phylogeny. I. Lower termites. *Ann. Soc. Entomol. Fr. (N.S.)* 31:197–226.
- Noirot C (2001). The gut of termites (*Isoptera*). Comparative anatomy, systematics, phylogeny. II. Higher termites (*Termitidae*). *Ann. Soc. Entomol. Fr. (N.S.)* 37:431–471.
- Ohkuma M and Brune A (2011). Diversity, structure, and evolution of the termite gut microbial community, pp. 413–438. *In*: Bignell DE, Roisin Y, Lo N (eds.), *Biology of Termites: A Modern Synthesis*. Springer, Dordrecht.
- Ohkuma M and Kudo T (1996). Phylogenetic diversity of the intestinal bacterial community in the termite *Reticulitermes speratus*. *Appl. Environ. Microbiol.* 62:461–468.
- Ohkuma M and Kudo T (1998). Phylogenetic analysis of the symbiotic intestinal microflora of the termite *Cryptotermes domesticus*. *FEMS Microbiol. Lett.* 164:389–395.
- Ottesen EA and Leadbetter JR (2011). Formyltetrahydrofolate synthetase gene diversity in the guts of higher termites with different diets and lifestyles. *Appl. Environ. Microbiol.* 77:3461–3467.
- Paul K, Nonoh JO, Mikulski L and Brune A (2012). "Methanoplasmatales," *Thermoplasmatales*-related archaea in termite guts and other environments, are the seventh order of methanogens. *Appl. Environ. Microbiol.* 78:8245–8253.
- Pester M and Brune A (2007). Hydrogen is the central free intermediate during lignocellulose degradation by termite gut symbionts. *ISME J.* 1:551–565.
- Poulsen M (2015). Towards an integrated understanding of the consequences of fungus domestication on the fungus-growing termite gut microbiota. *Environ. Microbiol.* (published online) <http://dx.doi.org/10.1111/1462-2920.12765>.
- Poulsen M, Hu H, Li C, Chen Z, Xu L, Otani S, Nygaard S, Nobre T, Klaubauf S, Schindler PM, Hauser F, Pan H, Yang Z, Sonnenberg A9, de Beer ZW, Zhang Y, Wingfield MJ, Grimmelikhuijzen CJ, de Vries RP, Korb J, Aanen DK, Wang J, Boomsma JJ, Zhang G (2014). Complementary symbiont contributions to plant decomposition in a fungus-farming termite. *Proc. Natl. Acad. Sci. USA* 111: 14500–14505.

- Radek R, Strassert JFH, Krüger J, Meuser K, Scheffrahn RH and Brune A (2014). Phylogeny and ultrastructure of *Oxymonas jouteli*, a rostellum-free species, and *Opisthomitius longiflagellatus* sp. nov., oxymonadid flagellates from the gut of *Neotermes jouteli*. *Protist* 165:384–399.
- Rosengaus RB, Schultheis K.F, Yalonetskaya A, Bulmer M.S, du Comb W, Benson RW, Thottam JP and Godoy-Carter V (2014). Symbiont-derived  $\beta$ -1,3-glucanases in a social insect: mutualism beyond nutrition. *Front. Microbiol.* 5:607.
- Rosenthal AZ, Zhang X, Lucey KS, Ottesen EA, Trivedi V, Choi HMT, Pierce NA and Leadbetter JR (2013). Localizing transcripts to single cells suggests an important role of uncultured *Deltaproteobacteria* in the termite gut hydrogen economy. *Proc. Natl. Acad. Sci. USA* 110:16163–16168.
- Schauer C, Thompson CL and Brune A (2012). The bacterial community in the gut of the cockroach *Shelfordella lateralis* reflects the close evolutionary relatedness of cockroaches and termites. *Appl. Environ. Microbiol.* 78:2758–2767.
- Schmitt-Wagner D, Friedrich MW, Wagner B and Brune A (2003). Phylogenetic diversity, abundance, and axial distribution of bacteria in the intestinal tract of two soil-feeding termites (*Cubitermes* spp.). *Appl. Environ. Microbiol.* 69:6007–6017.
- Shinzato N, Matsumoto T, Yamaoka I, Oshima T and Yamagishi A (1999). Phylogenetic diversity of symbiotic methanogens living in the hindgut of the lower termite *Reticulitermes speratus* analyzed by PCR and in situ hybridization. *Appl. Environ. Microbiol.* 65:837–840.
- Shinzato N, Muramatsu M, Matsui T and Watanabe Y (2005). Molecular phylogenetic diversity of the bacterial community in the gut of the termite *Coptotermes formosanus*. *Biosci. Biotechnol. Biochem.* 69:1145–1155.
- Stingl U, Radek R, Yang H and Brune A (2005). 'Endomicrobia': Cytoplasmic symbionts of termite gut protozoa form a separate phylum of prokaryotes. *Appl. Environ. Microbiol.* 71:1473–1479.
- Tokuda G, Nakamura T, Murakami R and Yamaoka I (2001). Morphology of the digestive system in the wood-feeding termite *Nasutitermes takasagoensis* (Shiraki) [Isoptera: Termitidae]. *Zool. Sci.* 18:869–877.
- Warnecke F, Luginbühl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, Cayouette M, McHardy AC, Djordjevic G, Aboushadi N, Sorek R, Tringe SG, Podar M, Martin HG, Kunin V, Dalevi D, Madejska J, Kirton E, Platt D, Szeto E, Salamov A, Barry K, Mikhailova N, Kyrpides NC, Matson EG, Ottesen EA, Zhang X, Hernández M, Murillo C, Acosta LG, Rigoutsos I, Tamayo G, Green BD, Chang C, Rubin EM, Mathur EJ, Robertson DE, Hugenholtz P and Leadbetter JR (2007). Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nat.* 450:560–565.
- Watanabe H and Tokuda G (2010). Cellulolytic systems in insects. *Annu. Rev. Entomol.* 55:609–632.
- Wertz JT and Breznak JA (2007). *Stenoxybacter acetivorans* gen. nov., sp. nov., an acetate-oxidizing obligate microaerophile among diverse O<sub>2</sub>-consuming bacteria from termite guts. *Appl. Environ. Microbiol.* 73:6819–6828.
- Wertz JT, Kim E, Breznak JA, Schmidt TM and Rodrigues JLM (2012). Genomic and physiological characterization of the *Verrucomicrobia* isolate *Diplosphaera colotermitum* gen. nov., sp. nov. reveals microaerophily and nitrogen fixation genes. *Appl. Environ. Microbiol.* 78:1544–1555.
- Yamada A, Inoue T, Noda Y, Hongoh Y and Ohkuma M (2007). Evolutionary trend of phylogenetic diversity of nitrogen fixation genes in the gut community of wood-feeding termites. *Mol. Ecol.* 16:3768–3777.
- Zheng, H, Bodington D, Zhang C, Miyanaga K, Tanji Y, Hongoh Y and Xing XH (2013). Comprehensive phylogenetic diversity of [FeFe]-hydrogenase genes in termite gut microbiota. *Microb. Environ.* 28:491–494.

# High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation in the gut of wood-feeding higher termites (*Nasutitermes* spp.)

Tim Köhler<sup>1</sup>, Carsten Dietrich<sup>1</sup>,  
Rudolf H. Scheffrahn<sup>2</sup> and Andreas Brune<sup>1</sup>

*Affiliations:* <sup>1</sup>Max Planck Institute for Terrestrial Microbiology, 35043 Marburg, Germany; <sup>2</sup>Fort Lauderdale Research and Education Center, University of Florida, Davie, FL 33314, USA | *This chapter is published in:* Köhler T, Dietrich C, Scheffrahn RH and Brune A (2012). High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation of the gut in wood-feeding higher termites (*Nasutitermes* spp.). *Appl. Environ. Microbiol.* 78:4691–4701 | *Contributions:* T.K. designed and carried out experiments, analyzed data and wrote the manuscript. C.D. designed and carried out experiments, analyzed data and contributed to the manuscript. R.H.F. provided termites and Figure 1. A.B. conceived the study, wrote the manuscript, and secured funding.

## Abstract

Higher termites are characterized by a purely prokaryotic gut microbiota and an increased compartmentation of their intestinal tract. In soil-feeding species, each gut compartment has different physicochemical conditions and is colonized by a specific microbial community. Although considerable information has been accumulated also reported for wood-feeding species of the genus *Nasutitermes*, including cellulase activities and metagenomic data, a comprehensive study linking physicochemical gut conditions with the structure of the microbial communities in the different gut compartments is lacking. In this study, we measured high-resolution profiles of H<sub>2</sub>, O<sub>2</sub>, pH, and redox potential in the gut of *Nasutitermes corniger*, determined the fermentation products accumulating in the individual gut compartments, and analyzed the bacterial communities in detail by pyrotag sequencing of the V3–V4 region of the 16S rRNA genes. The dilated hindgut paunch (P3) was the only anoxic gut region, showed the highest density of bacteria, and accumulated H<sub>2</sub> to high partial pressures (up to 12 kPa). Molecular hydrogen is apparently produced by a dense community of *Spirochaetes* and *Fibrobacteres*, which dominate also the gut of other *Nasutitermes* species. All other compartments, such as the alkaline P1 (average pH 10.0), showed high redox potentials and comprised small but distinct populations characteristic for each gut region. In the crop and the posterior hindgut compartments, the community was even more diverse than in the paunch. Similarities in the communities of the posterior hindgut and crop suggested that proctodeal trophallaxis or coprophagy occurs also in higher termites. The large sampling depths of pyrotag sequencing in combination with the determination of important physicochemical parameters allow to draw cautious conclusions concerning the functions of particular bacterial lineages in the respective gut sections.

## Introduction

Termites contribute substantially to the turnover of carbon and nitrogen in tropical ecosystems (Jouquet et al., 2011). Their diet consists exclusively of lignocellulose in various stages of decomposition, ranging from sound wood to humus (Bignell, 2011). The digestion of this recalcitrant diet relies on the metabolic activities of a dense and diverse intestinal microbiota (Brune and Ohkuma, 2011). In the evolutionarily lower termites, flagellate protists hydrolyze the wood and ferment the resulting monomers, but in higher termites, these cellulolytic symbionts are lacking (Brune and Ohkuma, 2011, and references therein). Although the endoglucanases in the midgut region are secreted by the host epithelium, the cellulolytic activities in the hindgut of higher termites seem to be of bacterial origin (Tokuda and Watanabe, 2007; Watanabe and Tokuda, 2010).

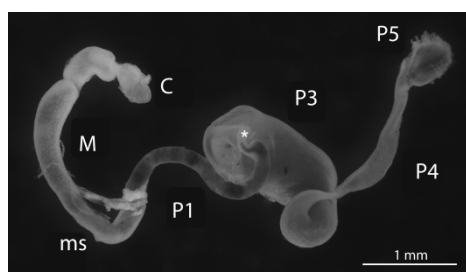
In many higher termites, the hindgut is strongly compartmentalized (Noirot, 2001), forming several consecutive microbial bioreactors, some of which are extremely alkaline (Bignell and Eggelton, 1995; Brune and Kühl, 1996). The hindgut microbiota of wood-feeding *Microcerotermes* and *Nasutitermes* spp. is dominated by *Spirochaetes*, *Fibrobacteres*, and members of the candidate phylum TG3 (Hongoh et al., 2005; Hongoh et al., 2006). A metagenomic analysis of the microbiota in the hindgut paunch (P3) of a *Nasutitermes* sp. implicated members of *Spirochaetes* and *Fibrobacteres* in the hydrolysis of wood (Warnecke et al., 2007). Although the presence of hydrogenase genes indicates the capacity of the gut microbiota to form or consume H<sub>2</sub>, the presence of H<sub>2</sub> in the paunch or other sections of *Nasutitermes* spp., particularly the alkaline gut region (Brune et al., 1995), remains to be elucidated. The individual gut compartments of soil-feeding *Cubitermes* spp. (subfamily Termitinae) are colonized by different communities of bacteria and archaea (Friedrich et al., 2001; Schmitt-Wagner et al. 2003a), and the alkaline P1 compartments of different higher termites harbor a similar bacterial microbiota (Thongaram et al., 2005). However, apart from a study of the bacteria colonizing the mixed segment of *Nasutitermes takasagoensis* (subfamily Nasutitermitinae; Tokuda et al. 2000), information about the microbial communities in the different hindgut sections of wood-feeding higher termites and a highly resolved analysis of important physicochemical parameters in the different gut regions are lacking.

In this study, we combine microsensor measurements of physicochemical conditions (oxygen and hydrogen partial pressure, redox potential, and pH) with high-resolution profiles of the bacterial microbiota and their fermentation products in the different gut compartments of *Nasutitermes corniger*.



## Materials and Methods

**Sample preparation.** *Nasutitermes corniger* (Motschulsky) was taken from a laboratory nest (University of Florida) collected in Dania Beach, Florida, from established field populations of this exotic arboreal termite, which is synonymous with *Nasutitermes costalis* (Scheffrahn et al. 2002; Scheffrahn et al., 2005). *Nasutitermes takasagoensis* was collected on Iriomote Island, Japan by Gaku Tokuda (University of the Ryukyus, Okinawa). Only worker caste termites were used for the experiments. After dissecting the termites with sterile, fine-tipped forceps, we used intact guts for microsensor studies of the individual compartments (Fig. 1). For metabolic profiles and pyrotag sequencing, we separated the guts placed under a stereomicroscope into six major sections, comprising the crop, the midgut, and the four major hindgut compartments (P1, P3, P4, and P5), and homogenized them with sterile micropestles (Eppendorf, Hamburg, Germany). Because they are difficult to delineate, the mixed segment (ms) was included with the P1. To increase sensitivity of detection and to account for intraspecific variations, we always pooled an indicated number of gut sections (see below).



**Fig. 1 | The intestinal tract of *Nasutitermes corniger*.** The gut includes crop (C), midgut (M), mixed segment (ms), and several hindgut segments (P1–P5); the asterisk marks the position of the P2 (enteric valve).

**Microsensor measurements.** All microsensors were purchased from Unisense (Aarhus, Denmark). Oxygen (OX-10) and hydrogen (H<sub>2</sub>-10) microsensors had tip diameters of ca. 10 μm and detection limits of ca. 0.02 and 0.04 kPa, and were polarized and calibrated as previously described (Brune et al., 1995; Ebert and Brune, 1997). The redox electrode (RD-10) had a tip diameter of ca. 10 μm, and the pH electrode (PH-10) had a tip diameter of 10–20 μm and a sensitive tip length of 100–150 μm; each was used together with a Ag–AgCl reference electrode and a high-impedance voltmeter. pH measurements were calibrated using standard curves obtained with commercial standard solutions of pH 4.0, 7.0, 9.0, and 11.0 (Carl Roth, Karlsruhe, Germany) as previously described (12). Redox measurements were calibrated using freshly prepared saturated quinhydrone solutions in pH standards at pH 4.0 and 7.0.

Microsensor profiles were measured in glass-faced microchambers as previously described (Brune et al., 1995). Freshly dissected guts of *N. corniger* were placed flat and fully extended onto a 4-mm layer of 2% agarose and covered with 0.5% agarose (w/v; both made up with Ringer's solution). Microsensors were positioned using a manual micromanipulator (Märzhäuser, Wetzlar, Germany), and tip position was visually controlled with a horizontally mounted stereomicroscope (Zeiss, Jena, Germany); measurement commenced ca. 10 min after embedding and lasted less than 1 h.

**Metabolite pools.** Forty gut sections each of *N. corniger* were homogenized in 80 μl NaOH (10 mM), and metabolites in the clarified supernatants were analyzed using a combination of gas chromatography (GC; 66) and high performance liquid chromatography (HPLC; Pester and Brune, 2007), as previously described in detail.

**Microbial cell counts.** Twenty gut sections each of *N. corniger* were homogenized in 0.5 ml phosphate-buffered saline (PBS; Pernthaler et al., 2004) and fixed with 4% (v/v) formaldehyde at 4 °C for 13 h. Microbial cells were counted using the procedure of Pernthaler et al. (2001) but excluding the sonication step. Samples were washed with PBS, and appropriate dilutions were filtered onto polycarbonate filters (0.2 µm; GTTP; Millipore, Schwalbach/Ts., Germany) and stored at –20 °C. For analysis, filters were stained with 4',6'-diamidino-2-phenylindole (DAPI), washed with sterile water and then with 70% (v/v) ethanol, and embedded in Citifluor AF1 (Citifluor, London, UK). Microbial cells were counted at 1,000-fold magnification using a Zeiss Axiophot epifluorescence microscope equipped as previously described (Schmitt-Wagner et al., 2003a).

**Primer design.** Primers 341F (Muyzer et al., 1993) and 787R (Hugenholtz et al., 2001) targeting the V3–V4 region of the bacterial 16S rRNA gene were modified on the basis of the sequence information in the SILVA 100 database (Pruesse et al., 2007), focusing on an optimal coverage of the taxa known to prevail in termite guts. Modifications were tested using the *probe match* function of the ARB software (version 5.1; Ludwig et al., 2004). The resulting primer set, 343Fmod (TACGGGWGGCWGCA) and 784Rmod (GGGTMCTAATCCBKTT), showed perfect matches to 87% of the sequences in the database (90.5% allowing one mismatch), and coverage was even higher in the phyla relevant to termite gut environment (see Fig. S1).

**Pyrotag sequencing.** Twenty sections of each gut compartment of *N. corniger*, ten hindguts (P1–P5) of both *N. corniger* and *N. takasagoensis*, and ten whole guts of *N. corniger* were each pooled and homogenized in PBS. DNA was extracted with phenol–chloroform using the bead-beating protocol as described in Henckel et al. (1999), precipitated with two volumes of polyethylene glycol, and amplified with the newly designed primers using a high-fidelity polymerase (Herculase II, Agilent, Waldbronn, Germany). The PCR conditions were: initial denaturation (3 min at 95 °C), 26 cycles of amplification (20 s at 95 °C, 20 s at 48 °C, and 30 s at 72 °C), and terminal extension (3 min at 72 °C). Both the forward and the reverse primers each had an additional, sample-specific 6-bp barcode at the 5' end that differed by at least 2 bp between samples and contained no homopolymers. The amplicons were quantified photometrically (NanoDrop; Thermo Fisher Scientific, Schwerte, Germany) and mixed in equimolar amounts before further analysis. Adaptor ligation, subsequent amplification, and pyrosequencing (454 GS FLX Titanium; Roche, Mannheim, Germany) were done by a commercial service (GATC Biotech; Konstanz, Germany).

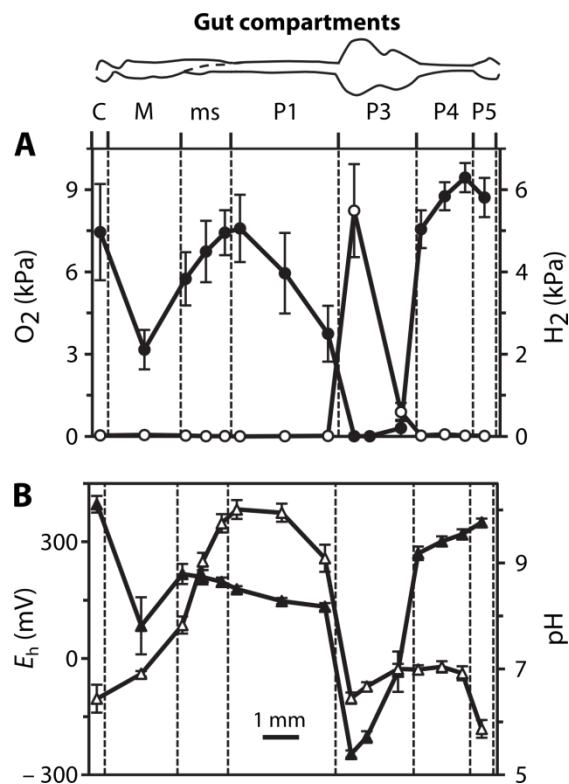
**Sequence processing and classification.** Pyrotag data were pre-processed using the *mothur* software suite (version 1.15.0; Schloss et al., 2009), following the strategy described by Schloss et al. (2011) with slight modifications. After sorting the sequences by their unique barcodes, all sequences that were shorter than 200 bp, contained ambiguous bases, had errors in the primer sequence, or showed homopolymer regions of more than 10 nucleotides were removed from the datasets. For phylotype analyses the remaining sequences were denoised as previously described (Schloss et al., 2011); for classification analyses the sequences were aligned against the SILVA 102 non-redundant database (Pruesse et al., 2007) using a stand-alone version of the SINA aligner (<http://www.arb-silva.de>).

The sequences were assigned to taxonomic groups with the *Naïve Bayesian Classifier* implemented in the *mothur* software using a manually curated reference database and a confidence threshold of 60%. The reference database consisted of the SILVA 102 non-redundant database amended with numerous unpublished sequences from termite and cockroach guts obtained in our laboratory. The existing classification of the SILVA database was extended and refined down to the genus level by

introducing additional, termite-specific groups and renaming redundant or uninformative taxa. To allow processing in the *mothur* software environment and to improve the speed of the classifier, uninformative sequences from those taxa that contained no gut-related sequences were removed. The resulting reference database (82,400 sequences) contained all bacterial isolates, all uncultivated bacteria from intestinal environments, and at least three representative sequences from every other lowest-level group in the SILVA database. It is available from the authors upon request.

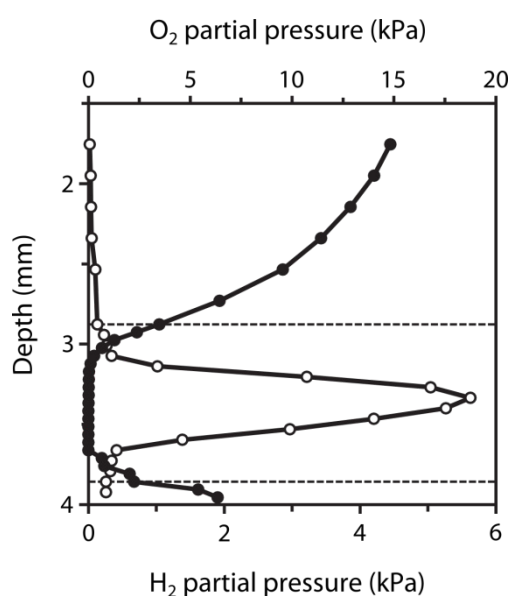
## Results

**Physicochemical gut conditions.** Microsensor profiles along the gut axis of *N. corniger* showed strong dynamics of oxygen concentration (Fig. 2A). Total anoxia was observed only in the highly dilated hindgut paunch (anterior P3), but not in the less dilated, posterior part of the P3 compartment, where traces of oxygen at its center were often found, which suggested that a complete removal of oxygen depends on the diameter of the respective gut region. Radial oxygen profiles of the P3 compartment showed that the gut periphery acts as an oxygen sink, with the microoxic zone typically extending 200–300  $\mu\text{m}$  below the gut wall (Fig. 3). However, the penetration depth of oxygen changed with the depth of embedding. If the agarose layer above the hindgut paunch was very shallow (< 2 mm), occasionally the entire compartment became oxic.



**Fig. 2 | Axial profiles of (A) oxygen (●) and hydrogen (○) partial pressure, and (B) redox potential (▲) and pH (△) along the gut of *Nasutitermes corniger*, measured at the gut center.** Values are means  $\pm$  standard errors obtained with 8–12 termites (except for the crop, which was lost in about half of the preparations). For abbreviations of gut compartments, see Fig. 1.

The oxygen status corresponded to the redox conditions in the respective compartments, i.e., only the anoxic region (P3) had a negative redox potential (Fig. 2B). Also the accumulation of H<sub>2</sub> was restricted to the P3 compartment, with maximal values in the anterior region (Fig. 2A). Radial hydrogen profiles of the anterior P3 revealed steep H<sub>2</sub> gradients from the gut center towards the gut wall (Fig. 3). However, hydrogen partial pressures in the P3 varied over a wide range (from 0.02 to 12 kPa) and were quite sensitive to the depth of embedding. When intestinal hydrogen partial pressures were measured *in situ* (by inserting the microsensor through the dorsal cuticle into the abdomen of decapitated termites), hydrogen partial pressures ranged between 0.1 and 2.4 kPa. However, these values have to be regarded with caution, because the intransparency of the cuticle did not allow us to determine the exact location of the microsensor tip or to assess the damage to the intestines caused by the sensor.



**Fig. 3 | Radial profiles of oxygen (●) and hydrogen (○) partial pressure in the agarose-embedded anterior P3 compartment of *Nasutitermes corniger*, relative to the agarose surface.** The dotted lines indicate the position of the proximal and distal gut wall. The profiles were selected as typical among six similar sets obtained with different termites.

Also the pH of the gut contents was found to be highly dynamic along the gut axis. The intestinal pH was slightly acidic in the crop, circumneutral in the midgut, and increased sharply in the mixed segment. The most alkaline values (pH 9.3–10.9) were found in the anterior P1. The pH decreased again in the P3 compartment and remained neutral in most of the posterior hindgut, and turned again slightly acidic in the P5 (Fig. 2B).

**Metabolite pools.** The metabolites accumulating in the different gut sections of *N. corniger* were determined by means of HPLC and GC (Table 1). Acetate was the predominant metabolite in all gut sections, except for the midgut, where succinate was more abundant. The highest proportion of acetate was present in the P3 section, which also contained the largest metabolite pool of all compartments. Lactate was detected only in the posterior gut, with highest concentration in the P5 section. Similar results have been previously reported for other *Nasutitermes* spp., except that the pool sizes of propionate, butyrate, and formate were lower (66).

**Bacterial diversity.** The microbial cell counts in homogenates of individual gut sections of *N. corniger* differed greatly (Table 1). The highest absolute number was found in the P3 compartment ( $1.5 \times 10^7$  cells), surpassing those in the other gut regions by more than two orders of magnitude. The microbiota of the crop consisted mostly of cocci, whereas the midgut microbiota was dominated by short rods; cells with spirochetal shape were rare in both compartments. In the P1 compartment, we observed mostly longer rods; cocci were less abundant, and the density of cells with spirochetal shape began to increase. The density of spirochetes was highest in the P3 compartment, but decreased again in the posterior sections, whose microbiota was dominated by coccoid cells.

Bacterial diversity in the different gut sections of *N. corniger* and in total hindguts of *N. corniger* and *N. takasagoensis* was determined by pyrotag sequencing of the V3–V4 region of the 16S rRNA genes in the DNA extracted from the different samples. Trimming and quality control removed 10–20% of the sequences from each dataset, resulting in sequence libraries of 3,200–26,000 reads per sample (Table S1).

**Tab. 1 | Pool sizes of major metabolites, fresh weight, and microbial cell counts for different gut sections of *Nasutitermes corniger*.** Values are averages  $\pm$  range of two independent experiments with 40 gut sections each.

Section	Pool size (nmol)						Fresh wt. (mg)	Cell counts ( $10^6$ section $^{-1}$ )	Cell density ( $10^9$ g $^{-1}$ ) <sup>a</sup>
	Acetate	Propionate	Butyrate	Succinate	Lactate	Formate			
C	0.7 $\pm$ 0.1	0.2 $\pm$ 0.2	— <sup>b</sup>	0.4 $\pm$ 0.1	—	0.3 $\pm$ 0.1	0.7 $\pm$ 0.2	0.15 $\pm$ 0.03	0.21 $\pm$ 0.07
M	0.9 $\pm$ 0.0	0.2 $\pm$ 0.2	—	2.0 $\pm$ 1.0	—	0.1 $\pm$ 0.1	0.6 $\pm$ 0.1	0.08 $\pm$ 0.02	0.13 $\pm$ 0.04
ms/P1	1.4 $\pm$ 0.2	0.1 $\pm$ 0.1	—	1.1 $\pm$ 0.7	—	0.7 $\pm$ 0.1	0.8 $\pm$ 0.2	0.10 $\pm$ 0.04	0.13 $\pm$ 0.06
P3	8.6 $\pm$ 1.8	0.7 $\pm$ 0.4	0.1 $\pm$ 0.1	1.0 $\pm$ 0.8	0.1 $\pm$ 0.1	0.5 $\pm$ 0.1	1.4 $\pm$ 0.3	15.2 $\pm$ 3.1	10.9 $\pm$ 3.2
P4	2.1 $\pm$ 1.0	0.6 $\pm$ 0.2	0.1 $\pm$ 0.1	0.3 $\pm$ 0.1	0.1 $\pm$ 0.1	0.7 $\pm$ 0.1	0.4 $\pm$ 0.2	0.08 $\pm$ 0.01	0.20 $\pm$ 0.10
P5	1.9 $\pm$ 1.2	0.4 $\pm$ 0.2	—	0.3 $\pm$ 0.2	0.7 $\pm$ 0.7	0.6 $\pm$ 0.2	0.6 $\pm$ 0.5	0.04 $\pm$ 0.01	0.07 $\pm$ 0.06
<b>Total gut<sup>c</sup></b>	<b>15.6 <math>\pm</math> 2.4</b>	<b>2.2 <math>\pm</math> 0.5</b>	<b>0.2 <math>\pm</math> 0.2</b>	<b>5.1 <math>\pm</math> 1.5</b>	<b>0.9 <math>\pm</math> 0.7</b>	<b>2.9 <math>\pm</math> 0.3</b>	<b>4.5 <math>\pm</math> 0.7</b>	<b>15.6 <math>\pm</math> 3.1</b>	<b>3.47 <math>\pm</math> 0.87</b>

<sup>a</sup> Based on fresh wt., using error propagation<sup>b</sup> Below detection limit (ca. 0.02 nmol)<sup>c</sup> Calculated from the amount in each compartment**Tab. 2 | Comparison of the classification success at different taxonomic levels using the RDP online platform (Release 10) and the curated reference database (this study).** Values (in %) are based on the total number of sequences in the sample.

Section	Phylum		Family		Genus	
	RDP	This study	RDP	This study	RDP	This study
C	89	98	79	90	49	76
M	96	99	17	95	10	89
ms/P1	88	98	88	80	46	67
P3	78	99	80	95	57	87
P4	85	99	76	83	46	72
P5	88	99	80	79	45	72

Preliminary analysis using the classifier of the online platform of the Ribosomal Database Project (RDP, Release 10; Wang et al., 2007) yielded large fractions of unclassified sequences at all taxonomic levels (Table 2). Since most of the unclassified sequences were termite-specific bacterial groups that were not represented or poorly resolved in the reference database used by RDP, we prepared a manually curated reference database specifically adapted to the bacterial diversity in termite guts (see Methods). Reclassification of the samples using the *mothur* software suite (version 1.15.0; Schloss et al., 2009) resulted in a significantly improved classification at the phylum level, reducing the fraction of unclassified sequences in the different samples from 4–22% to 1–2%. The effect was even stronger at lower taxonomic ranks; at the genus level, the fraction of unclassified sequences in the samples decreased from 43–90% to 11–33%. The remaining sequences could be assigned only to higher taxa mostly because closely related reference sequences were lacking. Closer inspection of 36 randomly selected sequences without phylum-level classification revealed that half of them were putative chimeras and the other half did not code for 16S rRNA.

The individual gut sections of *N. corniger* each contained sequences from 200–300 different taxa (genus level), with highest numbers in the crop, P3, and P4 (Table 3). Similarity-based clustering of the sequences into phylotypes with 5% (genus level) or 3% (species level) sequence divergence indicated that genus/species richness in each sample was considerably higher than indicated by hierarchical classification (many rare species were not classified in lieu of appropriate reference sequences); predictions of species richness based on the abundance of singletons in the different datasets (using the Chao1 estimator) were higher (Table 3).

Diversity and evenness of the bacterial community were lowest in the midgut, which harbored a few very abundant groups. In the other gut sections, diversity was much higher and community structure was more balanced (Table 3); the same trends were observed also with similarity-based classification (5% cutoff) and hierarchical classification (genus level; details not shown). The relatively small number of phylotypes agrees with the results of Engelbrektson et al. (20), who observed less than 1000 phylotypes in their rarefaction analyses when they tested several primer pairs for pyrotag sequencing using *N. corniger* luminal P3 hindgut compartment DNA as template. Nevertheless, the composition of the communities differed substantially between the compartments (Table 4). High similarities were observed between the crop and hindgut (P4–P5), whereas the midgut had only low similarities to other compartments.

**Community structure.** The major bacterial phyla consistently encountered in the different gut compartments of *N. corniger* were *Spirochaetes*, candidate phylum TG3, *Firmicutes*, *Fibrobacteres*, *Bacteroidetes*, *Proteobacteria*, and *Actinobacteria* (Fig. 4). *Spirochaetes* and members of the TG3 phylum were represented in all compartments, but were most abundant in the hindgut paunch (P3). The phylum-level patterns in the posterior hindgut sections (P4 and P5) were similar to that of the crop, except for an increased abundance of *Firmicutes* in all anterior sections.

At higher taxonomic resolution, it becomes apparent that most phyla are represented by various lineages that are often unevenly distributed among the compartments. Figure 5 summarizes the relative abundance of the 50 major families represented in the different samples; detailed results for all taxonomic ranks can be found in an interactive table included as supplementary material (Table S1). A prominent example is the *Firmicutes*. In the midgut, most of the sequences of this phylum (80% of all sequences) are members of the order *Clostridiales*, consisting almost exclusively of a particular group of *Lachnospiraceae* ("uncultured 67"; Table S1). Although present also in the other compartments, this group is outnumbered by other *Clostridiales* (*Ruminococcaceae* or Family XIII Incertae Sedis) in the posterior hindgut (P4) and by *Lactobacillales* in the crop (here mainly

*Streptococcaceae*) and in the anterior hindgut (P1; here mainly members of the insect group PeH08 and other, unclassified *Lactobacillales*). Many family-level taxa are highly represented in all gut sections (e.g., the termite clusters in *Fibrobacteres* and TG3 subphylum 1, or some *Bacteroidetes*). In some cases, the patterns in the posterior hindgut sections (P4 and P5) were similar to those of the crop, e.g., the *Ruminococcaceae* (*Clostridiales*), the *Acidobacteriaceae* (*Acidobacteria*), and the candidate divisions OP11, SR1, and TM7.

**Tab. 3 | Diversity and evenness of the bacterial communities in the different gut sections of *Nasutitermes corniger*. The number of taxa (hierarchical classification to genus level) is compared to the number of phylotypes (similarity-based classification, using 3% or 5% dissimilarity threshold).**

Section	Genus-level taxa <sup>a</sup>	Phylotypes		Diversity indices (based on 3% dissimilarity)		
		5%	3%	Expected phylotypes <sup>b</sup>	Diversity <sup>c</sup>	Evenness <sup>d</sup>
C	298	351	563	1174	3.39	0.45
M	217	285	511	1231	1.54	0.20
ms/P1	187	195	337	944	2.80	0.38
P3	264	360	653	1626	2.42	0.31
P4	307	411	726	1748	3.87	0.50
P5	173	167	275	494	3.77	0.57

<sup>a</sup> Lowest level of classification

<sup>b</sup> Chao1 estimator (Chao, 1984)

<sup>c</sup> Non-parametric Shannon index (Chao and Shen, 2003)

<sup>d</sup> Evenness (Legendre and Legendre, 1998)

Many of the sequences obtained from the gut of *N. corniger* represent termite-specific lineages that have already been encountered in clone-based inventories of the gut microbiota of other *Nasutitermes* species (e.g., Tokuda et al., 2000; Hongoh et al., 2006; Warnecke et al., 2007). However, deep sequencing of the communities in the individual gut regions also revealed the presence of many lineages hitherto undetected in termite guts (e.g., from the phyla *Lentisphaerae*, *Planctomycetes*, *Firmicutes*, and candidate divisions OP11, TM7, and SR1), which underlines the high diversity of the gut microbiota reflected also by the high Shannon indices for most compartments (Table 3). Although 75% of the families detected each represent less than 1% of the sequences obtained from the different sections (Table S1), many of these groups are numerically important because of the high density of the community (i.e., in the P3 compartment; Fig. 4) or because of their apparent specificity for termite guts. In any case, it should be considered that especially in the P3 section, taxa that are close to the detection limit of the pyrotag analysis still form substantial populations.

**Interspecific variation.** The bacterial community profiles of the P3 compartment are virtually identical to the artificial profiles for hindgut and whole gut generated using cell density and relative abundance of different families in the individual compartments (Fig. S2), which illustrates that community profiles of the total gut will always be dominated by the microbiota of the P3 compartment (Fig. 4). The high similarities of these profiles to the replicate profiles of the hindgut and total gut of *N. corniger* (Fig. S2), which were obtained with different batches of termites from the same nest, shared only 31% of the classified genera but corresponded to 95% of the total sequence abundance. This documents the reproducibility of the profiles and the noise in the low-abundant taxa/singletons that leads to high species-richness estimations (Table 3). The differences between the



samples were exclusively in the low-abundance taxa. The similarities between the hindgut profiles of *N. corniger* and the closely related *N. takasagoensis*, an allopatric species from Japan, were slightly lower (23% shared genera), but the profiles showed striking similarities in the presence and abundance of family-level taxa (Fig. 5).

**Tab. 4 | Similarity indices of the bacterial communities in different gut sections of *Nasutitermes corniger*.**

Section	Similarity <sup>a</sup>					
	C	M	P1	P3	P4	P5
C	1.00					
M	0.21	1.00				
ms/P1	0.39	0.22	1.00			
P3	0.36	0.13	0.23	1.00		
P4	0.39	0.14	0.33	0.48	1.00	
P5	0.35	0.17	0.33	0.20	0.20	1.00

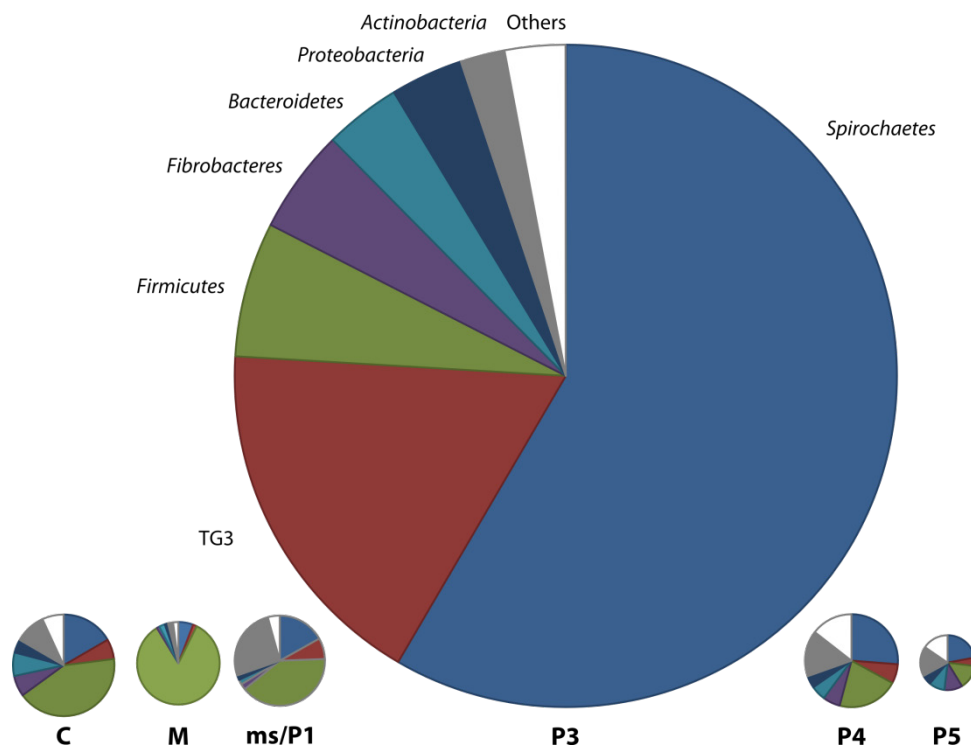
<sup>a</sup> Bray–Curtis coefficient (Bray and Curtis, 1957), based on sequence similarity (3% dissimilarity threshold for phylotypes).

**Comparison to clone libraries.** A comparison of the pyrotag datasets of *N. corniger* and *N. takasagoensis* to previous clone libraries of the bacterial 16S rRNA genes from the gut of *Nasutitermes* species showed that each of the major family-level lineages is represented in all *Nasutitermes* species, although their relative abundance differs (Fig. 5). An notable exception is a termite-specific lineage of *Bacteroidetes* (M2PB4-65) that is moderately abundant in the 454 datasets (0.6–1.7%) but not represented in the clone libraries. Strong differences between the datasets are encountered among the *Fibrobacteres*, TG3, *Firmicutes*, and the *Spirochaetes*, particularly in the virtual absence of *Fibrobacteres* from the hindgut of the batch of *N. takasagoensis* used in this study.

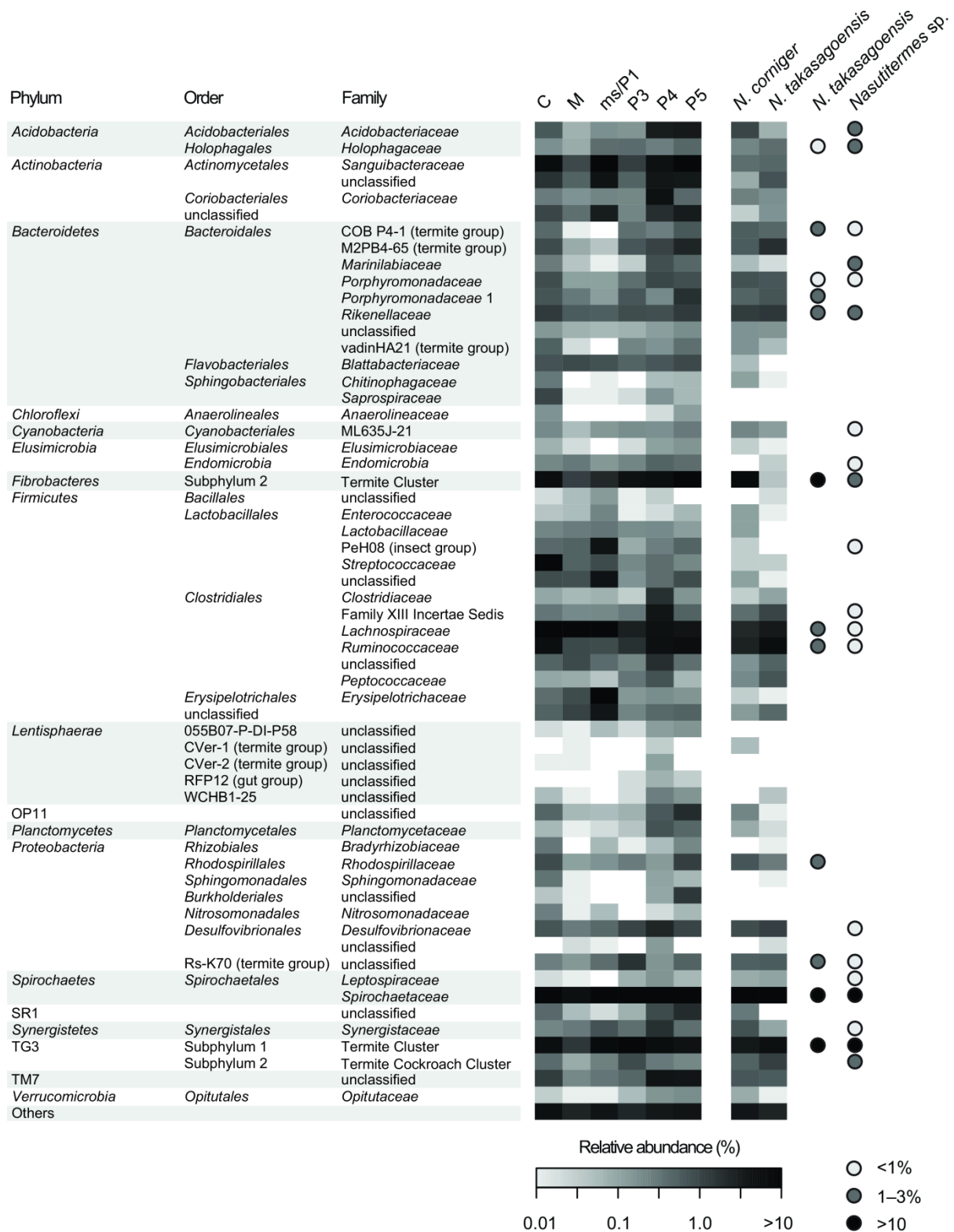
When we compared at the genus level the bacteria in the total P3 section of *N. corniger* to the bacteria detected in the lumen of this compartment of a *Nasutitermes* sp. collected in Costa Rica (73), we found that 79% of the taxa in the pyrotag libraries were represented, which indicated that the bulk of the P3 compartment gut microbiota was already detected by a clone library of 1252 sequences (73). However, the pyrotag library of the P3 (24,029 reads) comprised 217 additional taxa. Many of them were present also in the pyrotag library of *N. takasagoensis*, which indicated that they are likely to occur also in other *Nasutitermes* species.

An interesting aspect became apparent when we compared the two datasets in the opposite direction. Since the pyrotag dataset for *N. corniger* generated in this study was obtained from a homogenate of the complete P3 compartment and the clone library of *Nasutitermes* sp. was based only on its luminal content (73), any major taxa present in the analysis of the total compartment but missing from the luminal sample potentially represent wall-associated bacteria. To compensate the lower sequencing depth of the luminal sample, a threshold for the larger amounts of pyrotag sequences was set by taking the noise signal (i.e., one sequence) multiplied by three (i.e., three sequences in the luminal dataset corresponding to 0.24% in the pyrotag dataset). Taking this threshold, we discovered ten taxa that are strong candidates for gut-wall-associated bacteria (Table S1), including *Sanguibacter* spp. and other *Actinobacteria*, *Bacteroidetes* Cluster V

(*Porphyromonadaceae* 1), *Arthromitus* spp. (*Lachnospiraceae*), and some lineages of *Spirochaetaceae* specific for termite guts. Together, they formed 10% of the sequences from the P3 compartment. By contrast, taxa that were exclusively present in the luminal sample (Table S1) were only a small fraction (0.6%) of the clones in the library. Moreover, two of these groups, OPB56 (*Chlorobi*) and Rs-H88 (*Spirochaetes*) were present in the total hindgut sample of *N. corniger*.



**Fig. 4 | Relative abundance of major bacterial phyla in the different gut compartments of *Nasutitermes corniger*, based on pyrotag analysis of the V3–V4 region of the 16S rRNA genes.** The area of the circles reflects the microbial cell counts in the respective gut sections (see Table 1). For abbreviations, see Fig. 1.



**Fig. 5 | Relative abundance of the major bacterial taxa in the different gut sections of *Nasutitermes corniger* (for abbreviations, see Fig. 1) and in the total hindguts of *N. corniger* and *N. takasagoensis*.** Classification is shown down to the family level (for genus level, see Table S1). The heatmap uses a logarithmic scale to increase visibility of low abundance groups. The remaining sequences were extremely diverse (111 to 157 families) but represented each less than 1% of the community in the respective compartment (Table S1). The families represented in previously published clone libraries of *N. takasagoensis* (Hongoh et al., 2006; total gut) and a *Nasutitermes* sp. (Warnecke et al., 2007; P3 lumen), are shown for comparison (shading of circles indicates relative abundance).

## Discussion

This is the first comprehensive analysis of the digestive tract of a wood-feeding higher termite from a microbiological perspective. It combines a microsensor study of physicochemical gut conditions with a highly resolved analysis of the bacterial microbiota in the individual gut compartments. The results revealed that the gut is a highly structured microenvironment, with differences in metabolic activities and microbial community structure. The bulk of the microbiota is located in the dilated hindgut paunch (P3), but also the other compartments such as the alkaline P1 and the tubular P4 harbor communities that are each distinct from those in other gut regions. The differences are apparent already at the phylum level, and a detailed analysis of relative abundance indicates that individual lineages preferentially colonize particular niches. The results allow us to draw cautious conclusions concerning the functions of particular bacterial lineages in the respective sections.

**The hindgut paunch.** Since higher termites lack cellulolytic flagellates, wood fiber in the dilated hindgut paunch has to be digested by a purely prokaryotic microbiota (Brune and Ohkuma, 2011). The hindgut of *Nasutitermes takasagoensis* and *N. walkeri* contains substantial cellulolytic activity (Tokuda and Watanabe, 2007), and metagenomic analysis of the luminal contents of the P3 compartment of another *Nasutitermes* species has identified numerous glycosyl hydrolases putatively involved in the degradation of cellulose and hemicelluloses. The genes have been tentatively assigned to members of the phyla *Fibrobacteres* and *Spirochaetes* based on phylogenetic binning (Warnecke et al., 2007). Diverse members of these phyla and of the related TG3 phylum (included in the *Fibrobacteres* by Warnecke and colleagues; Warnecke et al., 2007) have been documented to occur abundantly in the hindgut of *Nasutitermes* species (Czolij et al., 1985; Paster et al., 1996; Ohkuma et al., 1999; Hongoh et al., 2006; Warnecke et al., 2007). In accordance with these reports, the mentioned phyla were highly represented also in the pyrotag sequences of the total hindgut of *N. corniger* and *N. takasagoensis*. They dominate the microbiota in the P3 compartment of *N. corniger* (Fig. 4), the major microbial bioreactor in terms of anoxic status, microbial cell count, and concentration of fermentation products.

Interestingly, the P3 compartment is also the only gut region where H<sub>2</sub> accumulated. Hydrogen partial pressures in the anterior P3 of *N. corniger* were in the same range as those in the paunch of *Reticulitermes flavipes* (Ebert and Brune, 1997), where H<sub>2</sub> production is attributed largely to the gut flagellates. Several higher termites, including *Nasutitermes triodiae*, have been reported to emit H<sub>2</sub> *in vivo* (Sugimoto et al., 1998), but microsensor profiles have so far been available only for soil-feeding *Cubitermes* spp., where the mixed segment and the P3 showed substantial accumulation of H<sub>2</sub> (Schmitt-Wagner and Brune, 1999). The bacterial populations responsible for hydrogen production have not been identified, but by means of phylogenetic analyses of conserved single-copy protein-coding genes, Warnecke et al. (2007) could link the iron-only hydrogenases in the metagenome of *Nasutitermes* sp. to members of the *Spirochaetes*. Molecular hydrogen is a major fermentation product of carbohydrates in many *Spirochaeta* spp. (Leschine et al., 2006) and also in *Treponema azotonutricium*, an isolate from the lower termite *Zootermopsis angusticollis* (Graber et al., 2004). All isolates of termite gut treponemes possess several [FeFe] hydrogenases (Ballor et al., 2011), and related hydrogenase genes are present also in other lower termites (Ballor and Leadbetter, 2011). It is therefore likely that spirochetes are—at least in part—also responsible for hydrogen production in *Nasutitermes* spp. It is possible that also members of *Fibrobacteres* and TG3, the other highly abundant bacterial phyla in the P3 compartment of *N. corniger* and hindguts of other *Nasutitermes* spp., contribute to hydrogen production. While nothing is known about these uncultivated lineages,

the genome of the distantly related *Fibrobacter succinogenes* does not encode any hydrogenases (Suen et al., 2011), and there were no hydrogenases binning with *Fibrobacteres* in the *Nasutitermes* sp. metagenome (Warnecke et al., 2007).

The steep radial profiles of H<sub>2</sub> in the P3 compartment of *N. corniger* indicate the presence of a strong hydrogen sink, which is in agreement with the high rates of reductive acetogenesis in hindgut homogenates of several *Nasutitermes* species (Brauman et al., 1992) and consolidates the large accumulation of H<sub>2</sub> within the lumen of the hindgut paunch with the low rates of hydrogen emission by living termites (Sugimoto et al., 1998). Analyses of the *fhs* gene, encoding formyl-tetrahydrofolate synthetase (FTHFS), a functional marker for the Wood–Ljungdahl pathway, has provided strong evidence that spirochetes are responsible for reductive acetogenesis from H<sub>2</sub> and CO<sub>2</sub> in the gut of lower termites (e.g. Leadbetter et al., 1992; Ottesen and Leadbetter, 2011; Pester and Brune, 2006; Salmassi and Leadbetter, 2003). A metagenomic survey of the hindgut microbiota of *Nasutitermes* sp. has indicated that *fhs* genes in the hindgut community are highly similar to those in the hindgut of lower termites, including that of the genuine homoacetogen *Treponema primitia* (Warenecke et al., 2007). Also the *cooS* genes in the metagenome, encoding a catalytic subunit of the carbon monoxide dehydrogenase, have been predicted to be encoded by treponemes (Warenecke et al., 2007).

It is not clear whether all spirochetal lineages present in the *Nasutitermes* gut are involved in reductive acetogenesis. Furthermore, not all *fhs* genes obtained by Warnecke et al. (2007) are clustered with treponemal sequences. It is possible that also members of the *Ruminococaceae* (Fig. 5) contribute to reductive acetogenesis in the hindgut because many *Ruminococcus* species are homoacetogenic (33, 54). The same argument can be made for the *Holophagaceae* (*Acidobacteria*, Fig. 5) present in all gut compartments of *N. corniger* and the hindgut sample of *N. takasagoensis*, which are closely related to the homoacetogenic *Holophaga foetida* (Liesack et al., 1994).

In view of the large surface-to-volume ratios of small guts (Brune, 1998), the gut wall emerges as an important microhabitat. Methanogenic archaea associated with the gut wall of lower termites have been implicated as a hydrogen sink, both in methanogenesis and owing to their capacity for hydrogen-dependent reduction of inflowing oxygen (Tholen et al., 2004). However, methanogenesis is not as important in *Nasutitermes* spp. as in other termite species (see Brune (2010) and references therein). The situation is a bit more ambiguous in the case of sulfate-reducing microorganisms. About 1% of the sequences in the P3 (*Desulfovibrio* 1, Table S1) represent sulfate-reducing *Deltaproteobacteria* related to *Desulfovibrio intestinalis*, which—like other *Desulfovibrio* spp. isolated from termite guts—exhibits high rates of hydrogen-dependent oxygen reduction (Fröhlich et al., 1999; Kuhnigk et al., 1996). However, it is not known whether the *Desulfovibrio* spp. in *Nasutitermes* are located at the hindgut wall.

Clearly, the radial organization of the microbiota and the location of individual populations with respect to the oxygen gradient are important issues. Although the 454 datasets of the gut sections do not contain direct information about the localization of microorganisms within the respective compartments, the obvious absence of some bacterial groups from the purely luminal sample of the P3 gut compartment (Warnecke et al., 2007) compared to the total P3 sample, allows us to make some careful inferences regarding peripheral localization. Among the possible gut-wall colonizers are *Sanguibacter* populations, a genus comprising aerobic and facultatively anaerobic isolates (e.g. Huang et al., 2005), and other unclassified lineages of *Actinobacteria*. Also several lineages of *Trinervitermes* Cluster A and several other termite-specific *Spirochaetaceae* groups (Table S1) are abundant in the total P3 sample of *N. corniger* but absent from the luminal sample of *Nasutitermes* sp. (Warnecke et al., 2007), which is in accordance with previous reports of an attachment of spirochetes to the gut wall of lower and higher termites (Czolij et al., 1985; Nakajima et al., 2005). An association with the gut

wall of many lower termites has been documented also for relatives of the *Bacteroidales* Cluster V (Nakajima et al., 2006), a group that is abundantly encountered also in gut homogenates of *Nasutitermes* spp. (Hongoh et al., 2006; this study). The frequent association of Cluster-V *Bacteroidales* also with the surface of cellulolytic protists in lower termites (Noda et al., 2006) suggests that the need for attachment is a strategy to prevent wash out. The presence of oxygen-removing mechanisms among obligate anaerobes, necessary for the colonization of the microoxic gut periphery, has been previously documented for *Methanobrevibacter* species in lower termites (Tholen et al., 2007), and is encountered also among the *Bacteroidales* (Baughn and Malamy, 2004).

**The posterior hindgut.** Microbial cell counts decrease by two orders of magnitude and cell density drops 50-fold between the P3 and P4 section (Table 1), which suggested that the microbial biomass produced in the P3 is digested after being transported into the posterior hindgut with the flow of the digesta. The distinct differences in community structure between the P3 and P4 sections (Fig. 5) indicate the presence of a microbiota specifically adapted to the environment of the posterior hindgut. Most obvious is the increase in relative abundance of *Acidobacteriaceae* and *Coriobacteriaceae* (Table S1), but also specific lineages of *Lentisphaerae* and members of candidate divisions OP11, SR1, and TM7 are enriched in the posterior hindgut. Distinct changes in diversity and community structure have been observed also between the alkaline P3 and the neutral P4 compartments of soil-feeding *Cubitermes* spp. (Schmitt-Wagner 2003a and 2003b). Since both gut regions are neutral in *Nasutitermes* spp. (Brune et al. 1995; this study), it is likely that factors other than pH are responsible for this shift. Rather, the forces driving community structure could be the increasing influence of oxygen in the tubular P4. The slightly acidic pH of the P5 compartment observed in *N. corniger* has also been found among several soil-feeding species (pH 5–6; Brune and K uhl, 1996).

**Crop and midgut.** Since sound wood is a highly nitrogen-deficient diet, termites have developed the strategy to exploit the assimilatory capacities of their gut microbiota to acquire essential amino acids and vitamins (Brune and Ohkuma, 2011). This is accomplished by digesting microbial biomass derived of the hindgut contents—either by coprophagy or by proctodeal trophallaxis. Although little is known about the behavior of *Nasutitermes* species, numerous similarities in the community patterns of the rectum (P5) and the anterior gut (crop) suggest that fecal material is consumed by the termites (Fig. 5). This agrees with observations of proctodeal feeding in *N. corniger* and other species (R.H.S., unpublished). The strong shift in the bacterial community profiles (Fig. 5) and the reduction of microbial density between crop and midgut indicate that bacteria are digested in the midgut, which is in agreement with the presence of lysozyme and protease activities in this gut region (Fujita and Abe, 2002). The community of the midgut is dominated by *Firmicutes*, particularly members of *Lachnospiraceae* (Fig. 5), which represent a lineage of uncultivated bacteria from intestinal environments, including termite guts ("uncultured 67" group, Table S1). The family *Lachnospiraceae* comprises many species with high proteolytic, xylanolytic, and also cellulolytic activities (e.g., *Butyrivibrio* and *Pseudobutyrvibrio* spp.; Cotta and Forster, 2006), but it remains to be clarified whether these bacteria contribute to the digestive capacities of the midgut of *Nasutitermes* spp. (Tokuda et al., 2012) or whether they are simply transient and inactive forms (e.g., spores) of bacteria residing in other gut regions.

**Alkaline gut regions.** The anterior hindgut of many higher termites is highly alkaline (Bignell and Egelton, 1995). In soil-feeding *Termitinae*, the pH increases sharply in the mixed segment and reaches its maximum (pH > 12) in the P1 compartment (Brune and Köhl, 1996). The alkalinity in the tubular P1 of *N. corniger* is considerably less pronounced (pH 10), and the pH returns to neutral already in the P3, a compartment that remains strongly alkaline in the soil feeders. Since the profiles obtained for *N. corniger* (this study) were almost identical to previous profiles of *N. nigriceps* (Brune et al., 1995), it seems safe to conclude that they are typical at least for the wood-feeding members of this genus.

A prevalent bacterial lineage in the alkaline P1 section of *N. corniger* is the genus *Turicibacter* (*Firmicutes*, *Erysipelotrichaceae*, 13%, Table S1); their relative abundance in all other compartments is less than 1%. Representatives of this cluster have also been detected in the (putatively) alkaline P1 regions of the soil-feeding *Pericapritermes latignathus* and a grass-feeding *Speculitermes* sp. (Thongaram et al., 2005), and in the alkaline midgut of the humivorous larva of the scarab beetle *Pachnoda ephippiata* (Egert et al., 2003), which indicates an adaptation to high pH. The occurrence of *Turicibacter* spp. in the gut of *Microcerotermes* sp. (Hongoh et al., 2005), which also comprises an alkaline P1 (Brune et al., 1995), is in agreement with this assumption. However, alkaliphily is not a typical trait for the whole genus; the next cultured relative, *Turicibacter sanguinis*, does not grow above pH 8 (Bosshard et al., 2002), and members of the *Turicibacter* clade are also present in mammals (Ley et al., 2008), which lack a highly alkaline gut.

Other bacterial groups prevailing in the P1 section are several lineages of *Lactobacillales* (*Firmicutes*). Sequences of cluster PeH08 (5% relative abundance) have been obtained from the alkaline compartments of other higher termites (Thongaram et al., 2005) and beetle larvae (Egert et al., 2003), but were also encountered in the posterior hindgut compartments of *N. corniger* (Fig. 5; Table S1). The P1 section also contains a small number of sequences (1%) from a termite-specific lineage of *Lachnospiraceae* that is related to the sequences NT-1 and NT-2, which have been previously assigned to rod-shaped bacteria predominantly colonizing the mixed segment of *N. takasagoensis* (Tokuda et al., 2000). The presence of these bacteria in this sample is explained by the inclusion of the mixed segment in the P1 section, which was not separated for technical reasons.

**Pyrotag sequencing of termite gut microbiota.** The diversity and community structure of the bacterial gut microbiota of termites has been addressed by numerous studies (see Ohkuma and Brune, 2011). Many of the more detailed analyses combined Sanger sequencing of 16S rRNA genes with T-RFLP profiling, thus compensating for the shortcomings of the individual approaches (Sanger sequencing is notoriously undersampled and T-RFLP analyses lack phylogenetic resolution). The application of high-throughput sequencing techniques in targeting the 16S rRNA gene opened a new dimension of detecting even very low-abundant, so-far undetected microorganisms in microbial ecology studies (Sogin et al., 2006).

As in all PCR-based approaches, primer bias is an important issue. Our pyrotag sequencing data provided a good coverage of all lineages previously discovered by Sanger sequencing (Warnecke et al., 2007), which documented that the primers for the V3–V4 region did not introduce a serious bias over the 27F–1492R primers. The striking difference in the abundance of *Fibrobacteres* sequences between the clone library (Hongoh et al., 2006) and pyrotag library (this study) of *N. takasagoensis* (Fig. 5) may be rooted in the different batches of termites used in the respective studies.

The results of our study underline that a comprehensive and well-curated database is crucial for a reliable sequence assignment (Werner et al., 2011). Refining the classification of the SILVA database by the introduction of additional, termite-specific groups significantly improved the assignment of

pyrotag reads especially at lower taxonomic levels (Table 2). This revealed the presence of rare taxa that were not discovered in previous, clone-based studies of *Nasutitermes* spp. An example are the sequences related to the fat-body-colonizing *Blattabacterium* (*Flavobacteria*, 0.5–0.9% relative abundance), which so far had been detected only in cockroaches and the primitive termite *Mastotermes darwiniensis* (Lo et al., 2003), and whose presence in higher termites requires further analysis.

The high resolution and fast sample treatment using pyrotag sequencing provides a perfect tool for community profiling, combining the virtues of fingerprinting approaches with the benefit of exact taxonomic classification. The large sampling depths of pyrotag sequencing also decreases the detection limit of the analysis, which will help to investigate the existence of a core microbiota and other important questions concerning the evolution of the termite gut microbiota from a putative dictyopteran ancestor.

## **Acknowledgements**

This study was supported by the Max Planck Society. Tim Köhler received a doctoral fellowship from the International Max Planck Research School for Environmental, Cellular, and Molecular Microbiology. We thank Gaku Tokuda for providing termites, and Yuichi Hongoh and Moriya Ohkuma for sharing unpublished details of a previous study.



## References

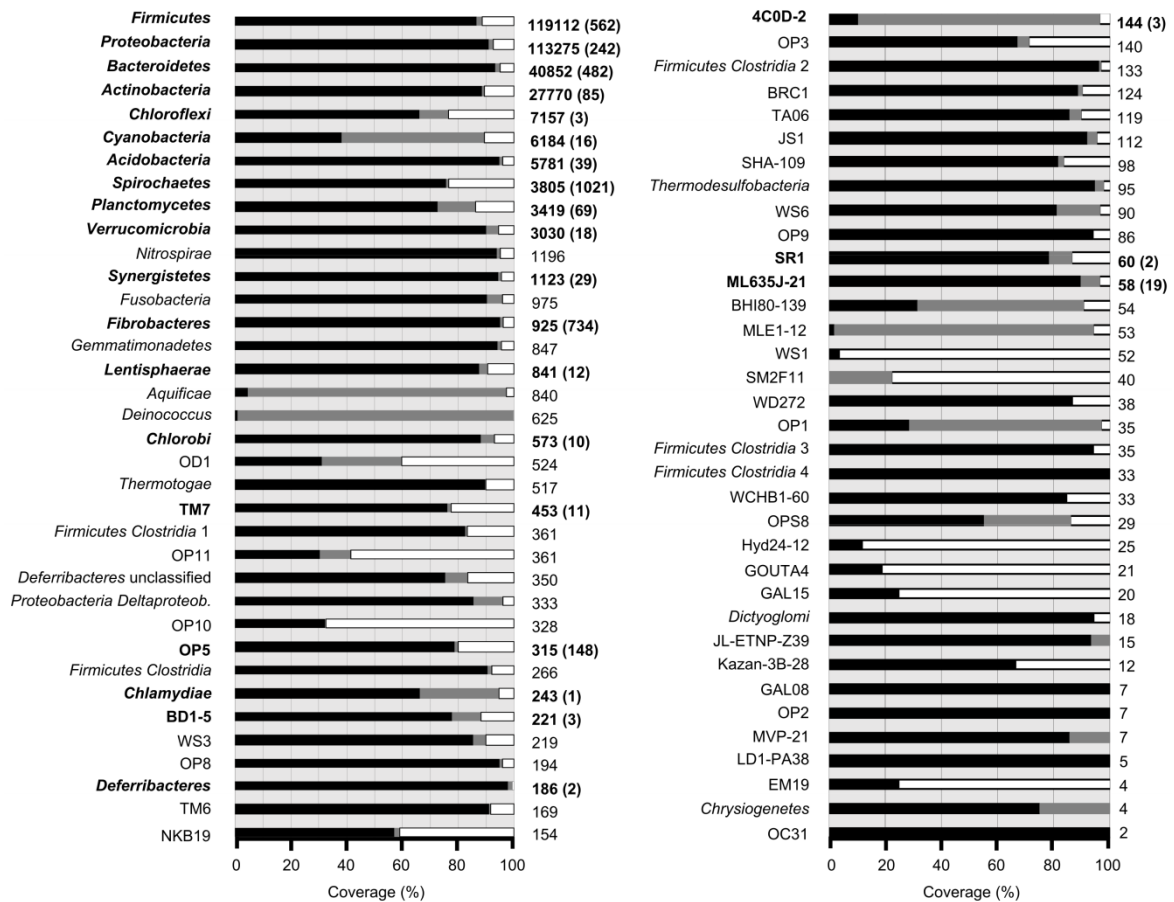
- Ballor NR and Leadbetter JR (2011).** Analysis of extensive [FeFe] hydrogenase gene diversity within the gut microbiota of insects representing five families of *Dictyoptera*. *Microb. Ecol.* doi:10.1007/s00248-011-9941-5.
- Ballor NR, Paulsen I and Leadbetter JR (2011).** Genomic analysis reveals multiple [FeFe] hydrogenases and hydrogen sensors encoded by treponemes from the H<sub>2</sub>-rich termite gut. *Microb. Ecol.* doi:10.1007/s00248-011-9922-8.
- Baughn AD and Malamy MH (2004).** The strict anaerobe *Bacteroides fragilis* grows in and benefits from nanomolar concentrations of oxygen. *Nature* 427:441–444.
- Bignell DE (2011).** Morphology, physiology, biochemistry and functional design of the termite gut: an evolutionary wonderland. p 375–412. *In* Bignell DE, Roisin Y, Lo N (ed), *Biology of Termites: A Modern Synthesis*, Springer, Dordrecht.
- Bignell DE and Eggleton P (1995).** On the elevated intestinal pH of higher termites (Isoptera: Termitidae). *Insect. Soc.* 42:57–69.
- Bosshard PP, Zbinden R and Altwegg M (2002).** *Turicibacter sanguinis* gen. nov., sp. nov., a novel anaerobic, Gram-positive bacterium. *Int. J. Syst. Evol. Microbiol.* 52:1263–1266.
- Brauman A, Kane MD, Labat M and Breznak JA (1992).** Genesis of acetate and methane by gut bacteria of nutritionally diverse termites. *Science* 257:1384–1387.
- Bray JR and Curtis JT (1957).** An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monogr.* 27:325–349.
- Brune A, Emerson D and Breznak JA (1995).** The termite gut microflora as an oxygen sink: microelectrode determination of oxygen and pH gradients in guts of lower and higher termites. *Appl. Environ. Microbiol.* 61:2681–2687.
- Brune A (1998).** Termite guts: the world's smallest bioreactors. *Trends Biotechnol.* 16:16–21.
- Brune A (2010).** Methanogenesis in the digestive tracts of insects. p 707–728. *In* Timmis KN (ed), *Handbook of Hydrocarbon and Lipid Microbiology*, vol 8. Springer, Heidelberg.
- Brune A and Kühl M (1996).** pH profiles of the extremely alkaline hindguts of soil-feeding termites (Isoptera: Termitidae) determined with microelectrodes. *J. Insect Physiol.* 42:1121–1127.
- Brune A and Ohkuma M (2011).** Role of the termite gut microbiota in symbiotic digestion. p 439–475. *In* Bignell DE, Roisin Y., Lo N (ed) *Biology of Termites: A Modern Synthesis*, Springer, Dordrecht.
- Chao A (1984).** Nonparametric estimation of the number of classes in a population. *Scan. J. Stat.* 11:265–270.
- Chao A and Shen TJ (2003).** Nonparametric estimation of Shannon's index of diversity when there are unseen species in sample. *Environ. Ecol. Stat.* 10:429–443.
- Cotta M and Forster R (2006).** The family Lachnospiraceae, including the genera *Butyrivibrio*, *Lachnospira* and *Roseburia*. p 1002–1021 *In* Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E (ed), *The Prokaryotes*, 3rd ed, vol 4. Springer, New York.
- Czolij R, Slaytor M and O'Brien RW (1985).** Bacterial flora of the mixed segment and the hindgut of the higher termite *Nasutitermes exitiosus* Hill (Termitidae, Nasutitermitinae). *Appl. Environ. Microbiol.* 49:1226–1236.
- Ebert A and Brune A (1997).** Hydrogen concentration profiles at the oxic-anoxic interface: a microsensor study of the hindgut of the wood-feeding lower termite *Reticulitermes flavipes* (Kollar). *Appl. Environ. Microbiol.* 63:4039–4046.
- Egert M, Wagner B, Lemke T, Brune A and Friedrich MW (2003).** Microbial community structure in midgut and hindgut of the humus-feeding larva of *Pachnoda ephippiata* (Coleoptera: Scarabaeidae). *Appl. Environ. Microbiol.* 69:6659–6668.
- Engelbrektson A, Kunin V, Wrighton KC, Zvenigorodsky N, Chen F, Ochman H, Hugenholtz P (2010).** Experimental factors affecting PCR-based estimates of microbial species richness and evenness. *Environ. Microbiol.* 4:642–647.
- Friedrich MW, Schmitt-Wagner D, Lueders T, Brune A (2001).** Axial differences in community structure of *Crenarchaeota* and *Euryarchaeota* in the highly compartmentalized gut of the soil-feeding termite *Cubitermes orthognathus*. *Appl. Environ. Microbiol.* 67:4880–4890.
- Fröhlich J, Sass H, Babenzien HD, Kuhnigk T, Varma A, Saxena S, Nalepa C, Pfeiffer P, König H (1999).** Isolation of *Desulfovibrio intestinalis* sp. nov. from the hindgut of the lower termite *Mastotermes darwiniensis*. *Can. J. Microbiol.* 45:145–152.

- Fujita A and Abe T (2002).** Amino acid concentration and distribution of lysozyme and protease activities in the guts of higher termites. *Physiol. Entomol.* 27:76–78.
- Grabner JR, Leadbetter JR and Breznak JA (2004).** Description of *Treponema azotonutricium* sp. nov. and *Treponema primitia* sp. nov., the first spirochetes isolated from termite guts. *Appl. Environ. Microbiol.* 70:1315–1320.
- Henckel T, Friedrich M and Conrad R (1999).** Molecular analyses of the methane-oxidizing microbial community in rice field soil by targeting the genes of the 16S rRNA, particulate methane monooxygenase, and methanol dehydrogenase. *Appl. Environ. Microbiol.* 65:1980–1990.
- Hongoh Y, Deevong P, Hattori S, Inoue T, Noda S, Noparatnaraporn N, Kudo T and Ohkuma M (2006).** Phylogenetic diversity, localization and cell morphologies of the candidate phylum TG3 and a subphylum in the phylum *Fibrobacteres*, recently found bacterial groups dominant in termite guts. *Appl. Environ. Microbiol.* 72:6780–6788.
- Hongoh Y, Deevong P, Inoue T, Moriya S, Trakulnaleamsai S, Ohkuma M, Vongkaluang C, Noparatnaraporn N and Kudo T (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.
- Huang Y, Dai X, He L, Wang YN, Wang BJ, Liu Z, Liu SJ (2005).** *Sanguibacter marinus* sp. nov., isolated from coastal sediment. *Int. J. Syst. Evol. Microbiol.* 55:1755–1758.
- Hugenholtz P and Goebel BM (2001).** The polymerase chain reaction as a tool to investigate microbial diversity in environmental samples. p 31–41. *In* Rochelle PA (ed), *Environmental molecular microbiology: Protocols and applications*, Horizon Press Inc., New York.
- Jouquet P, Traoré S, Choosai C, Hartmann C and Bignell D (2011).** Influence of termites on ecosystem functioning. Ecosystem services provided by termites. *Eur. J. Soil Biol.* 47:215–222.
- Kuhnigk T, Branke J, Krekeler D, Cypionka H and König H (1996).** A feasible role of sulfate-reducing bacteria in the termite gut. *System. Appl. Microbiol.* 19:139–149.
- Leadbetter JR, Schmidt TM, Grabner JR and Breznak JA (1999).** Acetogenesis from H<sub>2</sub> plus CO<sub>2</sub> by spirochetes from termite guts. *Science* 283:686–689.
- Leaphart AB and Lovell CR (2001).** Recovery and analysis of formyltetrahydrofolate synthetase gene sequences from natural populations of acetogenic bacteria. *Appl. Environ. Microbiol.* 67:1392–1395.
- Legendre P and Legendre L (1998).** *Numerical Ecology* 2nd ed. Elsevier Amsterdam, Montreal (Quebec).
- Leschine S, Paster B and Canale-Parola E (2006).** Free-living saccharolytic spirochetes: the genus *Spirochaeta*. p 195–210. *In* Dworkin M, Falkow S, Rosenberg E, Schleifer KH, Stackebrandt E (ed) *The Prokaryotes*, 3rd ed, vol 7. Springer, New York.
- Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, Schlegel ML, Tucker TA, Schrenzel MD, Knight R and Gordon JI (2008).** Evolution of mammals and their gut microbes. *Science* 320:1647–1651.
- Liesack W, Bak F, Kreft JU and Stackebrandt E (1994).** *Holophaga foetida* gen. nov., sp. nov., a new, homoacetogenic bacterium degrading methoxylated aromatic compounds. *Arch. Microbiol.* 162:85–90.
- Lo N, Bandi C, Watanabe H, Nalepa C and Beninati T (2003).** Evidence for co-cladogenesis between diverse dictyopteran lineages and their intracellular endosymbionts. *Mol. Biol. Evol.* 20:907–913.
- Ludwig W, Strunk O, Westram R, Richter L, Meier H, Yadhukumar, Buchner A, Lai T, Steppi S, Jobb G, Förster W, Brettske I, Gerber S, Ginhart AW, Gross O, Grumann S, Hermann S, Jost R, König A, Liss T, Lüssmann R, May M, Nonhoff B, Reichel B, Strehlow R, Stamatakis A, Stuckmann N, Vilbig A, Lenke M, Ludwig T, Bode A and Schleifer KH (2004).** ARB: a software environment for sequence data. *Nucleic Acids Res.* 32:1363–1371.
- Muyzer G, de Waal EC and Uitterlinden AG (1993).** Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl. Environ. Microbiol.* 59:695–700.
- Nakajima H, Hongoh Y, Usamib R, Kudo T and 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.
- Nakajima H, Hongoh Y, Noda S, Yoshida Y, Usami R, Kudo T and Ohkuma M (2006).** Phylogenetic and morphological diversity of *Bacteroidales* members associated with the gut wall of termites. *Biosci. Biotechnol. Biochem.* 70:211–218.

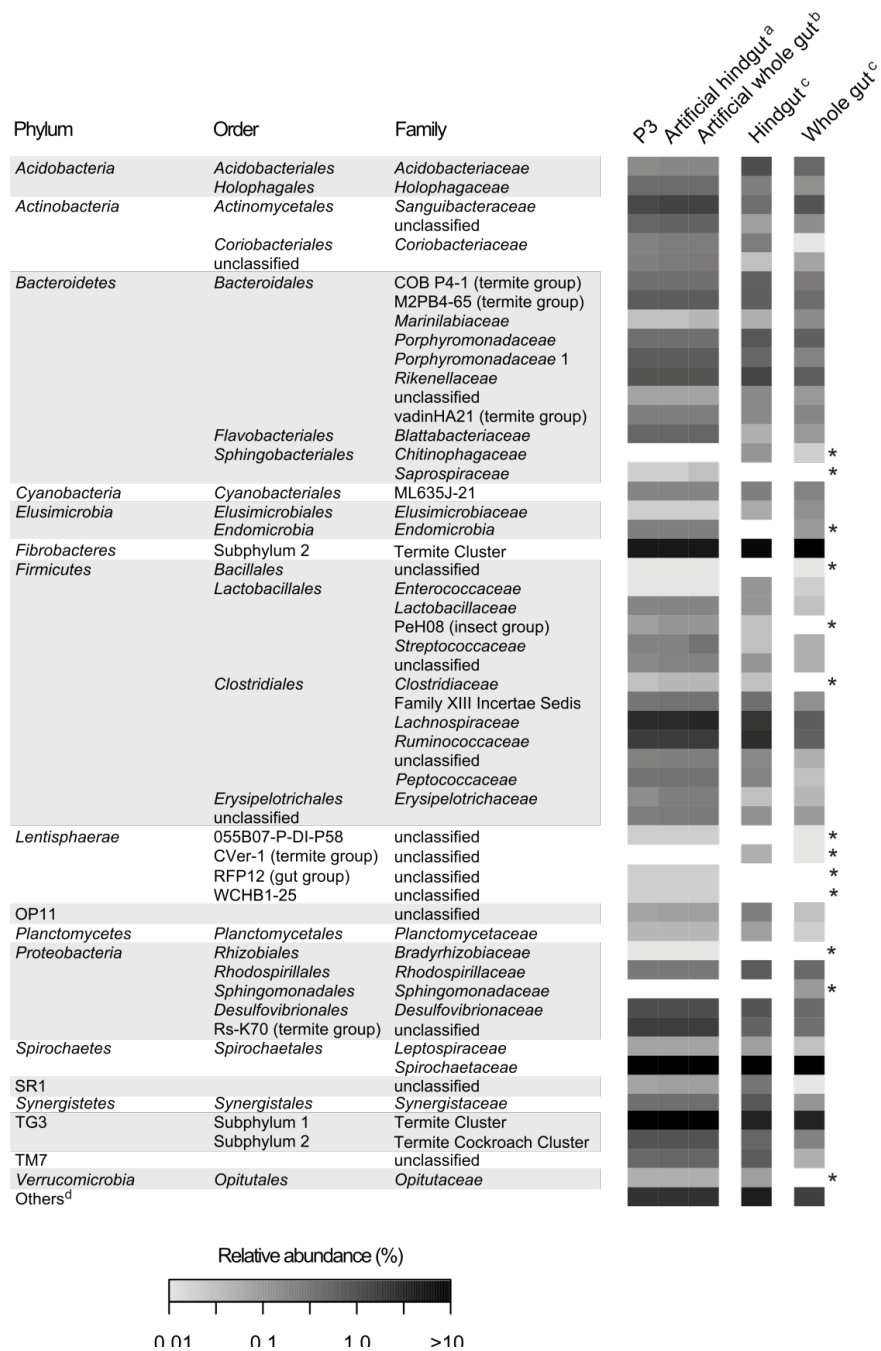
- Noda S, Inoue T, Hongoh Y, Nalepa CA, Vongkaluang C, Kudo T and Ohkuma M (2006). Identification and characterization of ectosymbionts of distinct lineages in *Bacteroidales* attached to flagellated protists in the gut of termites and a wood-feeding cockroach. *Environ. Microbiol.* 8:11–20.
- Noirot C (2001). The gut of termites (Isoptera). Comparative anatomy, systematics, phylogeny. II. Higher termites (Termitidae). *Ann. Soc. Entomol. Fr. (N.S.)* 37:431–471.
- Ohkuma M and Brune A (2011). Diversity, structure, and evolution of the termite gut microbial community. p 413–438 *In* Bignell DE, Roisin Y, Lo N (ed), *Biology of Termites: A Modern Synthesis*, Springer, Dordrecht.
- Ohkuma M, Iida T and Kudo T (1999). Phylogenetic relationships of symbiotic spirochetes in the gut of diverse termites. *FEMS Microbiol. Lett.* 181:123–129.
- Ottesen EA and Leadbetter JR (2011). Formyltetrahydrofolate synthetase gene diversity in the guts of higher termites with different diets and lifestyles. *Appl. Environ. Microbiol.* 77:3461–3467.
- Paster BJ, Dewhirst FE, Cooke SM, Fussing V, Poulsen LK and Breznak JA (1996). Phylogeny of not-yet-cultured spirochetes from termite guts. *Appl. Environ. Microbiol.* 62:347–352.
- Pernthaler A, Pernthaler J and Amann R (2004). Sensitive multi-color fluorescence *in situ* hybridization for the identification of environmental microorganisms. p 711–726. *In* Kowalchuk GA, de Bruijn FJ, Head IM, Akkermans ADL, van Elsas JD (ed), *Molecular Microbial Ecology Manual*, 2nd ed, vol 1. Kluwer Academic Publisher, Dordrecht.
- Pernthaler J, Glöckner FO, Schönhuber W and Amann R (2001). Fluorescence *in situ* hybridization with rRNA-targeted oligonucleotide probes. p 207–226. *In* Paul J (ed), *Methods in Microbiology: Marine Microbiology*. vol 30. Academic Press, London.
- Pester M and Brune A (2006). Expression profiles of *fhs* (FTHFS) genes support the hypothesis that spirochaetes dominate reductive acetogenesis in the hindgut of lower termites. *Environ. Microbiol.* 8:1261–1270.
- Pester M and Brune A (2007). Hydrogen is the central free intermediate during lignocellulose degradation by termite gut symbionts. *ISME J.* 1:551–565.
- Pruesse E, Quast C, Knittel K, Fuchs B, Ludwig W, Peplies J and Glöckner FO (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 35:7188–7196.
- Rieu-Lesme F, Morvan B, Collins MD, Fonty G and Willems A (1996). A new H<sub>2</sub>/CO<sub>2</sub>-using acetogenic bacterium from the rumen: Description of *Ruminococcus schinkii* sp. nov. *FEMS Microbiol. Lett.* 140:281–286.
- Salmassi TM and Leadbetter JR (2003). Molecular aspects of CO<sub>2</sub>-reductive acetogenesis in cultivated spirochetes and the gut community of the termite *Zootermopsis angusticollis*. *Microbiology* 149:2529–2537.
- Scheffrahn RH, Cabrera BJ, Kern JRW and Su NY (2002). *Nasutitermes costalis* (Isoptera: Termitidae) in Florida: first record of a non-endemic establishment by a higher termite. *Florida Entomol.* 85:273–275.
- Scheffrahn RH, Krecke J, Szalanski AL and Austin JW (2005). Synonymy of the neotropical arboreal termites *Nasutitermes corniger* and *N. costalis* (Isoptera: Termitidae: Nasutitermitinae), with evidence from morphology, genetics, and biogeography. *Ann. Entomol. Soc. Am.* 98:273–281.
- Schloss PD, Gevers D and Westcott SL (2011). Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS ONE* 6:e27310.
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ and Weber CF (2009). Introducing *mothur*: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75:7537–7541.
- Schmitt-Wagner D and Brune A (1999). Hydrogen profiles and localization of methanogenic activities in the highly compartmentalized hindgut of soil-feeding higher termites (*Cubitermes* spp.). *Appl. Environ. Microbiol.* 65:4490–4496.
- Schmitt-Wagner D, Friedrich MW, Wagner B and Brune A (2003a). Phylogenetic diversity, abundance, and axial distribution of bacteria in the intestinal tract of two soil-feeding termites (*Cubitermes* spp.). *Appl. Environ. Microbiol.* 69:6007–6017.
- Schmitt-Wagner D, Friedrich MW, Wagner B and Brune A (2003b). Axial dynamics, stability, and interspecies similarity of bacterial community structure in the highly compartmentalized gut of soil-feeding termites (*Cubitermes* spp.). *Appl. Environ. Microbiol.* 69:6018–6024.

- Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, Neal PR, Arrieta JM and Herndl GJ (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proc. Natl. Acad. Sci. USA* 103:12115–12120.
- Suen G, Weimer PJ, Stevenson DM, Aylward FO, Boyum J, Deneke J, Drinkwater C, Ivanova NN, Mikhailova N, Chertkov O, Goodwin LA, Currie CR, Mead D and Brumm PJ (2011). The complete genome sequence of *Fibrobacter succinogenes* S85 reveals a cellulolytic and metabolic specialist. *PLoS ONE* 6:e18814.
- Sugimoto A, Inoue T, Tayasu I, Miller L, Takeichi S and Abe T (1998). Methane and hydrogen production in a termite-symbiont system. *Ecol. Res.* 13:241–257.
- Tholen A and Brune A (2000). Impact of oxygen on metabolic fluxes and *in situ* rates of reductive acetogenesis in the hindgut of the wood-feeding termite *Reticulitermes flavipes*. *Environ. Microbiol.* 2:436–449.
- Tholen A, Pester M and Brune A (2007). Simultaneous methanogenesis and oxygen reduction by *Methanobrevibacter cuticularis* at low oxygen fluxes. *FEMS Microbiol. Ecol.* 62:303–312.
- Thongaram T, Hongoh Y, Kosono S, Ohkuma M, Trakulnaleamsai S, Noparatnaraporn N and Kudo T (2005). Comparison of bacterial communities in the alkaline gut segment among various species of higher termites. *Extremophiles* 9:229–238.
- Tokuda G, Watanabe H, Hojo M, Fujita A, Makiya H, Miyagia M, Arakawa G and Arioka M (2012). Cellulolytic environment in the midgut of the wood-feeding higher termite *Nasutitermes takasagoensis*. *J. Ins. Physiol.* 58:147–154.
- Tokuda G and Watanabe H (2007). Hidden cellulases in termites: revision of an old hypothesis. *Biol. Lett.* 3:336–339.
- Tokuda G, Yamaoka I and Noda H (2000). Localization of symbiotic clostridia in the mixed segment of the termite *Nasutitermes takasagoensis* (Shiraki). *Appl. Environ. Microbiol.* 66:2199–2207.
- Wang Q, Garrity GM, Tiedje JM and Cole JR (2007). Naive Bayesian Classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73:5261–5267.
- Warnecke F, Luginbühl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, Cayouette M, McHardy AC, Djordjevic G, Aboushadi N, Sorek R, Tringe SG, Podar M, Martin HG, Kunin V, Dalevi D, Madejska J, Kirton E, Platt D, Szeto E, Salamov A, Barry K, Mikhailova N, Kyrpides NC, Matson EG, Ottosen EA, Zhang X, Hernández M, Murillo C, Acosta LG, Rigoutsos I, Tamayo G, Green BD, Chang C, Rubin EM, Mathur EJ, Robertson DE, Hugenholtz P, Leadbetter JR (2007). Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* 450:560–565.
- Watanabe H and Tokuda G (2010). Cellulolytic systems in insects. *Annu. Rev. Entomol.* 55:609–632.
- Werner JJ, Koren O, Hugenholtz P, Desantis TZ, Walters WA, Caporaso JG, Angenent LT, Knight R and Ley RE (2011). Impact of training sets on classification of high-throughput bacterial 16S rRNA gene surveys. *ISME J.* 6:94–103.

## Supplementary Material



**Fig. S1 | Phylum-level coverage of the improved primer set used for 454 sequencing of the V3–V4 region based on all bacterial sequences > 1,200 bp in the SILVA 100 database.** Phyla containing clones from dictyopteran guts are marked in bold. The bars indicate the proportion of sequences without (black), with one (gray), and with two or more (white) mismatches and are followed by the total number of sequences in the dataset. Numbers of sequences obtained from dictyopteran guts are shown in parentheses.



**Fig. S2 | Comparison of the P3 compartment, hindgut, and whole gut of *Nasutitermes corniger*.** The artificial hindgut is calculated by the relative abundance of the P1 to P5 hindgut sections multiplied by the cell number of each section. All samples were from different batches of the same nest. To increase the sensitivity for low-abundant groups, the shading uses a logarithmic scale. Gut sections, hindgut, and whole gut were each derived from different batches of termites. <sup>a</sup> Calculated from all hindgut sections (P1–P5); <sup>b</sup> calculated from all gut sections; <sup>c</sup> derived from different batches of termites; <sup>d</sup> 111 to 157 remaining families, each representing less than 1% of the respective community (for details, see Table S1); \* taxa missing in one or more samples.

**Tab. S1 | Relative abundance of all genus-level groups in the gut sections of *Nasutitermes corniger* as xls file.** Please download from: <https://dx.doi.org/10.1128/AEM.00683-12>.

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

Carsten Dietrich\*, Tim Köhler\* and Andreas Brune

\* These authors contributed equally | *Affiliations*: Max Planck Institute for Terrestrial Microbiology, 35043 Marburg, Germany. | *This chapter is published in*: Dietrich C, Köhler T and Brune A (2014). The cockroach origin of the termite gut microbiota: patterns in bacterial community structure reflect major evolutionary events. *Appl. Environ. Microbiol.* 80:2261-2269 | *Contributions*: C.D. designed and carried out experiments, analyzed data and wrote the manuscript. T.K. designed and carried out experiments, analyzed data and contributed to the manuscript. A.B. conceived the study, wrote the manuscript, and secured funding.

### Abstract

---

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 (Brune and Ohkuma, 2011). The ability to mineralize lignocellulose and humus lends termites an important place in carbon and nitrogen cycling in tropical soils (Jouquet et al., 2011) and makes them promising models for the industrial conversion of lignocellulose into microbial products and the production of biofuels (Ni and Tokuda, 2013).

The ancestors of termites were presumably detritivorous, subsocial cockroaches (Nalepa et al., 2001; Inward et al., 2007). About 130 million years ago, they gained the ability to digest wood through acquisition of cellulolytic flagellates (Engel et al., 2009; Ohkuma et al., 2009). 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 (Brune and Ohkuma, 2011; Hongoh, 2011). 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 (Bignell and Eggleton, 2000; Engel et al., 2009). While the *Macrotermitinae* established a unique symbiosis with a lignocellulolytic fungus (Nobre et al., 2011), other lineages of higher termites started to exploit diets of increasing humification, a development accompanied by further differentiation of the hindgut (Bignell and Eggleton, 2000) and its entirely prokaryotic microbiota (Hongoh, 2011; Ohkuma and Brune, 2011).

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 (Brune and Ohkuma, 2011; Hongoh, 2011). Most importantly, the origin of the bacteria colonizing the hindgut bioreactor is entirely in the dark (Ohkuma and Brune, 2011).

Although the gut microbiota differs substantially between termite species, it comprises many phylogenetic clusters that are unique to termites (e.g. Lilburn et al., 1999; Hongoh et al., 2006a; Hongoh et al., 2006b). The origin of these lineages remains unclear, but their detection also in the gut of several cockroaches (Ohkuma et al., 2007; Geissinger et al., 2009; Schauer et al. 2012), the closest relatives of termites, together with occasional evidence of cocladogenesis with the termite hosts (Hongoh et al., 2005; Noda et al., 2009) 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 (Köhler et al., 2012). 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 (Schauer et al., 2012; Köhler et al. 2012). 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 (Deevong et al., 2006). 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 (Pester and Brune, 2006). 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 (Paul et al., 2012). 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 (Köhler et al., 2012). 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 (Schloss et al., 2009) (version 1.27.0) using stringent conditions (Schloss et al., 2011) (reads > 200 bp; no ambiguous bases, maximum number of homopolymers ≤ 8). The sequences in each sample were denoised with the *Acacia* program (Bragg et al., 2012) 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 (Köhler et al., 2012; 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 phylogenies of relevant groups (e.g. Hongoh, 2006a; Hongoh 2006b; Thompson et al., 2012) 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.

**Tab. 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.

Sample	Host species	Origin <sup>a</sup>	Total reads	Genus-level taxa	OTUs (3% dissim.)	Coverage (%) <sup>b</sup>	Diversity indices <sup>c</sup> (based on OTUs)			NCBI accession number <sup>d</sup>
							Richness	Diversity	Evenness	
<b>Cockroaches</b>										
1	<i>Ergaula capucina</i>	B1	6,020	232	891	70.5	1,266	5.52	0.84	068
2	<i>Symploce macroptera</i>	B1	5,045	135	499	80.9	431	4.70	0.82	069
3	<i>Rhyparobia maderae</i>	B1	12,164	268	1346	70.8	2,540	5.42	0.78	070
4	<i>Elliptorhina chopardi</i>	B1	6,794	200	663	79.6	798	5.25	0.83	071
5	<i>Panchlora</i> sp.	B1	11,889	212	2042	66.6	1,064	4.41	0.72	072
6	<i>Diploptera punctata</i>	B1	5,708	161	543	80.8	627	4.93	0.82	073
7	<i>Opisthopteria orientalis</i>	B1	11,707	291	1515	70.5	3,153	5.72	0.80	074
8	<i>Panesthia angustipennis</i>	B1	5,394	202	1141	72.5	1,710	6.01	0.88	075
9	<i>Salganea esakii</i>	B1	17,412	296	1916	80.8	2,955	6.27	0.84	076
10	<i>Eublabeus posticus</i>	B1	103,530	416	5743	79.9	12,034	5.34	0.64	077
11	<i>Schultesia lampyridiformis</i>	B1	5,085	217	857	70.3	1,482	5.42	0.83	078
12	<i>Eurycotis floridana</i>	B1	41,336	354	3410	77.2	6,855	5.80	0.75	079
13	<i>Shelfordella lateralis</i>	B1	6,226	186	714	82.4	674	5.30	0.86	080
14	<i>Blatta orientalis</i>	B1	8,024	246	1069	68.6	2,045	5.14	0.76	081
15	<i>Cryptocercus punctulatus</i>	F1	6,715	180	715	75.5	884	4.90	0.78	082
<b>Lower termites</b>										
16	<i>Mastotermes darwiniensis</i>	L1	7,596	137	398	86.3	583	3.94	0.68	083
17	<i>Zootermopsis nevadensis</i>	L2	6,129	278	1617	72.6	3,451	5.18	0.72	084
18	<i>Hodotermopsis sjoestedti</i>	L1	7,600	272	1584	73.9	3,569	5.25	0.73	085
19	<i>Hodotermes mossambicus</i>	F2	16,520	204	978	74.3	1,840	5.33	0.79	086
20	<i>Incisitermes marginipennis</i>	L1	16,491	299	2807	79.0	6,354	4.27	0.56	087
21	<i>Neotermes jouteli</i>	F3	6,256	276	2354	78.4	4,547	4.70	0.63	088
22	<i>Reticulitermes santonensis</i>	L2	48,066	112	427	85.1	602	3.92	0.67	089
23	<i>Coptotermes niger</i>	L1	53,003	91	166	87.2	202	2.26	0.45	090
<b>Higher termites</b>										
24	<i>Odontotermes</i> sp. <sup>e</sup>	F4	12,898	307	1005	63.1	1,391	5.77	0.86	091
25	<i>Macrotermes</i> sp.	L1	12,073	260	1358	69.5	2,790	5.34	0.76	092
26	<i>Macrotermes subhyalinus</i> <sup>e</sup>	F5	27,297	211	4805	68.4	1,182	5.23	0.84	093
27	<i>Alyscotermes trestus</i> <sup>e</sup>	F5	24,582	550	3203	78.6	5,940	6.57	0.82	094

Sample	Host species	Origin <sup>a</sup>	Total reads	Genus-level taxa	OTUs (3% dissim.)	Coverage (%) <sup>b</sup>	Diversity indices <sup>c</sup> (based on OTUs)			NCBI accession number <sup>d</sup>
							Richness	Diversity	Evenness	
28	<i>Cubitermes ugandensis</i>	F6	22,832	211	5413	49.7	2,020	6.49	0.97	095
29	<i>Ophiotermes</i> sp. <sup>e</sup>	F7	8,418	328	1336	76.0	2,026	6.13	0.85	096
30	<i>Amitermes meridionalis</i> <sup>e</sup>	F8	23,840	354	1556	85.5	2,246	5.04	0.70	097
31	<i>Microcerotermes</i> sp. <sup>e</sup>	F5	34,626	291	2358	79.1	4,407	4.55	0.61	098
32	<i>Nasutitermes corniger</i>	L3	10,363	175	1998	65.5	1,208	4.15	0.67	099
33	<i>Nasutitermes takasagoensis</i>	F9	16,619	198	1602	77.4	3,607	4.04	0.56	100
34	<i>Trinervitermes</i> sp.	F5	25,173	232	1103	84.2	1,943	4.68	0.67	101
<b>Others</b>										
35	<i>Pachnoda ehippiata</i>	B2	10,033	339	1325	80.2	1,335	5.77	0.85	102
36	<i>Acheta domesticus</i>	B2	5,326	104	241	84.2	276	4.14	0.80	103
37	<i>Gryllus assimilis</i>	B1	26,800	190	669	90.3	712	4.13	0.70	104

<sup>a</sup> Origin of samples: B, commercial breeders: B1, Jörg Bernhardt, Helbigsdorf, Germany (<http://www.schaben-spinnen.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. Plarre, Federal Institute for Materials Research and Testing, Berlin, Germany; L2 MPI Marburg; L3, R.H. Scheffrahn, University of Florida, Fort Lauderdale, Florida, USA.

<sup>b</sup> Good's coverage estimator (Good,1953)

<sup>c</sup> Richness: Chao 1 estimator (Chao, 1984); diversity: non-parametric Shannon index (Chao and Shen, 2003); evenness index (Legendre and Legendre, 1998).

<sup>d</sup> 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.

<sup>e</sup> Ethanol-preserved specimen; the entire gut was used for DNA extraction.

**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 (Bray and Curtis, 1957). 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 (Lozupone and Knight, 2005) 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 (Abdi and Williams, 2010). Multivariate statistics were carried out using the *R* software (version 2.15.1) with the *vcd* and *vegan* packages (R Development Core Team, 2013; Meyer et al., 2013; Oksanen et al., 2013). 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* (Price et al., 2010), transformed into ultrametric trees using *PATHd8* (Britton et al., 2007), and visualized using the *R* package *APE* (Paradis et al., 2004).

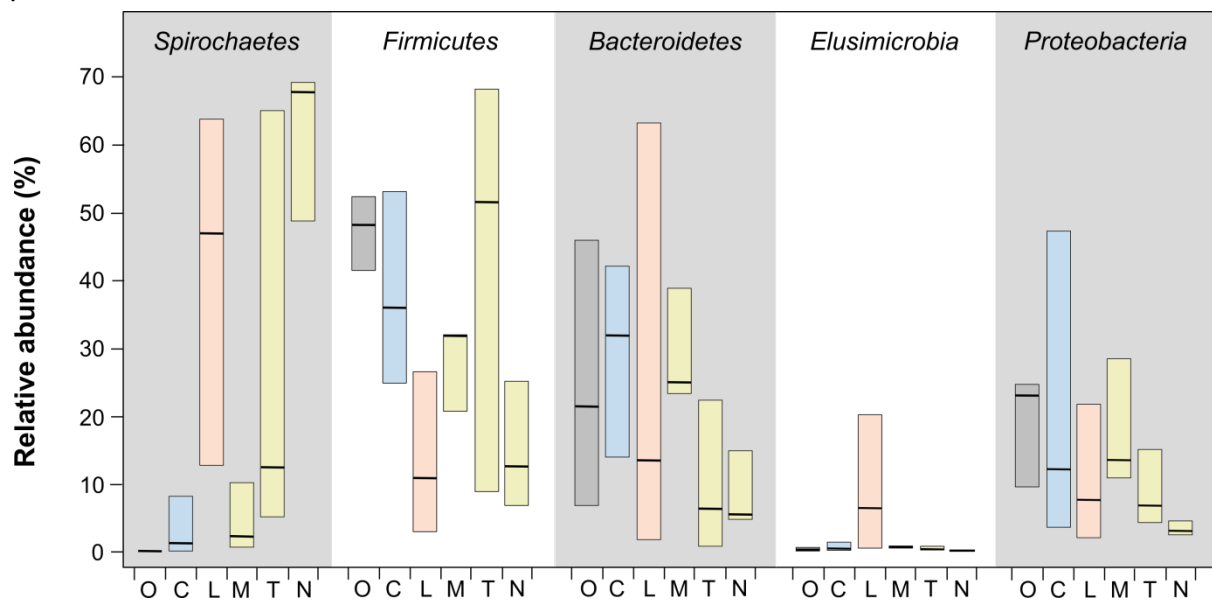
**Tab. 2 | Improvement of classification success using our curated reference database that included all homologs previously obtained from insect guts and optimized to resolve all termite- and cockroach-specific groups (Köhler et al., 2012) over that using the RDP database (Release 10, Update 30).** The 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.

Host species	Classification success (%)									
	Phylum		Class		Order		Family		Genus	
	RDP	This study	RDP	This study	RDP	This study	RDP	This study	RDP	This study
<b>Cockroaches</b>										
<i>Ergaula capucina</i>	94.9	99.3	79.3	96.1	73.4	93.1	63.3	88.0	33.2	61.6
<i>Panesthia angustipennis</i>	95.9	99.3	79.3	96.1	77.2	94.2	59.0	86.0	24.8	61.0
<i>Salganea esakii</i>	93.8	99.4	79.7	95.3	78.8	93.5	71.1	90.1	40.6	74.6
<i>Blatta orientalis</i>	96.2	99.3	79.8	97.7	77.9	96.3	70.5	93.4	48.0	66.6
<i>Cryptocercus punctulatus</i>	93.8	99.0	74.8	93.4	72.5	90.2	59.0	82.8	33.1	67.7
<b>Termites</b>										
<i>Reticulitermes santonensis</i>	85.5	99.2	80.9	96.2	79.2	95.6	76.1	94.5	68.0	93.3
<i>Cubitermes ugandensis</i>	92.0	98.7	84.4	97.4	78.4	94.9	64.7	83.3	24.3	69.4
<i>Nasutitermes corniger</i>	87.8	98.9	83.0	97.4	81.6	96.5	77.9	95.3	40.9	91.0

## Results and Discussion

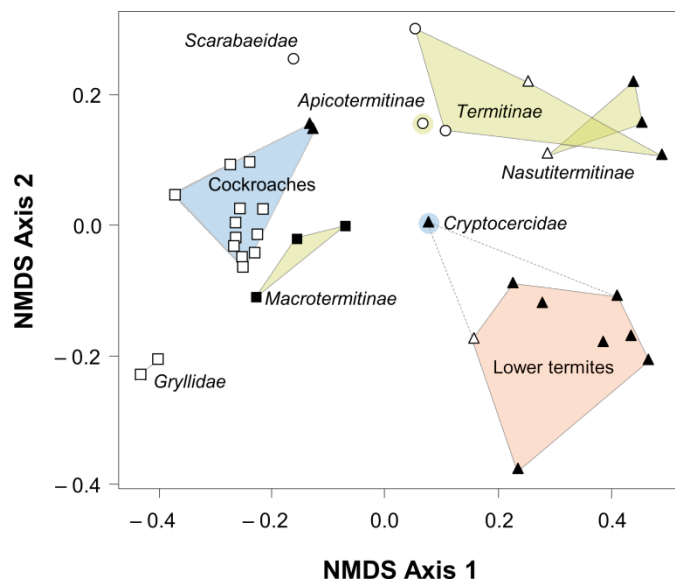
The bacterial 16S rRNA genes in hindgut DNA of 34 termites and cockroaches and a few other insects were amplified with universal, bar-coded primers for the V3–V4 region. For each host species, we obtained an average of 10,000 high-quality sequence reads (Table 1). This is the first such data for most of these species, and classification against the RDP database (Cole et al., 2009) yielded large fractions of unclassified sequences, particularly at lower taxonomic ranks (Table 2). Our curated reference database (Köhler et al., 2012) 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 underline that also 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)



**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 *Macrotermitinae*, *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.

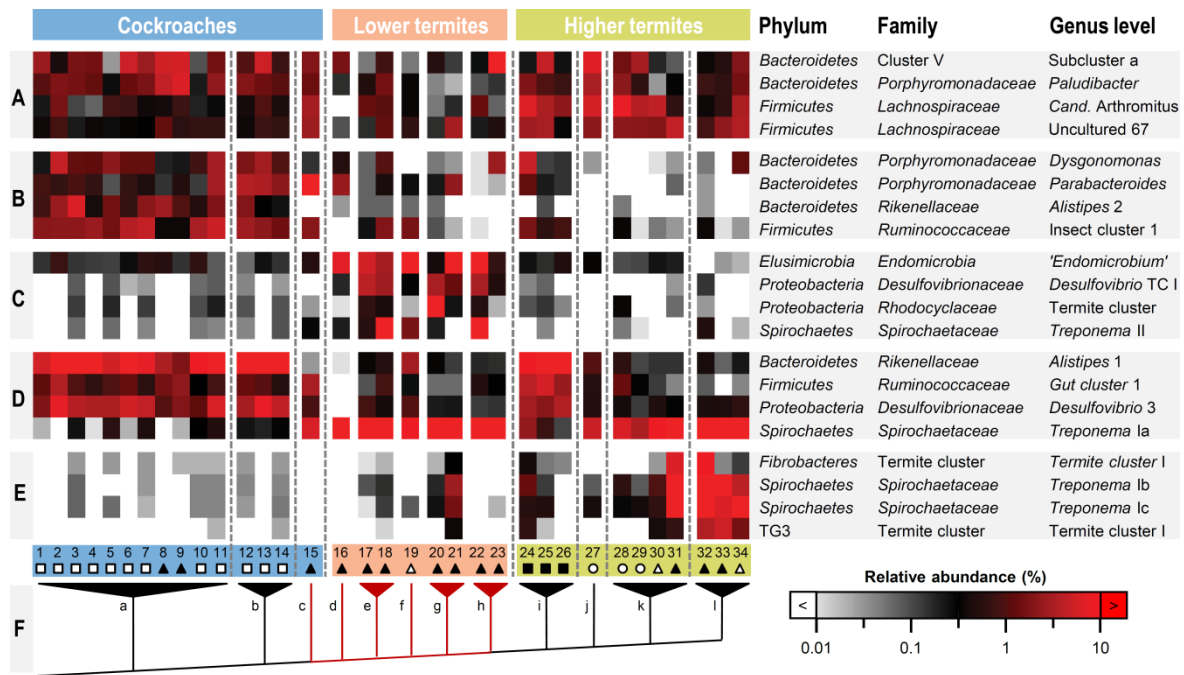
The bacterial communities of each host group differed strongly already at the phylum level (Fig. 1). *Spirochaetes* were rare in cockroaches but abundant in lower termites and wood-feeding higher termites, often representing the majority of the reads, which is in agreement with the general notion that spirochaetes are the most characteristic element of the termite gut microbiota (Brenak, 2000). *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 (Schmitt-Wagner et al., 2003; Noda et al., 2005; Thongaram et al. 2005; Schauer et al., 2012). Members of *Elusimicrobia* were highly represented only in lower termites.



**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 out-groups. For the species behind each data point, see (Fig. 3). The clusters were supported by both ADONIS and MRPP analyses ( $p < 0.001$ ).

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 *Macrotermitinae* showing a strong affinity to the cockroaches. A notable exception was the wood-feeding cockroach *Cryptocercus punctulatus*, the closest relative of termites (Inward et al., 2007; Engel et al., 2009). 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.



**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.

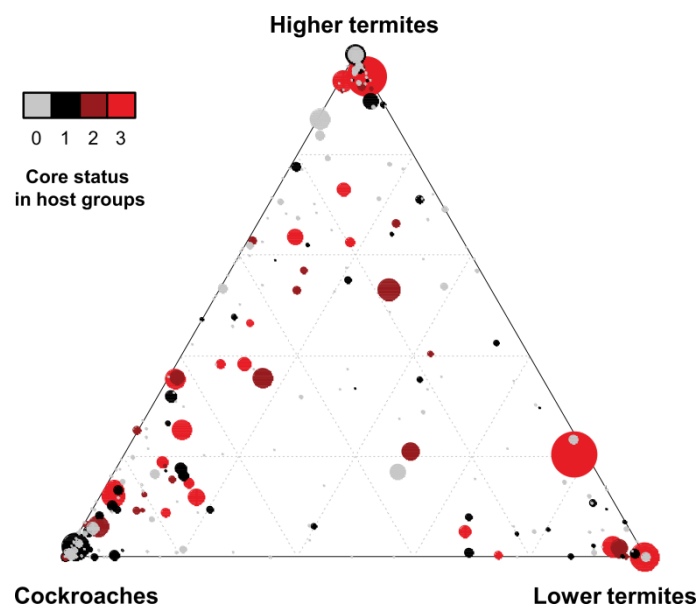
When we ranked all 884 genus-level taxa in the dataset according to their contribution to the ordination results, it became apparent that already the top 100 genera were responsible for almost 70% of the pattern and represented 90% of the sequences in the dataset (Table S3). Many genus-level taxa occurred in all major host lineages, extending the previously postulated presence of termite-specific bacterial lineages to all cockroaches (Schauer et al., 2012; Noda et al., 2009), 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 (Breznak, 2000; Ohkuma and Brune, 2011). Since this cluster comprises the homoacetogenic *Treponema primitia* (Graber et al., 2004), its upshift is also consistent with changes in the distribution of several functional marker genes (formyltetrahydrofolate synthetase, CO dehydrogenase and hydrogenase) (e.g. Matson et al., 2011; Ottesen and Leadbetter, 2011; Ballor and Leadbetter, 2012), 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) (Iida et al., 2000) and endosymbiotic '*Endomicrobium*' (Stingl, et al., 2005;



Ohkuma et al., 2007) and *Desulfovibrio* spp. (TC I) (Sato et al., 2009; Strassert et al., 2012) 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 (Ikeda-Ohtsubo et al., 2010) that were recruited as endosymbionts presumably long after the flagellates had established their symbiosis with lower termites (Ikeda-Ohtsubo and Brune, 2009). 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 (Noda et al., 2009, Desai et al., 2010), is in agreement with their recruitment from free-living relatives that are present but of low abundance in termites lacking these flagellates (Noda et al., 2009).

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 (Hongoh et al., 2006a). There is strong evidence from enzymatic (Tokuda and Watanabe, 2007) and metagenomic (Warnecke et al., 20007; He et al., 2013) 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.

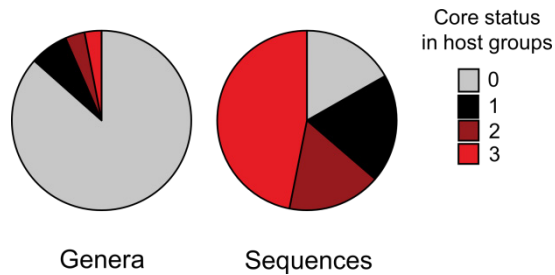


**Fig. 4 | Ternary plot of the distribution of genus-level taxa across the major host groups.** The area of the circles represents the relative abundance of the reads in the entire dataset, the position specifies their average abundance in the respective host groups, and the colors indicate the number 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. See Fig. S3 at <http://www.termites.de/brune/publ/suppl/AEM04206-13 Figure S3.html>.

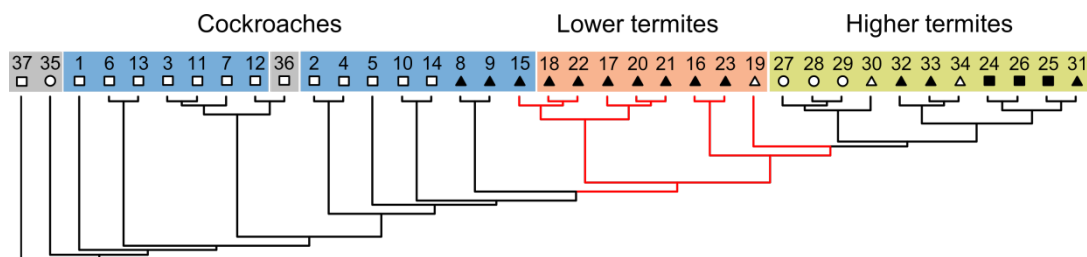
Also the resurgence in the *Macrotermitinae* of taxa that prevail in cockroaches may be related to the dietary diversification of higher termites following the loss of flagellates. The high similarity between the gut microbiota of omnivorous cockroaches and *Macrotermitinae*, first discovered in a study of the gut microbiota of the cockroach *S. lateralis* (Schauer et al., 2012), 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 (Eckburg et al., 2005; Leser et al., 2002); their abundance in the *Macrotermitinae* may be caused by the protein-rich fungal biomass included in the diet of the fungus-cultivating species (Hyodo et al., 2003).

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 ehippiata*, 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. ehippiata*) 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.



**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 (△), soil/humus feeding (○), and fungus cultivating (■). The branches connecting species that harbor gut flagellates are marked in red.

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; Carlseña et al., 2012) has to be considered.

It remains to be investigated whether traces of host phylogeny can be found also in the archaeal microbiota in the guts of termites and cockroaches. Although archaea are much less abundant than bacteria (0.1–3% of prokaryotes in termite guts; Hackstein et al., 2006), methanogens seem to be present in all dictypteran lineages (Brune, 2010). The diversity of the archaeal community is much

smaller than that of bacteria and comprises both termite-specific clusters and lineages with representatives from many environments (e.g., Paul et al., 2012; Brune, 2010).

**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 (Noda et al., 2009) 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 (He et al., 2013), microbial interactions in hydrogen metabolism and methanogenesis (Pester and Brune, 2007), and the emerging role of the flagellate symbionts in the nitrogen economy of the digestive symbiosis (Hongoh, 2011). Following Dobzhansky's famous dictum (Dobzhansky, 1973), 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

- Abdi H and Williams LJ (2010).** Principal component analysis. *Wiley Int. Rev. Com. St.* 2:433–459.
- Ballor NR and Leadbetter JR (2012).** Analysis of extensive [FeFe] hydrogenase gene diversity within the gut microbiota of insects representing five families of Dictyoptera. *Microb. Ecol.* 63:586–595.
- Bignell DE, Eggleton P (2000).** Termites in ecosystems, pp. 363–387. *In* Abe T, Bignell DE, Higashi M (eds.). *Termites: Evolution, Sociality, Symbiosis, Ecology.* Kluwer Academic Publishers, Dordrecht.
- Bragg L, Stone G, Imelfort M, Hugenholtz P and Tyson GW (2012).** Fast, accurate error-correction of amplicon pyrosequences using Acacia. *Nat. Methods* 9:425–426.
- Brauman A, Dore J, Eggleton P, Bignell D, Breznak JA and Kane MD (2001).** Molecular phylogenetic profiling of prokaryotic communities in guts of termites with different feeding habits. *FEMS Microbiol. Ecol.* 35:27–36.
- Bray JR and Curtis JT (1957).** An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monogr.* 27:325–349.
- Breznak JA (2000).** Ecology of prokaryotic microbes in the guts of wood- and litter-feeding termites, pp. 209–231. *In* Abe, T.; Bignell, D.E.; Higashi, M. (ed.), *Termites: Evolution, Sociality, Symbiosis, Ecology.* Kluwer Academic Publishers, Dordrecht.
- Britton T, Anderson C, Jacquet D, Lundqvist S and Bremer K (2007).** Estimating divergence times in large phylogenetic trees. *Syst. Biol.* 56:741–752.
- Brune A (2010).** Methanogens in the digestive tract of termites, pp. 81–100. *In* Hackstein, J.H.P. (ed.), *(Endo)symbiotic Methanogenic Archaea.* Springer, Heidelberg.
- Brune A and Ohkuma M (2011).** Role of the termite gut microbiota in symbiotic digestion, pp. 439–475. *In* Bignell DE, Roisin Y, Lo N (eds.). *Biology of Termites: A Modern Synthesis,* Springer, Dordrecht.
- Carlsena T, Aasa AB, Lindnerb D, Vrålstada T, Schumachera T and Kauseruda H (2012).** Don't make a mista(g)ke: is tag switching an overlooked source of error in amplicon pyrosequencing studies? *Fung. Ecol.* 5:747–749.
- Chao A (1984).** Nonparametric estimation of the number of classes in a population. *Scan. J. Stat.* 11:265–270.
- Chao A and Shen TJ (2003).** Nonparametric estimation of Shannon's index of diversity when there are unseen species in sample. *Environ. Ecol. Stat.* 10:429–443.
- Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM and Tiedje JM (2009).** The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Res.* 37:D141–D145.
- Deevong P, Hongoh Y, Inoue T, Trakulnaleamsai S, Kudo T, Noparatnaraporn N and Ohkuma M (2006).** Effect of temporal sample preservation on the molecular study of a complex microbial community in the gut of the termite *Microcerotermes* sp. *Microbes Environ.* 21:78–85.
- Desai M, Strassert JFH, Meuser K, Hertel H, Ikeda-Ohtsubo W, Radek R and Brune A (2010).** Strict cospeciation of devescovinid flagellates and *Bacteroidales* ectosymbionts in the gut of dry-wood termites (*Kalotermitidae*). *Environ. Microbiol.* 12:2120–2132.
- Dobzhansky T (1973).** Nothing in biology makes sense except in the light of evolution. *Am. Biol. Teacher* 35:125–129.
- Eckburg PB, Bik EM, Bernstein CN, Purdom, E, Dethlefsen L, Sargent M, Gill SR, Nelson KE and Relman DA (2005).** Diversity of the human intestinal microbial flora. *Science* 308:1635–1638.
- Engel MS, Grimaldi DA and Krishna K (2009).** Termites (Isoptera): their phylogeny, classification, and rise to ecological dominance. *American Museum Novitates* 3650:1–27.
- Geissinger O, Herlemann DPR, Mörschel E, Maier UG and Brune A (2009).** The ultramicrobacterium "*Elusimicrobium minutum*" gen. nov., sp. nov., the first cultivated representative of the Termite Group 1 phylum. *Appl. Environ. Microbiol.* 75:2831–2840.
- Good IJ (1953).** The population frequencies of species and the estimation of population parameters. *Biometrika* 40:237–264.

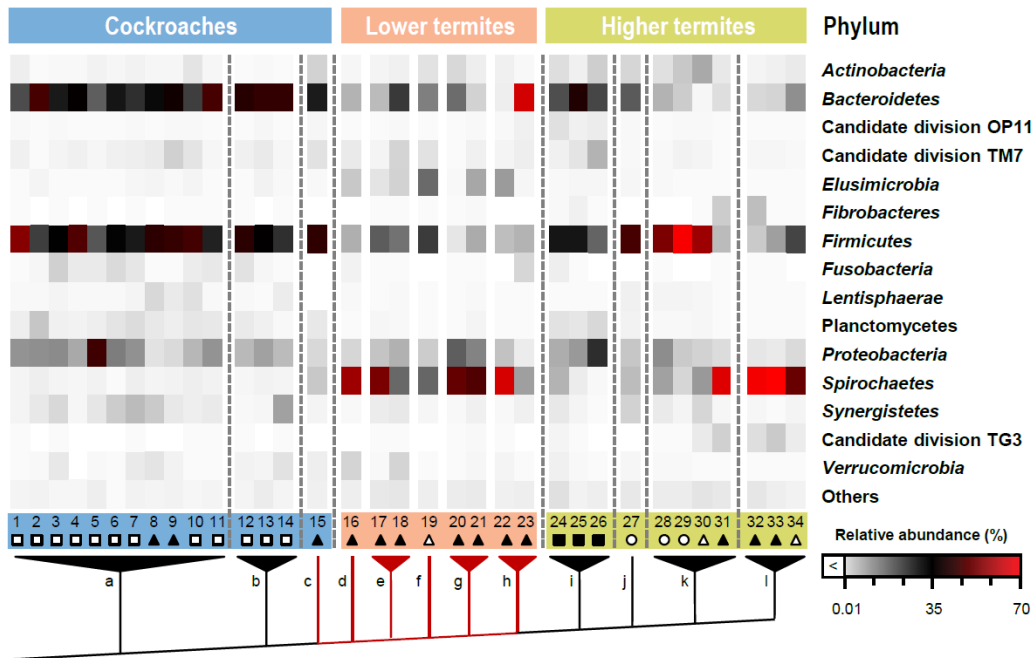
- Graber JR, Leadbetter JR and Breznak JA (2004).** Description of *Treponema azotonutricium* sp. nov. and *Treponema primitia* sp. nov., the first spirochetes isolated from termite guts. *Appl. Environ. Microbiol.* 70:1315–1320.
- Hackstein JHP, van Alen TA and Rosenberg J (2006).** Methane production by terrestrial arthropods, pp. 155–180. In König, H.; Varma, A. (ed.), *Intestinal Microorganisms of Termites and Other Invertebrates*. Springer, Berlin.
- He S, Ivanova N, Kirton E, Allgaier M, Bergin C, Scheffrahn RH, Kyrpidis NC, Warnecke, F, Tringe SG and Hugenholtz P (2013).** Comparative metagenomic and metatranscriptomic analysis of hindgut paunch microbiota in wood- and dung-feeding higher termites. *PLoS ONE* 8:e61126.
- Hongoh Y (2011).** Toward the functional analysis of uncultivable, symbiotic microorganisms in the termite gut. *Cell. Mol. Life Sci.* 68:1311–1325.
- Hongoh Y, Deevong P, Hattori S, Inoue T, Noda S, Noparatnaraporn N, Kudo T and Ohkuma M (2006a).** Phylogenetic diversity, localization and cell morphologies of the candidate phylum TG3 and a subphylum in the phylum *Fibrobacteres*, recently found bacterial groups dominant in termite guts. *Appl. Environ. Microbiol.* 72:6780–6788.
- Hongoh Y, Deevong P, Inoue T, Moriya S, Trakulnaleamsai S, Ohkuma M, Vongkaluang C, Noparatnaraporn N and Kudo T (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.
- Hongoh Y, Ekpornprasit L, Inoue T, Moriya S, Trakulnaleamsai S, Ohkuma M, Noparatnaraporn and Kudo T (2006b).** Intracolony variation of bacterial gut microbiota among castes and ages in the fungus-growing termite *Macrotermes gilvus*. *Mol. Ecol.* 15:505–516.
- Hyodo F, Tayasu I, Inoue T, Azuma JI, Kudo T and Abe T (2003).** Differential role of symbiotic fungi in lignin degradation and food provision for fungus-growing termites (Macrotermitinae: Isoptera). *Funct. Ecol.* 17:186–193
- Iida T, Ohkuma M, Ohtoko K and Kudo T (2000).** Symbiotic spirochetes in the termite hindgut: phylogenetic identification of ectosymbiotic spirochetes of oxymonad protists. *FEMS Microbiol. Ecol.* 34:17–26.
- Ikeda-Ohtsubo W and Brune A (2009).** Cospeciation of termite gut flagellates and their bacterial endosymbionts: *Trichonympha* species and 'Candidatus Endomicrobium trichonymphae'. *Mol. Ecol.* 18:332–342.
- Ikeda-Ohtsubo W, Faivre N and Brune A (2010).** Putatively free-living 'Endomicrobia' — ancestors of the intracellular symbionts of termite gut flagellates? *Environ. Microbiol. Rep.* 2:554–559.
- Inward D, Beccaloni G and Eggleton P (2007).** Death of an order: a comprehensive molecular phylogenetic study confirms that termites are eusocial cockroaches. *Biol. Lett.* 3:331–335.
- Jouquet P, Traoré S, Choosai C, Hartmann C, Bignell D (2011).** Influence of termites on ecosystem functioning. Ecosystem services provided by termites. *Eur. J. Soil Biol.* 47:215–222.
- Köhler T, Dietrich C, Scheffrahn RH and Brune A (2012).** High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation of the gut in wood-feeding higher termites (*Nasutitermes* spp.). *Appl. Environ. Microbiol.* 78:4691–4701.
- Legendre P and Legendre L (1998).** *Numerical Ecology* 2nd Edition. Elsevier Amsterdam, Montreal (Quebec).
- Leser TD, Amenuvor JZ, Jensen TK, Lindecrona RH, Boye M and Møller K (2002).** Culture-Independent Analysis of Gut Bacteria: the Pig Gastrointestinal Tract Microbiota Revisited. *Appl. Environ. Microbiol.* 68:673–690.
- Lilburn TG, Schmidt TM and Breznak JA (1999).** Phylogenetic diversity of termite gut spirochaetes. *Environ. Microbiol.* 1:331–345.
- Lozupone C and Knight R (2005).** UniFrac: a New Phylogenetic Method for Comparing Microbial Communities. *Appl. Environ. Microbiol.* 71:8228–8235.
- Matson EG, Gora KG and Leadbetter JR (2011).** Anaerobic carbon monoxide dehydrogenase diversity in the homoacetogenic hindgut microbial communities of lower termites and the Wood Roach. *PLoS ONE* 6:e19316.
- Meyer D, Zeileis A and Hornik K (2013).** vcd: Visualizing Categorical Data. R package version 1.2–13 URL: <http://cran.r-project.org/package=vcd>.
- Nalepa CA, Bignell DE and Bandi C (2001).** Detritivory, coprophagy, and the evolution of digestive mutualisms in Dictyoptera. *Insect. Soc.* 48:194–201.

- Ni J and Tokuda G (2013). Lignocellulose-degrading enzymes from termites and their symbiotic microbiota. *Biotechnol. Adv.* 31:838–850.
- Nobre T, Rouland-Lefèvre C and Aanen DK (2011). Comparative biology of fungus cultivation in termites and ants, pp. 193–210. *In* Bignell, D.E.; Roisin, Y., Lo, N. (eds.), *Biology of Termites: A Modern Synthesis*. Springer, Dordrecht.
- Noda S, Hongoh Y, Sato T and Ohkuma M (2009). Complex coevolutionary history of symbiotic Bacteroidales bacteria of various protists in the gut of termites. *BMC Evol. Biol.* 9:158.
- Noda S, Iida T, Kitade O, Nakajima H, Kudo T and Ohkuma M (2005). Endosymbiotic *Bacteroidales* bacteria of the flagellated protist *Pseudotriconympha grassii* in the gut of the termite *Coptotermes formosanus*. *Appl. Environ. Microbiol.* 71:8811–8817.
- Ohkuma M and Brune A (2011). Diversity, structure, and evolution of the termite gut microbial community, pp. 413–438. *In* Bignell DE, Roisin Y, Lo N (eds.). *Biology of Termites: A Modern Synthesis*, Springer, Dordrecht.
- Ohkuma M, Noda S, Hongoh Y, Nalepa CA and Inoue T (2009). Inheritance and diversification of symbiotic trichonymphid flagellates from a common ancestor of termites and the cockroach *Cryptocercus*. *Proc. R. Soc. B* 276:239–245.
- Ohkuma M, Sato T, Noda S, Ui S, Kudo T and Hongoh Y (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.
- Oksanen J, Blanchet JG, Kindt R, Legendre P, Minchin PR, O’Hara RB, Simpson GL, Solymos P, Stevens MHH and Wagner H (2013). *vegan: Community Ecology Package*. <http://CRAN.R-project.org/package=vegan>.
- Ottesen EA and Leadbetter JR (2011). Formyltetrahydrofolate synthetase gene diversity in the guts of higher termites with different diets and lifestyles. *Appl. Environ. Microbiol.* 77: 3461–3467.
- Paradis E, Claude J and Strimmer K (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290.
- Paul K, Nonoh JO, Mikulski L and Brune A. 2012. "*Methanoplasmatales*," *Thermoplasmatales*-related archaea in termite guts and other environments, are the seventh order of methanogens. *Appl. Environ. Microbiol.* 78:8245–8253.
- Pester M and Brune A (2006). Expression profiles of *fhs* (FTHFS) genes support the hypothesis that spirochaetes dominate reductive acetogenesis in the hindgut of lower termites. *Environ. Microbiol.* 8:1261–1270.
- Pester M and Brune A (2007). Hydrogen is the central free intermediate during lignocellulose degradation by termite gut symbionts. *ISME J.* 1:551–565
- Price MN, Dehal PS and Arkin AP (2010). FastTree 2 – approximately maximum-likelihood trees for large alignments.. *PLoS ONE* 5:e9490.
- R Development Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: <http://www.r-project.org>.
- Sato T, Hongoh Y, Noda S, Hattori S, Ui S and Ohkuma M (2009). *Candidatus* *Desulfovibrio trichonymphae*, a novel intracellular symbiont of the flagellate *Trichonympha agilis* in termite gut. *Environ. Microbiol.* 11:1007–1015.
- Schauer C, Thompson CL and Brune A (2012). The bacterial community in the gut of the cockroach *Shelfordella lateralis* reflects the close evolutionary relatedness of cockroaches and termites. *Appl. Environ. Microbiol.* 78:2758–2767.
- Schloss PD, Gevers D and Westcott SL (2011). Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. *PLoS ONE* 6:e27310.
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ and Weber CF (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75:7537–7541.
- Schmitt-Wagner D, Friedrich MW, Wagner B and Brune A (2003). Phylogenetic diversity, abundance, and axial distribution of bacteria in the intestinal tract of two soil-feeding termites (*Cubitermes* spp.). *Appl. Environ. Microbiol.* 69:6007–6017

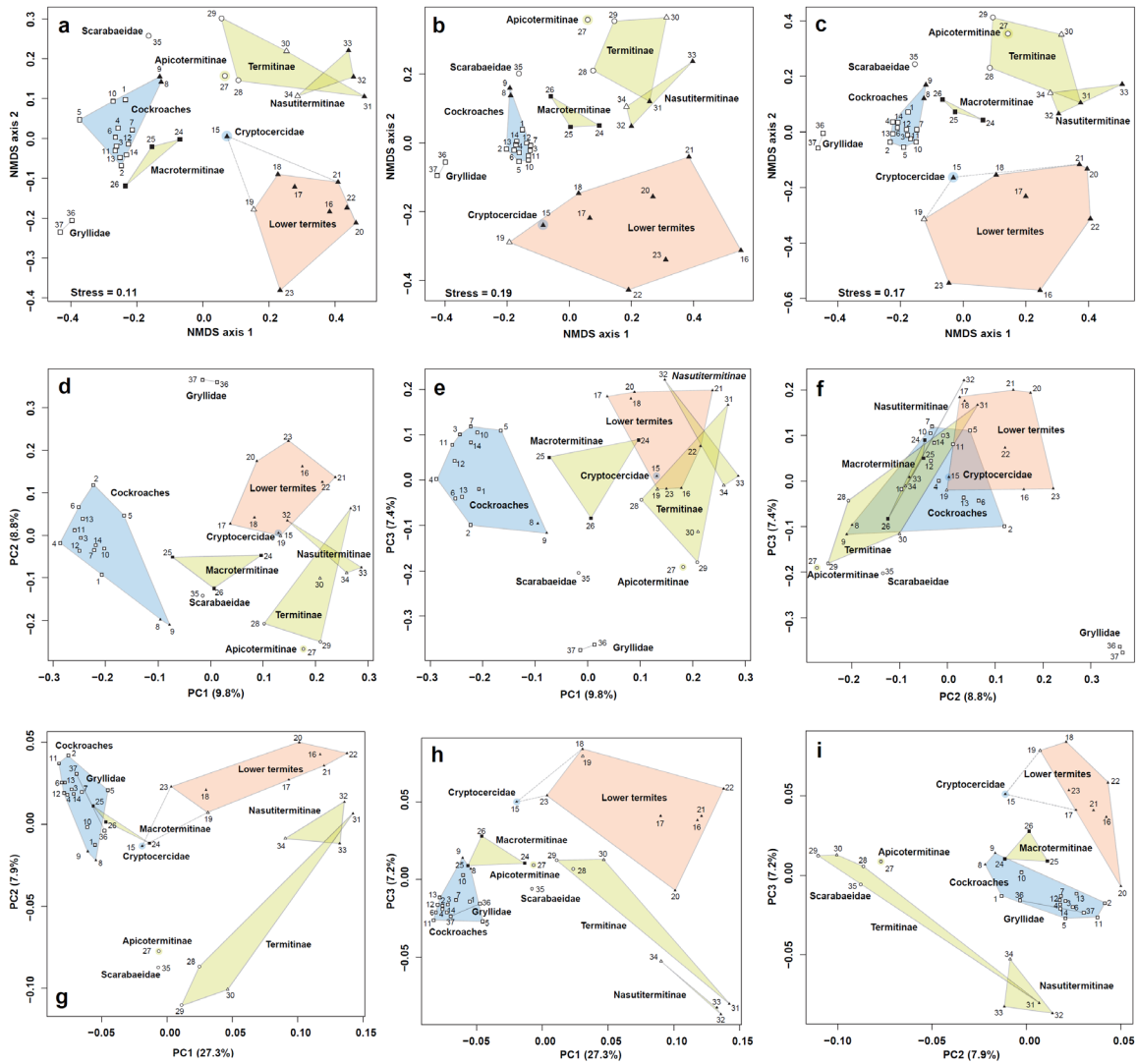
- Stingl U, Radek R, Yang H and Brune A (2005).** 'Endomicrobia': Cytoplasmic symbionts of termite gut protozoa form a separate phylum of prokaryotes. *Appl. Environ. Microbiol.* 71:1473–1479.
- Strassert JFH, Köhler T, Wienemann THG, Ikeda-Ohtsubo W, Faivre N, Franckenberg S, Plarre R, Radek R and Brune A (2012).** 'Candidatus Ancillula trichonymphae', a novel lineage of endosymbiotic *Actinobacteria* in termite gut flagellates of the genus *Trichonympha*. *Environ. Microbiol.* 14:3259–3270.
- Thompson CL, Vier R, Mikaelyan A, Wienemann T and Brune A (2012).** 'Candidatus Arthromitus' revised: Segmented filamentous bacteria in arthropod guts are members of *Lachnospiraceae*. *Environ. Microbiol.* 14:1454–1465.
- Thongaram T, Hongoh Y, Kosono S, Ohkuma M, Trakulnaleamsai S, Noparatnaraporn N and Kudo T (2005).** Comparison of bacterial communities in the alkaline gut segment among various species of higher termites. *Extremophiles* 9:229–238.
- Tokuda G and Watanabe H (2007).** Hidden cellulases in termites: revision of an old hypothesis. *Biol. Lett.* 3:336–339
- Warnecke F, Luginbühl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, Cayouette M, McHardy AC, Djordjevic G, Aboushadi N, Sorek R, Tringe SG, Podar M, Martin HG, Kunin V, Dalevi D, Madejska J, Kirton E, Platt D, Szeto E, Salamov A, Barry K, Mikhailova N, Kyrpides NC, Matson EG, Ottesen EA, Zhang X, Hernández M, Murillo C, Acosta LG, Rigoutsos I, Tamayo G, Green BD, Chang C, Rubin EM, Mathur EJ, Robertson DE, Hugenholtz P, Leadbetter JR (2007).** Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nat.* 450:560–565.



## Supplementary Material



**Fig. S1 | Relative abundance of sequence reads assigned to the major bacterial phyla in the gut microbiota of cockroaches, lower termites, and higher termites.** The color scale is linear. Numbers indicate host species (see Table 1 or Table S2). Symbols indicate lifestyle: (□), wood feeding (▲), grass feeding (Δ), soil/humus feeding (○), and fungus cultivating (■). The schematic tree shows a simplified host phylogeny of the 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 red branches mark the presence of cellulose-digesting flagellates.



**Fig. S2 | Comparison of bacterial community structure in cockroaches and termites using various ordination techniques and taxon-independent (OTU-based) approaches.** Non-metric multidimensional scaling (NMDS) of a pairwise Bray-Curtis dissimilarity distance matrix based on genus-level taxa (a) (relative abundance of 16S rRNA genes) or on the distribution of operational taxonomic units (OTUs) with 3% (b) and 5% sequence divergence (c). Principal coordinate analysis of the unweighted (d–f) and weighted (g–i) UniFrac distance matrix (normalized dataset); the variance explained by the three principal coordinates (PC) is indicated on the respective axes. Numbers indicate host species (see Table 1 or Table S2). Clustering of cockroaches, lower termites and different subfamilies of higher termites was significant in all analyses ( $p < 0.001$ ).

**Tab. S1 | Excel table containing information with the classification results for all samples (relative abundance of reads at different phylogenetic levels).** Please download the interactive spreadsheet Table\_S1.xlsx from <http://aem.asm.org/content/early/2014/01/27/AEM.04206-13/suppl/DCSupplemental>.

**Tab. S2 | Classification success at different taxonomic levels for all insects included in this study.**

No.	Host species	Classified (%)				
		Phylum	Class	Order	Family	Genus
<b>Cockroaches</b>						
1	<i>Ergaula capucina</i>	99.3	96.1	93.1	88.0	61.6
2	<i>Symploce macroptera</i>	99.5	98.7	90.1	82.5	70.8
3	<i>Rhyparobia maderae</i>	99.4	97.2	94.2	83.5	67.4
4	<i>Elliptorhina chopardi</i>	99.6	97.2	95.3	87.3	65.1
5	<i>Panchlora</i> sp.	99.1	97.2	92.7	86.5	73.5
6	<i>Diploptera punctata</i>	99.8	97.9	94.4	88.1	74.1
7	<i>Opisthoptlatia orientalis</i>	99.0	95.8	89.6	83.9	63.3
8	<i>Panesthia angustipennis</i>	99.3	96.1	94.2	86.0	61.0
9	<i>Salganea esakii</i>	99.6	92.4	90.7	85.7	63.5
10	<i>Eublaberus posticus</i>	99.2	94.0	90.1	82.3	68.5
11	<i>Schultesia lampyridiformis</i>	99.5	97.8	94.3	89.2	73.7
12	<i>Eurycotis floridana</i>	99.4	96.4	94.0	90.1	66.2
13	<i>Shelfordella lateralis</i>	99.4	95.3	93.5	90.1	74.6
14	<i>Blatta orientalis</i>	99.3	97.7	96.3	93.4	66.6
15	<i>Cryptocercus punctulatus</i>	99.0	93.4	90.2	82.8	67.7
<b>Lower termites</b>						
16	<i>Mastotermes darwiniensis</i>	98.6	94.8	91.9	84.1	78.5
17	<i>Zootermopsis nevadensis</i>	97.9	90.1	88.3	76.0	70.0
18	<i>Hodotermopsis sjoestedti</i>	98.5	96.4	95.3	92.9	86.3
19	<i>Hodotermes mossambicus</i>	98.8	93.9	90.5	86.8	72.4
20	<i>Incisitermes marginipennis</i>	99.2	98.3	97.3	94.9	91.4
21	<i>Neotermes jouteli</i>	99.3	95.7	92.2	86.8	81.8
22	<i>Reticulitermes santonensis</i>	99.2	96.2	95.6	94.5	93.3
23	<i>Coptotermes niger</i>	98.2	94.2	93.8	92.7	91.0
<b>Higher termites</b>						
24	<i>Odontotermes</i> sp.	99.0	93.2	89.9	83.9	76.3
25	<i>Macrotermes</i> sp.	99.2	90.7	87.4	83.7	76.7
26	<i>Macrotermes subhyalinus</i>	99.9	86.4	81.0	72.6	66.1
27	<i>Alyscotermes trestus</i>	99.7	95.7	93.0	87.1	68.0
28	<i>Cubitermes ugandensis</i>	98.7	97.4	94.9	83.3	69.4
29	<i>Ophiotermes</i> sp.	99.8	97.4	95.2	90.6	73.7
30	<i>Amitermes meridionalis</i>	97.8	93.3	89.6	81.6	68.7
31	<i>Microcerotermes</i> sp.	99.2	98.3	96.8	93.2	87.7
32	<i>Nasutitermes corniger</i>	98.9	97.4	96.5	95.3	91.0
33	<i>Nasutitermes takasagoensis</i>	98.8	97.7	97.0	94.9	88.9
34	<i>Trinervitermes</i> sp.	99.8	98.9	98.2	95.7	87.3
<b>Others</b>						
35	<i>Pachnoda ehippiata</i>	99.8	98.6	95.3	85.9	69.1
36	<i>Acheta domesticus</i>	99.7	99.4	98.0	96.0	79.8
37	<i>Gryllus assimilis</i>	99.9	99.7	97.0	94.9	75.4

**Tab. S3 | Contribution of the 100 most important genus-level taxa to the separation of cockroaches, lower termites, and higher termites in the ordination analyses.** The taxa are ranked in order of their contribution, and their cumulative abundance among the total reads of all samples and their representation among the three major host lineages are given. Taxa represented by > 70% of the species in a host group were considered core lineages (marked in bold). Taxa with core status for all hosts groups that were included in the community similarity analysis (Fig. 5) are indicated with grey shading.

Rank	Phylum	Family	Genus level	No. of OTUs (3%)	Contribution (%) <sup>a</sup>		Read abundance (%) <sup>b</sup>	Representation in hosts (%) <sup>c</sup>		
					Individual	Cumulative		Cockroaches	Lower termites	Higher termites
1	<i>Spirochaetes</i>	<i>Spirochaetaceae</i>	<i>Treponema la</i>	3907	5.5	5.5	9.6	<b>87</b>	<b>100</b>	<b>100</b>
2	<i>Spirochaetes</i>	<i>Spirochaetaceae</i>	<i>Treponema lc</i>	901	3.6	9.1	3.6	47	63	<b>100</b>
3	<i>Bacteroidetes</i>	<i>Rikenellaceae</i>	<i>Alistipes 2</i>	969	3.5	12.6	7.4	<b>100</b>	<b>100</b>	<b>100</b>
4	<i>Bacteroidetes</i>	Cluster V	Subcluster Va	9	3.5	16.1	1.8	13	13	0
5	<i>Spirochaetes</i>	<i>Spirochaetaceae</i>	<i>Treponema ld</i>	42	2.2	18.3	1.2	0	50	18
6	<i>Firmicutes</i>	<i>Ruminococcaceae</i>	Uncultured 24	326	2.0	20.3	1.4	<b>87</b>	63	<b>100</b>
7	<i>Elusimicrobia</i>	<i>Endomicrobia</i>	Endomicrobium	306	1.9	22.3	2.4	<b>100</b>	<b>100</b>	<b>91</b>
8	<i>Spirochaetes</i>	<i>Spirochaetaceae</i>	<i>Treponema lf</i>	981	1.8	24.1	1.9	40	63	<b>100</b>
9	<i>Firmicutes</i>	<i>Lachnospiraceae</i>	Uncultured 65	1268	1.8	25.9	3.9	<b>100</b>	<b>88</b>	<b>100</b>
10	<i>Proteobacteria</i>	<i>Campylobacteraceae</i>	<i>Arcobacter</i>	226	1.6	27.5	1.1	53	50	55
11	<i>Bacteroidetes</i>	<i>Blattabacteriaceae</i>	<i>Blattabacterium</i>	665	1.5	29.0	2.2	<b>100</b>	<b>75</b>	64
12	<i>Spirochaetes</i>	<i>Spirochaetaceae</i>	<i>Treponema ll</i>	185	1.5	30.6	1.2	47	<b>75</b>	55
13	<i>Bacteroidetes</i>	<i>Bacteroidaceae</i>	<i>Bacteroides</i>	286	1.3	31.9	1.5	<b>100</b>	63	<b>73</b>
14	<i>Bacteroidetes</i>	<i>Rikenellaceae</i>	<i>Alistipes 1</i>	422	1.3	33.2	1.7	<b>93</b>	<b>75</b>	<b>73</b>
15	<i>Bacteroidetes</i>	<i>Porphyromonadaceae</i>	<i>Parabacteroides</i>	439	1.2	34.4	2.0	<b>100</b>	<b>100</b>	64
16	<i>Bacteroidetes</i>	<i>Porphyromonadaceae</i>	<i>Dysgonomonas</i>	628	1.2	35.5	2.1	<b>100</b>	<b>75</b>	<b>91</b>
17	<i>Firmicutes</i>	<i>Lachnospiraceae</i>	<i>Cand. Arthromitus</i>	1647	1.0	36.5	2.1	<b>100</b>	<b>88</b>	<b>100</b>
18	<i>Bacteroidetes</i>	Insect cluster I	Insect Cluster	2	1.0	37.5	0.7	53	38	27
19	<i>Bacteroidetes</i>	Cluster V	Subcluster Ve	1	1.0	38.4	1.5	<b>100</b>	50	<b>73</b>
20	<i>Proteobacteria</i>	<i>Desulfovibrionaceae</i>	<i>Desulfovibrio 3</i>	769	1.0	39.4	2.1	<b>100</b>	<b>88</b>	<b>100</b>
21	<i>Firmicutes</i>	<i>Ruminococcaceae</i>	Insect Cluster 1	84	0.9	40.3	1.9	<b>100</b>	63	<b>91</b>
22	<i>Bacteroidetes</i>	<i>Porphyromonadaceae</i>	<i>Paludibacter</i>	98	0.8	41.1	1.8	<b>100</b>	<b>100</b>	<b>100</b>
23	<i>Firmicutes</i>	<i>Lachnospiraceae</i>	Gut Cluster 2	1728	0.8	41.9	1.8	<b>100</b>	<b>100</b>	<b>100</b>
24	<i>Actinobacteria</i>	<i>Corynebacteriaceae</i>	<i>Corynebacterium</i>	8	0.8	42.7	0.5	53	25	18
25	<i>Firmicutes</i>	<i>Ruminococcaceae</i>	Gut Cluster 1	61	0.8	43.5	1.7	<b>100</b>	<b>88</b>	<b>100</b>
26	TG3	Termite cluster	Termite Cluster I	286	0.8	44.3	0.6	47	38	<b>82</b>
27	<i>Firmicutes</i>	<i>Ruminococcaceae</i>	Uncultured 23	407	0.7	45.0	0.6	67	50	<b>82</b>

Rank	Phylum	Family	Genus level	No. of OTUs (3%)	Contribution (%) <sup>a</sup>		Read abund- ance (%) <sup>b</sup>	Representation in hosts (%) <sup>c</sup>		
					Indivi- dual	Cumula- tive		Cock- roaches	Lower termites	Higher termites
28	<i>Synergistetes</i>	<i>Synergistaceae</i>	Uncultured 6	297	0.7	45.7	1.2	<b>100</b>	<b>100</b>	<b>100</b>
29	<i>Firmicutes</i>	<i>Ruminococcaceae</i>	<i>Anaerotruncus</i>	110	0.7	46.3	0.8	<b>100</b>	<b>75</b>	<b>100</b>
30	<i>Proteobacteria</i>	<i>Rhodocyclaceae</i> 3	Termite cluster	157	0.6	47.0	0.4	40	<b>75</b>	55
31	<i>Firmicutes</i>	Family XIII	Uncultured 5	269	0.6	47.6	1.2	<b>100</b>	<b>100</b>	<b>100</b>
32	<i>Firmicutes</i>	<i>Lachnospiraceae</i>	<i>Robinsoniella</i>	1	0.6	48.2	0.4	20	25	64
33	<i>Firmicutes</i>	<i>Ruminococcaceae</i>	Uncultured 12	529	0.6	48.8	1.3	<b>93</b>	<b>100</b>	<b>100</b>
34	<i>Firmicutes</i>	<i>Streptococcaceae</i>	<i>Lactococcus</i> 1	10	0.6	49.3	0.5	<b>73</b>	50	64
35	<i>Bacteroidetes</i>	<i>Flavobacteriaceae</i> 2	Uncultured a	113	0.5	49.9	0.4	40	50	18
36	<i>Firmicutes</i>	<i>Ruminococcaceae</i>	Uncultured 30	762	0.5	50.4	0.9	<b>93</b>	<b>88</b>	<b>100</b>
37	<i>Bacteroidetes</i>	Cluster V	Subcluster Vb	7	0.5	50.9	0.5	7	13	<b>91</b>
38	<i>Fibrobacteres</i>	Termite cluster	Termite cluster II	181	0.5	51.4	0.3	40	50	<b>73</b>
39	<i>Firmicutes</i>	<i>Ruminococcaceae</i>	Uncultured 36	632	0.5	51.8	0.9	<b>100</b>	50	<b>100</b>
40	<i>Proteobacteria</i>	<i>Acetobacteraceae</i>	<i>Acidisoma</i>	4	0.5	52.3	0.2	20	0	9
41	<i>Fusobacteria</i>	<i>Leptotrichiaceae</i>	<i>Sebaldella</i>	103	0.4	52.7	0.5	<b>80</b>	63	55
42	<i>Bacteroidetes</i>	<i>Rikenellaceae</i>	<i>Alistipes</i>	42	0.4	53.2	0.4	<b>73</b>	25	55
43	<i>Bacteroidetes</i>	<i>Marinilabiaceae</i>	Uncultured 1	74	0.4	53.6	0.5	<b>93</b>	38	27
44	<i>Firmicutes</i>	<i>Lactobacillaceae</i>	<i>Lactobacillus</i> 5	43	0.4	54.0	0.3	60	38	0
45	<i>Proteobacteria</i>	<i>Neisseriaceae</i> 1	Uncultured a	1	0.4	54.4	0.2	7	0	0
46	<i>Firmicutes</i>	<i>Veillonellaceae</i>	Uncultured 7	155	0.4	54.7	0.8	<b>100</b>	<b>75</b>	<b>73</b>
47	<i>Spirochaetes</i>	<i>Spirochaetaceae</i>	<i>Spirochaeta</i> 1	73	0.4	55.1	0.4	<b>73</b>	50	55
48	<i>Firmicutes</i>	<i>Enterococcaceae</i>	<i>Enterococcus</i> 2	24	0.4	55.5	0.4	<b>80</b>	50	9
49	<i>Firmicutes</i>	<i>Veillonellaceae</i>	<i>Phascolarctobacterium</i>	30	0.4	55.9	0.4	67	38	18
50	<i>Proteobacteria</i>	<i>Desulfovibrionaceae</i>	<i>Desulfovibrio</i> TC I	78	0.4	56.3	0.4	47	<b>100</b>	36
51	<i>Fibrobacteres</i>	Termite cluster	Termite cluster I	166	0.4	56.6	0.3	13	25	<b>73</b>
52	<i>Bacteroidetes</i>	<i>Porphyromonadaceae</i>	Cockroach cluster	1	0.4	57.0	0.4	<b>80</b>	38	27
53	<i>Proteobacteria</i>	Cluster 1	Insect cluster	1	0.4	57.3	0.6	<b>100</b>	38	55
54	<i>Firmicutes</i>	<i>Lactobacillaceae</i>	<i>Lactobacillus</i> 2	12	0.3	57.7	0.3	53	25	55
55	<i>Verrucomicrobia</i>	<i>Verrucomicrobiaceae</i>	<i>Akkermansia</i>	19	0.3	58.0	0.4	60	25	18
56	<i>Bacteroidetes</i>	<i>Rikenellaceae</i>	Insect cluster	2	0.3	58.3	0.4	<b>100</b>	<b>100</b>	<b>73</b>
57	<i>Synergistetes</i>	<i>Synergistaceae</i>	Uncultured 2	14	0.3	58.7	0.2	47	13	0
58	<i>Bacteroidetes</i>	<i>Porphyromonadaceae</i>	<i>Odoribacter</i>	26	0.3	59.0	0.2	40	25	27
59	<i>Firmicutes</i>	<i>Lactobacillaceae</i>	<i>Lactobacillus</i> 4	17	0.3	59.3	0.4	67	25	9
60	<i>Firmicutes</i>	<i>Lachnospiraceae</i>	Gut cluster 1	61	0.3	59.6	0.5	<b>100</b>	63	55
61	<i>Firmicutes</i>	<i>Ruminococcaceae</i>	<i>Papillibacter</i>	143	0.3	59.8	0.5	<b>87</b>	<b>75</b>	<b>91</b>

Rank	Phylum	Family	Genus level	No. of OTUs (3%)	Contribution (%) <sup>a</sup>		Read abund- ance (%) <sup>b</sup>	Representation in hosts (%) <sup>c</sup>		
					Indivi- dual	Cumula- tive		Cock- roaches	Lower termites	Higher termites
62	<i>Actinobacteria</i>	<i>Coriobacteriaceae</i>	Uncultured 10	170	0.3	60.1	0.6	<b>93</b>	<b>100</b>	<b>100</b>
63	<i>Bacteroidetes</i>	<i>Porphyromonadaceae</i>	Higher termite cluster	2	0.3	60.4	0.2	13	0	55
64	TG3	Termite cluster	Termite cluster II	152	0.3	60.6	0.2	7	13	<b>73</b>
65	<i>Proteobacteria</i>	<i>Campylobacteraceae</i>	<i>Sulfurospirillum</i>	125	0.3	60.9	0.3	53	50	55
66	<i>Bacteroidetes</i>	<i>Porphyromonadaceae</i>	<i>Tannerella</i>	209	0.2	61.1	0.4	<b>93</b>	63	<b>100</b>
67	<i>Bacteroidetes</i>	<i>Porphyromonadaceae</i>	<i>Bacteroidales</i> Cluster Vg	23	0.2	61.4	0.5	<b>100</b>	<b>88</b>	<b>82</b>
68	<i>Firmicutes</i>	<i>Enterococcaceae</i>	<i>Enterococcus</i> 7	2	0.2	61.6	0.1	20	0	0
69	<i>Firmicutes</i>	<i>Peptococcaceae</i>	Uncultured 2	10	0.2	61.9	0.4	<b>93</b>	50	<b>82</b>
70	<i>Proteobacteria</i>	<i>Enterobacteriaceae</i>	<i>Pragia</i>	8	0.2	62.1	0.1	7	13	0
71	<i>Firmicutes</i>	<i>Ruminococcaceae</i>	Uncultured 35	380	0.2	62.3	0.4	<b>87</b>	<b>88</b>	<b>100</b>
72	<i>Proteobacteria</i>	<i>Desulfobulbaceae</i>	<i>Desulfobulbus</i>	80	0.2	62.5	0.4	<b>93</b>	38	<b>82</b>
73	<i>Firmicutes</i>	<i>Ruminococcaceae</i>	Incertae Sedis 1	1	0.2	62.7	0.2	60	<b>88</b>	<b>100</b>
74	<i>Firmicutes</i>	<i>Erysipelotrichaceae</i>	<i>Turicibacter</i>	59	0.2	62.9	0.3	0	13	<b>91</b>
75	<i>Firmicutes</i>	<i>Ruminococcaceae</i>	Uncultured 28	97	0.2	63.1	0.4	<b>87</b>	50	<b>73</b>
76	<i>Firmicutes</i>	<i>Ruminococcaceae</i>	Uncultured 25	115	0.2	63.4	0.4	<b>80</b>	<b>75</b>	<b>100</b>
77	<i>Proteobacteria</i>	<i>Enterobacteriaceae</i>	Enteric cluster	38	0.2	63.6	0.2	60	63	45
78	<i>Firmicutes</i>	<i>Streptococcaceae</i>	<i>Lactovum</i>	77	0.2	63.7	0.2	7	25	<b>73</b>
79	<i>Actinobacteria</i>	<i>Coriobacteriaceae</i>	Marine group	84	0.2	63.9	0.2	13	38	64
80	<i>Bacteroidetes</i>	Insect cluster II	Uncultured a	79	0.2	64.1	0.4	<b>87</b>	50	<b>91</b>
81	<i>Proteobacteria</i>	<i>Desulfobacteraceae</i>	<i>Desulfatiferula</i>	52	0.2	64.3	0.3	<b>93</b>	63	36
82	<i>Spirochaetes</i>	<i>Spirochaetaceae</i>	Higher termite cluster	40	0.2	64.5	0.2	47	<b>75</b>	<b>91</b>
83	<i>Firmicutes</i>	<i>Lachnospiraceae</i>	Incertae sedis 34	9	0.2	64.7	0.2	<b>87</b>	63	36
84	<i>Firmicutes</i>	<i>Clostridiaceae</i> 1	<i>Cand. Savagella</i> -related	56	0.2	64.9	0.2	67	38	<b>82</b>
85	<i>Proteobacteria</i>	<i>Rhodocyclaceae</i> 2	Uncultured	36	0.2	65.0	0.2	33	<b>88</b>	<b>91</b>
86	<i>Firmicutes</i>	<i>Ruminococcaceae</i>	Termite cluster	84	0.2	65.2	0.3	60	<b>88</b>	<b>73</b>
87	<i>Proteobacteria</i>	Cluster 1	Higher termite cluster	129	0.2	65.4	0.2	7	50	<b>100</b>
88	<i>Bacteroidetes</i>	<i>Rikenellaceae</i>	Insect Cluster	2	0.2	65.5	0.3	53	<b>100</b>	<b>91</b>
89	<i>Bacteroidetes</i>	<i>Rikenellaceae</i>	RC9	25	0.2	65.7	0.2	<b>80</b>	38	18
90	<i>Actinobacteria</i>	<i>Nakamurellaceae</i>	<i>Nakamurella</i>	2	0.2	65.9	0.1	0	0	9
91	<i>Firmicutes</i>	<i>Ruminococcaceae</i>	Incertae Sedis 4	1	0.2	66.0	0.4	<b>93</b>	<b>88</b>	<b>91</b>
92	<i>Planctomycetes</i>	<i>Planctomycetaceae</i>	Termite cluster	85	0.2	66.2	0.3	<b>87</b>	50	<b>73</b>
93	<i>Lentisphaerae</i>	<i>Victivallaceae</i>	<i>Victivallis</i>	70	0.2	66.4	0.3	<b>87</b>	38	36
94	<i>Firmicutes</i>	<i>Peptococcaceae</i>	<i>Desulfitibacter</i>	35	0.2	66.5	0.1	33	50	64
95	<i>Bacteroidetes</i>	<i>Rikenellaceae</i>	Gut cluster c	89	0.2	66.7	0.3	<b>93</b>	38	55

Rank	Phylum	Family	Genus level	No. of OTUs (3%)	Contribution (%) <sup>a</sup>		Read abund- ance (%) <sup>b</sup>	Representation in hosts (%) <sup>c</sup>		
					Indivi- dual	Cumula- tive		Cock- roaches	Lower termites	Higher termites
96	<i>Firmicutes</i>	<i>Leuconostocaceae</i>	<i>Weissella</i> 1	5	0.2	66.8	0.1	40	25	9
97	<i>Proteobacteria</i>	<i>Enterobacteriaceae</i>	<i>Morganella</i>	1	0.2	67.0	0.1	0	0	0
98	<i>Firmicutes</i>	<i>Leuconostocaceae</i>	<i>Leuconostoc</i>	2	0.2	67.2	0.1	20	13	0
99	<i>Bacteroidetes</i>	<i>Rikenellaceae</i>	Insect cluster 2	2	0.2	67.3	0.3	67	<b>75</b>	<b>91</b>
100	<i>Bacteroidetes</i>	Insect cluster II	uncultured b	81	0.2	67.5	0.2	13	25	64
119	<i>Firmicutes</i>	<i>Enterococcaceae</i>	<i>Enterococcus</i> 4	2	0.1	70.0	0.1	47	25	9
320	<i>Acidobacteria</i>	<i>Acidobacteriaceae</i>	Uncultured 5	1	<0.1	80.0	0.0	53	<b>100</b>	55
598	<i>Proteobacteria</i>	<i>Comamonadaceae</i>	<i>Ottowia</i>	2	<0.1	90.0	0.0	47	<b>75</b>	<b>91</b>
884	<i>Proteobacteria</i>	<i>Desulfarculaceae</i>	<i>Desulfarculus</i>	21	<0.1	100.0	0.0	0	13	9

<sup>a</sup>The contribution to the overall clustering in a PCA analysis, expressed as the fraction of the total variance explained by this particular genus.

<sup>b</sup>Fraction of reads in all samples belonging to a particular taxon (normalized samples).

<sup>c</sup>To increase sensitivity, datasets were not subsampled for the core analysis





## Dietary and phylogenetic determinants of bacterial gut community structure in higher termites

Aram Mikaelyan<sup>1</sup>, Carsten Dietrich<sup>1</sup>, Tim Köhler<sup>1</sup>,  
Katja Meuser<sup>1</sup>, David Sillam-Dussès<sup>2</sup> und Brune A<sup>1</sup>

*Affiliations:* <sup>1</sup>Max Planck Institute for Terrestrial Microbiology, 35043 Marburg, Germany; <sup>2</sup>Université Paris 13, Sorbonne Paris Cité, Laboratoire d'Éthologie Expérimentale et Comparée, Villetaneuse, France | *This Manuscript is submitted.* | *Contributions:* A.M. designed and carried out experiments, analyzed data and wrote the manuscript. C.D. designed and carried out experiments, analyzed data and contributed to the manuscript. T.K. designed experiments. K.M. carried out experiments. A.B. conceived the study, wrote the manuscript, and secured funding.

### Abstract

---

The microbial symbionts of termites play critical roles in the digestion of lignocellulose. Unlike wood-feeding lower termites, which are associated with cellulolytic flagellates, higher termites harbor an entirely prokaryotic gut microbiota and have extended the range of their diet to lignocellulosic plant litter in various stages of humification. Since the same dietary specializations have evolved in different evolutionary lineages of higher termites, these insects offer the unique opportunity to study potential drivers of microbial community structure in the intestinal environment. We assessed the influence of host phylogeny and diet on the composition of the termite gut microbiota by analyzing the bacterial microbiota in the hindgut of 19 higher termite species from different feeding guilds. Amplified 16S rRNA genes were sequenced with Illumina; sequence reads were taxonomically classified using a curated reference database, and were phylogenetically and statistically analyzed. The high similarity in the bacterial gut microbiota among the wood-feeding and humivorous members of different host lineages identified diet as a strong determinant of microbial community structure in the guts of higher termites. At higher taxonomic resolution, however, individual bacterial taxa showed a strong specificity for certain host groups, which suggests that they are coevolving with their respective hosts. Nevertheless, evidence of cocladogenesis was scarce and most bacterial lineages did not appear to be cospeciating with their respective hosts over a longer evolutionary time. Instead, the observed patterns of host restriction seem to be enforced by a combined selection by both microhabitat and ecological niche, and enhanced by the vertical transmission of symbionts facilitated by the social lifestyle.

## Introduction

Termites are highly derived eusocial cockroaches that specialize on a diet of lignocellulose. The historical transition from less-specialized cockroaches seems to have involved several digestive modifications, including a distinctive enlargement of the hindgut (Noirot 1995) and the acquisition of a highly specialized gut microbiota, both of which play critical roles in the digestion of plant material in the guts of termites (Brune 2014).

The primitive families of termites, collectively called the “lower” termites, are primarily wood-feeding specialists that harbor oxymonad and hypermastigid flagellates in their guts (Ohkuma et al. 2000). The termite gut communities also comprise several bacterial lineages, some of which are associated with flagellates as secondary symbionts (Ohkuma 2008), while others are free-living in the lumen or attached to the gut wall (Breznak & Pankratz 1977; Yang et al. 2005).

The most highly evolved family of termites is the *Termitidae*, referred as “higher” termites, which lost these protozoan symbionts and possess an entirely prokaryotic gut community (Brune 2014). Since their separation from lower termites 80 million years ago, higher termites have evolved in many important aspects, evidenced by their phylogenetic diversity and multitude of dietary specializations. Unlike lower termites, which comprise wood feeders, the different species of higher termites specialize on diets of wood, humus, leaf litter, grass, dung, or fungal mycelia (Donovan et al. 2001).

The tremendous diversity in feeding strategies of higher termites is correlated with stark differences not only in gut anatomy and physiology, but also in the composition and structure of the gut bacterial communities (Dietrich et al. 2014). More importantly, several phylogenetically distinct lineages of higher termites have independently evolved the ability to digest remarkably similar diets, such as wood or soil, and therefore, higher termites offer rare opportunities to assess the impact of factors such as host phylogeny and diet on the diversification of their gut communities.

Clone libraries constructed in past studies using near-full-length 16S rRNA genes obtained from the same or related hosts have identified the presence of phylogenetically related bacterial lineages in related species of higher termites. However, in addition to the poor sequencing depth afforded by the traditional Sanger sequencing approach, these surveys of gut community structure suffer from poor host taxon sampling. More recently, deep-sequencing technologies have paved the way for cost-effective and highly parallelized analysis of gut communities in multiple termite hosts (Dietrich et al. 2014; Otani et al. 2014). Furthermore, because of the greater sequencing depth afforded by this technology, several bacterial lineages that were previously thought to be restricted to certain termite hosts could be detected as rare inhabitants of other hosts (Dietrich et al. 2014). This implies that the detection of rare lineages in any given host is a function of sampling effort and could contribute to a finer understanding of the factors shaping bacterial community structure in higher termites. However, deep sequencing studies of higher termites conducted to date are characterized by inadequate termite taxon sampling and have been unsuccessful at disentangling the evolutionary effects of host phylogeny and diet on the gut microbiota.

In the current study, we expand upon existing surveys of the termite gut microbiota by using the most comprehensive taxon sampling of higher termites undertaken to date. We characterized the composition of these gut communities by Illumina sequencing of amplified V3–V4 regions of the 16S rRNA genes, and taxonomically analyze the reads using a phylogenetically curated reference database (DictDB), tailor-made for accurate identification of bacterial lineages specific to termite guts. We critically assessed the distribution of these important bacterial lineages among different dietary and phylogenetic groups of higher termites. Additionally, we analyzed the phylogenetic patterns within these bacterial lineages to enable a better understanding of potential constraints imposed by host diet and phylogeny on gut community structure.

## Materials and Methods

**Sampling and dissection.** Termites (Tab. 1; 19 species) were sampled either from colonies maintained in the laboratory or collected in the field and were kept in plastic containers at room temperature and maintained with their nest material on specific diets for several months.

Insects were degutted with fine-tipped forceps, and the hindguts were separated from the rest of the gut with a scalpel. Hindguts (10–20) from each termite species were pooled in 2-ml tubes containing 750  $\mu$ l sodium phosphate buffer (120 mM; pH 8.0) and homogenized. DNA from the pooled samples was extracted and purified using a bead-beating protocol as previously described (Paul et al. 2012).

**Tab. 1 | Gut microbiota samples used in this study.**

ID	Host species	Dietary group	Reference
<b>Macrotermitinae</b>			
1	<i>Macrotermes</i> sp.	Fungus	Dietrich et al. (2014)
2	<i>Macrotermes subhyalinus</i>	Fungus	Dietrich et al. (2014)
3	<i>Odontotermes</i> sp.	Fungus	Dietrich et al. (2014)
<b>Syntermitinae</b>			
4	<i>Cornitermes</i> sp.	Soil/wood	This study <sup>a</sup>
<b>Apicotermitinae</b>			
5	<i>Apicotermes trestus</i>	Soil	Dietrich et al. (2014)
<b>Termitinae</b>			
6	<i>Termes hospes</i>	Soil	This study <sup>a</sup>
7	<i>Cubitermes</i> sp.	Soil	This study <sup>b</sup>
8	<i>Cubitermes ugandensis</i>	Soil	Dietrich et al. (2014)
9	<i>Amitermes</i> sp.	Grass	Dietrich et al. (2014)
10	<i>Microcerotermes parvus</i>	Wood	This study <sup>a</sup>
11	<i>Microcerotermes</i> sp.	Wood	Dietrich et al. (2014)
12	<i>Neocapritermes taracua</i>	Soil	This study <sup>a</sup>
13	<i>Ophiotermes</i> sp.	Soil	Dietrich et al. (2014)
14	<i>Proboscitermes</i> sp.	Soil	This study <sup>a</sup>
<b>Nasutitermitinae</b>			
15	<i>Atlantitermes</i> sp.	Soil/wood	This study <sup>a</sup>
16	<i>Nasutitermes corniger</i>	Wood	Dietrich et al. (2014)
17	<i>Nasutitermes takasagoensis</i>	Wood	Dietrich et al. (2014)
18	Unclassified <i>Nasutitermitinae</i>	Soil	This study <sup>a</sup>
19	<i>Trinervitermes</i> sp.	Grass	Dietrich et al. (2014)

<sup>a</sup> Lab colony of D. Sillam-Dusseze, Bondy, France.

<sup>b</sup> Field collection in South Africa by M. Poulsen.

**iTag library sequencing.** 16S rRNA genes in each sample were amplified using the universal bacterial primers 343Fmod and 784Rmod (Köhler et al. 2012). Purified PCR products were mixed in equimolar amounts and sequenced commercially (paired-end; 2 × 350 nt) (Illumina MiSeq; GATC Biotech, Konstanz, Germany).

**Processing of sequences.** The sequences from the iTag libraries and the downloaded sequences from Dietrich et al. (2014) were processed as previously described by Edgar (2013). Briefly, reads with a minimum length of 250 bp and a maximum expected error of 0.5 were selected and demultiplexed using their barcode sequences (no mismatch allowed). After removal of regions containing the barcodes and primers, the sequence pairs were merged when applicable by the supplied scripts and software of Edgar (2013). Sequences in each sample were clustered at a threshold of 99% similarity (most abundant sequence in cluster as preferred representative) using *UPARSE* (Edgar, 2013). Sequences were then dereplicated and aligned using the *mothur* aligner.

**Classification of sequences.** Aligned sequences were assigned to taxonomic groups using the naïve Bayesian classifier implemented in *mothur* (Schloss et al., 2009) at a confidence threshold of 80% in combination with a manually curated reference database DictDb v.3.5 of bacterial lineages specific to termite guts (Mikaelyan et al., 2015).

**Ordination analyses of community similarity.** The community similarity among all 19 species was calculated at the genus level based on the Soergel metric, which has proven to be a good metric for microbial ecology (Parks and Beiko, 2013). The pairwise distances were then subjected to principal coordinate analysis (PCoA) using the *vegan* package (Oksanen et al., 2013). To identify bacterial genera contributing most to the dissimilarities observed among communities, we subjected the communities to principal component analysis (PCA), and all taxa were ranked by their cumulative contribution to the components obtained in the PCA (Dietrich et al., 2014).

**Phylogenetic analysis of iTag sequences affiliated to *Spirochaetes*, *Fibrobacteres*, and *Candidate division Termite group 3 (TG3)*.** For the phylogenetic analyses of iTAG sequences, those sequences that were classified as *Fibrobacteres*, TG3, or *Spirochaetes* at the phylum level were imported into the DictDb reference database in the ARB software environment, and phylogenetically placed into the guide tree using the quick-add-by-parsimony tool. *FastTree* was used to estimate a maximum-likelihood tree for the selected iTAG sequences. Full-length 16S rRNA gene sequences from publicly available clone libraries were also included in the analysis as reference sequences.

## Results and Discussion

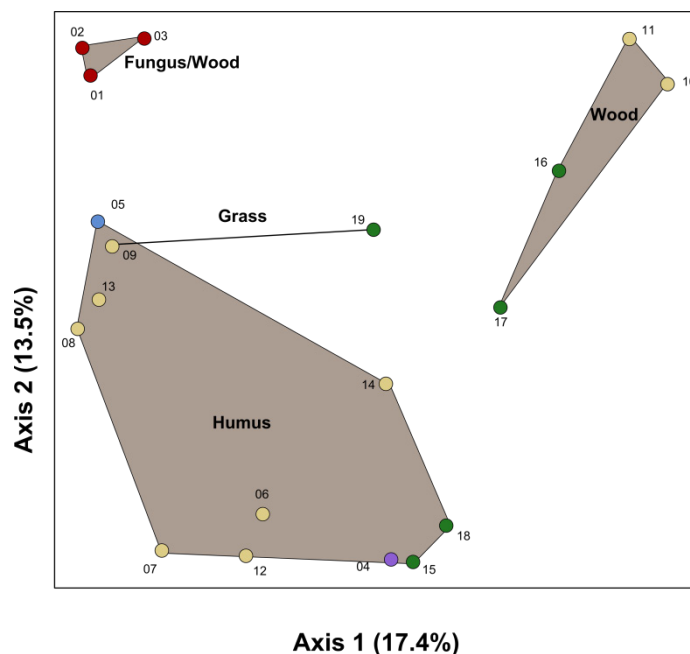
Using the Illumina sequencing platform for 16S rRNA amplicon analysis, we obtained over 250,000 high-quality reads from 19 different host species, yielding approximately 13,400 sequences per sample for downstream analysis (Tab. S1). For taxonomic analysis of the datasets, we utilized the RDP classifier to assign the quality-filtered sequences to the genus level in combination with DictDb, a reference training set developed specifically for the accurate classification of short bacterial 16S rRNA reads from termite guts (Mikaelyan et al., 2015). In total, 99% of the reads could be assigned to 22 phyla in the DictDb taxonomy. The dominant phyla included *Actinobacteria*, *Bacteroidetes*, *Fibrobacteres*, *Firmicutes*, *Spirochaetes*, *Synergistetes*, and the *Candidate divisions* TM7 and TG3, which represented on average 4, 15, 2, 36, 20, 3, 5, and 4% of the bacterial communities, respectively (Tab. S1).

**Host diet as a major driver of gut community structure.** Many gross differences in community structure observed among the hosts were apparent already at the phylum level, and primarily showed diet-specific trends (Tab. S1). Among the fungus-feeding termites, the majority of the reads could be assigned to *Firmicutes* and *Bacteroidetes*, which were relatively rare in the other diet groups. These phyla have been reported to characteristically dominate the gut communities of most fungus-feeding termites (Otani et al. 2014). Recent studies have revealed a stark similarity in gut community structure between fungus-feeding higher termites and their primitive relatives, the cockroaches — an unexpected similarity that possibly represents a convergent adaptation of the microbiota to the protein-rich diet of cockroaches and macrotermitine termites.

The datasets obtained from the wood-feeding termites revealed striking similarities in the distribution of abundant phyla. Despite their belonging to phylogenetically distinct subfamilies, the gut microbiotas of wood feeders from *Termitinae*, *Nasutitermitinae*, and *Syntermitinae* were dominated by members of *Spirochaetes*, *Fibrobacteres*, and TG3. The phylum *Fibrobacteres* was relatively abundant in the library from *Nasutitermes corniger* (forming 20.8% of the community), but it was considerably rarer in the library from *Cornitermes* sp. (subfamily *Syntermitinae*) and *Microcerotermes* spp. (subfamily *Termitinae*). By contrast, in wood-feeding termites where members of *Fibrobacteres* were only low in abundance, members of the TG3 were highly abundant (e.g., *Microcerotermes parvus*; Tab. S1). However, the basis of this variation among the wood feeders in the relative proportions of *Fibrobacteres* and TG3 is unclear. A recent deep-sequencing analysis of the fiber-associated communities in both *Nasutitermes corniger* and *Nasutitermes takasagoensis* revealed that wood particles serve as abundant microhabitats for a specific cellulolytic bacterial community. Aside from the *Spirochaetes* that formed about 50% of the fiber-associated community, it was shown that the phyla *Fibrobacteres* and TG3 in the fiber fraction are almost ninefold as abundant as in the fiber-free fraction (Mikaelyan et al. 2014). The distribution of *Spirochaetes*, *Fibrobacteres*, and TG3 among the higher termites illustrates how a wood-feeding diet, by providing colonizable wood particles in the gut, can profoundly modulate overall gut community structure.

*Firmicutes* were relatively abundant among humus feeders (mean 41%) compared to termites specializing on wood (mean 10%), grass (mean 40%), and fungus (mean 28%). These results are in agreement with the predominance of the phylum in the alkaline guts of humivorous *Cubitermes* (Schmitt-Wagner et al. 2003a; Dietrich et al. 2014) and *Ophiotermes* (Dietrich et al. 2014) in previous studies. Moreover, although non-humivorous higher termites, such as the wood-feeding *Nasutitermes* spp. (Brune et al. 1995; Köhler et al. 2012) and *Microcerotermes parvus* (Brune et al. 1995), are not characterized by generally alkaline guts as is found in *Cubitermes* spp. (Brune & Köhl 1996), they do

possess alkaline first proctodeal (P1) compartments. Previous surveys of the bacterial microbiota of both hyperalkaline and alkaline P1 compartments of several higher termites have revealed a correlation between alkalinity and high proportions of *Firmicutes* (Schmitt-Wagner et al. 2003b; Thongaram et al. 2005; Köhler et al. 2012), which suggests that gut alkalinity might strongly select for a high relative abundance of certain *Firmicutes*. However, unlike the gut communities of many previously examined humus-feeding *Cubitermes* spp., the gut communities of *Proboscitermes* and *Termes hospes* (both members of *Termitinae*) and the more distantly related humus-feeding nasute *Atlantitermes* sp. showed a higher abundance of *Spirochaetes*. This is not entirely unexpected because *Cubitermes* and *Ophiotermes* are close relatives, and although extensively studied, represent only a small portion of the entire diversity of humivorous termites. Moreover, gut content analysis suggests that in contrast to true soil feeders like *Cubitermes* spp. and *Ophiotermes* spp., *Termes hospes* specializes on a diet of wood that has decayed to a friable and soil-like state (Sleaford et al. 1996; Donovan et al. 2001), and the same has been suggested for species of the genus *Atlantitermes* (Ackerman et al. 2009). Given that the abundance of *Spirochaetes* in wood-feeding termites (Hongoh et al. 2006; Köhler et al. 2012; Dietrich et al. 2014) is explained by their colonization of wood particles in the gut (Mikaelyan et al. 2014), their relative presence in humivorous termites could be explained by the proportion of sound wood in their diet.



**Fig. 1 | Dietary patterns of the higher termite gut microbiota.** Principal coordinate analysis of the pairwise Soergel metric between gut microbiotas. Original genus-level data can be found in Tab. S1. Polygons visually separate termite gut microbiota based on host diet. Colors of the points indicate the higher termite subfamily: red, *Macrotermitinae*; purple, *Syntermitinae*; blue, *Apicotermitinae*; yellow, *Termitinae*; green, *Nasutitermitinae*. The numbering of samples indicates the termite host species (see Tab. 1).

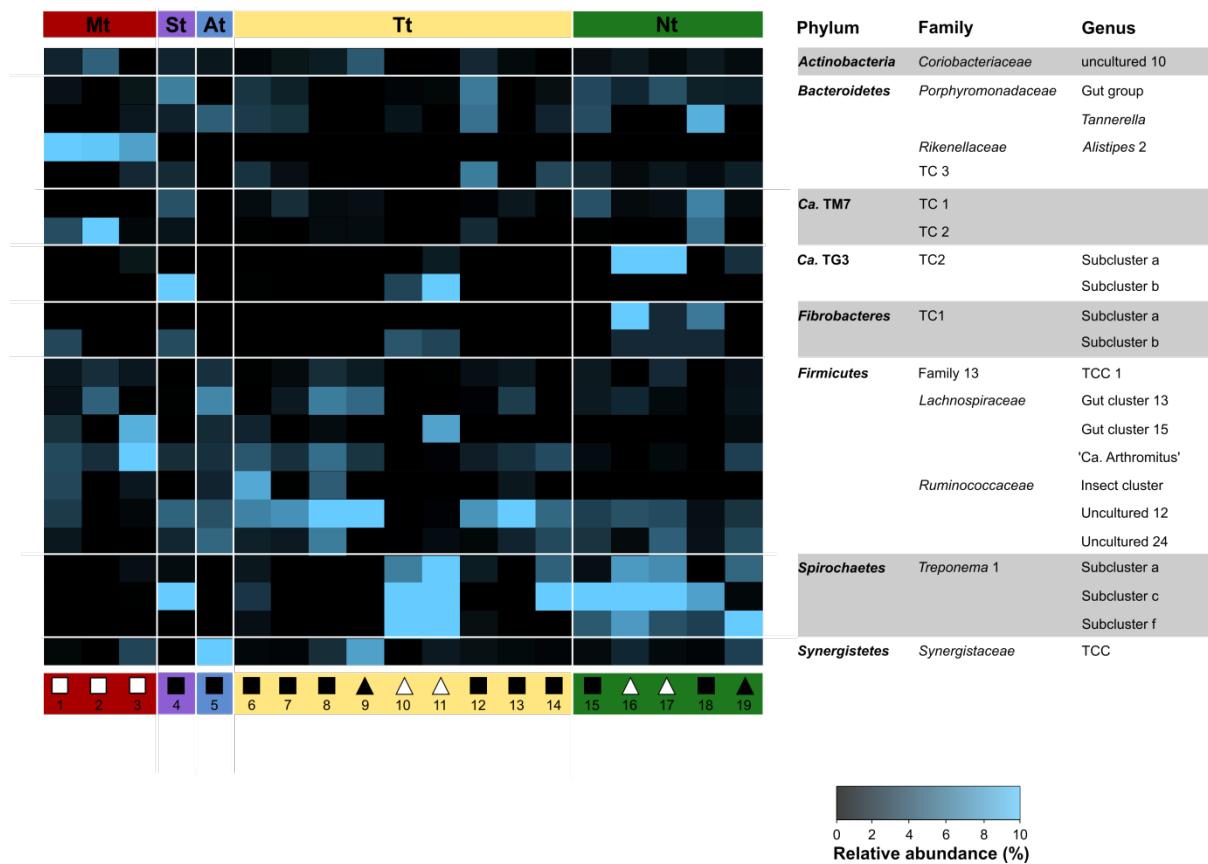
To determine how the diet-specific trends observed in the bacterial communities at the phylum level change with taxonomic resolution, we determined the overall dissimilarity among the 19 gut communities at the genus level. The RDP classifier confidently assigned 70% of the total high-quality reads to 267 genus-level groups (Tab. S1). We used the Soergel metric, which takes taxon abundances into account, to calculate pairwise distances among the different samples, and then visualized the community dissimilarity using PCoA (Fig. 1). In order to identify the genus-level taxa that contribute most to the observed clustering pattern shown in Fig. 1, we additionally conducted a PCA and ranked the genus-level taxa in descending order of their cumulative contribution to the loading factors. The relative abundances of the 20 most important taxa are summarized in a heatmap (Fig. 2).

Our ordination analyses also showed that the shared similarities, particularly in the distribution of genus-level lineages affiliated with *Bacteroidetes*, *Firmicutes*, and *Spirochaetes*, account for the diet-specific clustering observed in the PCoA analysis (Fig. 1). Among wood-feeding termites, gut communities were characterized by a higher richness (Tab. S1) and were dominated by a few genus-level groups, e.g., *Treponema* Ic and If (Family: *Spirochaetaceae*) and those affiliated to *Fibrobacteres* and TG3. In comparison, humivorous termites were characterized by a more even distribution of genus-level groups. Certain taxa, e.g., “Gut cluster 13” and ‘*Ca. Arthromitus*’ (both Family: *Lachnospiraceae*) and “uncultured 12” (Family: *Ruminococcaceae*), although ubiquitously distributed among the hosts, were relatively more abundant among the gut communities of humivorous *Termitinae*. The gut communities of the fungus-feeding termites (subfamily *Macrotermitinae*) also clustered firmly in a distinct group and were dominated by genus-level groups, such as “*Alistipes* 2”. However, because the *Macrotermitinae* are entirely composed of specialized fungus feeders (with the exception of *Sphaerothermes*), it is difficult to disentangle the influence of diet on the gut microbiota from the influence of host phylogeny. A recent and more detailed community analysis of fungus-feeding higher termites compared the community structure of several representatives with that of other termites and cockroaches, and suggested that diet could have played a role in the development of a unique gut microbiota, distinct from that of other higher termites (Otani et al. 2014).

The gut communities of the two grass-feeding termites did not cluster as firmly together as observed for the other diet groups. Both *Amitermes meridionalis* and *Trinervitermes* sp. clustered with their phylogenetic relatives in *Termitinae* and *Nasutitermitinae*, respectively. This pattern of clustering among the grass feeders could be suggestive of phylogenetic inertia, but could also be the result of subtle differences in the grass-based diets of the two termite species.

**Host-specificity of bacterial clades with higher termite subfamilies.** Major differences in the distribution of bacterial lineages at coarser taxonomic levels such as at the phylum-level explain simple several differences at finer taxonomic resolution, which contribute to the ordered separation of samples by diet in our PCoA analysis (Fig. 1). However, these large-scale differences that contribute to the separation among diet-specific clusters (Fig. 1) overshadow smaller, yet critical, taxonomic differences within each cluster of samples.

This phenomenon is illustrated well by the extremely host-restricted distribution of bacterial genus-level clades affiliated to phyla *Fibrobacteres* and TG3 among different higher termites. Hongoh et al. (2006) were the first to observe the exclusive association of monophyletic clades of *Fibrobacteres* and TG3 in their 16S rRNA clone libraries of higher termite genera. Our results showed that these genus-level bacterial clades are not exclusively restricted to termite genera but instead extend to termite subfamilies, which suggests that the pattern of host specificity is more diffuse than previously thought.



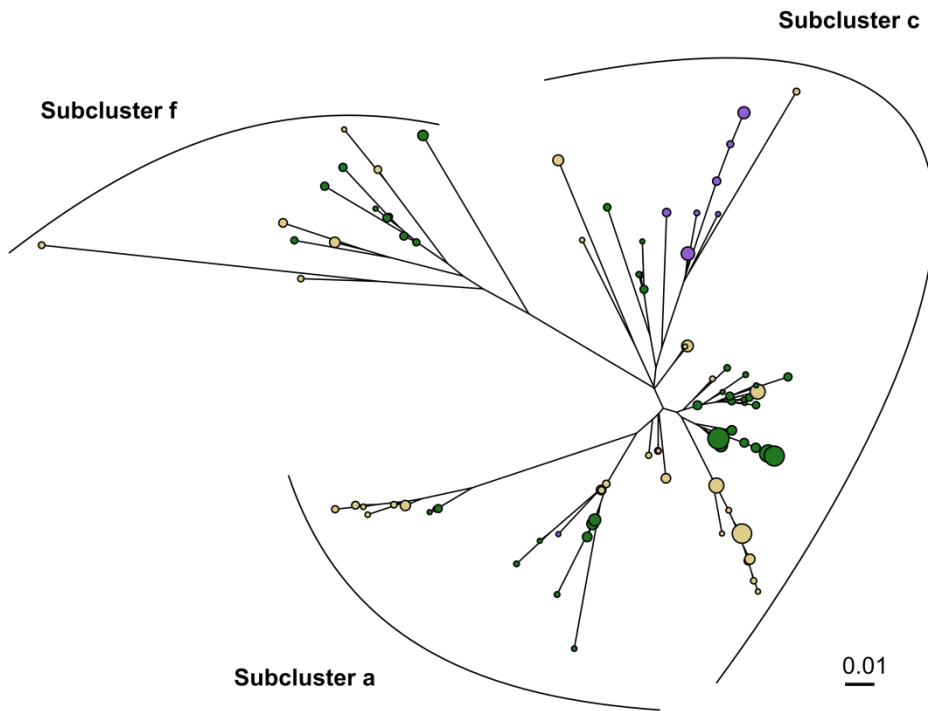
**Fig. 2 | Abundance of selected genus-level groups of the higher termite gut microbiota.** Sequences were classified with DictDB (Mikalelyan et al., unpublished results). Termite subfamily abbreviations: Mt, *Macrotermitinae*; St, *Syntermitinae*; At, *Apicotermitinae*; Tt, *Termitinae*; Nt, *Nasutitermitinae*. Symbols below the heatmap indicate diet: open rectangle, fungus; filled rectangle, soil; filled triangle, grass; open triangle, wood. The numbering of samples indicates the termite host species (see Tab. 1).

**A signal of host phylogeny in the gut community structure of higher termites.** The key question is thus whether the observed patterns of host specificity in *Fibrobacteres* and TG3 indicate that the bacterial clades are coevolving with their termite hosts. It is important to note that while host specificity is indicative of host specialization, it does not necessarily imply coevolution. These host-symbiont pairs could also be the result of an uptake of two distinct bacterial lineages adapted to different physicochemical environments in the different termite subfamilies. To explore the possibility of coevolution between the bacterial symbionts and higher termites, we constructed phylogenetic trees for select bacterial genus-level taxa of *Spirochaetes*, *Fibrobacteres*, and TG3 using representative operational taxonomic units (99%) from the Illumina libraries (Fig. 3).

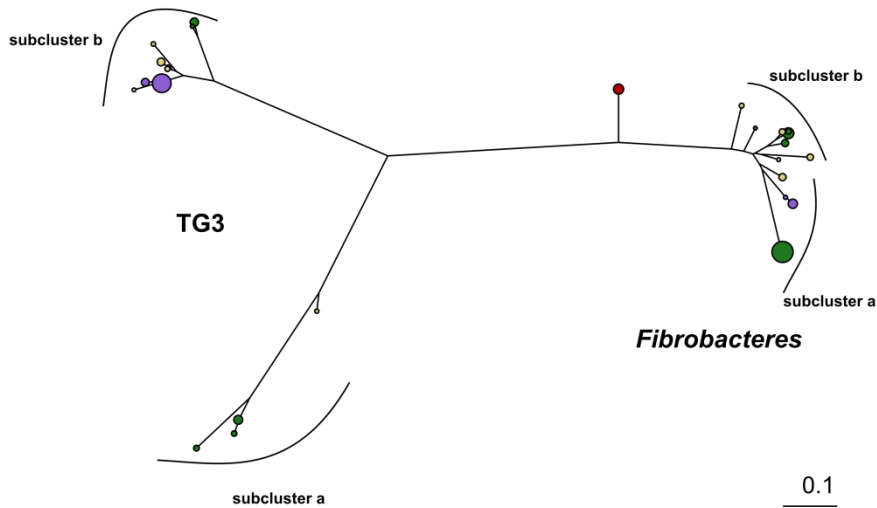
Our phylogenetic analyses revealed tremendous diversity underlying each of the different bacterial genera examined. In the case of the treponemes (Fig. 3) as well as the *Fibrobacteres*/TG3 clade, the phylotypes generally clustered by termite subfamily and not by diet. The terminal topology of the trees, however, did not always exactly mirror host phylogeny (Fig. 3). These results clearly showed a signal of host phylogeny and a role of phylogenetic inertia in the determination of the gut community composition in higher termites. This imprint of host phylogeny on gut community structure may be interpreted as a form of diffuse coevolution of the bacterial lineages with their host subfamilies.



**a *Treponema* I**



**b *Fibrobacteres* and TG3**



**Fig. 3 | 16S rRNA gene phylogeny of abundant higher termite gut bacterial lineages clustered at 99%.** (a) *Spirochaetes* genus-level groups *Treponema* I subcluster a,c and f. (b) Phylum *Fibrobacteres* and TG3. Size of the circles at the tips of the tree indicates relative abundance of the respective lineage. Color of the circle area indicates higher termite subfamily: red, *Macrotermitinae*; purple, *Syntermitinae*; blue, *Apicotermitinae*; yellow, *Termitinae*; green, *Nasutitermitinae*). Fig. S1 is the alternative of this figure and contains a more resolved termite host annotation.

## Conclusion

Our study is the first large-scale comparison of bacterial community structure in the guts of higher termites that covers a wide range of different higher termite genera and attempts to separate and understand the confounding roles of the host diet and phylogeny on the evolution of termite gut microbiota. While phylogenetic inertia clearly plays a role in the determination of community composition at finer taxonomic levels, the relative abundances of the bacterial lineages are greatly determined by the availability of microhabitats in the gut, which in turn is strongly influenced by host diet and the physiological conditions in the gut (e.g., pH, redox potential, and hydrogen partial pressure).

The question of whether the signal of host phylogeny observed in higher termites represents a form of diffuse coevolution and/or habitat selection still remains open. Nevertheless, the presence of phylogenetically related bacterial lineages in the gut communities of geographically isolated but phylogenetically related hosts strongly suggests that host genetic determinants play critical roles in constraining the composition of the communities in higher termites.

The findings of the current study expand our understanding of how host diet and phylogenetic history can impact the evolution of gut microbial communities in higher termites, and open the field for other important evolutionary and ecological questions, such as those concerning the role of the host in determining its gut microbiota, or the challenge of assigning roles to the different members of the gut communities. Future studies will also have to explore the extent of reciprocal adaptations between the host and its microbial lineages.

## Acknowledgements

This study was funded by a grant of the Deutsche Forschungsgemeinschaft (DFG) in the Collaborative Research Center SFB 987 and the LOEWE program of the state of Hessen (Synmikro). AM and CD each received a scholarship from the International Max Planck Research School for Environmental, Cellular and Molecular Microbiology (IMPRS-MIC). We thank Katja Meuser for technical assistance.

## References

- Ackerman IL, Constantino R, Gauch HG, Lehmann J, Riha SJ and Fernandes ECM (2009). Termite (Insecta: Isoptera) Species Composition in a Primary Rain Forest and Agroforests in Central Amazonia. *Biotrop*. 41:226–233.
- Breznak JA and Pankratz HS (1977). In Situ Morphology of the Gut Microbiota of Wood-Eating Termites [*Reticulitermes flavipes* (Kollar) and *Coptotermes formosanus* Shiraki]. *Appl. Environ. Microbiol.* 33:406–426.
- Brune A (2014). Symbiotic digestion of lignocellulose in termite guts. *Nat. Rev. Microbiol.* 12:168–80.
- Brune A and Kühl M (1996). pH profiles of the extremely alkaline hindguts of soil-feeding termites (Isoptera: Termitidae) determined with microelectrodes. *J. Ins. Physiol.* 42:1121–1127.
- Brune A, Emerson D and Breznak JA (1995). The termite gut microflora as an oxygen sink: microelectrode determination of oxygen and pH gradients in guts of lower and higher termites. *Appl. Environ. Microbiol.* 61: 2681–2687.
- Dietrich C, Köhler T and Brune A (2014). The cockroach origin of the termite gut microbiota: patterns in bacterial community structure reflect major evolutionary events. *Appl. Environ. Microbiol.* 80:2261–2269.
- Donovan SE, Eggleton P and Bignell DE (2001). Gut content analysis and a new feeding group classification of termites. *Ecolog. Entomol.* 26:356–366.
- Hongoh Y, Deevong P, Hattori S, Inoue T, Noda S, Noparatnaraporn N, Kudo T and Ohkuma M (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.
- Köhler T, Dietrich C, Scheffrahn RH and Brune A (2012). High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation of the gut in wood-feeding higher termites (*Nasutitermes* spp.). *Appl. Environ. Microbiol.* 78, 4691–4701.
- Mikaelyan A, Lampert N, Köhler T, Rohland J, Boga H, Meuser K and Brune A (2015). DictDb: an expanded reference database for the highly resolved classification of the bacterial gut microbiota of termites and cockroaches. Submitted manuscript.
- Mikaelyan A, Strassert JFH, Tokuda G and Brune A (2014). The fibre-associated cellulolytic bacterial community in the hindgut of wood-feeding higher termites (*Nasutitermes* spp.). *Environ. Microbiol.* 16:2711–2722.
- Noirot C (1995). The gut of termites (Isoptera). Comparative anatomy, systematics, phylogeny. *Ann. Soc. Entomol. Fra.* 31:197–226.
- Ohkuma M (2008). Symbioses of flagellates and prokaryotes in the gut of lower termites. *Trend. Microbiol.* 16:345–52.
- Ohkuma M, Ohtoko K, Iida T, Tokura M, Moriya S, Usami R, Horikoshi K and Kudo T (2000). Phylogenetic Identification of Hypermastigotes, Pseudotriconympha, Spirotriconympha, Holomastigotoides, and Parabasalian Symbionts in the Hindgut of Termites. *J. Euk. Microbiol.* 47:249–259.
- Oksanen J, Blanchet JG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH and Wagner H (2013). *vegan*: Community Ecology Package. <http://CRAN.R-project.org/package=vegan>.
- Otani S, Mikaelyan A, Nobre T, Hansen LH, Koné NA, Sørensen SJ, Aanen DK, Boomsma JJ, Brune A and Poulsen M (2014). Identifying the core microbial community in the gut of fungus-growing termites. *Mol. Ecol.* 23:4631–4644.
- Parks DH and Beiko RG (2013). Measures of phylogenetic differentiation provide robust and complementary insights into microbial communities. *ISME J.* 7:173–183.
- Paul K, Nonoh JO, Mikulski L and Brune A (2012). “Methanoplasmatales,” Thermoplasmatales-related archaea in termite guts and other environments, are the seventh order of methanogens. *Appl. Environ. Microbiol.* 78:8245–8253.
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski, RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn, DJ and Weber CF (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75:7537–7541.

- Schmitt-Wagner D, Friedrich MW, Wagner B and Brune A (2003a).** Phylogenetic diversity, abundance, and axial distribution of Bacteria in the intestinal tract of two soil-feeding termites (*Cubitermes* spp.). *Appl. Environ. Microbiol.* 69: 6007–6017.
- Schmitt-Wagner D, Friedrich MW, Wagner B, Brune A (2003b).** Axial dynamics, stability, and interspecies similarity of bacterial community structure in the highly compartmentalized gut of soil-feeding termites (*Cubitermes* spp.). *Appl. Environ. Microbiol.* 69:6018–6024.
- Sleaford F, Bignell DE and Eggleton P (1996).** A pilot analysis of gut contents in termites from the Mbalmayo Forest Reserve. *Cam. Ecol. Entomol.* 21:279–288.
- Thongaram T, Hongoh Y, Kosono S, Ohkuma M, Trakulnaleamsai S, Noparatnaraporn N, Kudo T (2005).** Comparison of bacterial communities in the alkaline gut segment among various species of higher termites. *Extremophyl.* 9:229–38.
- Yang H, Schmitt-Wagner D, Stingl U and Brune A (2005).** Niche heterogeneity determines bacterial community structure in the termite gut (*Reticulitermes santonensis*). *Environ. Microbiol.* 7: 916–932.

## Supplementary Material

**Tab. S1 | Interactive spreadsheet Tab. S1.** This table contains the genus-level classification, diversity data, and library sizes and can be downloaded from:

[https://dl.dropboxusercontent.com/u/50542577/Chapter\\_3\\_Table\\_S1.xlsx](https://dl.dropboxusercontent.com/u/50542577/Chapter_3_Table_S1.xlsx)

**Fig. S1 | Detailed trees of selected bacterial genus-level lineages.**

Please download from:

[https://dl.dropboxusercontent.com/u/50542577/Chapter\\_3\\_Short\\_Read\\_Trees.pdf](https://dl.dropboxusercontent.com/u/50542577/Chapter_3_Short_Read_Trees.pdf)

## Structure, spatial distribution, and succession of the bacterial community in the hindgut and fungus comb of the termite *Odontotermes formosanus* (Macrotermitinae)

Hongjie Li<sup>\*1</sup>, Carsten Dietrich<sup>\*2</sup>, Na Zhu<sup>1</sup>, Aram Mikaelyan<sup>2</sup>, Bin Ma<sup>3</sup>,  
Ruoxi Pi<sup>4</sup>, Yu Liu<sup>5</sup>, Mengyi Yang<sup>6</sup>, Andreas Brune<sup>2</sup> and Jianchu Mo<sup>1</sup>

\* These author contributed equally | *Affiliations:* <sup>1</sup>Ministry of Agriculture Key Laboratory of Molecular Biology of Crop Pathogens and Insects, Institute of Insect Sciences, Zhejiang University, Zhejiang, PR China; <sup>2</sup>Max Planck Institute for Terrestrial Microbiology, 35037 Marburg, Germany; <sup>3</sup>Yantai Institute of Coastal Zone Research, Chinese Academy of Sciences, Yantai, PR China; <sup>4</sup>Program in the Biological and Biomedical Sciences and Department of Microbial Pathogenesis, Yale University, New Haven, USA; <sup>5</sup>College of Life Sciences, Zhejiang University; <sup>6</sup>Xiaoshan Institute of Termite Control, Xiaoshan, Zhejiang, PR China. | *This manuscript is submitted.* | *Contributions:* H.L. conceived and designed the study, carried out experiments, analyzed data, discussed the results and wrote the manuscript. C.D. analyzed data, discussed the results and contributed to the manuscript. N.Z. carried out experiments. A.M. analyzed data and discussed the results. B.M. carried out experiments. R.P. carried out experiments. Y.L. carried out experiments. M.Y. carried out experiments. A.B. discussed the results and contributed to the manuscript. J.M. contributed to the manuscript.

### Abstract

The fungus cultivating termites are members of the subfamily *Macrotermitinae*, the most basal subfamily of higher termites. In addition to their obligate nutritional dependence on *Termitomyces* fungi, macrotermitine termites also harbor dense and diverse communities of bacterial symbionts in their hindguts that have been extensively investigated in the past. However, the hindguts of fungus-feeding termites are characterized by several radial physicochemical gradients that potentially describe distinct microhabitats, each supporting distinct bacterial communities and associated processes. Yet, almost nothing is known about how this environmental variation influences the local distribution of bacterial species in the gut. In this study, we investigated the bacterial communities associated with the major intestinal microhabitats, namely, the hindgut wall and hindgut fluid, of *Odontotermes formosanus* workers at different ages using pyrotag sequencing of 16S rRNA genes, and electron microscopy. We observed microhabitat differences in the distribution of bacterial lineages in the hindgut of *O. formosanus*, and that these microhabitat-associated communities mature with the age of the termites. These important structural dissimilarities among the communities were consistent with differences in density, morphology and spatial distribution of bacterial cells observed with electron microscopy, and well-supported by the differences observed in metabolite pools. Overall, our results clearly demonstrate that environmental heterogeneity strongly influences the preferential distribution of bacterial lineages and, by extension, the distribution of core metabolic processes in the gut of *O. formosanus*.

## Introduction

Termite guts provide distinctive niches for a diverse intestinal microbiota (Brune and Friedrich, 2000). In return, the symbionts contribute to the ecological success of the termite hosts by helping them gain access to recalcitrant diets, including lignocellulosic plant materials, dung, and humus (Brune and Ohkuma, 2011). This elaborate symbiosis in the tiny termite guts has long fascinated microbiologists. Research of the past decades has yielded systematic information on physicochemical conditions, microbial community structures, and metabolic and metagenomic profiles of guts of several termite species (Brune et al., 1995; Warnecke et al., 2007; Liu et al., 2013). Termite gut systems are highly compartmentalized environments with axial and radial physicochemical gradients as well as structurally and functionally different microbial populations, rendering them excellent models for microbial ecology studies (Brune, 2014).

In lower termites, the gut microbiota is spatially distributed in four microhabitats: midgut, intestinal protozoa (inside and attached to the surface), gut epithelium, and hindgut lumen (Brune, 2014). Higher termites lack intestinal protozoa but show an increased gut compartmentation which results in new habitats that colonized by a specific microorganism (Schmitt-Wagner and Brune, 1999). In soil-feeding *Cubitermes* species (Schmitt-Wagner et al., 2003a, b) and wood-feeding *Nasutitermes* species (Koehler et al., 2012), the microbial community structures of these compartments differ profoundly. Additional niches in higher termites are created by free wood particles, resulting in specific fiber-associated bacterial communities in the microbe-dominated P3 compartment (Mikaelyan et al., 2014).

Fungus-cultivating termites rely on a multi-partner microbial symbiosis consisting of the microbiota of the gut and a basidiomycete fungus (*Termitomyces* spp.) and a bacterial community on the fungus comb (Nobre et al., 2011). The gut physicochemical conditions of fungus-cultivating termites largely resemble those of lower termites, whereas the bacterial community is similar to that of omnivorous cockroaches (Hongoh et al., 2005; Long et al., 2010; Li et al., 2012; Liu et al., 2013; Dietrich et al., 2014; Otani et al., 2014). However, nothing is known about the spatial distribution of microbial populations in guts of fungus-cultivating termites, and the establishment of microbial populations has not been studied in any termite species. Fungus-cultivating termites provide an excellent system for studying establishment, succession, and spatial distribution of microbial populations as workers have different diets at different ages, i.e., age polyethism (Hinze et al., 2002b; Li et al., 2015), and the different age groups can be easily distinguished.

Here we combined scanning and transmission electron microscopy of midgut and hindgut paunch sections and pyrotag sequencing of bacterial 16S RNA genes to investigate the bacterial community structure in the luminal hindgut fluid and attached to the gut epithelium of newly molted, young, and old worker termites of *Odontotermes formosanus*, as well as in their new and old fungus combs, and determined fermentation products in the different gut compartments and in fungus combs.

## Materials and Methods

**Fungus-cultivating termites and fungal combs.** One colony of *Odontotermes formosanus* harboring the king and queen were collected in March 2013 from a forest area in Hangzhou City, Zhejiang Province, P.R. China. The entire colony containing fungus combs was wrapped in plastic film and transported to the laboratory within 6 h after excavation and maintained in complete darkness at  $27 \pm 1$  °C and 85% relative humidity. Colony was used for experiments for no more than one week.

*O. formosanus* colony had only one worker caste, in contrast to *Macrotermes* spp., which has two worker castes (major and minor) (Huang et al. 2000). The adult worker termites of *O. formosanus* were differentiated and sorted into three ages groups based on the color of their abdomen cuticles (Hinze et al., 2002a; Li et al., 2015); newly molted workers were light yellow, young workers were reddish brown, and old workers were dark brown (Fig. 1). Based on our previous observation data (Li et al., 2015), fungus combs of the termite colonies were separated into three age groups according to clearly distinguishable colors. New fungus combs were dark brown, middle-aged fungus combs were yellowish brown, and aged fungus combs were gray.

### Measurement of metabolite pools

After dissecting worker termites with sterile forceps, we carefully separated intact guts into three sections comprising the crop, midgut, and hindgut compartments. Forty gut sections of each age group of worker termites were pooled and homogenized with sterile micropestles in 10 mM NaOH. Additionally, 0.5 g of randomly collected new, middle-aged, and old fungus combs were each homogenized in 10 mM NaOH. Organic acids were measured using the ion chromatography system ICS-3000 (Dionex, USA) equipped with a Dionex IonPac AS11-HC column (Dionex, USA).

**Preparation of hindgut wall and hindgut lumen fractions and fungus comb for bacterial community analyses.** After worker termites were surface sterilized with 70% ethanol and washed in phosphate-buffered saline (PBS), intact guts of each age group were immediately dissected using sterile, fine-tipped forceps. The foregut and midgut sections were carefully removed, the remaining hindguts were sliced open using a sterile razor, and the gut contents were gently squeezed into PBS. The hindguts of each age group of workers were placed on a nylon mesh (100- $\mu$ m mesh size) and hindguts were washed at least three times with PBS as described previously (Nakajima et al., 2005). The washed hindguts comprised the hindgut fraction. The bacterial cells in the filtrate were collected by centrifugation and comprised the hindgut lumen fraction. Fifty hindguts of each age group were pooled, as were the bacterial cells from the filtrate of 50 hindguts. Samples of new and old fungus combs were randomly collected from each age fungus comb and pooled into age groups (0.2 mg total, respectively).

**Gut volume and morphology, and electron microscopy of hindgut fractions.** The volume of the crop, midgut, and hindgut sections was measured using a digital microscope (VHX-2000; KEYENCE, Osaka, Japan) and the geometrical shapes were estimated as previously described (Li et al. 2015). To determine the distribution of bacteria in the midgut and hindgut, gut sections of newly molted, young, and old workers were examined by scanning and transmission electron microscopy.

For samples for scanning electron microscopy, midgut wall fractions from five individuals and hindgut wall sections from five individuals of each age group of worker termites were rinsed with PBS and then fixed with 2.5% glutaraldehyde in 0.1 M phosphate buffer (pH 7.0) overnight. Samples were washed three times in 0.1 M phosphate buffer (pH 7.0) for 15 min each and then dehydrated with a graded series of ethanol (50%, 70%, 80%, 90% and 95%) for 15 min each, followed by 100% ethanol for

20 min. To each sample, a 1:1 (v/v) mixture of ethanol and isoamyl acetate was added, and the mixture was incubated for 30 min; each sample was then transferred to pure isoamyl acetate and the mixture was incubated for 1 h. Samples were dehydrated in a Hitachi Model HCP-2 critical point dryer with liquid CO<sub>2</sub>. The luminal side of the gut wall was placed face up and coated with gold-palladium. The samples were observed with a Siron 200 scanning electron microscope (FEI). Each micrograph shown is one of five intestinal tracts observed.

For transmission electron microscopy, midgut and hindgut paunch sections of each age group of worker termites were fixed with 2.5% glutaraldehyde in 0.1 M phosphate buffer (pH 7.0) overnight. Samples were then washed three times in 0.1 M phosphate buffer (pH 7.0) for 15 min each, followed by postfixing with 1% OsO<sub>4</sub> in 0.1 M phosphate buffer (pH 7.0) for 45 min. Samples were washed again as described above and then dehydrated with a graded series of ethanol (50%, 70%, 80%, 90%, 95% and 100%) for 15 min each and then incubated in absolute acetone for 20 min. Samples were infiltrated at room temperature successively with 1:1 (v/v) mixture of absolute acetone and Spurr resin for 1 h, 1:3 (v:v) mixture of absolute acetone and Spurr resin for 3 h, and Spurr resin overnight. For embedding and ultrathin sectioning, samples were placed vertically in 0.5 ml Eppendorf tubes containing Spurr resin and heated at 70 °C for 9 h. Samples were sectioned with a thickness of 5–10 µm, stained successively with uranyl acetate and alkaline lead citrate for 15 min each, and observed with a JEM-1230 transmission electron microscopy (JEOL). Each micrograph shown is one of five intestinal tracts observed.

**DNA extraction.** DNA was extracted from the pooled hindgut wall fractions and hindgut lumen fractions of 50 individuals of each age group. The pooled fractions were homogenized using plastic pestles until no obvious gut particles were observed and then subjected to the bead-beating protocol as described in Henckel et al. (Henckel et al., 1999). The lysate was extracted with phenol-chloroform-isoamyl alcohol (25:24:1, by vol.), and the extracted DNA was further purified with isopropanol and ethanol precipitation steps.

The new and aged fungus comb samples were ground to fine powder using a prechilled mortar and pestle. The ground fungus comb samples were subjected to bacterial cell lysis with bead-beating as described above. Humic substances in the lysis mixtures were removed by incubating in 10% cetyltrimethyl ammonium bromide (CTAB) and 0.7 M NaCl at 65 °C for 10 min (Hongoh et al., 2006). Then purify the mixtures with a Qiagen DNeasy column using the method described for crude lysate purification as recommended by the manufacturer. The mixture was further subjected to successively phenol-chloroform-isoamyl alcohol extraction, isopropanol and ethanol precipitation steps. Purified DNA from hindgut fractions and fungus combs were resuspended in 100 µl AE elution buffer (Qiagen) and quantified photometrically using NanoDrop ND-1000 (Thermo Scientific, Germany).

**Pyrotag sequencing.** The bacterial community in each hindgut wall and hindgut lumen fraction from workers of each age group and from new and old fungus combs was analyzed using 454 pyrotag sequencing. The V3–V4 region of bacterial 16S rRNA genes was amplified using the pre-optimized primer pair 343Fmod and 784Rmod, which enhances coverage of the taxa known to prevail in termite and cockroach guts, as described in (Koehler et al., 2012). For samples of old fungus combs, 30 PCR cycles were performed; for all other samples, 26 PCR cycles were performed. Amplicons were quantified photometrically (NanoDrop; Thermo Fisher Scientific, USA), adjusted to equimolar amounts, sequenced on a 454 Life Sciences Genome Sequencer GS FLX (Roche Diagnostics; BGI-Shenzen, China).



**Sequence processing and classification.** Erroneous reads in the pyrotag sequences were corrected using *Acacia* (version 1.52). The quality of the sequences was then improved by filtering under stringent conditions (reads > 200 bp, no ambiguous bases, and a maximum number of homopolymers  $\leq 8$ ) (Schloss et al., 2011). The resulting high-quality sequences were aligned using *mothur* software (version 1.29.0). Aligned sequences were taxonomically classified using the naïve Bayesian classifier with a bootstrap value of 60% as cutoff and the customized reference database DictDb v. 2.3 (Koehler et al., 2012). This reference database was built on the basis of the Silva database with additional sequences of termite and cockroach guts from published studies and unpublished data from the Brunei laboratory, which highly improves taxonomic resolution in the groups represented in termites and cockroaches (Dietrich et al., 2014; Mikaelyan et al., 2014). DictDb v. 2.3 is available upon request.

**Statistical analyses.** The similarity of bacterial communities between samples was estimated with weighted UniFrac distances and then visualized with principal coordinates analysis (PCoA) using *R* software (version 3.0.1). To determine the contributions of taxa to community dissimilarities, principal component analysis (PCA) was used for measuring occurrence and abundance of genus-level taxa in the entire data set using *R* software, and then taxa were ordered based on the contribution metric proposed by (Abdi and Williams, 2010).

**Tab. 1 | Pool size of major metabolites in gut fractions of worker termites of different ages and in fungus combs of different age.**

Sample	Fraction	Volume ( $\mu\text{l}$ ) <sup>a</sup>	Pool size (nmol section <sup>-1</sup> ) <sup>b</sup>					
			Acetate	Propionate	Butyrate	Succinate	Lactate	Formate
Newly molted worker	Crop	0.05 ± 0.01	0.6 ± 0.1	— <sup>d</sup>	0.1 ± 0.0	—	0.3 ± 0.1	0.1 ± 0.0
	Midgut	0.10 ± 0.04	1.2 ± 0.0	— <sup>d</sup>	0.1 ± 0.0	0.1 ± 0.0	0.3 ± 0.1	0.1 ± 0.0
	Hindgut	0.21 ± 0.08	3.2 ± 0.5	0.8 ± 0.4	0.5 ± 0.2	0.6 ± 0.1	0.3 ± 0.0	0.7 ± 0.2
	Total gut <sup>c</sup>	0.37 ± 0.12	4.9 ± 0.7	0.9 ± 0.4	0.7 ± 0.2	0.7 ± 0.2	0.9 ± 0.0	0.9 ± 0.2
Young worker	Crop	0.07 ± 0.04	1.0 ± 0.1	—	0.2 ± 0.0	0.1 ± 0.0	0.4 ± 0.1	0.1 ± 0.0
	Midgut	0.44 ± 0.33	2.1 ± 0.1	—	0.4 ± 0.1	0.1 ± 0.0	0.3 ± 0.1	0.1 ± 0.1
	Hindgut	1.24 ± 0.28	3.1 ± 0.1	0.2 ± 0.0	1.4 ± 0.2	0.3 ± 0.1	0.2 ± 0.0	0.1 ± 0.0
	Total gut	1.74 ± 0.49	6.1 ± 0.2	0.2 ± 0.0	1.9 ± 0.3	0.6 ± 0.1	0.9 ± 0.0	0.3 ± 0.1
Old worker	Crop	0.05 ± 0.02	0.9 ± 0.2	— <sup>d</sup>	0.2 ± 0.0	0.1 ± 0.0	0.4 ± 0.2	0.1 ± 0.0
	Midgut	0.79 ± 0.54	2.2 ± 0.1	— <sup>d</sup>	0.7 ± 0.1	0.1 ± 0.0	0.3 ± 0.0	0.2 ± 0.0
	Hindgut	1.20 ± 0.34	11.3 ± 2.4	1.3 ± 0.4	7.8 ± 0.4	0.4 ± 0.1	0.8 ± 0.1	1.8 ± 0.6
	Total gut	2.07 ± 0.76	14.5 ± 2.3	1.4 ± 0.5	8.7 ± 0.5	0.5 ± 0.0	1.6 ± 0.0	2.0 ± 0.6
Fungus comb	New		2.9 ± 0.07	0.1 ± 0.0	12.5 ± 0.2	0.1 ± 0.0	1.7 ± 0.0	1.0 ± 0.1
	Middle-aged		8.2 ± 0.4	0.1 ± 0.0	6.5 ± 0.2	0.4 ± 0.0	2.0 ± 0.7	2.2 ± 0.0
	Old		4.5 ± 0.0	0.1 ± 0.0	1.2 ± 0.1	1.0 ± 0.1	1.6 ± 0.1	1.8 ± 0.1
	Total		15.6 ± 0.5	0.2 ± 0.0	20.2 ± 0.0	1.5 ± 0.0	5.2 ± 0.8	5.0 ± 0.1

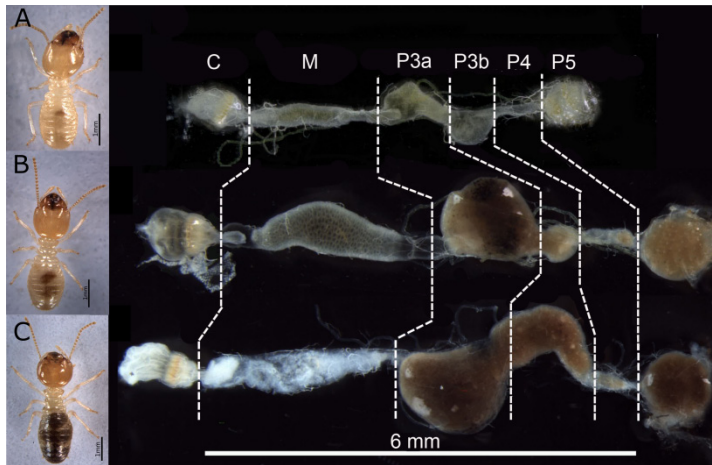
<sup>a</sup> Values are averages ( $\pm$  s.d.) of 15 independent measurements calculated by geometric approximation.

<sup>b</sup> Values are averages ( $\pm$  mean deviation) of two homogenates of 40 gut sections each (nmol section<sup>-1</sup>) or of two independent measurements of 0.5 g fungus comb each (nmol mg<sup>-1</sup>).

<sup>c</sup> Calculated from amount in each gut section or in each fungus comb fraction.

<sup>d</sup> Below detection limit (ca. 0.02 nmol).

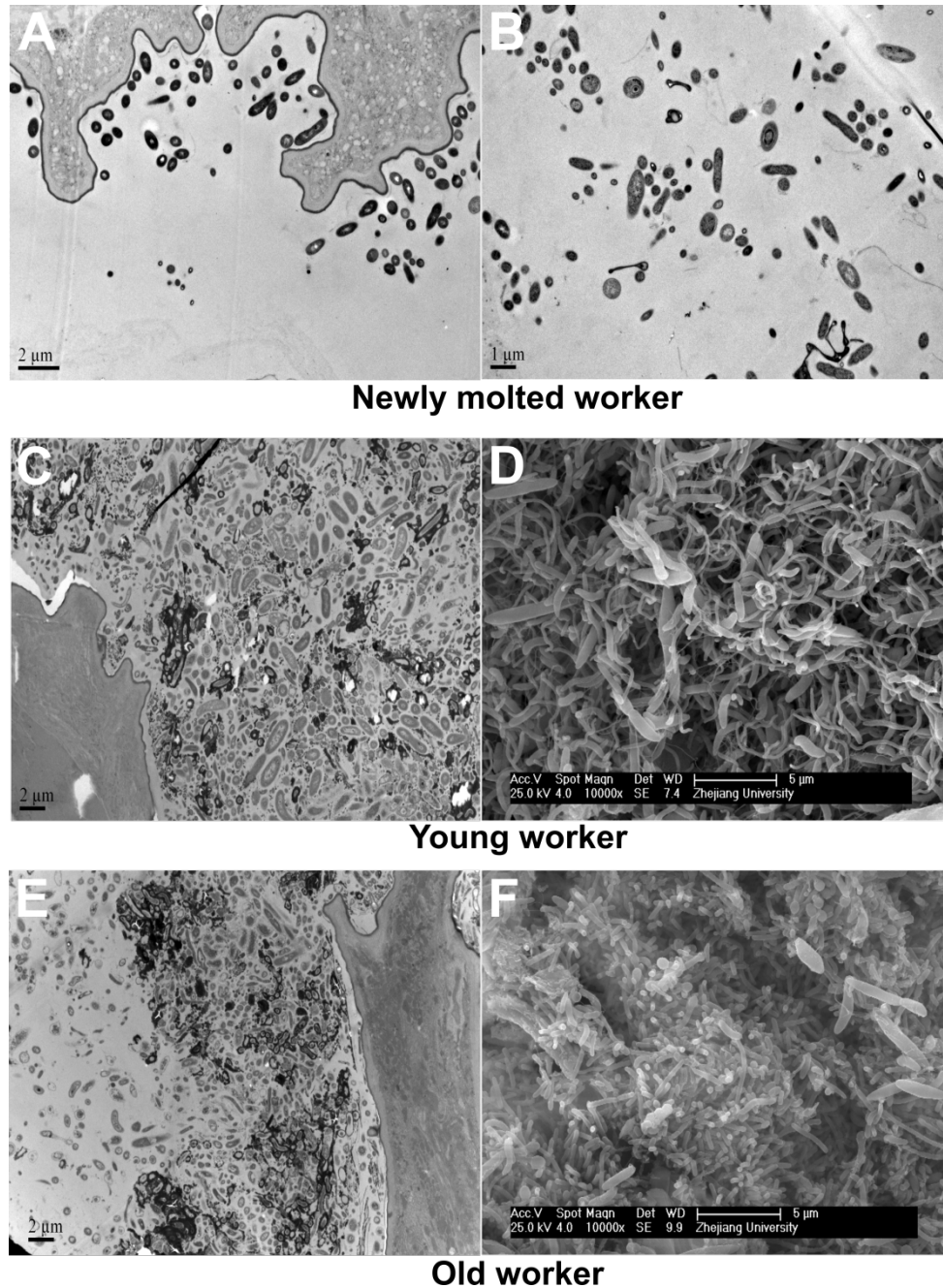
## Results



**Fig. 1 | Morphological characteristics of the cuticles and intestinal tract of different age groups of *Odontotermes formosanus*.** (A) a newly molted worker, (B) a young worker, and (C) an old worker. The typical worker termite gut consists of the crop (C), midgut (M), dilated part of paunch (P3a), more tubular part of paunch (P3b), colon (P4), and rectum (P5).

**Gut development in various age groups of worker termites.** The general arrangement of the alimentary tract of all three age groups of the worker caste of *Odontotermes formosanus* is similar (Fig. 1) and is composed of foregut, midgut, paunch, colon, and rectum. However, the dimensions and volumes of gut sections of workers of different ages clearly differed. The gut volume of newly molted workers ( $0.37 \pm 0.12 \mu\text{l}$ ) was much larger than that of young workers ( $1.74 \pm 0.49 \mu\text{l}$ ); the mean volume of individual gut compartments increased to different extents (Tab. 1): crop, 1.4-fold; midgut, 4.4-fold; hindgut, 5.9-fold; and paunch 11-fold; Tab. 1. The crop volume of young workers ( $0.07 \pm 0.04 \mu\text{l}$ ) was higher than that of both newly molted workers ( $0.05 \pm 0.01 \mu\text{l}$ ) and old workers ( $0.05 \pm 0.02 \mu\text{l}$ ). Also the volume percentage of the crop compared to the total gut volume of old workers (2.42%) was degenerated compared to that of young workers (4.02%). Young and old workers also differed in behavior and diet; young workers masticate plant material, while old workers feed on mature fungal combs, which might account for crop degeneration as the workers age.

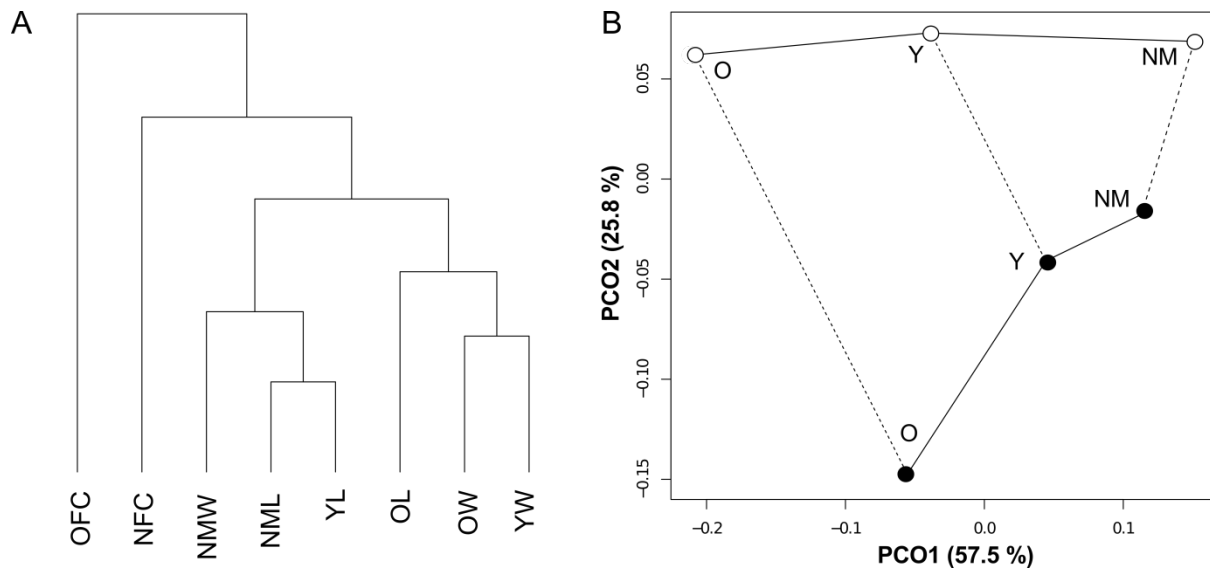
**Population size and morphology of the bacterial community in the hindgut paunch of workers of different ages.** Scanning and transmission electron microscopy of microbial communities in the hindgut paunch of termite workers of each age group revealed changes in density, morphology, and distribution of microbial cells. As the worker termites aged, the density of microbial cells residing in the hindgut paunch increased greatly from the newly molted workers to young workers, but then decreased from the young workers to old workers (Fig. 2). The hindgut paunch of young workers is populated by various types of readily identifiable filamentous or spiral-shaped bacterial cells (Fig. 2D), while that of old workers is densely inhabited by various types of cocci or rod-shaped bacterial cells (Fig. 2F), which indicates that as worker termites age, different groups of microbes dominate the hindgut paunch. In newly molted workers, bacterial cells are loosely associated with the hindgut wall (Fig. 2A), in young workers, the cells are evenly distributed on the wall and in the lumen (Fig. 2C), and in old workers, more bacterial cells are attached to or closely associated with the gut wall than are free in the lumen (Fig. 2E), which indicates that a large portion of the bacteria colonize the gut wall to form a biofilm as worker termites age. In the midgut lumen and on the midgut epithelium of both young workers and old workers, bacterial cells were absent or present only in low numbers (Fig. S1).



**Fig. 2 | Transmission and scanning electron micrographs of *Odontotermes formosanus* gut microbiota in the three different age groups.** (A, B) Newly molted worker; transmission electron micrographs of transverse section through (A) the peripheral paunch and (B) paunch lumen. (C, D) Young worker: (C) transmission electron micrograph of transverse section through the peripheral paunch and (D) scanning electron micrograph of the paunch epithelium showing attached microbiota. (E, F) Old worker: (E) transmission electron micrograph of transverse section through the peripheral hindgut paunch and (F) scanning electron micrograph of the hindgut paunch showing microbiota attached to the gut epithelium.

**Metabolite analysis.** We determined the metabolites that accumulated in different gut compartments and in different aged parts of the fungus comb of *O. formosanus* using ion chromatography (Tab. 1). Among all metabolites detected, acetate was predominant in all gut sections of all ages of workers. Among all gut sections sampled, the hindgut had the largest volume and also contained the largest pool of all metabolites detected, except for lactate in newly molted and young workers. In newly molted workers, equal amounts of lactate were detected in all three gut sections, and in young workers, lactate was slightly more abundant in the crop. The highest concentration of a metabolite in the fungus comb was that of butyrate in new fungus comb; its concentration decreased as the comb aged. The second-highest concentration of a metabolite was that of acetate in middle-aged combs. In old fungus combs, the concentration of acetate decreased, but was higher than that of all other metabolites.

**High-throughput sequencing of bacterial communities.** We determined the bacterial diversity in the hindgut lumen and on the hindgut wall of *O. formosanus* of different ages as well as on fungus combs of different ages by amplifying the V3–V4 region (about 450 bp) of 16S rRNA genes of extracted DNA and analyzing the products by 454 pyrosequencing. Trimming and quality control removed 10–20% of the sequences from each data set, resulting in sequence libraries of 15,020 to 59,149 reads per sample (Tab. S1). The obtained sequences were classified using the three databases, namely the Ribosomal Database Project (RDP; release 10), Silva database, and DictDB database; the latter supplements the Silva database with all homologous sequences previously obtained from termite and cockroach guts. The online RDP classifier yielded large fractions of unclassified sequences at all taxonomic levels for all samples except old fungus comb (Tab. S2). Most of these unclassified sequences were termite-specific bacterial groups that were unrepresented or poorly resolved in the RDP database. Classification using the Silva database greatly increased the fraction of sequences assigned for all gut samples and new fungus combs. Analysis using DictDB proved to be the most informative and allowed identification of groups not yet resolved in public databases.

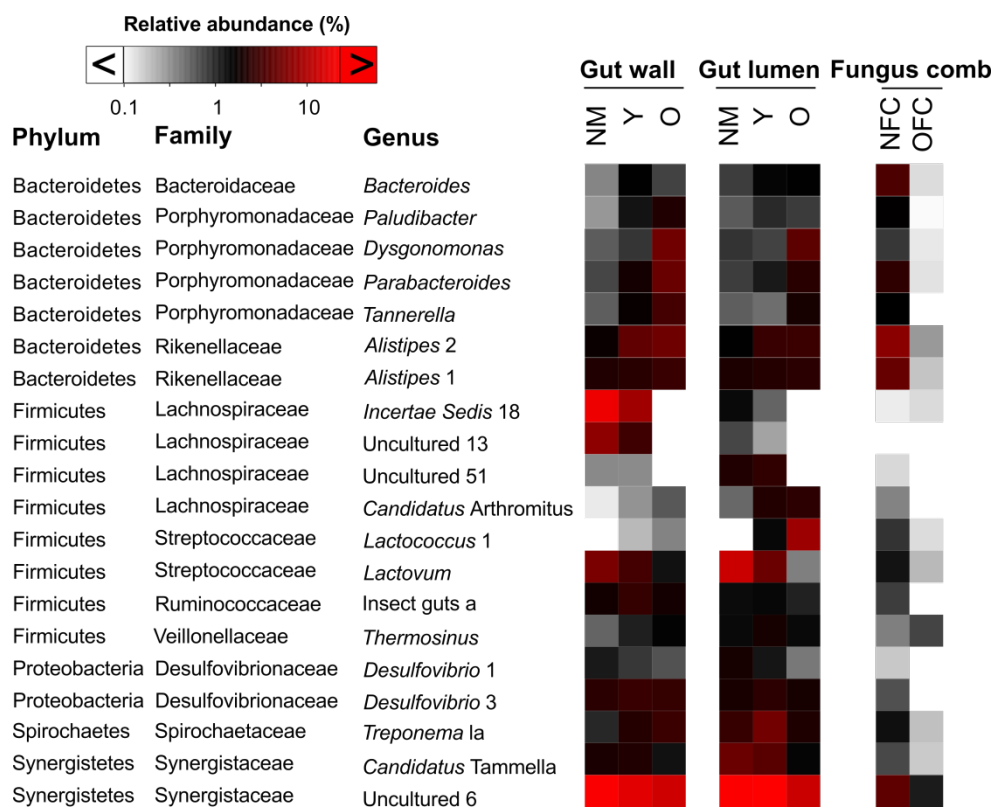


**Fig. 3 | Bacterial community profiles in the hindgut lumen and on the hindgut wall of *Odontotermes formosanus* workers of three age groups as well as new and old parts of their fungus comb.** (A) Dendrogram based on weighted UniFrac metric of bacterial communities across hindgut wall of newly molted workers (NMW), hindgut lumen of newly molted worker (NML), hindgut wall of young worker (YW), hindgut lumen of young worker (YL), hindgut wall of old worker (OW), hindgut lumen of old worker (OL), and new fungus comb (NFC) and old fungus comb (OFC). (B) Principal coordinates analysis (PCoA) visualizing bacterial community similarities across hindgut lumen (●) and hindgut wall (○) fractions from newly molted worker (NM), young workers (Y), and old workers (O).

**Microbial community structure.** Bacterial community similarity of the hindgut of different aged workers and of new and old fungus combs was calculated using weighted UniFrac distances and visualized by both a dendrogram (Fig. 3A) and dimensional reduction with principal coordinates analysis (PCoA; Fig. 3B). The dendrogram showed that the two fungus comb samples, especially the old fungus comb, were dissimilar to the six gut samples (Fig. 3A). The bacterial community of the hindgut lumen of young workers was most similar to that of newly molted workers, while the bacterial community of the hindgut wall of young workers was most similar to that of old workers (Fig. 3A). The PCoA plot (Fig. 3B) showed that the bacterial communities associated with the hindgut wall fraction and lumen fraction were highly dissimilar, a trend that developed further as the termites aged, i.e., the bacterial communities of the gut wall and lumen of newly molted workers were less dissimilar to each other than those of young workers, which in turn were less dissimilar to each other than those of old workers.

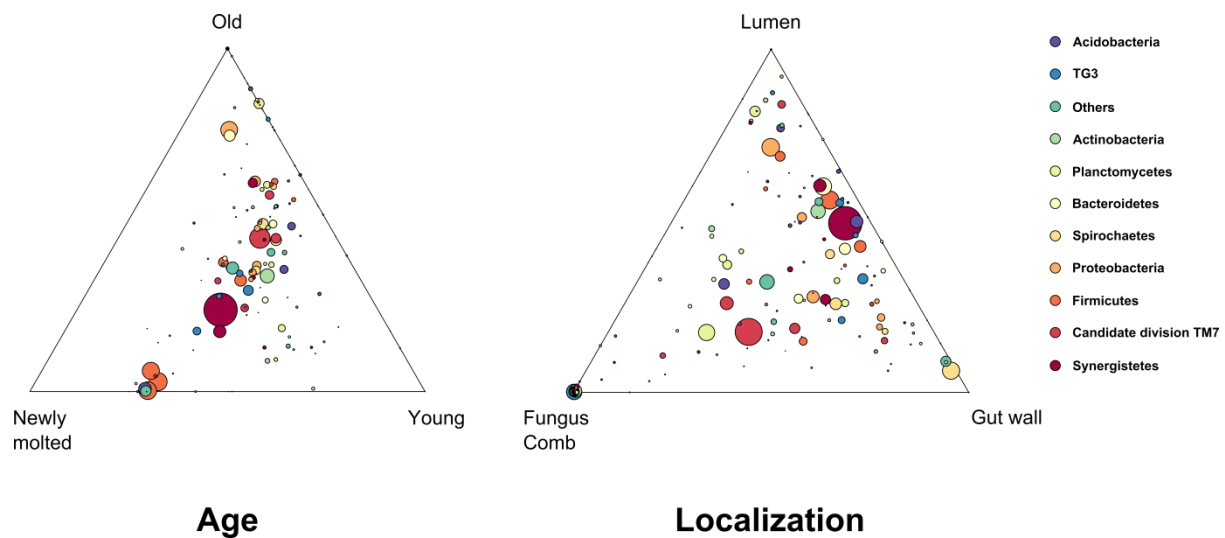
An analysis of the most variable and abundant genus-level taxa in the total data set revealed that the 20 top taxa accounted for half of the total reads. Among those 20 taxa, we found 8 members of *Firmicutes*, 7 of *Bacteroidetes*, 2 of *Proteobacteria*, 2 of *Synergistetes*, and 1 of *Spirochaetes* (Fig. 4). Of the 8 genus-level clusters in *Firmicutes*, Incertae Sedis 18 and uncultured 13 were enriched in the gut wall fraction of newly molted workers (19.0% and 6.8%, respectively) but sharply decreased and were even not detectable in both the gut wall and lumen fractions as the worker hosts aged. Abundance of *Lactovum* also decreased as the worker termites aged, but mainly from the gut lumen fraction (from 13.0% in the gut lumen of newly molted workers to 0.4% in the lumen of old workers). Relative abundances of *Candidatus* Arthromitus and *Lactococcus* 1 increased as the termites aged, mainly in

the lumen fraction (increased from almost absent in the lumen of newly molted workers to 2.4% and 7.9%, respectively, in the lumen of old workers). The abundance of the seven *Bacteroidetes* genus level taxa mostly increased as the worker termites aged and were generally similarly distributed in the gut wall and lumen fractions in the newly molted workers, but all except *Bacteroides* became more enriched in the gut wall fraction in both young and old workers. For genus-level taxa of phylum *Synergistetes*, sequences classified as uncultured 6 were relatively evenly distributed in both the gut lumen and wall fraction, and formed the most abundant group among workers of all ages, even though the abundance was much lower in the gut of old workers. Sequences classified as *Candidatus Tammella* were more abundant in the gut lumen fraction of newly molted and young workers than in the lumen fraction of old workers.



**Fig. 4 | Comparison of relative abundance of selected bacterial lineages in different localizations of *Odontotermes formosanus*.** Data is divided into hindgut lumen and on the hindgut wall fraction of newly molted workers (NM), young workers (Y), and old workers (O) of *Odontotermes formosanus* as well as new fungus comb (NFC) and old fungus comb (OFC). Detailed classification results at all taxonomic levels are provided in Supplementary material Tab. S3.

**Localization specificity and age specificity.** Most genus-level taxa were consistently represented across newly molted, young, and old workers, but many were unevenly distributed (Fig. 5A). Some extreme examples, such as some Lachnospiraceae genus-level taxa of the phylum *Firmicutes* (also Incertae Sedis 18, uncultured 13, and uncultured 51), were found only in the gut of newly molted and young workers, while *Lactococcus* 1 (*Firmicutes*, Bacilli, Lactobacillales, Streptococcaceae) was found only in the gut of old workers. Similarly, most genus-level taxa were consistently represented across fungus comb and gut wall and lumen, but many were more unevenly distributed among these fractions (Fig. 5B). Some extreme examples, such as the genus *Acetonema* a and the class Bacilli in *Firmicutes*, as well as the genus *Bradyrhizobium* 12 in Alphaproteobacteria, were found only in the fungus comb (mainly in old fungus comb). Taxa that were more prevalent in the fungus comb and gut wall samples mostly originated only from the new fungus comb (e.g., *candidate division*\_OP11 and *candidate division* TM7, *Alistipes* 1 and *Alistipes* 2 in phylum Bacteroidetes), which was to be expected considering that the new fungus comb was inoculated by the feces of the young workers. This also indicated that taxa that were transmitted from the gut wall, and not from the lumen, of young workers to the fungus comb were more easily sustained on the fungus comb as it matured.



**Fig. 5 | Ternary plot of the distribution of genus-level taxa across different termite age and localizations.** (A) *Odontotermes formosanus* workers of three age groups and (B) three *Odontotermes formosanus* fractions. The area of each circle represents the relative abundance of the reads in the entire data set, the position specifies their average abundance in the respective groups, and the colors indicate the phylum of taxa. Data are from Tab. S3 in the Supplementary material). An interactive version that allows identification of the genus behind each data point of the figure is given in Fig. S3 in the Supplementary material.



## Discussion

The termite gut is a small but very complex ecosystem with an astonishing diversity of niches and spatial heterogeneity. Only little is known about the establishment, succession, and spatial structure of the bacterial community. This study of three age groups of workers of the fungus-cultivating termite *O. formosanus*, combining electron microscopy of *in situ* gut microbiota with a highly resolved analysis of the bacterial communities in the individual gut fractions, revealed that the spatial organization of different bacterial guilds dynamically shifts as the workers age. Considering that in older compared to younger *O. formosanus* workers, both the conditions in the gut and their role in lignocellulose digestion and processing change (Li et al., 2015), the results suggest that the spatial arrangement of the different populations of gut microbiota may be largely driven by dietary substrates and gut physiology, such the oxygen penetration depths.

**Axial distribution patterns of bacterial communities and enzyme activity in the gut of *O. formosanus*.** We observed that midguts of the fungus-growing termite contain almost no microbes, as in lower termites and in contrast to higher wood-feeding termites (Breznak and Pankratz, 1977; Slaytor and Brien, 1985). An explanation for this characteristic could be that fungus-growing termites are the most basal lineage of higher termites and both the anatomy and physicochemical parameters of the intestine still largely resemble those of lower termites (Brune, 2014). Since the highest cellulase activity of fungus-cultivating termites has been detected in the midgut of workers, it is widely accepted that the midgut is important for lignocellulose digestion (Veivers et al., 1991). The recent discovery that the host endogenous cellulases endo- $\beta$ -1,4-glucanase and  $\beta$ -1,4-glucosidase are expressed in the midgut of fungus-cultivating termites indicates that these enzymes play an important role in lignocellulose digestion (Wu et al., 2012; Ni et al., 2014). The almost complete absence of microbes in the midgut of *O. formosanus* indicated that not microbial but rather endogenous digestion occurs in the midgut of these termites.

The midgut morphology of young and old workers of fungus-cultivating termites differs greatly (Fig. 1). The midgut of young workers is engorged, whereas that of old workers is shrunken. This difference could be related to the age-related polyethism of *Macrotermitinae* termites. Young workers feed primarily on plant materials; as it is difficult to obtain energy from this substrate, the young workers must consume large amounts to survive, thereby leading to an engorged midgut (Hinze et al., 2002b). Old workers, in contrast, feed on old fungus combs mixed with the nutritious fungal mycelium, which can be easily digested; therefore, smaller amounts of food should satisfy their energy requirements, and an engorged midgut is not necessary. Since morphology usually correlates tightly with function, the more developed midgut of young workers compared to that of old workers should indicate that midgut digestion is more important for young workers than for old workers.

**Transmission, succession, and establishment of gut bacterial communities in *O. formosanus*.** To better understand the transmission, succession, and establishment of gut bacterial communities in an *O. formosanus* colony, we characterized the microbiota in the hindguts of three age groups of adult worker termites and in new and old fungus combs. We found evidence for transmission of gut bacteria within the colony; the bacterial community profiles in the lumen of newly molted and young workers were highly similar. This similarity could be related to the stomodeal trophallaxis behavior of *O. formosanus*, in which young workers feed newly molted workers (Li et al., 2015), which, as is generally believed, lost all intestinal microbiota during the last ecdysis from the larval stage (May et al.,

1941). Stomodaeal trophallaxis would lead to the transmission of the hindgut microbiota from young workers to the newly molted workers.

Young workers are also responsible for building new fungus combs using their feces, and old workers feed on old fungus combs; thus, one might expect that the bacterial community of each pair, i.e., young workers and new fungus combs, and old workers and aged fungus combs, would be similar because of transmission. However, our results indicated that the bacterial communities of each of these pairs are not closely clustered, although some genus-level taxa are shared between new fungus combs and gut lumen fraction of young workers. This lack of transmission in these cases could be caused by greatly different physicochemical conditions, e.g., oxygen pressure, in the gut and fungus comb. Indeed, the intestinal microbial community of the soil-feeding higher termite *Cubitermes niokoloensis* even differs greatly from that of its fresh feces (Fall et al., 2007).

Diet is also a major external force shaping both the microbial communities of the termite gut and the corresponding metabolic pool (Hongoh et al., 2006; Boucias et al., 2013; Dietrich et al., 2014). We found a high concentration of butyrate in the hindgut of old workers (Tab. 1). Analysis of metagenomic data has revealed four main microbial butyrate-producing pathways (Marius et al., 2014). That the most prevalent is the acetyl-CoA pathway, for which major substrates include polysaccharides derived both from plants and from cross-feeding with lactate-synthesizing bacteria. In the other three pathways, i.e., the glutarate, 4-aminobutyrate, and lysine pathways, amino acids serve as major substrates. It also revealed that the polysaccharide-degrading acetyl-CoA pathway is predominately used by the Firmicute families *Lachnospiraceae* and *Ruminococcaceae*, whereas the amino-acid-degrading pathways are predominately used by the Firmicutes family *Veillonellaceae* and by the Bacteroidetes families *Porphyromonadaceae* and *Rikenellaceae* (specifically by *Alistipes putredinis*) (Marius et al., 2014). We showed that the relative abundance of members of the *Lachnospiraceae* and *Ruminococcaceae* decreased as the workers aged, and that of *Veillonellaceae*, *Porphyromonadaceae*, and *Rikenellaceae* increased as the workers aged. These results as well as the observed higher butyrate pool in the hindgut of older workers may be explained by the shift from a polysaccharide-rich diet of young workers to a protein-rich diet of old workers.

**Distinct spatial and temporal distribution of the hindgut microbiota of *O. formosanus* workers.** Steep radial gradients of oxygen and hydrogen partial pressure between the hindgut wall and lumen of flagellate-harboring lower termites of the genus *Reticulitermes* (Ebert and Brune, 2000) lead to a radially heterogeneous niche colonized by specific microbial communities (Tokura et al., 2000; Nakajima et al., 2005; Yang et al., 2005). Our results from the fungus-cultivating termite *O. formosanus* not only extend these findings to a flagellate-free higher termite, but also document the differences among different age groups of worker termites.

The differences in bacterial community profiles between the hindgut wall and lumen fractions of *O. formosanus* workers increased with age (Fig. 3B), which suggests that the radial organization of the different bacterial communities established gradually with age. The lowest dissimilarity index between hindgut wall and lumen was observed in newly molted workers, for which transmission electron microscopy showed low bacterial populations and few morphotypes (Fig. 2A and B). In young workers, the dissimilarity index between the hindgut wall and lumen fractions was slightly higher. The increase may be only slight because the gut microbiota increased in both abundance and diversity in both fractions but the populations were evenly distributed between the two fractions (Fig. 2C). The highest dissimilarity index between hindgut wall and lumen was observed in old workers, which could be largely caused by the clear uneven distribution of bacterial populations between the two fractions (Fig. 2E), with a higher density of bacteria at the hindgut wall (Fig. 2F). The gradual shaping of the

termite–microbiota symbiosis in the hindgut with age may be related to the development of the gut itself (Fig. 1), as shown for the *Drosophila*–microbiota symbiosis, where the bacterial symbionts in the gut might promote the growth and development of the host’s intestinal cells (Broderick et al., 2014) .

The divergence in the bacterial communities of the hindgut wall and lumen fractions increased with the age of the worker termites. For example, in the hindgut wall fraction, earlier colonizers belonged to the family *Lachnospiraceae* of the order *Clostridiales* in phylum *Firmicutes* were most abundant in newly molted workers (34.8% of all bacteria). However, as the hindgut wall developed as the workers aged, the relative abundance of this family in this fraction significantly decreased to 16.7% in young workers down to 3.5% old workers. Members of the family *Lachnospiraceae* are also early colonizers in the intestine of chickens; these bacteria inoculated into newly hatched chickens easily colonize the ileum and may affect host gene expression there (Yin et al., 2010).

In the gut lumen fraction, members of the family *Enterobacteriaceae* in the *Gammaproteobacteria* were in extremely low abundance in newly molted workers (0.5%), increased in abundance in young workers (2.1%), and greatly increased in old workers (18.6%). Two representatives of this family that degrade benzoic acid have been isolated from a fungus-cultivating termite of the genus *Macrotermes* (Ngugi et al., 2007); therefore, the enrichment of this bacterial group in the hindgut lumen of old workers of *O. formosanus* observed here points toward benzoate metabolism in this fraction in old workers. This is feasible considering the diet of old workers, which is composed of high concentrations of aromatic compounds derived from lignin after degradation by the symbiotic fungus on the comb and is consistent with a recent metagenomic study that showed that the microbial community encodes for more genes of degradation of aromatic compound are in old workers of the fungus cultivator *Odontotermes yunnanensis* (Liu et al., 2013).

The highest abundance of *Treponema* subcluster la was found in the hindgut lumen of young workers (5.1%), followed by that of newly molted workers (2.7%) and old workers (2.1%). Members of *Treponema* subcluster la are found in high abundance in the hindgut of wood-feeding lower termites and of wood-feeding and grass-feeding higher termites (Dietrich et al., 2014). It is believed that these bacteria are responsible for acetogenesis in these hindguts because cultured representatives of *Treponema* subcluster la utilize mono- and di- and/or oligo-saccharides as energy sources for fermentative growth and acetogenesis (Graber and Breznak, 2004; Graber et al., 2004; Dröge et al., 2008). Young workers of *O. formosanus* primarily ingest wood, which resembles the diet of the wood- and grass-feeding termites mentioned above; this could explain the increase in abundance of *Treponema* subcluster la in the hindgut lumen of the young workers of *O. formosanus*. The hindgut of old workers had the highest concentration of acetate yet had the lowest abundance of *Treponema* subcluster la, which indicates that most of the acetogenesis in the hindgut of old workers of *O. formosanus* is not mainly carried out by this group of bacteria.

**Concluding remarks.** This is the first study that investigates the gut microbiota of a fungus-cultivating termite over space and time. Our results demonstrate that there are large differences in microbial community structure between the gut wall and fluid of *O. formosanus*. However, substantial changes in community structure could also be observed over time. Since the hindgut is less dilated in younger workers, these changes were found be most likely due to the relative oxygen penetration depths. Surprisingly, the microbial community structure of *O. formosanus* is very similar to the gut microbiotas of other fungus-cultivating termites (Dietrich et al., 2014; Otani et al., 2014) which indicates that a climax community was established towards the end of the experiment. These results underline how microhabitats and the temporal development of termite workers shape microbial communities.

## References

- Abdi H and Williams LJ (2010). Principal component analysis. Wiley Int. Rev. Com. St 2: 433–459.
- Badertscher S, Gerber C and Leuthold RH (1983). Polyethism in food supply and processing in termite colonies of *Macrotermes subhyalinus* (Isoptera). Behav Ecol Sociobiol 12:115–119.
- Boucias DG, Cai Y, Sun Y, Lietze V-U, Sen R, Raychoudhury R and Scharf ME (2013). The hindgut lumen prokaryotic microbiota of the termite *Reticulitermes flavipes* and its responses to dietary lignocellulose composition. Mol. Ecol. 22: 1836–1853.
- Breznak J and Pankratz H (1977). In situ morphology of the gut microbiota of wood-eating termites [*Reticulitermes flavipes* (Kollar) and *Coptotermes formosanus* Shiraki]. Appl. Environ. Microbiol. 33: 406–426.
- Broderick NA, Buchon N and Lemaitre B (2014). Microbiota-induced changes in *Drosophila melanogaster* host gene expression and gut morphology. mBio 5: e01117-01114.
- Brune A (2014). Symbiotic digestion of lignocellulose in termite guts. Nat. Rev. Microbiol. 12: 168–180.
- Brune A and Friedrich M (2000). Microecology of the termite gut: structure and function on a microscale. Curr. Opin. Microbiol. 3: 263–269.
- Brune A and Ohkuma M (2011). Role of the termite gut microbiota in symbiotic digestion. In *Biology of Termites: a Modern Synthesis*. Bignell DE, Roisin Y and Lo N (eds.): Springer Netherlands, pp. 439–475.
- Brune A, Emerson, D, and Breznak JA (1995). The termite gut microflora as an oxygen sink: microelectrode determination of oxygen and pH gradients in guts of lower and higher termites. Appl. Environ. Microbiol. 61: 2681–2687.
- Dietrich C, Köhler, T, and Brune A (2014). The cockroach origin of the termite gut microbiota: patterns in bacterial community structure reflect major evolutionary events. Appl. Environ. Microbiol. 80: 2261–2269.
- Dröge S, Rachel R, Radek R and König H (2008). *Treponema isoptericolens* sp. nov., a novel spirochaete from the hindgut of the termite *Incisitermes tabogae*. Int J Syst Evol Micr 58: 1079–1083.
- Ebert A and Brune A (1997). Hydrogen concentration profiles at the oxic-anoxic interface: a microsensor study of the hindgut of the wood-feeding lower termite *Reticulitermes flavipes* (Kollar). Appl. Environ. Microbiol. 63:4039–4046.
- Eggleton P (2011). An introduction to termites: biology, taxonomy and functional morphology. In *Biology of Termites: a Modern Synthesis*. Bignell DE, Roisin Y, and Lo N (eds.): Springer Netherlands, pp. 1–26.
- Graber JR and Breznak JA (2004). Physiology and Nutrition of *Treponema primitia*, an H<sub>2</sub>/CO<sub>2</sub>-acetogenic spirochete from termite hindguts. Appl. Environ. Microbiol. 70: 1307–1314.
- Graber JR, Leadbetter JR and Breznak JA (2004). Description of *Treponema azotonutricium* sp. nov. and *Treponema primitia* sp. nov., the first spirochetes isolated from termite guts. Appl. Environ. Microbiol. 70: 1315–1320.
- Henckel T, Friedrich M and Conrad R (1999). Molecular analyses of the methane-oxidizing microbial community in rice field soil by targeting the genes of the 16S rRNA, particulate methane monooxygenase, and methanol dehydrogenase. Appl. Environ. Microbiol. 65:1980–1990.
- Hinze B, Crailsheim K, and Leuthold R (2002a). Polyethism in food processing and social organisation in the nest of *Macrotermes bellicosus* (Isoptera, Termitidae). Insect. Soc. 49: 31–37.
- Hinze B, Crailsheim K, and Leuthold R.H. (2002b). Polyethism in food processing and social organisation in the nest of *Macrotermes bellicosus* (Isoptera, Termitidae). Insect Soc 49: 31–37.
- Hongoh Y, Deevong P, Inoue T, Moriya S, Trakulnaleamsai S, Ohkuma M, Vongkaluang C, Noparatnaraporn N and Kudo T (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.
- Hongoh Y, Ekpornprasit L, Inoue T, Moriya S, Trakulnaleamsai S, Ohkuma M, Noparatnaraporn N and Kudo T (2006). Intracolony variation of bacterial gut microbiota among castes and ages in the fungus-growing termite *Macrotermes gilvus*. Mol. Ecol. 15: 505–516.

- Huang F, Zhu S, Ping Z, He X, Li G and Gao D (2000). Fauna Sinica. Insecta Isoptera. Vol.17. China Beijing: Science Press, pp. 566–568. (In Chinese)
- Kamanda Ngugi D, Khamis Tsanuo M, and Iddi Boga H (2007). Benzoic acid-degrading bacteria from the intestinal tract of *Macrotermes michaelseni* Sjöstedt. J. Basic Microbiol. 47: 87–92.
- Koehler T, Dietrich C, Scheffrahn R.H, and Brune A (2012). High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation of the gut in wood-feeding higher termites (*Nasutitermes* spp.). Appl. Environ. Microbiol. 78: 4691–4701.
- Li H, Sun J, Zhao J, Deng T, Lu J, Dong Y, Deng W and Mo J (2012). Physicochemical conditions and metal ion profiles in the gut of the fungus-growing termite *Odontotermes formosanus*. J Insect. Physiol. 58: 1368–1375.
- Li H, Yang M, Chen Y, Na Z, Lee C, Wei J and MO J (2015). Investigation of age polyethism in food processing of the fungus-growing termite *Odontotermes formosanus* (Blattodea: Termitidae) using a laboratory artificial rearing system. J. Econ. Entomol., doi: 10.1093/jee/tou005, e-pub ahead of print 22 January 2015.
- Liu N, Zhang L, Zhou H, Zhang M, Yan X, Wang Q, Long Y, Xie L, Wang S, Huang Y and Zhou Z (2013). Metagenomic insights into metabolic capacities of the gut microbiota in a fungus-cultivating termite (*Odontotermes yunnanensis*). Plos One 8:e69184.
- Long Y-H, Xie L, Liu N, Yan X, Li M-H, Fan M-Z, and Wang Q (2010). Comparison of gut-associated and nest-associated microbial communities of a fungus-growing termite (*Odontotermes yunnanensis*). Insect. Sci. 17: 265–276.
- Marius V, Howe AC and Tiedje JM (2014). Revealing the bacterial butyrate synthesis pathways by analyzing (Meta)genomic Data. mBio 5:e00889
- Mikaelyan A, Strassert J.F, Tokuda G and Brune A (2014). The fiber-associated cellulolytic bacterial community in the hindgut of wood-feeding higher termites (*Nasutitermes* spp.). Environ. Microbiol. 16: 2711–2722.
- Nakajima H, Hongoh Y, Noda S, Yoshida Y, Usami R, Kudo T and Ohkuma M (2006). Phylogenetic and morphological diversity of Bacteroidales members associated with the gut wall of termites. Biosci. Biotech. Bioch. 70: 211–218.
- Nakajima H, Hongoh Y, Usami R, Kudo T, and 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.
- Ni J, Wu Y, Yun C, Yu M, and Shen Y (2014). cDNA cloning and heterologous expression of an endo-beta-1,4-glucanase from the fungus-growing termite *Macrotermes barneyi*. Arch. Insect. Biochem. 86: 151-164.
- Nobre T, Rouland-Lefèvre C, and Aanen D (2011). Comparative biology of fungus cultivation in termites and ants. In Biology of Termites: a Modern Synthesis. Bignell DE, Roisin Y, and Lo N (eds): Springer Netherlands, pp. 193–210.
- Otani S, Mikaelyan A, Nobre T, Hansen LH, Kone NGA, Sorensen SJ, Aanen K, Boomsma JJ, Brune A and Poulsen M (2014). Identifying the core microbial community in the gut of fungus-growing termites. Mol. Ecol. 23: 4631–4644.
- Schloss PD, Gevers D, and Westcott SL (2011). Reducing the effects of PCR amplification and sequencing artifacts on 16S rRNA-based studies. Plos One 6: e27310.
- Schmitt-Wagner D and Brune A (1999). Hydrogen profiles and localization of methanogenic activities in the highly compartmentalized hindgut of soil-feeding higher termites (*Cubitermes* spp.). Appl. Environ. Microbiol. 65: 4490–4496.
- Schmitt-Wagner D, Friedrich M.W, Wagner B and Brune A (2003a). Phylogenetic diversity, abundance, and axial distribution of bacteria in the intestinal tract of two soil-feeding termites (*Cubitermes* spp.). Appl. Environ. Microbiol. 69: 6007–6017.
- Schmitt-Wagner D, Friedrich MW, Wagner B and Brune A (2003b). Axial Dynamics, Stability, and Interspecies Similarity of Bacterial Community Structure in the Highly Compartmentalized Gut of Soil-Feeding Termites (*Cubitermes* spp.). Appl. Environ. Microbiol. 69: 6018-6024.
- Slaytor R and Brien R. (1985). Bacterial Flora of the Mixed Segment and the Hindgut of the Higher Termite *Nasutitermes exitiosus* Hill (Termitidae, Nasutitermitinae). Appl. Environ. Microbiol. 49:1226–1236.
- Tholen A and Brune A (2000). Impact of oxygen on metabolic fluxes and in situ rates of reductive acetogenesis in the hindgut of the wood-feeding termite *Reticulitermes flavipes*. Environ. Microbiol. 2:436–449.

- Tokura M Ohkuma M and 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.
- Veivers PC, Mühlemann R, Slaytor M, Leuthold RH and Bignell DE (1991).** Digestion, diet and polyethism in two fungus-growing termites: *Macrotermes subhyalinus* Rambur and *M. michaelsoni* Sjøstedt. *J. Insect. Physiol.* 37: 675–682.
- Warnecke F, Luginbühl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, Cayouette M, McHardy AC, Djordjevic G, Aboushadi N, Sorek R, Tringe SG, Podar M, Martin HG, Kunin V, Dalevi D, Madejska J, Kirton E, Platt D, Szeto E, Salamov A, Barry K, Mikhailova N, Kyrpides NC, Matson EG, Ottesen EA, Zhang X, Hernández M, Murillo C, Acosta LG, Rigoutsos I, Tamayo G, Green BD, Chang C, Rubin EM, Mathur EJ, Robertson DE, Hugenholtz P and Leadbetter JR (2007).** Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nat.* 450:560–565.
- Wu Y, Chi S, Yun C, Shen Y, Tokuda G and Ni J (2012).** Molecular cloning and characterization of an endogenous digestive  $\beta$ -glucosidase from the midgut of the fungus-growing termite *Macrotermes barneyi*. *Insect. Mol. Biol.* 21: 604–614.
- Yang H, Schmitt-Wagner D, Stingl U and Brune A (2005).** Niche heterogeneity determines bacterial community structure in the termite gut (*Reticulitermes santonensis*). *Environ. Microbiol.* 7:916–932.

## Supplementary Material

**Tab. S1 | Characteristics of 16S rRNA gene amplicon libraries of bacterial communities in hindgut fractions of workers of different ages and in the fungus comb.**

Sample	Fraction <sup>a</sup>	Reads (10 <sup>3</sup> )	Genus-level taxa	OTUs (3% dissim.)	Diversity <sup>b</sup>
<b>Termite worker</b>					
Newly molted	GW	15.0	177	232	3.6
	GL	20.2	197	317	4.0
Young	GW	35.5	263	403	4.5
	GL	59.1	281	475	4.7
Old	GW	27.0	249	386	4.7
	GL	25.2	289	420	4.3
<b>Fungus comb</b>					
	New	27.3	385	487	4.7
	Old	22.9	343	310	3.8

<sup>a</sup>GW, Gut wall; GL, Gut lumen.

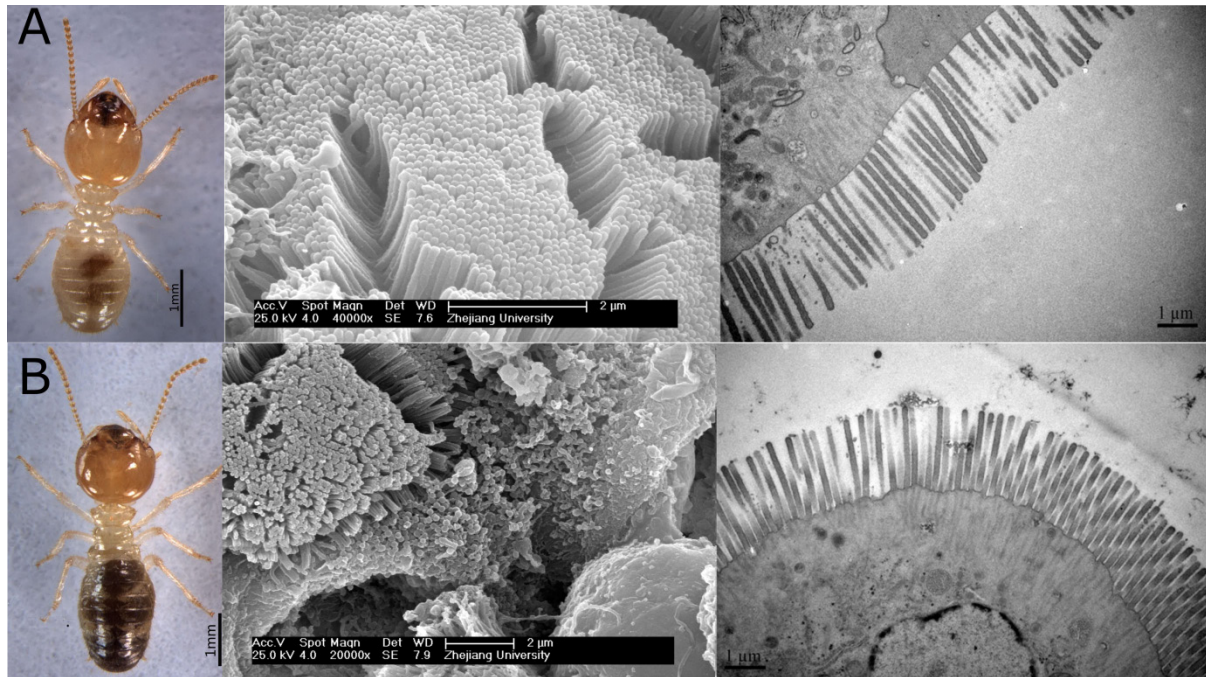
<sup>b</sup>Nonparametric Shannon index.

**Tab. S2 Comparison of classification success at different taxonomic levels using the RDP, Silva, and DictDB.**

Sample	Fraction	Classification success (%)								
		Phylum			Family			Genus		
		RDP	Silva	DictDB	RDP	Silva	DictDB	RDP	Silva	DictDB
<b>Termite worker</b>										
Newly molted	GW	79.3	100.0	100.0	70.5	94.1	95.6	41.7	62.4	76.2
	GL	75.4	100.0	100.0	64.5	90.2	94.1	35.9	77.3	76.5
Young	GW	81.0	99.9	100.0	58.4	82.3	86.7	32.9	63.1	75.9
	GL	73.5	100.0	100.0	57.6	87.7	92.5	30.2	76.4	81.2
Old	GW	83.6	100.0	100.0	50.7	74.8	82.0	30.6	65.5	75.8
	GL	84.1	100.0	99.9	70.6	86.7	93.0	52.1	60.0	69.3
<b>Fungus comb</b>										
New	—	84.0	99.9	100.0	46.4	60.5	66.0	32.9	53.9	59.1
Old	—	98.3	99.8	99.8	91.0	93.0	98.0	87.2	82.0	87.5



**Tab. S3 | Relative read abundance in the pyrotag libraries of the bacterial communities in the different hindgut fractions of newly molted workers, young workers, and old workers as well as in the fungus comb.** This interactive table allows the classification results at different taxonomic levels to be shown. Please download from: [https://dl.dropboxusercontent.com/u/50542577/Chapter\\_5\\_Tab\\_S3.xlsx](https://dl.dropboxusercontent.com/u/50542577/Chapter_5_Tab_S3.xlsx)



**Fig S1 | Scanning and transmission electron micrograph of midgut epithelium and midgut lumen from (A) young workers and (B) old workers of *Odontotermes formosanus*.**



## Habitat selection and vertical inheritance drive archaeal community structure in arthropod guts

Carsten Dietrich<sup>\*1</sup>, James Nonoh<sup>\*1</sup>, Kristina Lang<sup>1</sup>, Lena Mikulski<sup>1</sup>,  
Katja Meuser<sup>1</sup>, Tim Köhler<sup>1</sup>, Hamadi I. Boga<sup>2</sup>, David K. Ngugi<sup>3</sup>,  
David Sillam-Dussès<sup>4</sup> and Andreas Brune<sup>1</sup>

\* These authors contributed equally | *Affiliations*: <sup>1</sup>Max Planck Institute for Terrestrial Microbiology, Department of Biogeochemistry, Marburg, Germany; <sup>2</sup>Jomo Kenyatta University of Agriculture and Technology, Department of Botany, Nairobi, Kenya; <sup>3</sup>King Abdullah University of Science and Technology, Red Sea Research Center 4700 KAUST, Saudi Arabia; <sup>4</sup>Université Paris 13, Sorbonne Paris Cité, Laboratoire d'Ethologie Expérimentale et Comparée, Villetaneuse, France | *This manuscript is submitted.* | *Contributions*: C.D. designed and conducted the high-throughput sequencing analysis; created a reference database including all previously published data; conducted the final phylogenetic analysis, all bioinformatics, and statistical analyses; and wrote the manuscript. J.O.N. created the majority of the clone libraries, and contributed to the phylogenetic analysis and the evaluation of the results. K.L. created clone libraries and contributed to the phylogenetic analysis. L.M. created clone libraries. K.M. prepared libraries for the high-throughput sequencing analysis. T.K. improved the primer set for the high-throughput sequencing analysis. H.I.B. collected termites and created clone libraries. D.K.N. collected termites and prepared DNA. D.S.D. collected termites. A.B. conceived the study, discussed the results, and secured funding.

### Abstract

The arthropod gut is streamlined to anaerobically mineralize mostly plant-derived organic matter. In only five major arthropod groups (millipedes, scarab beetle larvae, cockroaches, and lower and higher termites) are considerable amounts of methane formed by methanogenic archaea at the end of this process. Bacterial communities in the guts of termites and cockroaches mirror major events in the evolutionary history of their host. Whether this is also true for archaeal communities or whether diet is the key factor is unknown. Here, we used both clone libraries and high-throughput sequencing to document that the archaeal community structure in arthropod guts and the phylogeny of archaeal lineages is dependent on the host group and to a lesser extent on diet. With the exception of lower termites, all major arthropod groups contained at least one group each of hydrogenotrophic and methylotrophic methanogens regardless of the host diet. Hydrogenotrophic methanogenesis is almost exclusively carried out by members of the genus *Methanobrevibacter*, whereas methyl reduction is accomplished by different genera of the orders *Methanomassiliicoccales* and *Methanosarcinales*. The occurrence of specific genera of these obligately hydrogen-dependent methylotrophs differs among the hosts, which indicated that host habitat selection is the major driving force for arthropod archaea. Analysis of the phylogeny of the most abundant archaeal lineages in the arthropod host gut revealed host-group-specific clusters of archaeal lineages. Since cladogenesis was absent in the resulting phylogenetic trees, cospeciation could be excluded what indicates that archaea only coevolve with their hosts. This underlines that the mechanisms for selection of archaeal lineages must be host-habitat-specific, as, e.g., the highly alkaline gut compartment of most higher termites, which selects for alkali-tolerant strains. In contrast to other studies, we did not find a uniform archaeal community in the guts of flagellate-containing lower termites; the archaeal community was similar to that of other host groups. Therefore, in lower termites, other mechanisms must select for archaeal lineages, e.g., microhabitats provided by flagellates. This hypothesis was supported by phylogenetic analysis of the corresponding representative operational taxonomic units, which often have long branches — an indication of a different rate of evolution that is frequently observed in endosymbionts.

## Introduction

The gut of terrestrial arthropods is a diverse and important ecosystem, where mostly plant-derived organic matter is transformed and mineralized. This breakdown is mostly facilitated in concert with their complex gut microbiota in a series of fermentative processes (for a detailed description, see Brune, 2014 and references therein). The terminal processes include acetogenesis and methanogenesis from a variety of substrates. Although the generally accepted scheme of anaerobic breakdown suggests a steady ratio of acetate formed to methane produced, the dominance of one of these two processes in termites varies (Pester and Brune, 2007). Methane emission is restricted to five groups of terrestrial arthropods: millipedes, scarab beetle larvae, cockroaches, lower termites, and higher termites (Hackstein and Stumm, 1994; Hackstein et al., 2006). These groups differ greatly in their methane emission rates (summarized in Brune, 2010), but the driving forces behind this phenomenon are mostly unidentified.

Methane is formed exclusively by methanogenic archaea, which in termites may account for up to 3% of the total microbial community (Brauman et al., 2001). Archaeal lineages detected in arthropod guts can be classified to either the Miscellaneous Crenarchaeotic Group (MCG), the Soil Crenarchaeotic Group (SCG) of the archaeal phylum *Thaumarchaeota*, or one of the four methanogenic orders *Methanobacteriales*, *Methanomassiliicoccales*, *Methanomicrobiales*, and *Methanosarcinales*. The distribution of these lineages differs in the different hosts. For millipedes, no clone library data have been obtained, but recently a DGGE analysis and sequencing revealed the presence of *Methanosarcinales*, *Methanobacteriales*, *Methanomicrobiales*, and some unclassified archaea (Sustr et al., 2014). The archaeal community of the scarab beetle *Pachnoda ephippiata* larva is dominated by members of the order *Methanobacteriales* (Egert et al., 2003). In the cockroach gut, members of the *Methanosarcinales* are the most abundant archaea (Hara et al., 2002). In lower termites, almost exclusively members of the *Methanobacteriales* were identified (Ohkuma and Kudo, 1998; Shinzato et al., 1999; Tokura et al., 2000; Shinzato et al., 2001). In higher termites, in contrast, the most abundant methanogenic groups differ among the host subfamilies: *Methanobacteriales* (*Termitinae*), *Methanosarcinales* (*Macrotermitinae*), or *Methanomicrobiales* (*Nasutitermitinae*). Members of the non-methanogenic phylum of *Thaumarchaeota* were detected only in the hindguts of higher termites and the midgut of *Pachnoda ephippiata* (Egert et al., 2003). However, these results cannot be generalized as most of the earlier studies suffer from low sequencing depths (in some cases, five clones per sample), which does not highly support confidence. The still hidden diversity could be in the worst case as high as 20% of the archaeal community.

Earlier studies have also addressed the distribution of archaea along the different gut compartments of arthropods (Friedrich et al., 2001; Egert et al., 2003), possibly caused by the availability of different substrates and by differences in microenvironments (Schmitt-Wagner and Brune, 1999). Within the gut compartments, archaea (especially methanogens) can occur on the gut wall (or cuticular hair), associated with protists (Leadbetter et al., 1998), or in the lumen (Hackstein et al., 2006); in lower termites and cockroaches, they can also occur as symbionts of protists (Odelson and Breznak, 1985; Gijzen et al., 1991). Especially protists could influence the archaeal community structure in lower termites since the presence of endosymbiotic methanogens favors the protist host and is therefore beneficial for the termite (Odelson and Breznak, 1985; Messer and Lee, 1989). The success of this lifestyle in lower termites is also supported by the high numbers of methanogens associated with protists (10–50 per protist) and the high number of protists per gut (sometimes > 100,000) (Tokura et al., 2000).

Recently, it has been shown that bacterial communities mirror major events in the evolutionary history of termites and cockroaches (Dietrich et al., 2014). Whether also the archaeal community is determined by the host evolutionary history is unknown, but host diet has been proposed (Hackstein et al., 2006). However, the resemblance of microbial communities in related arthropods does not necessarily result from coevolution or cospeciation between arthropod hosts and their microbial lineages. Instead, the selection of certain archaeal lineages could be influenced by the specific gut habitat within the host (Rawls, 2006).

To identify the mechanisms that drive the methanogenic archaeal community structure in terrestrial arthropods and archaeal phylogeny, we used a hybrid approach consisting of high-throughput sequencing and clone libraries to profile the communities across a wide range of all major groups of methane-emitting hosts.

## Materials and Methods

**Insect samples.** Termites were from laboratory colonies or field collections or purchased from commercial breeders (Tab. 1 and Tab. 2). Insect hindguts were dissected immediately upon arrival or collection (Schauer et al., 2012; Köhler et al., 2012). Samples were identified by sequencing the cytochrome oxidase subunit II gene (COII) (Pester and Brune, 2006). COII genes not represented in public databases were submitted to NCBI GenBank.

**DNA extraction.** Hindguts were dissected with sterile forceps. Owing to the large differences in size of the different insect hosts, the number of animals of each host group used differed; for millipedes, scarab beetles and cockroaches, one animal was used, and for termites, 3–10 animals were used. Guts were homogenized and DNA was extracted using a bead-beating protocol with subsequent phenol–chloroform purification (Paul et al., 2012).

**PCR amplification and cloning.** Archaeal 16S rRNA genes were amplified according to Paul et al. (2012) using the primer set Ar109f (5'-AMDGCTCAGTAACACGT-3') of Imachi et al. (2006) and Ar912r (5'-CTCCCCCGCCAATTTCCTTTA-3') of Lueders and Friedrich (2000) or the primer set Ar109f and 1490R with the modification of Hatamoto et al. (2007) (5'-GGHTACCTTGTTACGACTT-3'). Briefly, each PCR mixture (50 µl) contained reaction buffer, 2.5 mM MgCl<sub>2</sub>, 1 U Taq DNA polymerase (all Invitrogen, Carlsbad, CA), 50 µM deoxynucleoside triphosphates, 0.3 µM each primer, 0.8 mg ml<sup>-1</sup> bovine serum albumin, and 20 ng DNA. PCR was carried out with an initial denaturation step (94 °C for 3 min), followed by 30 cycles of denaturation (94 °C for 20 s), annealing (52 °C for 30 s), extension (72 °C for 45 s), and a final extension step (72 °C for 7 min).

**Sequence data from published studies.** For the reanalysis of published clone library sequences (Tab. 1) in the community structure analysis, the respective data were downloaded from NCBI GenBank. In some cases, only a representative phylotypes were available; therefore, quantitative information on these phylotype sequences was taken from the respective publications. Clone libraries were recreated by creating count tables, which used the *deunique.seqs* command in the *mothur* software suite version 1.33.3 (Schloss et al., 2009). The resulting data were treated the same way as the clone library data obtained in this study. Other available clone libraries that did not contain information about arthropod gut community structure (such as clone libraries from picked flagellates) were not used to visualize archaeal order-level differences in the arthropod hosts, but were included in the construction of the phylogenetic trees.

**Phylogenetic analysis of sequence data.** Raw sequences derived from clone libraries were analyzed and edited using *Seqman* (DNASTar) software. After importing the sequence into the current ARB-SILVA database (version 119, used throughout; Pruesse et al., 2007; <http://www.arb-silva.de>) using the *ARB* software package tool (Ludwig et al., 2004), sequences were aligned against the current SILVA alignment (Pruesse et al., 2007). If necessary, sequences were corrected manually. Location of sequences in the main SILVA tree was checked using the *ARB* parsimony tool. Afterwards, sequences belonging to different archaeal order or class levels were exported with the respective and adequate outgroup sequences. To conservatively exclude highly variable columns in the alignment, sequences were first clustered at 97% using the *usearch* software version 7.0.1090 (Edgar, 2010). The resulting representative sequences were

used to construct a 30% mask for the alignment, which was applied on all sequences of interest. Sequences with no ambiguous positions were used because of phylogenetic resolution, filtered by the respective mask, and analyzed phylogenetically using the 16-state GTR- $\Gamma$  model with 1,000 bootstraps in *RAxML* v8.1.3 (Stamatakis, 2014). Sequences that did not fit the quality criteria were inserted after treeing using the *ARB* parsimony tool by applying the same filter used to create the phylogenetic tree. Trees were rooted using type strains of the other methanogenic orders; for trees of archaea related to *Thaumarchaeota*, type strains from all methanogenic orders were employed.

**Primer design for high-throughput sequencing.** For high-throughput profiling of the archaeal community in the major arthropod groups, a primer set was needed that fit the requirements of the Illumina Miseq platform, i.e., that can deliver up to 300 nt paired-end reads. We aimed at an overlap of 100–150 nt for the paired-end reads. For this purpose, we slightly modified the primer pair A533b/A934b described by Grosskopf et al. (1998) to maximize the number of sequences bound in SILVA database and to better bind the termite-specific sequences from clone libraries of this study and other published sequences from arthropod guts. The new forward primer A533f\_mod (5'-TTACCGCGGCGGCTGVCA-3') was modified at position 16, where an ambiguity character replaces the former G. The reverse primer was changed by introducing the ambiguity character Y at positions 5 and 7, resulting in the reverse primer A934b\_mod (5'-GTGCYCYCCCGCCAATTCCT-3'). The resulting primer pair targets the V4–V5 region of the archaeal 16S rRNA gene. The performance of the primer was tested against SILVA database using TestPrime version 1.0 (Klindworth, 2013; <http://www.arb-silva.de/search/testprime/>). We followed the strategy of Daigle et al. (2011), which allowed us to multiplex. Briefly, primers A533f\_mod and A934b\_mod were flanked by universal M13 primers, which allows a very specific addition of multiplex identifiers (MIDs). The final primers consists of A533\_M13f\_mod (5'-cgccagggttttcccagtcacgacTTACCGCGGCGGCTGVCA-3') and A934\_M13b\_mod (5'-tcacacggaaacagctatgacGTGCYCYCCCGCCAATTCCT-3').

**High-throughput sequencing.** The V4–V5 region of the archaeal 16S rRNA gene was amplified using the flanked primer set A533\_M13f\_mod and A934\_M13b\_mod. For this step, 20 ng DNA was prepared as recommended by the Herculase II Fusion DNA Polymerase Kit (Agilent Technologies, USA) and amplified with an initial denaturation step (94 °C for 3 min), followed by 28 cycles of denaturation (94 °C for 20 s), annealing (58 °C for 20 s), and extension (72 °C for 50 s). The quality of the final products was by gel electrophoresis. To allow multiplexing in the sequencing run, we used the decamers as MIDs, as recommended by Roche (2009), flanked by the universal M13 primer, and again followed the protocol of Daigle et al. (2011). Final amplicons were mixed in equimolar amounts and commercially sequenced (Illumina Miseq; GATC Biotech, Konstanz, Germany). The resulting reads were processed according to the *UPARSE* pipeline (Edgar, 2013) by applying very stringent quality criteria (reads > 400 nt, no ambiguous bases, and maximum expected error rate 0.5). Subsequently, reads were clustered at different OTU dissimilarity levels (1%, 3%, and 5%) to obtain a classification-independent estimate of diversity.

**Classification.** Sequence reads were classified with the Naive Bayesian Classifier implemented in *mothur*, using a bootstrap value of 60% as cutoff. Since the classification success with public reference databases for arthropod clusters was limited owing to a lack of both annotation and sequences, we slightly modified the SILVA database by adding relevant published studies and sequences obtained in this study. The taxonomy

of relevant lineages was refined by adding or renaming groups that have been identified either in published phylogenies or by groups found in this study. The resulting reference database is available upon request.

**Statistics and visualization of the data.** For all statistical analyses, *R* version 3.0.1 (*R*, 2013) was used. Some graphics were produced using the *ggplot2* package (Wickham, 2009). Trees were exported with the node and bootstrap information in Newick format and plotted with all meta data using the *APE* package (Paradis, 2004) with some customized functions for plotting clusters and annotations. The ecological analysis was analyzed with the *vegan* package (Oksanen et al., 2013) in conjunction with the Soergel distance function written in pure *R*. For cluster analysis of genus-level groups, abundance data was normalized approximately to the smallest sample size (1,000 seqs), and the Soergel distance was calculated sample wise. We chose the Soergel distance as  $\beta$ -diversity measure since it performed well in a recent comparison study (Parks, 2013). For the logarithmic version of the Soergel distance (which was only used for the Supplementary Information), we expressed the data in per mill to circumvent negative values after the log transformation since many  $\beta$ -diversity measures rely on minimum, maximum, and sum terms, which would result in non-interpretable distances. When the data was finally logarithmized, we followed the recommended procedure of Costea et al. (2014). Briefly, a pseudo-count just a bit smaller than the smallest value of the dataset was added to circumvent  $\log(0)$ . To visualize the resulting distance matrices, we carried out a neighbor-joining analysis using the *bionj* implementation in *ape* since the classical hierarchical clustering is not well suited for biological (Rajaram, 2010) and especially compositional data sets (Friedman, 2012).

**Phylogenetic analysis of high-throughput sequencing derived sequences.** Sequences were clustered sample wise into operational taxonomic units (OTUs) at the 1% level using the *Uparse* strategy (Edgar, 2013) and classified into genus-level bins. All sequences in the same genus-level bin were aligned, and filtered to remove potentially poorly aligned columns (20% gap criterion). Sequences were subjected to a maximum-likelihood analysis using *RAxML* (Stamatakis, 2014) and rooted with a sequence from the methanogenic order *Methanopyrales*. Data were visualized using the *R* package *APE* (Paradis, 2004).

**Correlation of archaeal with bacterial genus-level groups.** In order to find possible dependencies of the archaeal genus-level groups on certain bacterial genus-level groups, a correlation analysis based on the SparCC algorithm was carried out since classical correlation analyses are not designed or even valid for compositional data (Friedman and Alm, 2012). For this purpose, we used the data set of this study and the data sets of the bacterial community structure (mostly taken from Dietrich et al., 2014). Both the archaeal and the bacterial data sets were classified into genus-level bins and exported in two ratios of 50:50 and 3:97 (archaea:bacteria) in the input format of sparCC. The choice of the latter ratio is natural as it has been reported to occur in termite guts (Brauman et al., 2001). The different ratios were used to test whether the ratio has an effect on the result, but no obvious differences were found in the genera combinations finally picked (minimum occurrence of a minimum of one  $\times|0.4$  per genus). The resulting SparCC  $-r$  values were visualized using the *R* software package.



## Results

**Distribution of major archaeal groups in the clone libraries.** We used a total of 31 clone libraries of archaeal 16S rRNA genes for the phylogenetic analysis of the archaeal communities of arthropods (Tab. 1). Seventeen libraries stemming from published studies and having a size ranging between 5 and 341 clones were downloaded from NCBI GenBank. These clone libraries included 1 scarab beetle larva, 2 cockroaches, and 13 termites. If necessary, the community structure of the phylotypes was recreated using tables in the respective publications.

The remaining 14 clone libraries consisting of 11 to 221 clones were created from a variety of different host groups, including 4 different millipede genera from 3 different families, 1 new cockroach species, and clone libraries of higher termites belonging to 4 subfamilies of higher termites. One of these subfamilies (*Apicotermitinae*) has never been investigated before. In addition to the subfamilies that have already been investigated in the literature, we added information about new dietary groups such as grass- and wood-feeding *Termitinae* and *Nasutitermitinae* (Tab. 1).

After classification into four methanogenic orders and the phylum *Thaumarchaeota*, differences between the compositions of the archaeal communities were obvious already at the order level (Fig. 1). *Methanobacteriales* was the predominant order, followed by *Methanosarcinales*, *Methanomicrobiales*, and *Methanomassiliicoccales*. Millipedes harbored either a combination of *Methanobacteriales* together with *Methanomicrobiales* and *Methanosarcinales* in high abundance (Aphco and Micun) or *Methanosarcinales* together with *Methanomassiliicoccales* and *Methanomicrobiales* (Anamo), or a relatively even community of all four methanogenic orders in the dataset (Harsp).

Similar to the gut community of the millipedes, the scarab beetle larvae hindgut (Pacep) was dominated by *Methanobacteriales*, and *Methanomassiliicoccales* and *Methanosarcinales* were almost equivalently abundant. When Egert et al. (2003) performed this analysis; they found only two methanogenic orders (*Methanobacteriales* and *Methanosarcinales*) and members that were then assigned to the order of *Thermoplasmatales*. It is now known that these sequences belong to the recently discovered methanogenic order *Methanomassiliicoccales* (Paul et al., 2012; Iino et al., 2013). Clones from wood-feeding cockroaches (Panan, Sales, Salta) were classified to all four methanogenic orders. In the data sets of *Panesthia angustipennis* (Panan) and *Salganea taiwanensis* (Salta) published by Hara et al. (2002), members of *Methanosarcinales* and *Methanomassiliicoccales* (in their article called cluster XSAT3A and XSAT3B, respectively, of *Thermoplasmatales*) dominate. Our clone library of *Salganea esakii* (Sales) differed from the other two clone libraries. Whether this difference is due to the use of different primers or whether this reflects true biological variability cannot be discerned. In the clone libraries of lower termites (all obtained from literature), mostly the order *Methanobacteriales* was detected. However, the clone library of *Reticulitermes speratus* (Retspe1) from Shinzato et al. (1999) also shows the presence of members of the orders *Methanomicrobiales* and *Methanomassiliicoccales*. Notably, in that study, the archaeal communities of six different colonies from different regions in Japan (data not shown) varied. The authors concluded that the orders *Methanomassiliicoccales* and *Methanomicrobiales* form only minor and variable fractions of the community in *R. speratus*. Earlier studies of lower termites used a different primer set that has been recently identified as having a mismatch against *Methanomassiliicoccales*-related sequences (Paul et al., 2012). This indicates that lower termites might not consist exclusively of a *Methanobacteriales*-dominated archaeal community, as the respective studies concluded.

**Tab. 1 | Characteristics of the 16S rRNA gene clone libraries of the archaeal hindgut microbiota of each host species.** The same identifiers are used to identify the samples in all tables and figures of clone libraries. For the origin of samples that were not part of a previous publication, see the legend of Tab. 2.

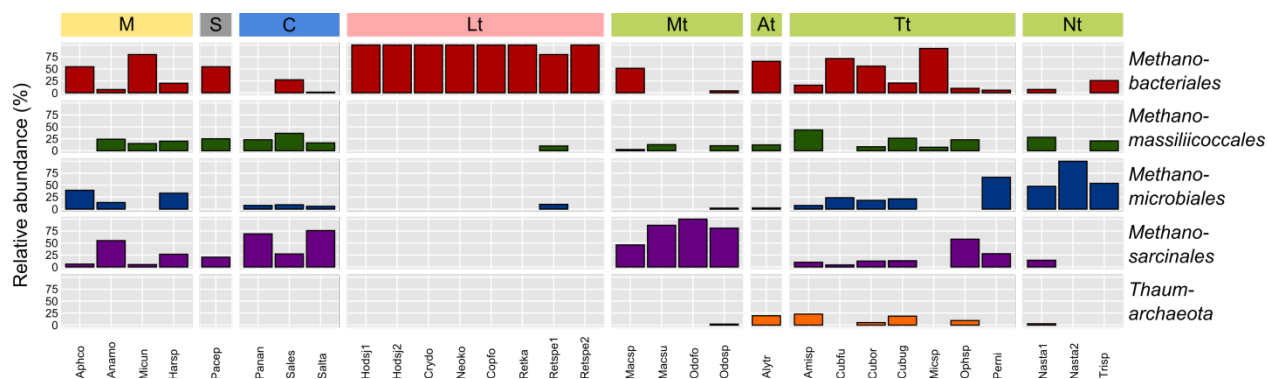
Host species	Compartment <sup>a</sup>	Identifier	Clones	Diet group	Origin/Reference <sup>b</sup>
<b>Millipedes</b>					
<b>Pachybolidae</b>					
<i>Aphistogonoiulus corallipes</i>	Wg	Aphco	33	Litter	B1
<b>Rhinocricidae</b>					
<i>Anadenobolus monilicornis</i>	Hg	Anamo	29	Litter	B1
<b>Spirostreptidae</b>					
<i>Microtrullius uncinatus</i>	Hg	Micun	22	Litter	B1
<b>Harpagophoridae</b>					
<i>Harpagophorida</i> sp.	Wg	Harsp	30	Litter	B1
<b>Scarab beetle larvae</b>					
<b>Scarabaeidae</b>					
<i>Pachnoda ephippiata</i>	M,Hg	Pacep	68 <sup>c</sup>	Humus	(Egert et al., 2003)
<b>Cockroaches</b>					
<b>Blaberidae</b>					
<i>Panesthia angustipennis</i>	Hg	Panan	27	Wood	(Hara et al., 2002)
<i>Salganea esakii</i>	Hg	Sales	11	Wood	B3
<i>Salganea taiwanensis</i>	Hg	Salta	69	Wood	(Hara et al., 2002)
<b>Lower termites</b>					
<b>Hodotermitidae</b>					
<i>Hodotermopsis sjoestedti</i>	Wg	Hodsj1	12	Wood	(Tokura et al., 2000)
<i>Hodotermopsis sjoestedti</i>	Wg	Hodsj2	5	Wood	(Shinzato et al., 2001)
<b>Kalotermitidae</b>					
<i>Cryptotermes domesticus</i>	Lumen	Crydo	37 <sup>c</sup>	Wood	(Ohkuma and Kudo, 1998)
<i>Neotermes koshunensis</i>	Wg	Neoko	5	Wood	(Shinzato et al., 2001)
<b>Rhinotermitidae</b>					
<i>Coptotermes formosanus</i>	Wg	Copfo	5	Wood	(Shinzato et al., 2001)
<i>Reticulitermes kanmonensis</i>	Wg	Retka	5	Wood	(Shinzato et al., 2001)
<i>Reticulitermes speratus</i>	Wg	Retspe1	60 <sup>c</sup>	Wood	(Shinzato et al., 1999)
<i>Reticulitermes speratus</i>	Wg	Retspe2	24	Wood	(Tokura et al., 2000)
<b>Higher termites (Termitidae)</b>					
<b>Macrotermitinae</b>					
<i>Macrotermes</i> sp.	Hg	Macsp	39	Fungus	L1
<i>Macrotermes subhyalinus</i>	Hg	Macsu	39	Fungus	F5
<i>Odontotermes formosanus</i>	Wg	Odofo	20	Fungus	(Ohkuma et al., 1999)
<i>Odontotermes</i> sp.	Hg	Odosp	48	Fungus	F4
<b>Apicotermitinae</b>					
<i>Alyscotermes trestus</i>	Hg	Alytr	41	Humus	F5
<b>Termitinae</b>					
<i>Amitermes</i> sp.	C—P5	Amisp	164 <sup>c</sup>	Interface	F8
<i>Cubitermes fungifaber</i>	Hg	Cubfu	50	Humus	(Donovan et al., 2004)

Host species	Compartment <sup>a</sup>	Identifier	Clones	Diet group	Origin/Reference <sup>b</sup>
<i>Cubitermes orthognathus</i>	P1—P5	Cubor	110 <sup>a</sup>	Humus	(Friedrich et al., 2001)
<i>Cubitermes ugandensis</i>	C—P5	Cubug	190 <sup>c</sup>	Humus	F6
<i>Microcerotermes</i> sp.	Wg	Micsp	41	Wood	F5
<i>Ophiotermes</i> sp.	C—P5	Ophsp	221 <sup>c</sup>	Humus	F7
<i>Pericapritermes nitobei</i>	Wg	Perni	18	Humus	(Ohkuma et al., 1999)
<b><i>Nasutitermitinae</i></b>					
<i>Nasutitermes takasagoensis</i>	Wg	Nasta1	12	Wood	(Ohkuma et al., 1999)
<i>Nasutitermes takasagoensis</i>	Wg	Nasta2	341 <sup>c</sup>	Wood	(Miyata et al., 2007)
<i>Trinervitermes</i> sp.	Hg	Trisp	39	Grass	F5

<sup>a</sup> Gut compartment: Wg, Whole gut; Hg, Hindgut; M, Midgut; C–P5, all adjacent compartments from crop to P5 were separately analyzed; P1–P5, all adjacent compartments from P1 to P5 were separately analyzed.

<sup>b</sup> Origins of samples: B, commercial breeders (B1: b.t.b.e. Insektenzucht, Schnürpflingen, Germany); B3, Jörg Bernhardt, Helbigsdorf, Germany [<http://www.schaben-spinnen.de>]; F, field collections (F4, near Kajiado, Kenya; F5, near Nairobi, Kenya [by J. O. Nonoh]; F6, Lhiranda Hill, Kakamega, Kenya [by J. O. Nonoh]; F8, near Eldoret, Kenya [by D.K. Ngugi]). L, laboratory colonies (L1, R. Plarre, Federal Institute for Materials Research and Testing, Berlin, Germany).

<sup>c</sup> Clones were distributed over either different compartments or samples of the same species. *Pachnoda ephippiata* (Pacep): M: 24, Hg: 44; *Cryptotermes domesticus* (Crydo), 23 of 37 clones were classified as archaeal; *Reticulitermes speratus* (Retspe1), sample RS1–RS6, each with 10 clones (RS1 was chosen); *Amitermes* sp. (Amisp), C: 2, M: 26, P1: 26, P3: 41, P4: 41, P5: 28; *Cubitermes orthognathus* (Cubor), P1: 27, P3: 26, P4: 26, P5: 31; *Cubitermes ugandensis*, (Cubug), C: 20, M: 15, P1: 35, P3: 48, P4: 42, P5: 30; *Ophiotermes* sp. (Ophsp), C: 26, M: 29, P1: 39, P3: 45, P4: 39, P5: 43; *Nasutitermes takasagoensis* (Nasta2), 341 clones in total from six libraries (The control group, wood-fed, had a total of 71 clones).



**Fig. 1 | Structure of the archaeal communities in millipedes, scarab beetle larvae, cockroaches, and termites based on phylogenetic analysis of the 16S rRNA gene.** The bar plots represent the order-level classification of the clone library sequences of this study and those obtained from various publications (see Tab. 1). The data is split vertically according to host group (M, Millipedes; S, Scarab beetles; C, Cockroaches; Lt: Lower termites; Mt, Macrotermitinae; At, Apicotermitinae; Tt, Termitinae; Nt, Nasutitermitinae) and horizontally according to the phylogenetic result at the order level (except Thaumarchaeota). Sample names are abbreviated as in Tab. 1.

The archaeal communities in higher termites showed similarities at the subfamily level. For example, the *Macrotermitinae* revealed a high abundance of *Methanosarcinales*, followed by *Methanomassiliicoccales*. In the *Macrotermes* sp. (Macsp), we also found a high abundance of *Methanobacteriales*, which indicated that variation within the termite genus *Macrotermes* can since the other *Macrotermes subhyalinus* (Macsu) has a different composition of methanogens. The archaeal communities of *Macrotermitinae* were relatively similar to those of wood-feeding cockroaches. Also, the subfamilies *Apicotermitinae* and *Termitinae* had a similar composition of archaeal order-level taxa. The analysis of several clone libraries indicated that members of the *Methanobacteriales* are the most abundant group of methanogens in these termites, but also other orders are represented. A clear change in community structure is evident in the higher termite subfamily *Nasutitermitinae*, in which the archaeal communities were dominated by the order *Methanomicrobiales*.

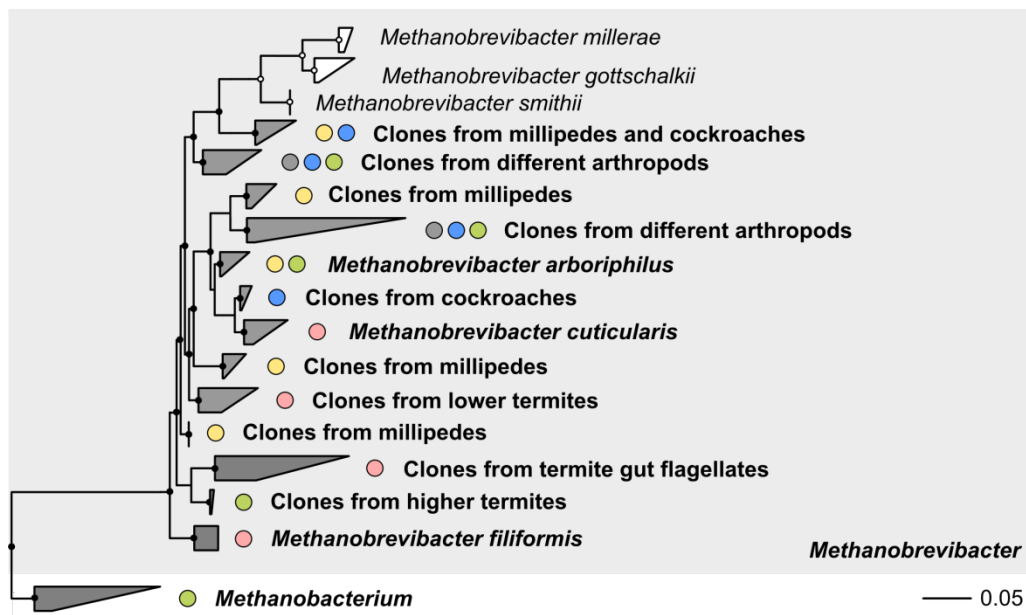
Sequences classified as the phylum *Thaumarchaeota* were only detected in higher termites, making up 2–20% of the sequences in the respective clone libraries. Compared to the abundance of the methanogenic orders, members of *Thaumarchaeota* were never the predominant archaeal group.

**Phylogenetic positions of arthropod archaeal sequences.** Sequences derived from the clone libraries were classified at the order level and phylogenetically analyzed, yielding calculated phylogenetic trees of the methanogenic orders found in arthropods (Figs. 2–4). The phylum *Thaumarchaeota* was analyzed only superficially as these sequences in the SILVA database were not of high quality. Therefore, these sequences were analyzed only at the class level. An in-depth phylogenetic analysis is planned for an upcoming publication.

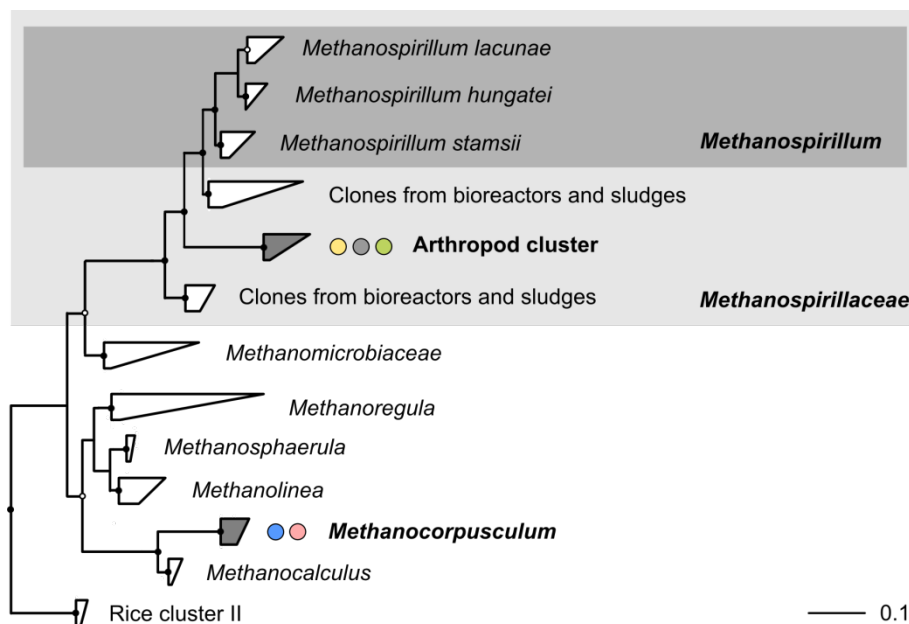
In general, sequences belonging to the order *Methanobacteriales* are clearly placed within the genus *Methanobrevibacter*. The closest cultured representatives include *Methanobrevibacter arboriphilus*, *Methanobrevibacter cuticularis*, *Methanobrevibacter filiformis*, and *Methanobrevibacter smithii*. All sequences from this study form either monophyletic clusters that consist of only sequences from specific

host groups (e.g., higher termites) or from picked flagellates, or form intermixed clusters with sequences originating from many arthropod hosts. This indicates a specificity of the lineages in question to the arthropod gut systems in general, but cospeciation was not detected. Some clones from the study of Deevong et al. (2004) were found to be in radiated in the genus *Methanobacterium*, as was reported in that study.

Sequences that clustered within the order Methanomicrobiales are less diverse than the Methanobacteriales-related sequences. To date, only two clusters have been identified as containing arthropod sequences (Fig. 3). One cluster is within the genus *Methanocorpusculum*. The closest cultured representative of the arthropod sequences belonging to this genus is *Methanocorpusculum parvum*, as has also been documented by Shinzato et al. (1999). This type strain uses either hydrogen and carbon dioxide or 2-propanol and carbon dioxide for methanogenesis (Zellner et al., 1987). The second cluster – into which most of the Methanomicrobiales-related sequences of arthropod guts fell – was a monophyletic cluster consisting only of sequences from higher termites. This cluster is basal to the sequences of the genus *Methanospirillum*, which indicated that sequences of this cluster form a new genus-level group within the Methanospirillaceae. This is also underlined by the high degree of dissimilarity of 9–12% to *Methanospirillum stamsii*, compared to the distances within this cluster (up to 5%). Therefore, we tentatively name the cluster ‘Methanospirillaceae arthropod cluster’.



**Fig. 2 | Phylogenetic tree based on 16S rRNA genes of *Methanobacteriales*-related sequences from clone libraries of this study and previously published sequences.** Colored circles indicate the origin of sequences in the clusters; the same color code as in Fig. 1 is used. Filled and unfilled dots indicate bootstrap support of the nodes (< 70%, no dot; ≥ 70%, ○; ≥ 90%, ●).



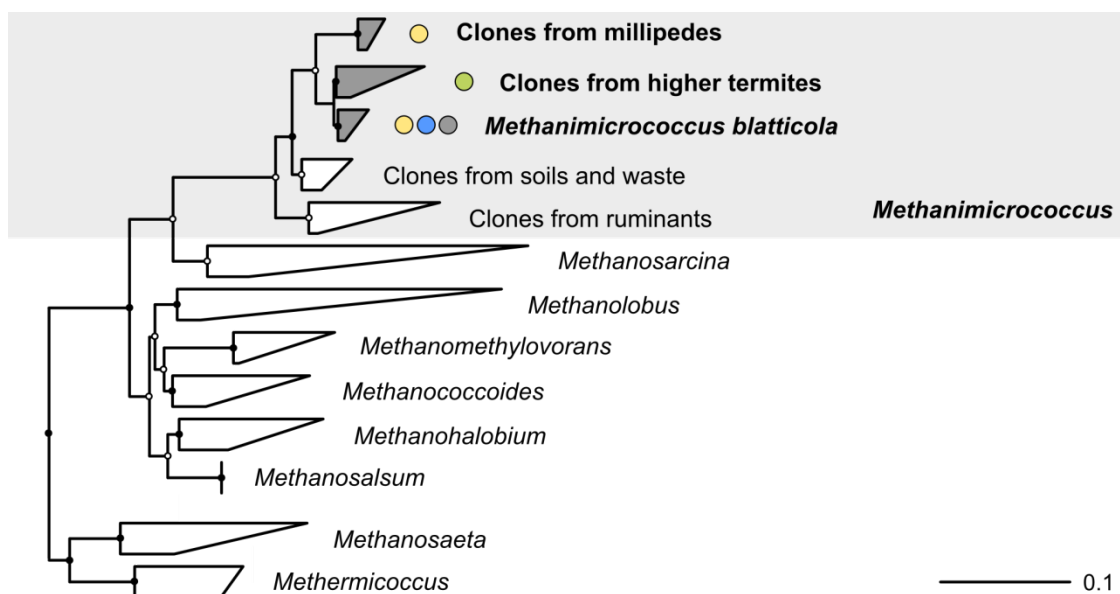
**Fig. 3 | Phylogenetic tree based on 16S rRNA genes of *Methanomicrobiales*-related sequences from clone libraries of this study and previously published sequences.** Colored circles indicate the origin of sequences in the clusters; the same color code as in Fig. 1 is used. Filled and unfilled dots indicate bootstrap support of the nodes (< 70%, no dot;  $\geq 70\%$ ,  $\circ$ ;  $\geq 90\%$ ,  $\bullet$ ).

Within the *Methanosarcinales*, all sequences from clone libraries were phylogenetically located in the radiation of *Methanomicrococcus* (Fig. 4). *Methanomicrococcus blatticola* is the only cultured representative of this phylogenetic group and is able to reduce methanol and different methylamines with hydrogen as external electron donor (Sprenger et al., 2000). Arthropod-gut-derived sequences were located in three clusters. One group lay within a 3% radius around *M. blatticola* (mostly cockroach-gut-derived sequences). Another cluster consisted only of sequences originating from higher termite gut samples (1–2% difference to *M. blatticola*). The third group was apical to the other two groups and consisted only of sequences from millipede guts (2–4% dissimilarity to the groups of *M. blatticola* and the higher termite cluster). Since all arthropod-gut-derived sequences were highly similar to that of the type strain *M. blatticola*, we grouped these clusters within the genus-level group of *Methanomicrococcus*.

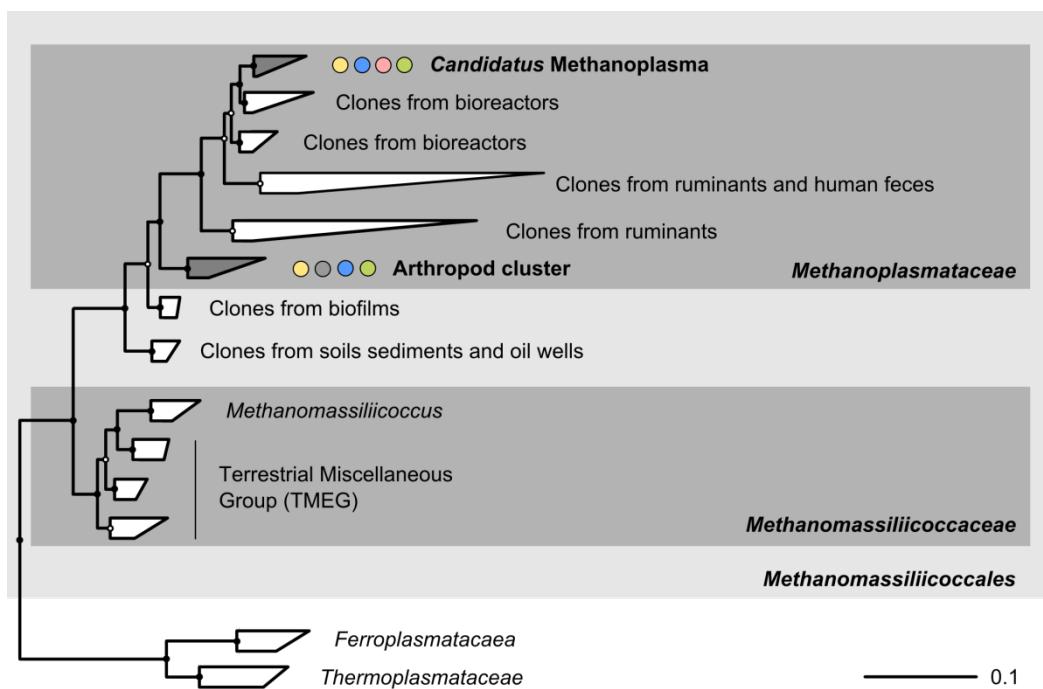
Recently, a monophyletic group within the class *Thermoplasmata* was found to represent methanogenic archaea (Dridi et al., 2012; Paul et al., 2012). Past studies of arthropod guts always identified a moderate proportion of sequences highly similar to the *Thermoplasmatales*. These studies included scarab beetle larvae (Egert et al., 2003), cockroaches (Hara et al., 2002), lower termites (Shinzato et al., 1999), and different higher termites (Ohkuma et al., 1999; Friedrich et al., 2001; Donovan et al., 2004; Miyata et al., 2007) and document that the novel order (*Methanomassiliicoccales*) is a widespread group of methanogens within arthropod guts. This is also reflected in the phylogenetic placement of the sequences from arthropod-gut-derived clones (Fig. 5). We found a total of ten minor clusters. The most apical group contained sequences highly similar to the enrichment culture “*Candidatus Methanoplasma termitum*” (clustered at a radius of 3%), consisting of sequences from millipedes, higher termites, and wood-feeding cockroaches.

The latter sequences were formerly described as the XSAT3A cluster (Hara et al., 2002). Since these sequences showed a difference of maximally 5% to “*Candidatus Methanoplasma termitum*”, we considered the whole cluster to be a genus-level group and name it ‘*Candidatus Methanoplasma*’. Basal to this group, a cluster containing sequences from ruminants, bioreactors, and human feces was located, including 16S rRNA sequences of the recently enriched methanogens “*Candidatus Methanomethylophilus alvus*” (Borrel et al., 2012) and “*Candidatus Methanogranum caenicola*” (Iino et al., 2013). The second major group of arthropod sequences was basal to the first group and contained also clones from different arthropod hosts. Each host group was represented by at least one cluster. The clones from wood-feeding cockroaches were formerly classified as the XSAT3B cluster (Hara et al. 2002). Since the sequences in this cluster differed by 1% to 6%, we defined this group as a new genus-level group ‘arthropod cluster’ within the *Methanoplasmataceae* (see Fig. 5). The least common node of all enrichment cultures and the isolate *Methanomassiliicoccus (Mmc.) luminyensis* (Dridi et al., 2012) were used to conservatively define the order *Methanomassiliicoccales*. Recently, Iino et al. (2013) proposed the family *Methanomassiliicoccaceae* based on an analysis of the gene encoding the alpha-subunit of methyl-coenzyme M reductase (*mcrA*) and the 16S rRNA gene. In both analyses, the authors find their enrichment culture *Ca. Methanogranum caenicola* clearly separated from the first isolate, *Mmc. luminyensis* (Dridi et al., 2012). We decided to place the cluster in which *Mmc. luminyensis* is located at the family level *Methanomassiliicoccaceae*. Subsequently, the least common node of the *Methanoplasmataceae* arthropod cluster and of the enrichment cultures *Ca. Methanogranum caenicola*, *Ca. Methanoplasma termitum*, and *Ca. Methanomethylophilus alvus* was used to define the *Methanoplasmataceae* at the family level.

Sequences that were classified to the phylum Thaumarchaeota are not described or discussed in detail as these will be part of another study.



**Fig. 4 | Phylogenetic tree based on 16S rRNA genes of *Methanosarcinales*-related sequences from clone libraries of this study and previously published sequences.** Colored circles indicate the origin of sequences in the clusters; the same color code as in Fig. 1 is used. Filled and unfilled dots indicate bootstrap support of the nodes (< 70%, no dot; ≥ 70%, ○; ≥ 90%, ●).



**Fig. 5 | Phylogenetic tree based on 16S rRNA genes of *Methanomassiliococcales*-related sequences from clone libraries of this study and previously published sequences.** Colored circles indicate the origin of sequences in the clusters; the same color code as in Fig. 1 is used. Filled and unfilled dots indicate bootstrap support of the nodes (< 70%, no dot; ≥ 70%, ○; ≥ 90%, ●).



**Primer design.** Based on the existing primer set A533b/A934b from Grosskopf et al. (1998), we designed new primers for the analysis of the archaeal community structure. We slightly modified the primers by maximizing the number of sequences targeted, with emphasis on arthropod gut archaeal groups and methanogens. In general, the primer set had a high specificity towards the domain Archaea (Tab. 2). In both scenarios, with and without a mismatch, no sequences originating from the bacterial domain were bound *in silico*. With one mismatch, 0.3% of the eukaryotic sequences were bound, but we excluded those sequences belonging to phylum *Arthropoda*. All archaeal phylum-level groups are well covered by this primer set, except the Ancient Archaeal Group (AAG), the Marine Hydrothermal Vent Groups I and II (MHVG-1 and MHVG-2), and the phylum-level group *Nanoarchaeota*. Since these phylum-level groups are not reported to occur in arthropod gut systems, we considered our primer set appropriate for the profiling of the archaeal communities in arthropod guts.

**Tab. 2 | Coverage of the primer set A533f\_mod A934b\_mod of major archaeal groups in the SILVA database. Data differentiates between the coverage obtained with no mismatch or one mismatch allowed.**

Taxonomic level	Coverage (%)	
	No mismatch	One mismatch
<b>Archaea</b>	<b>81.7</b>	<b>90.8</b>
Ancient Archaeal Group (AAG)	0.0	0.0
<i>Crenarchaeota</i>	86.2	95.3
<i>Euryarchaeota</i>	81.0	92.3
<i>Methanobacteriales</i>	90.9	96.8
<i>Methanococcales</i>	1.4	86.1
<i>Methanocellales</i>	83.3	86.7
<i>Methanomicrobiales</i>	89.5	94.1
<i>Methanosarcinales</i>	86.5	92.7
<i>Methanopyrales</i>	100.0	100.0
<i>Methanomassiliicoccales</i>	95.2	98.6
<i>Korarchaeota</i>	3.9	90.2
Marine Hydrothermal Vent Group 1 (MHVG-1)	7.7	7.7
Marine Hydrothermal Vent Group 2 (MHVG-2)	0.0	0.0
<i>Nanoarchaeota</i>	2.9	5.7
<i>Thaumarchaeota</i>	83.9	88.9
<b>Bacteria</b>	<b>0.0</b>	<b>0.0</b>
<b>Eukaryota</b>	<b>0.0</b>	<b>0.3</b>

**Reference database.** The current SILVA database contains 184 archaeal genus-level groups. We gave a name to the order *Methanomassiliicoccales*, which was not described in the database. Our phylogenetic analysis led to the addition of the following new genus-level groups to the database: arthropod cluster (order *Methanomassiliicoccales* and family *Methanoplasmataceae*) and arthropod cluster (order *Methanomicrobiales* and family *Methanospirillaceae*). The performance of this database is evaluated in the next subsection.

**Profiling of the archaeal communities in arthropods.** Since a comparison of the different clone libraries is challenging due to the use of different primer sets and sampling efforts (sometimes only 5 clones per clone library), we carried out a large-scale sequencing experiment with representatives of all major arthropod groups known to emit methane (Hackstein and Stumm, 1994). We collected 48 samples and sequenced DNA using the newly designed primer pair A533f\_mod and A934b\_mod. We obtained 1,026–16,745 sequences per sample (Tab. 3). Sequences were first clustered into OTUs at the 3% level, resulting in minimally 2 and up to 23 archaeal OTUs per sample. Subsequently, OTUs were classified into genus-level bins with the reference database created in this study. We were able to classify between 92.4% (*Ergaula capucina*, No. 9) and 100% (almost all samples) of all sequences at the genus level (Tab. S1). This indicated that almost all genus-level groups in the data set could be identified by the classification approach.

Eleven different genera were found in the total dataset. Each sample had a minimum of two and a maximum of nine different genus levels (Tab. 3). These numbers were in agreement with the number of OTUs at the 5% level, which showed that the genus-level groups of the reference database reflect the natural diversity well. The distribution of the archaeal genus-level groups indicated that most of the sequences were classified as methanogenic genera. The distribution of the genus-level groups in all samples revealed that the predominant lineages in this dataset are *Methanobrevibacter*, *Methanomicrococcus*, *Ca. Methanoplasma*, *Methanoplasmataceae* arthropod cluster, and the Miscellaneous Crenarchaeotic Group of the *Thaumarchaeota*. The archaeal community structure was consistent within each major arthropod group (Fig. 6).

Millipedes harbored mainly *Methanomicrococcus*, followed by *Methanobrevibacter*. The other genus-level groups were present only in low abundance. These sequencing results are in agreement with the results of the millipede clone libraries.

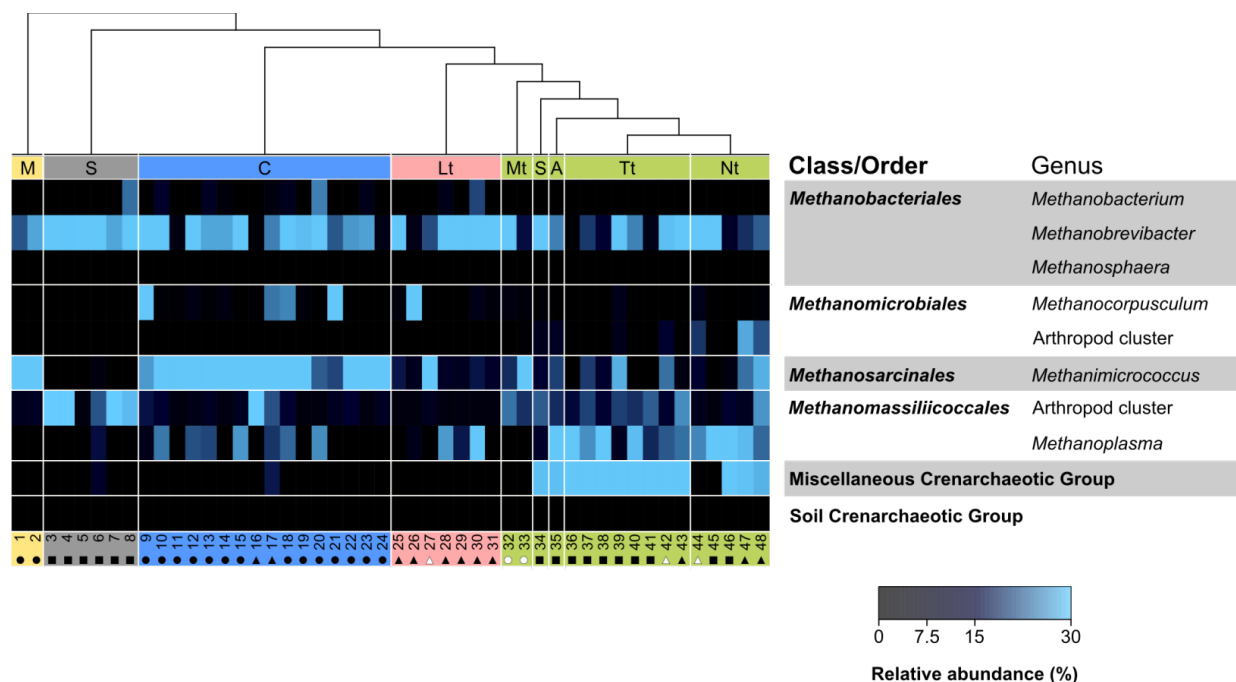
The archaeal communities of scarab beetle larvae hindguts revealed a high abundance of both *Methanobrevibacter* and the *Methanoplasmataceae* arthropod cluster or only *Methanobrevibacter*. This is not in agreement with the results of Egert et al. (2003) (Fig. 1), who found additionally members of the order *Methanosarcinales* in high abundance in the hindgut of *Pachnoda ephippiata*. Since the order *Methanosarcinales* was one of the dominating orders in the total data set, although not in any scarab beetle larvae in this sequencing study (No. 3–8), and since the archaeal communities of other soil-feeding larvae sampled in this study also showed variation, we concluded that the archaeal community of scarab beetle larvae is variable.

The archaeal community of most cockroaches was dominated by *Methanomicrococcus* and *Methanobrevibacter*, like the millipede samples. Major exceptions were the leaf-feeding cockroach *Ergaula capucina* (No. 9), the wood-feeding cockroaches *Panesthia angustipennis* and *Salganea esakii* (No. 16 and 17), and the generalist blattid cockroach *Blatta orientalis* (No. 21), which either had a high abundance of the archaeal genus *Methanocorpusculum* (No. 9 and 21) or a high abundance of the *Methanoplasmataceae*

arthropod cluster (No. 16 and 17). These results are in good agreement with the only other data available on archaeal community structure of cockroaches, namely *Salganea taiwanensis*, and *P. angustipennis* (Hara et al., 2002), and with the clone library data of our study.

The archaeal communities in most of the lower termites consisted of *Methanobrevibacter*, which is in agreement with results in the literature (Ohkuma and Kudo, 1998; Shinzato et al., 1999; Tokura et al., 2000; Shinzato et al., 2001). The only exceptions were *H. sjoestedti* (No. 26), *H. mossambicus* (No. 27), and *R. santonensis* (No. 30). *H. sjoestedti* had a high abundance of *Methanocorpusculum*, *H. mossambicus* had a high abundance of *Methanomicrococcus*, and *R. santonensis* had a mixture of both *Methanobrevibacter* and *Candidatus Methanoplasma* in high abundance. The results of *H. sjoestedti* contradict results of Tokura et al. (2000) and Shinzato et al. (2001); however, these studies used different primers and were at a different resolution. The archaeal community of *R. santonensis* had not yet been studied, but in studies of other species of the same genus (Shinzato et al., 1999; Tokura et al., 2000; Shinzato et al., 2001), *Methanobrevibacter* was the predominant order. However, Shinzato et al. (1999) revealed substantial differences in the community structure of different *R. speratus* samples depending on the sampling location and the termite colonies; members of the *Methanomassiliicoccales* were sometimes found. This suggests that low resolution of the analysis and the choice of primer sets in the past led to an underestimation of certain members of the archaeal communities.

The most typical characteristic of higher termite archaeal communities was the presence of group MCG of the phylum *Thaumarchaeota*. Except for *Macrotermitinae*, all higher termite subfamilies had a high abundance of this class, followed by *Ca. Methanoplasma* and *Methanobrevibacter*. Also other methanogenic genus-level groups were detected in small amounts. The *Macrotermitinae* had either a high abundance of *Methanobrevibacter* (*Odontotermes* sp.) or *Methanomicrococcus* (*Macrotermes* sp.). The community structure of the two *Nasutitermes* species (No. 47 and 48) are in good agreement with the results of Miyata et al. (2007).



**Fig. 6 | Distribution of archaeal genus-level groups in the arthropod gut samples from this study.** The two panels show the same dataset but are differently horizontally arranged. The left panel shows the dataset ordered according to the taxonomy of the hosts (M, millipedes; S, scarab beetle larvae; C, cockroaches; Lt, lower termites; Mt, *Macrotermitinae*; A, *Apicotermitinae*; S (in green), *Syntermitinae*; Tt, *Termitinae*; Nt, *Nasutitermitinae*). The numbers below each panel are the sample identifiers used in Tab. 3. The symbols below the sample identifiers indicate the dietary specialization of the arthropod host (●, litter; ■, humus; ▲, wood; △, grass; ○, fungus/wood).

**Tab. 3 | Characteristics of the high-throughput sequencing libraries used in this study.** The same identifiers are used to identify the samples in all tables and figures of the high-throughput sequencing data.

Insect species	No.	Diet group	Origin <sup>a</sup>	Sequences	No. of 5% OTUs	No. of genus-level groups	No. of 3% OTUs	Diversity
<b>Millipedes</b>								
<b><i>Spirostreptidae</i></b>								
<i>Anadenobolus monilicornis</i>	1	Litter	B1	14,804	9	6	15	1.1
<b><i>Harpagophoridae</i></b>								
<i>Harpagophorida</i> sp.	2	Litter	B1	13,886	9	6	15	1.1
<b>Scarab beetle larvae</b>								
<b><i>Cetoniidae</i></b>								
<i>Dicronorhina derbyana</i>	3	Humus	B2	3,390	4	2	4	1.0
<i>Genyodonta lequexi</i>	4	Humus	B2	5,881	3	3	3	0.6
<b><i>Scarabaeidae</i></b>								
<i>Pachnoda aemula</i>	5	Humus	B2	2,689	4	3	5	0.1
<i>Pachnoda ephippiata falkei</i>	6	Humus	B2	3,632	10	6	14	1.2
<i>Gnorimus tibialis</i>	7	Humus	B2	5,221	3	2	4	0.8
<i>Xylotrupes gideon</i>	8	Humus	B2	2,212	3	3	5	1.4
<b>Cockroaches</b>								
<b><i>Polyphagidae</i></b>								
<i>Ergaula capucina</i>	9	Litter	B3	4,030	12	8	22	1.9
<b><i>Blaberidae</i></b>								
<i>Elliptorhina chopardi</i>	10	Litter	B3	2,168	9	7	13	2.0
<i>Panchlora</i> sp.	11	Litter	B3	4,101	7	5	10	0.3
<i>Opisthoptlatia orientalis</i>	12	Litter	B3	2,940	10	6	16	1.7
<i>Nauphoeta cinerea</i>	13	Litter	B3	10,653	13	8	23	1.6
<i>Gromphadorhina portentosa</i>	14	Litter	B3	2,196	8	6	15	0.9
<i>Diploptera punctata</i>	15	Litter	B3	1,760	8	5	10	1.6

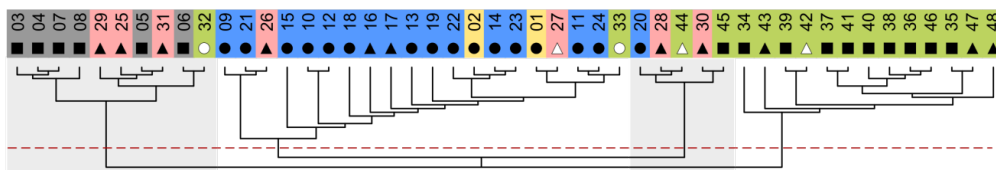
Insect species	No.	Diet group	Origin <sup>a</sup>	Sequences	No. of 5% OTUs	No. of genus-level groups	No. of 3% OTUs	Diversity
<i>Panesthia angustipennis</i>	16	Wood	B3	8,010	11	5	13	0.9
<i>Salganea esakii</i>	17	Wood	B3	4,515	12	7	20	2.0
<i>Eublaberus posticus</i>	18	Litter	B3	2,560	9	7	14	1.9
<i>Schultesia lampyrodiformis</i>	19	Litter	B3	13,700	11	5	17	1.5
<i>Henschoutedenia flexivitta</i>	20	Litter	B3	2,750	9	7	15	1.8
<b>Blattidae</b>								
<i>Blatta orientalis</i>	21	Litter	B3	3,330	8	6	16	1.4
<i>Eurycotis floridiana</i>	22	Litter	B3	5,025	12	3	19	1.8
<i>Shelfordella lateralis</i>	23	Litter	B3	16,745	10	6	17	0.9
<b>Cryptocercidae</b>								
<i>Cryptocercus punctulatus</i>	24	Wood	F1	14,294	10	5	13	0.3
<b>Lower termites</b>								
<b>Mastotermitidae</b>								
<i>Mastotermes darwiniensis</i>	25	Wood	L1	3,129	9	5	16	1.1
<b>Hodotermitidae</b>								
<i>Hodotermopsis sjoestedti</i>	26	Wood	L1	15,463	10	5	17	0.8
<i>Hodotermes mossambicus</i>	27	Grass	F2	5,844	10	6	16	0.7
<b>Termopsidae</b>								
<i>Zootermopsis nevadensis</i>	28	Wood	L2	2,269	5	4	8	0.9
<b>Kalotermitidae</b>								
<i>Neotermes jouteli</i>	29	Wood	F3	1,5181	10	7	16	0.8
<b>Rhinotermitidae</b>								
<i>Reticulitermes santonensis</i>	30	Wood	L2	2,067	9	8	14	1.4
<i>Coptotermes niger</i>	31	Wood	L1	1,575	8	5	11	0.5

Insect species	No.	Diet group	Origin <sup>a</sup>	Sequences	No. of 5% OTUs	No. of genus-level groups	No. of 3% OTUs	Diversity
<b>Higher termites (<i>Termitidae</i>)</b>								
<b><i>Macrotermitinae</i></b>								
<i>Odontotermes</i> sp.	32	Litter/Fungus	F4	1,533	8	7	14	1.4
<i>Macrotermes</i> sp.	33	Litter/Fungus	L1	3,231	7	7	16	0.9
<b><i>Apicotermitinae</i></b>								
<i>Alyscotermes trestus</i>	34	Humus	F5	2,188	11	9	17	1.7
<b><i>Syntermitinae</i></b>								
<i>Cornitermes</i> sp.	35	Humus	L3	2,415	7	6	8	1.7
<b><i>Termitinae</i></b>								
<i>Cubitermes</i> sp.	36	Humus	F6	3,932	5	5	8	1.0
<i>Cubitermes</i> sp.	37	Humus	F7	2,799	7	5	6	1.3
<i>Cubitermes</i> sp.	38	Humus	F8	5,181	7	5	7	1.1
<i>Ophiotermes</i> sp.	39	Humus	F9	3,402	12	8	18	2.2
<i>Neocapritermes taracua</i>	40	Humus	L3	4,801	10	6	11	1.5
<i>Proboscitermes</i> sp.	41	Humus	L3	3,745	7	6	10	1.3
<i>Amitermes meridionalis</i>	42	Grass	F10	1,345	12	7	16	1.7
<i>Microcerotermes parvus</i>	43	Wood	L3	2,995	8	6	11	1.7
<b><i>Nasutitermitinae</i></b>								
<i>Trinervitermes</i> sp.	44	Grass	F5	1,571	9	7	15	1.5
<i>Atlantitermes</i> sp.	45	Humus	L3	4,383	2	2	2	0.7
Unclassified <i>Nasutitermitinae</i>	46	Humus	L3	4,654	6	5	6	1.0
<i>Nasutitermes corniger</i>	47	Wood	L4	1,457	9	7	12	1.8
<i>Nasutitermes takasagoensis</i>	48	Wood	F11	1,026	10	8	12	2.2

<sup>a</sup> Origin of samples: B, commercial breeders (B1: b.t.b.e. Insektenzucht, Schnürpflingen, Germany; B2: Wirbellosen Welt, Rödinghausen, Germany; B3, Jörg Bernhardt, Helbigsdorf, Germany [<http://www.schaben-spinnen.de>]); F, field collections (F1, Heywood County, NC, USA by C. Nalepa; F2, near Pretoria, South Africa by J. Rohland; F3, Fort Lauderdale, FL, USA by R. H. Scheffrahn; F4, near Kajiado, Kenya; F5, near Nairobi, Kenya by J. O. Nonoh; F6, Lhiranda Hill, Kakamega, Kenya by J. O. Nonoh; F7, South Africa, by M. Poulsen; F8, near Eldoret, Kenya by D.K. Ngugi; F9, Kalunja Glade, Kakamega, Kenya by D. K. Ngugi; F10, near Darwin, Australia by A. Brune; F11, near Nishihara, Japan by G. Tokuda); L, laboratory colonies (L1, R. Plarre, Federal Institute for Materials Research and Testing, Berlin, Germany; L2, MPI Marburg; L3, D. Sillam-Dusez, Bondy, France; L4, R. H. Scheffrahn, University of Florida, Fort Lauderdale, FL, USA).

**Cluster analysis and correlation of archaeal with bacterial genus-level groups.** When the archaeal communities of the arthropod gut samples were subjected to neighbor-joining cluster analysis based on the Soergel distance, specific archaeal community clusters became apparent (Fig. 7). Four major clusters were found. In order to discuss the results, we partitioned the data set into four clusters (red dashed line in the dendrogram in Fig. 7), which also illustrates the clear separation of the data into four clusters.

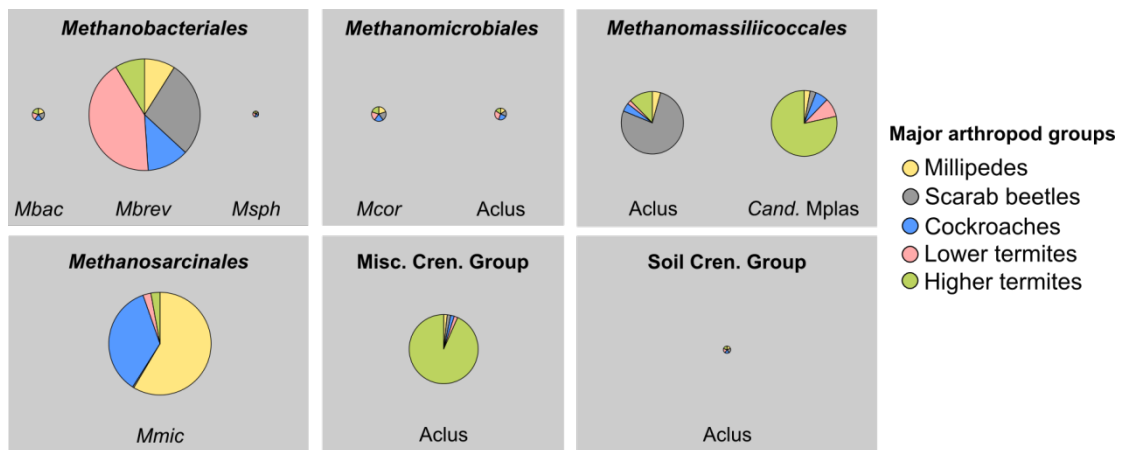
One major cluster was formed by the archaeal communities of higher termites, which documents the distinctness of their archaeal communities. Another cluster was formed by most cockroaches, which indicated a high similarity in their composition. However, within the cockroach cluster, also non-cockroach samples often clustered, including the lower termites *H. sjoestetdi* and *Hodotermes mossambicus*, the higher termite *Macrotermes* sp., and the two millipedes. This indicated that millipedes have a cockroach-like archaeal community. Also fungus-cultivating termites had an atypical archaeal community structure in contrast to other higher termites, as has already been documented for the bacterial community (Dietrich et al., 2014). The fourth cluster was formed by all scarab beetles. The only non-scarab beetle larvae archaeal communities were some lower termites and a fungus-cultivating termite. The archaeal community structures of the lower termites were not similar to each other since they did not show a specific clustering. When the lower termite samples were removed from the analysis (Fig. S1b), each archaeal community mostly reflected the membership of their host in a major host group.



**Fig. 7 | Community structure based host specificity. Unrooted BioNJ tree displayed as cladogram of the pairwise Soergel distances of the archaeal community structure in arthropod guts.** The dashed line aids in distinguishing the data set into four clusters (see also Fig. S1). The color code of hosts is the same as in Fig. 1; numbering is the same as in Tab. 3; and diet symbols are the same as in Fig. 6.

**Host specificity of archaeal lineages.** Since we found similarities in the archaeal community structure of arthropods belonging to the same major host group, we suspected also an abundance-based specificity of single archaeal genus-level groups for the major host groups. Therefore, we used the median abundance of each genus-level per host group and compared them with each other. The 5 most abundant genus-level groups belonged to the three different orders *Methanobacteriales* (*Mb*), *Methanomassiliicoccales* and *Methanosarcinales* and the class MCG (Fig. 8). The genus-level groups were *Methanobrevibacter*, *Methanomassiliicoccales* Arthropod cluster, *Candidatus* Methanoplasma, *Methanomicrococcus*, and the Miscellaneous Crenarchaeotic Group (MCG) Arthropod cluster. Each of the five groups occurs mostly only in certain host groups in high abundance. *Methanobrevibacter* was highly abundant in all host groups. In contrast, *Methanomicrococcus* occurred in high abundance only in cockroaches and millipedes, underlining the similarity of these host groups; The *Methanomassiliicoccales* Arthropod cluster clearly was mostly highly abundant in the guts of scarab beetle larvae, and both *Candidatus* Methanoplasma and the MCG Arthropod cluster occurred in high abundance in higher termites.





**Fig. 8 | Abundance-based host specificity of archaeal lineages illustrated as pie charts of different genus-level groups that occur in the high-throughput sequencing data set.** The area of the charts is scaled by the abundance in the total normalized dataset. The fractions of the different pie charts illustrate the median abundance in the different major arthropod groups. Genus-level group abbreviations: *Mbac*, *Methanobacterium*; *Mbrev*, *Methanobrevibacter*; *Msph*, *Methanosphaera*; *Aclus*, *Arthropod cluster*; *Cand. Mplas*, *Candidatus Methanoplasma*; *Mcor*, *Methanocorpusculum*; *Mmic*, *Methanomicrococcus*.

**Phylogenetic analysis of the short reads.** To elucidate not only whether archaeal lineages are preferentially abundant in certain host groups but also whether distinct phylotypes belonging to the different genus level groups are host specific, short reads were clustered sample-wise in 1%-level OTUs and were also subjected genus-level-wise to maximum-likelihood analyses. The resulting trees show representatives of OTUs that clustered at the 1%-level in the host sample (Fig. 9; see Fig. S2a–h for fully annotated trees). The genus *Methanobacterium* was represented by only a small number of sequences and was basically a trifurcation, which does not allow any conclusions about host-specific clusters and/or co-cladogenesis. The genus *Methanobrevibacter* was represented by enough OTU sequences to conclude that clusters exclusively consisted of sequences belonging to either the host group cockroaches or to lower termites. Lower termite sequences showed long branches compared to the sequences that originated from sample of other host groups might be caused by a different rate of evolution (Fig. S2b), which would indicate that these sequences might stem from flagellate symbionts, since the endosymbiotic lifestyle is often associated with increased mutation rates. However, in the lower termite cluster and the cockroach cluster of *Methanobrevibacter*, no evidence of co-cladogenesis could be found. *Methanobrevibacter*-related OTUs from the other host groups formed mainly clusters with OTU sequences from many host groups, which indicated that these *Methanobrevibacter* sequences are not shaped by host-specific mechanisms.

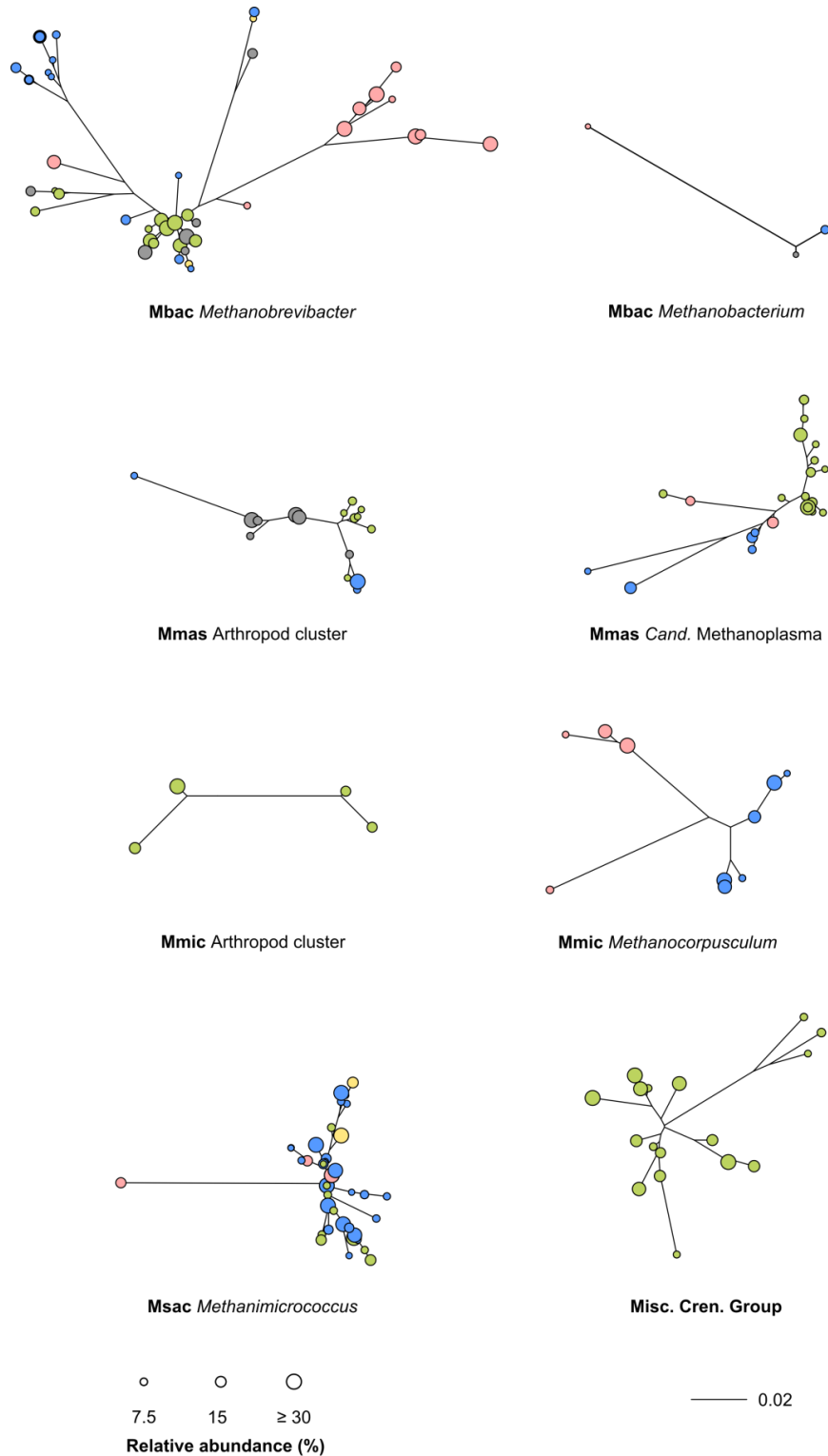
We observed similar phenomena for the order *Methanomassiliicoccales*. In the *Methanomassiliicoccales* arthropod cluster, a monophyletic group of sequences that originated only from higher termites samples occurred. The remaining two major clusters were consisting by either sequences from scarab beetle larvae or from the wood-feeding cockroach *Panesthia angustipennis* (No. 16). This indicates that scarab beetles and higher termites have host-specific lineages within the *Methanomassiliicoccales* Arthropod cluster with the exception of one OTU of *Panesthia angustipennis*. However, the more abundant and perhaps more important OTU from this host is located in another cluster. Notably, the OTU of *Macrotermes* sp. (No. 33) is located in the proximity of those of *Panesthia angustipennis* and *Pachnoda ephippiata falkei*. Co-cladogenesis could not be observed also in this genus-level group.

In the *Ca. Methanoplasma* genus-level group (Fig. S2d), a higher termite cluster and a cockroach cluster were clearly evident. Neither of these showed a cocladogenesis signal. Interestingly, *Reticulitermes santonensis* (No. 30) is represented by two OTUs in this tree, the first as a very basal lineage and the second in a deep-branching bifurcation with an OTU from *Nasutitermes corniger*.

The *Methanospirillaceae* arthropod cluster of the order *Methanomicrobiales* contained only sequences from higher termites; hence, the genus-level group itself seems to be a host-specific cluster (Fig. S2f). This is also supported by the clone library analysis in this study. The genus *Methanocopusculum* contains a monophyletic group that only consists of cockroach gut derived OTUs. Sequences originating from *Hodotermopsis sjoestedi* (No. 26) showed that within the community of *Methanocopusculum*, more than one abundant OTU is present. Therefore, it is not surprising that these sequences cluster together.

The order *Methanosarcinales* is only represented by a single genus, *Methanomicrococcus*. No large cluster was detected that only contained sequences originating from guts of a particular host group (Fig. S2g). Surprisingly, OTU sequences from millipedes and cockroaches often represented a monophyletic group, which was reflected by the community structure analysis. In addition, OTU sequences from cockroaches and higher termites often formed a cluster, which documented also similarities between OTU sequences from these two host groups.

OTU representatives from the *Thaumarchaeota* group Miscellaneous Crenarchaeotic Group contained exclusively sequences from abundant OTUs from higher termites, with the exception of one cockroach-derived OTU sequence. Based on the distances in the tree, we conclude that sequences fall within one major genus-level group. This is supported by the sequences that showed a maximal dissimilarity of 6% from each other.



**Fig. 9 | Phylogenetic host specificity of arthropod archaeal lineages.** Shown are phylogenetic trees of the OTUs at the 1%-level for the most abundant archaeal genus-level groups in arthropod guts (Sequences that show a minimum relative abundance of 1% are taken). The area of the circles indicates the relative abundance of the OTU in the host sample. Color coding of the host-group membership is the same as in Fig. 1. Archaeal group name abbreviations: *Mbac*, *Methanobacteriales*; *Mmas*, *Methanomassilliicoccales*; *Mmic*, *Methanomicrobiales*; *Msac*, *Methanosarcinales*; *Misc. Cren. Group.*, Miscellaneous Crenarcheotic Group. Note that the *Methanomicrobiales* arthropod cluster tree branch lengths were magnified by a factor of 10.

**Possible interactions with bacteria.** Another driver of the archaeal community structure could be the dependence on the availability of substrates. The most important methanogenic substrates are molecular hydrogen, methanol (or methyl derivatives), and acetate. Although the acetate is found in high concentrations in arthropods (Egert et al., 2003; Pester and Brune, 2007; Köhler et al., 2012; Schauer et al., 2012), aceticlastic methanogenesis has never been detected. Our study confirmed that since no genus-level groups were detected potentially carry out this pathway. It is believed that aceticlastic methanogenesis might not occur in termites owing to the low growth rates of the responsible organisms, which would not allow the organisms to cope with the short retention time of the whole digestion process (Brune, 2010). However, methanogens depending on only hydrogen and/or hydrogen together with methanol as substrates have been reported in arthropods and were identified in all major host groups sampled in this study. These substrates are released during the serial breakdown of biomass, which is carried out almost exclusively by the bacterial microbiota. Therefore, we asked whether archaeal genus-level groups show a dependency on certain bacterial lineages.

For this purpose, we used previously published sequencing results (Dietrich et al., 2014) and correlated both fractions using the SparCC algorithm (Fig. 10). We found six archaeal genus-level groups with high correlations with bacterial microbiota. We identified 41 bacterial genus-level groups belonging to ten different phyla that showed a high correlation to at least one archaeal genus-level group. First, we carried out a classical correlation analysis using the Spearman correlation coefficient. This analysis was not successful which underlines that this analysis is not useful or valid for compositional such as relative abundances (Friedman and Alm, 2012). This underlines the need for the SparCC-algorithm-based correlation that can deal with compositional data.

Based on the SparCC-algorithm we found a high number of potential correlations between members of the bacterial and the archaeal microbiota. The genus *Methanobrevibacter* had fewer potential dependencies on bacterial groups than the other archaeal genus-level groups. The largest correlation was negative, which means that these bacterial genus-level groups might only occur when the relative abundance of *Methanobrevibacter* is low. The bacterial genus-level that are correlated with *Methanobrevibacter* originated from the phyla *Bacteroidetes*, *Firmicutes*, and *Proteobacteria*.

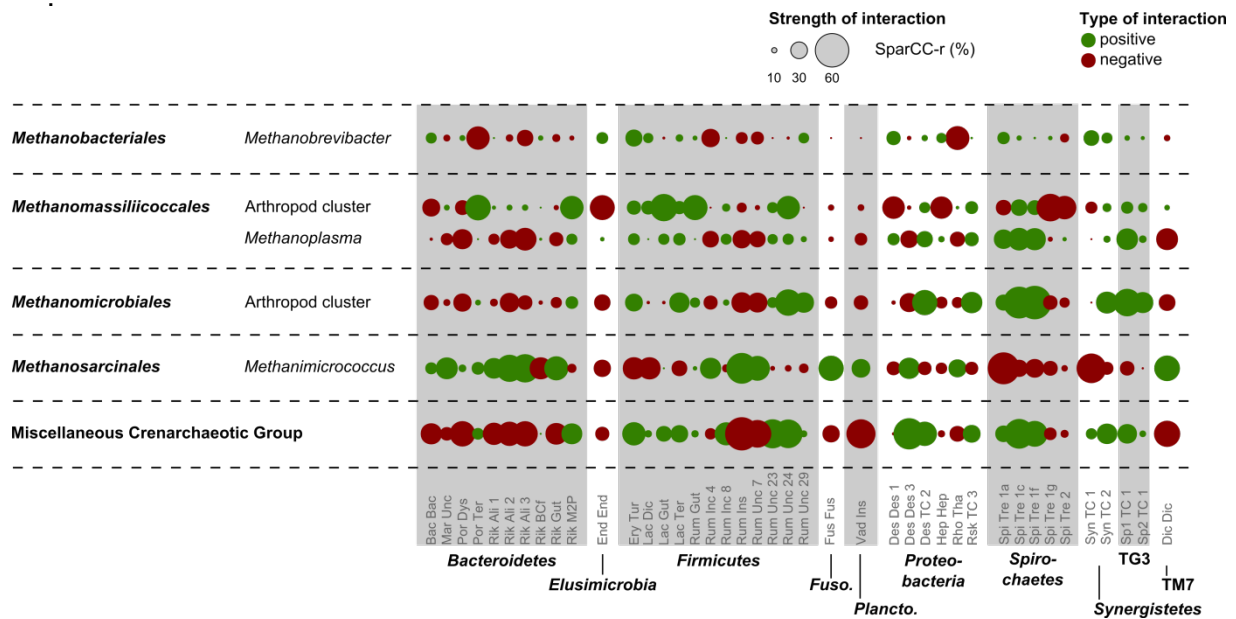
We found that the two genus-level groups of the methanogenic order *Methanomassiliicoccales*, Arthropod cluster and *Ca. Methanoplasma*, showed each as highly abundant, but in different host groups. This difference between both genus-level groups also visible for the bacterial genus-level lineages, which showed a high correlation with those genus-level groups.

Whereas the arthropod cluster mostly correlated with genus-level groups of the phyla *Bacteroidetes* and *Firmicutes*, *Ca. Methanoplasma* had the strongest correlations with *Treponema* clusters 1a, 1c, and 1f and the termite cluster of subphylum 2 of the *Candidate Division* TG3. *Treponema* cluster 1a is the only *Treponema* cluster that contains isolated representatives: *Treponema primitia* and *Treponema azotonutricium* (Graber et al., 2004). Both isolates are reported to be homoacetogenic and to metabolize to a limited extent also oligosaccharides (Graber and Breznak, 2004). However, as neither of these isolates is known to produce methanol or hydrogen, the causal links between these groups is unclear. *Ca. Methanoplasma* reduces methyl groups with external hydrogen, but it is not clear whether organisms of the arthropod cluster have this same metabolism.

In the methanogenic order *Methanomicrobiales*, only the *Methanospirillaceae* arthropod cluster showed noteworthy correlations with the bacterial microbiota. The *Methanospirillaceae* arthropod cluster strongly correlated with *Treponema* clusters 1a, 1c, and 1f and the *Ruminococcaceae* clusters uncultured 24 and 29. Little is known about the metabolic potential of the *Methanospirillaceae* arthropod cluster. The distantly related species *Methanospirillum hungatei* carries out methanogenesis from formate or hydrogen and carbon dioxide as substrates (Ferry, 1974). We wonder whether the *Methanospirillaceae* arthropod cluster might also be formicotropic or hydrogenotrophic.

Interestingly, the genus-level group of *Methanomicrococcus* (*Methanosarcinales*) highly correlated with genus-level groups of the phyla *Bacteroidetes*, *Firmicutes*, and the *Candidate division* TM7. The *Bacteroidetes* genus-level groups *Alistipes* clusters 1, 2, and 3 and the *Candidate division* TM7 are able or have the potential to metabolize oligosaccharides (Mishra, 2012; Albertsen, 2013). However, strains of *Alistipes* are often associated with carnivores, and this implies also proteolytic abilities. In a small genome survey of the genus *Alistipes*, we found genes for pectin esterases in many genomes; therefore, these organisms might be able to break down pectin and release methanol.

Members of the Miscellaneous Crenarchaeotic Group (MCG) showed the highest number of correlations, both negative and positive, within the data set. Positive correlations were found with the putative proteolytic and homoacetogenic genus-level groups from the phyla *Spirochaetes* and *Firmicutes*. Knowledge about the metabolism of this MCG is minimal. A recent metagenomic study implicated this group in the breakdown of aromatic compounds (Meng, 2014). If the members of the MCG are involved in the breakdown of aromatic compounds in higher termites, they would be placed at the top of the food chain in contrast to methanogens that usually occur at the end of the food chain.



**Fig. 10 | Potential dependencies between the archaeal and bacterial members of the gut microbiota of arthropods shown as a correlogram of the SparCC correlation results.** Data were filtered by applying a minimum threshold of  $|r|=0.4$  for each archaeal and bacterial genus-level group. Colors of the points indicate directionality of the possible interaction: red, negative; green, positive. Point size indicates the strength of interaction as measured by SparCC-r. Bacterial genus-level groups are abbreviated according to the first three letters of the family and the genus-level group abbreviation (Tab. S2).

## Discussion

Although the abundance of the archaeal community in arthropods is usually below 3% of the total microbial community (Brauman et al., 2001), the functions carried out by this group are very unique. However, these functions do not seem to be essential for the arthropod host (Messer and Lee, 1989). Since the archaeal community does not seem to be essential, we concluded that the adaptation of the archaeal community is not as strong as the adaptation of the bacterial microbiota. In this study, we used a curated reference database to classify reads derived from high-throughput sequencing to profile the archaeal community of representatives of all known major methane-emitting terrestrial arthropod groups. The results indicated that the archaeal communities of these insects have a high specificity for their hosts through host group membership on all possible levels.

**Host-specific and diet-specific community structure.** The archaeal community structure followed to a large extent the major host groups, although that of millipedes was always similar to that of cockroaches, and those of lower termites had high similarities to those of other groups. The similarity of the archaeal community structure of millipedes to that of cockroaches (Fig. 7 and Fig. S1) might be explained by the similar, litter-feeding diet of the two groups. Interestingly, when low-abundant genus-level groups were emphasized in the analysis (Fig. S1c–d), the archaeal community structure of the humus-feeding scarab beetle larvae was more similar to that of higher termites, which also consisted of many soil-feeding taxa. We concluded that community structure is shaped by host-group membership and to a lesser extent by diet. This also becomes apparent when the two wood-feeding cockroaches were studied. Both wood feeders clustered with other cockroaches but were always (with and without emphasis on rare taxa) next to each other, indicating their similarity.

Lower termites in general did not show a consistent archaeal community structure. This clearly contradicts earlier studies that almost exclusively found only *Methanobrevibacter* sp. (Ohkuma and Kudo, 1998; Shinzato et al., 1999; Tokura et al., 2000; Shinzato et al., 2001). However, several of these studies consist only of a few clone libraries, and the utilized primer sets employed in some of these studies have a mismatch towards at least one archaeal group (Paul et al., 2012). Our results suggest that the lower termite archaeal communities are in general not very similar and consistent. We believe that the main reason for the large heterogeneity of community structure in lower termites is the presence of flagellates. These symbiotic protists can be an important microhabitat for methanogens. However, the number of flagellate cells harboring methanogenic endosymbionts varies substantially. In *Reticulitermes speratus*, only 4% of the flagellate cells contain methanogens, whereas in *Hodotermopsis sjoestedti*, this association is found in 42% of the flagellate cells (Tokura et al., 2000). Furthermore, we showed that the lower termite *Reticulitermes santonensis* has a high abundance of strains belonging to the genus *Ca. Methanoplasma* (*Methanomassiliicoccales*). Interestingly, Shinzato et al. (1999) also found this group to be at least a variable part of the archaeal community in *R. speratus*. It has been already shown that *R. santonensis* has a different archaeal community than *R. speratus*. In *R. speratus*, a small number of flagellates harbor methanogens (Tokura et al., 2000), whereas in *R. santonensis*, no archaeal lineages associated with flagellates have been identified (Leadbetter and Breznak, 1996; Pester and Brune, 2006). Furthermore, our analysis revealed other genera present in high abundance in the lower termites sampled, such as *Methanocorpusculum* and *Methanomicrococcus*. These results challenge the concept of a purely hydrogenotrophic methanogenesis in lower termites, and this finding should be studied at the metabolic level in future studies.

Another characteristic of the archaeal communities in arthropod guts is the presence of at least two genus-level groups that carry out hydrogenotrophic and methyl-reducing methanogenesis. Surprisingly, the latter seems to be carried out by different genus-level groups that differ depending on

the membership of the host to a particular major host group. In scarab beetle larvae, the *Methanoplasmataceae* arthropod cluster seems to carry out this pathway, whereas in millipedes and cockroaches, *Methanomicrococcus* was highly abundant. In higher termites, in contrast, the methyl reduction is most likely performed by the members of the genus-level group *Ca. Methanoplasma*. Hydrogenotrophic methanogenesis seem to be almost exclusively linked to *Methanobrevibacter*, which occurred in all arthropod groups.

**Lineage specificity to major arthropod host groups.** The most predominant archaeal groups were in high abundance only in certain host groups, which indicated a clear specificity of these groups to specific hosts. An exception was *Methanobrevibacter*. Although this genus was very abundant in the total data set, it was present in all major arthropod groups in high abundance. In contrast, the genus *Methanomicrococcus* clearly showed a preference for millipedes and cockroaches, which is also one of the reasons why millipedes and cockroaches formed a cluster in the cluster analysis (Fig. 7). Members of this genus utilize both methanol and hydrogen for methanogenesis (Sprenger et al., 2000). Since these host groups mainly feed on litter that might contain sources of pectin (e.g., leaves), the required methanol most likely stems from pectin esterase activities in both host groups. Interestingly *Methanomicrococcus* also had a high correlation with members of the bacterial genus-level groups *Alistipes* clusters 1, 2, and 3. In public genome databases, genes for pectin methyl esterases are commonly found in these organisms. Both *Alistipes* and *Methanomicrococcus*, might therefore be linked via methanol.

*Ca. Methanoplasma* and the *Methanoplasmataceae* arthropod cluster require the same substrate combination as *Methanomicrococcus*, hydrogen and methanol, which seems to be a typical feature of the whole order (Dridi et al., 2012; Paul et al., 2012; Iino et al., 2013). Interestingly, the arthropod cluster was mainly specific to scarab beetles, whereas *Candidatus Methanoplasma* was more specific to higher termites. Both groups also differed in their correlation pattern, which indicated that these two lineages might have different niches or even different metabolic requirements. *Ca. Methanoplasma* is mostly correlated with different genus-level groups of *Treponema*, including *Treponema* cluster 1a, which contains isolates that are hydrogenotrophic acetogens (Leadbetter et al., 1999). This is interesting since both groups use the same substrate, hydrogen. Notably, *Ca. Methanoplasma* correlates with *Treponema* cluster 1c, which is part of the fiber fraction in *Nasutitermes corniger* (Mikaelyan et al., 2014). However, it has not been shown whether *Treponema* cluster 1c is able to produce hydrogen.

Compared to genus-level groups of methanogenic orders, virtually nothing is known about the Miscellaneous Crenarchaeotic Group. A recent publication revealed that this group is involved in the degradation of aromatic compounds (Meng, 2014). In the hindguts of arthropods, this group only occurs in high abundance in the non-fungus-cultivating higher termites, which indicated a high specificity of this class/genus-level group for higher termites. This group showed the most and strongest correlations with bacterial genus-level groups. A positive correlation was found for some of the *Ruminococcaceae* genus-level groups, which are known for their proteolytic abilities. Soil-feeding termites and their microbiota are reported to most likely feed on polypeptide residues conserved in humus and their subsequent intermediates (Ji et al., 2000). How these energy-rich nitrogenous compounds are broken down is still unknown and would require also the polyphenol lattice of soil organic matter to be degraded. Members of the MCG might act in concert with main organisms or use the released phenol monomers that might be highly solubilized in the alkaline gut compartments of higher termites. However, since the MCG is a very diverse group, it is also questionable whether this group generally shares a common metabolism.

**Host-specific phylogenetic clusters of major archaeal lineages.** Our results showed that the archaeal community structure mostly mirrors major host groups and is dependent to a lesser extent on diet. This is apparent in the strong similarity of the archaeal community of millipedes and cockroaches, and indicated by an emphasis on lower abundant genus-level groups leads to a higher similarity between humus-feeding scarab beetle larvae and humus-feeding higher termites (Fig. S1). These findings lead us to question whether coevolution between archaeal lineages and their hosts could be a feasible scenario, as it has been proposed earlier for bacterial termite lineages (Hongoh et al., 2005). When we constructed phylogenetic trees of the genus-level groups with representative OTUs from the high-throughput sequencing reads, we recognized that many trees clearly contained clusters of OTUs from the same host group. This indicated that archaeal lineages found in arthropod hosts form unique host-specific clusters. A prominent example was the lower termite cluster of *Methanobrevibacter* (Fig. 9 and Fig S2b). With a closer look, no cocoladogenesis with the respective hosts was visible, which indicated that cospeciation does not seem occur but the host-specific groups indicate a diffuse coevolution, at least for *Methanobrevibacter*. However, members of *Methanobrevibacter* have been reported to be endosymbionts of termite gut flagellates (Tokura et al., 2000), which might cause a cospeciation between *Methanobrevibacter* and the flagellate host. Therefore, such a tripartite symbiosis could mask a coevolution signal between the *Methanobrevibacter* lineages and the arthropod host. Yet, also the other lineages do not show cocoladogenesis, and hence cospeciation between the archaeal lineages and their hosts can generally be excluded as a driving force of the phylogeny of arthropod archaeal lineages, or, alternatively, these associations have not yet had enough time to show a cospeciation signal.

**Determinants of arthropod gut archaeal communities.** The peculiarity of host-specific clusters, an indicating of coevolution, without obvious cospeciation is puzzling. On one hand, there are large host-specific clusters in the phylogenetic trees, but on the other hand, cocoladogenesis is missing. This raises the question which mechanisms shape the arthropod archaeal communities. Also selection by host habitat has been previously considered to be a determinant of the community structure (Hongoh et al., 2005). Although Dietrich et al. (2014) showed that major evolutionary events are mirrored by the bacterial community structure of cockroaches and termites, the causes of these patterns remain unknown. We found a large number of correlations between archaeal and bacterial genus-level groups. Some of these correlations pointed out possible metabolic links, such as the correlation between *Methanomicrococcus* and *Alistipes* clusters 1, 2, and 3. Close relatives of this bacterial genus carry genes for pectin methyl esterases in their genome and produce methanol. This indicates that the availability of certain substrates plays a large role. At least for higher termites, it has been shown that the addition of different substrates alters the methane emission rate of gut compartments and whole guts (Schmitt-Wagner and Brune, 1999). This indicates the presence of a strong dependency of the methanogenic community on substrates and shows that these substrates are even limited for methanogens. Furthermore, the different microenvironmental conditions found in the compartments of higher termites (Brune, 2014 and references therein) might select strains that are adopted to and specialized on these conditions. A very important condition would definitely be the alkalinity found in the P1 compartment of higher termites, especially in the soil-feeding *Cubitermes* species (Brune and Kühl, 1996).



## **Acknowledgements**

This study was funded by a grant of the Deutsche Forschungsgemeinschaft (DFG) in the Collaborative Research Center SFB 987 and the LOEWE program of the state of Hessen (Synmikro). Carsten Dietrich received a scholarship from the International Max Planck Research School for Environmental, Cellular and Molecular Microbiology (IMPRS-MIC). James O. Nonoh received a scholarship of the Deutscher Akademischer Austauschdienst (DAAD). We thank Kiyoto Maekawa (University of Toyama), Christine Nalepa (North Carolina State University), Rudy Plarre (Federal Institute for Materials Research and Testing, Berlin), Rudolf Scheffrahn (University of Florida), and Gaku Tokuda (University of the Ryukyus) for insect samples and Karen Brune for correcting the manuscript.

## References

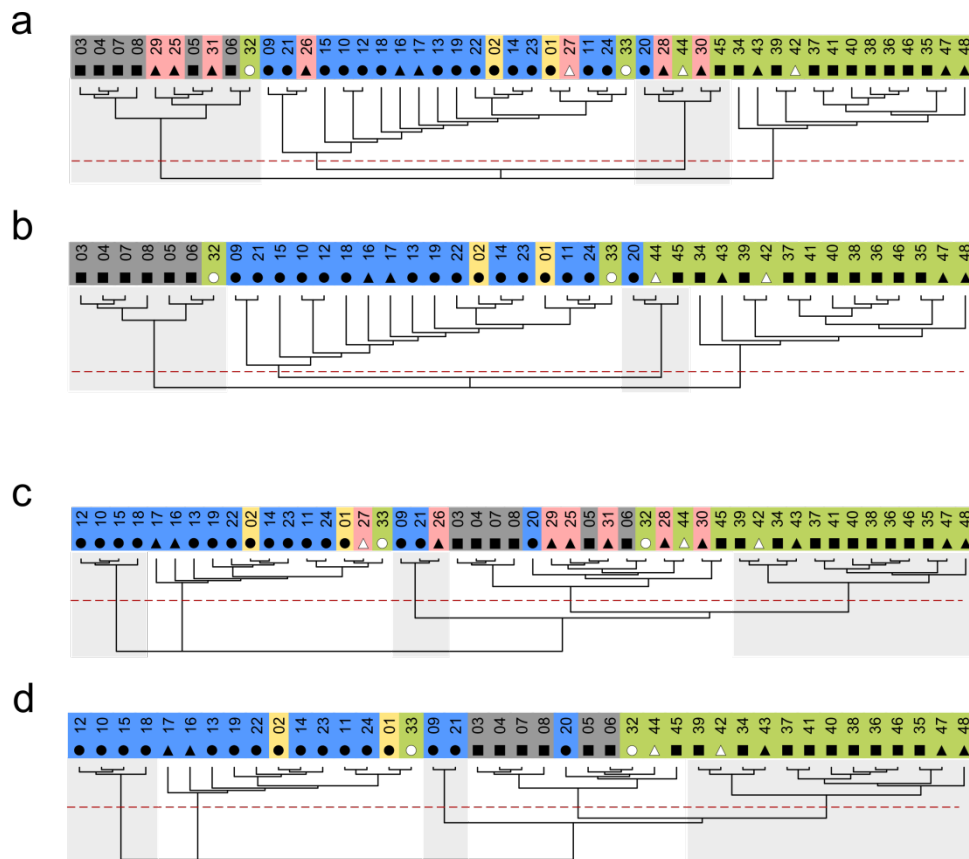
- Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW and Nielsen PH (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat. Biotechnol.* 31:533–538.
- Borrel G, Harris HMB, Tottey W, Mihajlovski A, Parisot N, Peyretailade E, Peyret P, Gribaldo S, O'Toole PW and Brugère JF (2012). Genome sequence of "*Candidatus Methanomethylophilus alvus*" Mx1201, a methanogenic archaeon from the human gut belonging to a seventh order of methanogens. *J. Bacteriol.* 194:6944–6945.
- Brauman A, Dore J, Eggleton P, Bignell D, Breznak JA and Kane MD (2001). Molecular phylogenetic profiling of prokaryotic communities in guts of termites with different feeding habits. *FEMS Microbiol. Ecol.* 35:27–36.
- Brune A and Kühl M (1996). pH profiles of the extremely alkaline hindguts of soil-feeding termites (Isoptera: Termitidae) determined with microelectrodes. *J. Insect Physiol.* 42:1121–1127.
- Brune A (2010). Methanogenesis in the digestive tracts of insects, pp. 707–728. *In* Timmis KN. (ed.), *Handbook of hydrocarbon and lipid microbiology*, vol. 8. Springer, Heidelberg, Germany.
- Brune A (2014). Symbiotic digestion of lignocellulose in termite guts. *Nat. Rev. Microbiol.* 12:168–180.
- Costea PL, Zeller G, Sunagawa S and Bork P (2014). A fair comparison. *Nat. Methods* 11:359.
- Daigle D, Simen BB, and Pochart P (2011). High-throughput sequencing of PCR products tagged with universal primers using 454 Life Sciences Systems. *Curr. Prot. Mol. Bio.* 96:7.5.1–7.5.14.
- Deevong P, Hattori S, Yamada A, Trakulnaleamsai S, Ohkuma M, Noparatnaraporn N and Kudo T (2004). Isolation and detection of methanogens from the gut of higher termites. *Microb. Environ.* 19:221–226.
- Dietrich C, Köhler T and Brune A (2014). The cockroach origin of the termite gut microbiota: patterns in bacterial community structure reflect major evolutionary events. *Appl. Environ. Microbiol.* 80:2261–2269.
- Donovan SE, Purdy KJ, Kane MD and Eggleton P (2004). Comparison of *Euryarchaea* strains in the guts and food-soil of the soil-feeding termite *Cubitermes fungifaber* across different soil types. *Appl. Environ. Microbiol.* 70:3884–3892.
- Dridi B, Fardeau ML, Ollivier B, Raoult D and Drancourt M (2012). *Methanomassiliicoccus luminyensis* gen. nov., sp. nov., a methanogenic archaeon isolated from human faeces. *Int. J. Syst. Evol. Microbiol.* 62:1902–1907.
- Edgar RC (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26:2460–2461.
- Edgar RC (2013). UPARSE: Highly accurate OTU sequences from microbial amplicon reads. *Nat. Meth.* 10:996–998.
- Egert M, Wagner B, Lemke T, Brune A, and Friedrich MW (2003). Microbial community structure in midgut and hindgut of the humus-feeding larva of *Pachnoda ehippiata* (Coleoptera: Scarabaeidae). *Appl. Environ. Microbiol.* 69:6659–6668.
- Ferry JG, Smith PH and Wolfe RS (1974). *Methanospirillum*, a new genus of methanogenic bacteria, and characterization of *Methanospirillum hungatii* sp. nov. *Int. J. Syst. Evol. Microbiol.* 24:465–469.
- Friedman J and Alm EJ (2012). Inferring correlation networks from genomic survey data. *PLoS Comput. Biol.* 8:e1002687.
- Friedrich MW, Schmitt-Wagner D, Lueders T and Brune A (2001). Axial differences in community structure of *Crenarchaeota* and *Euryarchaeota* in the highly compartmentalized gut of the soil-feeding termite *Cubitermes orthognathus*. *Appl. Environ. Microbiol.* 67:4880–4890.
- Gijzen HJ, Broers CA, Barugahare M and Stumm CK (1991). Methanogenic bacteria as endosymbionts of the ciliate *Nyctotherus ovalis* in the cockroach hindgut. *Appl. Environ. Microbiol.* 57:1630–1634.
- Graber JR and Breznak JA (2004). Physiology and nutrition of *Treponema primitia*, an H<sub>2</sub>/CO<sub>2</sub>-acetogenic spirochete from termite hindguts. *Appl. Environ. Microbiol.* 70:1307–1314.
- Graber JR, Leadbetter JR and Breznak JA (2004). Description of *Treponema azotonutricium* sp. nov. and *Treponema primitia* sp. nov., the first spirochetes isolated from termite guts. *Appl. Environ. Microbiol.* 70:1315–1320.
- Grosskopf R, Janssen PH and Liesack W (1998). Diversity and structure of the methanogenic community in anoxic rice paddy soil microcosms as examined by cultivation and direct 16S rRNA gene sequence retrieval. *Appl. Environ. Microbiol.* 64:960–969.

- Hackstein JHP and Stumm CK (1994).** Methane production in terrestrial arthropods. *Proc. Natl. Acad. Sci. USA* 91:5441–5445.
- Hackstein JHP, van Alen TA and Rosenberg J (2006).** Methane production by terrestrial arthropods, pp. 155–180. *In* König H, Varma A. (ed.), *Intestinal microorganisms of termites and other invertebrates*, Springer, Berlin, Germany.
- Hara K, Shinzato N, Seo M, Oshima T, Yamagishi A (2002).** Phylogenetic analysis of symbiotic archaea living in the gut of xylophagous cockroaches. *Microb. Environ.* 17:185–190.
- Hatamoto M, Imachi H, Ohashi A, Hara H (2007).** Identification and cultivation of anaerobic, syntrophic long-chain fatty acid-degrading microbes from mesophilic and thermophilic methanogenic sludges. *Appl. Environ. Microbiol.* 73:1332–1340.
- Hongoh Y, Deevong P, Inoue T, Moriya S, Trakulnaleamsai S, Ohkuma M, Vongkaluang C, Noparatnaraporn N and Kudo T (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.
- Iino T, Tamaki H, Tamazawa S, Ueno Y, Ohkuma M, Suzuki K, Igarashi Y and Haruta S (2013).** *Candidatus* Methanogranum caenicola: a novel methanogen from the anaerobic digested sludge, and proposal of *Methanomassiliicoccaceae* fam. nov. and *Methanomassiliicoccales* ord. nov., for a methanogenic lineage of the class *Thermoplasmata*. *Microbes Environ.* 28:244–250.
- Imachi H, Sekiguchi Y, Kamagata Y, Loy A, Qiu YL, Hugenholtz P, Kimura N, Wagner M, Ohashi A and Harada H (2006).** Non-sulfate-reducing, syntrophic bacteria affiliated with *Desulfotomaculum* cluster I are widely distributed in methanogenic environments. *Appl. Environ. Microbiol.* 72:2080–2091.
- Ji R, Kappler A and Brune A (2000).** Transformation and mineralization of synthetic <sup>14</sup>C-labeled humic model compounds by soil-feeding termites. *Soil Biol. Biochem.* 32:1281–1291.
- Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M and Glöckner FO (2013).** Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res* 41:e1.
- Köhler T, Dietrich C, Scheffrahn RH and Brune A (2012).** High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation of the gut in wood-feeding higher termites (*Nasutitermes* spp.). *Appl. Environ. Microbiol.* 78:4691–4701.
- Leadbetter JR and Breznak JA (1996).** Physiological ecology of *Methanobrevibacter cuticularis* sp. nov. and *Methanobrevibacter curvatus* sp. nov., isolated from the hindgut of the termite *Reticulitermes flavipes*. *Appl. Environ. Microbiol.* 62:3620–3631.
- Leadbetter JR, Crosby LD and Breznak JA (1998).** *Methanobrevibacter filiformis* sp. nov., a filamentous methanogen from termite hindguts. *Arch. Microbiol.* 169:287–292.
- Leadbetter JR, Schmidt TM, Graber JR and Breznak JA (1999).** Acetogenesis from H<sub>2</sub> plus CO<sub>2</sub> by spirochetes from termite guts. *Science* 283:686–689.
- Ludwig W, Strunk O, Westram R, Richter L, Meier H et al. (2004).** ARB: a software environment for sequence data. *Nucleic Acids Res.* 32:1363–1371.
- Lueders T and Friedrich M (2000).** Archaeal population dynamics during sequential reduction processes in rice field soil. *Appl. Environ. Microbiol.* 66:2732–2742.
- Meng J, Xu J, Qin D, He J, Xiao X and Wang F (2014).** Genetic and functional properties of uncultivated MCG archaea assessed by metagenome and gene expression analyses. *ISME J.* 8:650–659.
- Messer AC and Lee MJ (1989).** Effect of chemical treatments on methane emission by the hindgut microbiota in the termite *Zootermopsis angusticollis*. *Microb. Ecol.* 18:275–284.
- Mikaelyan A, Strassert JFH, Tokuda G and Brune A (2014).** The fiber-associated cellulolytic bacterial community in the hindgut of wood-feeding higher termites (*Nasutitermes* spp.). *Environ. Microbiol.* 16:2711–2722.
- Mishra AK, Gimenez G, Lagier JC, Robert C, Didier R and Fournier PE (2012).** Genome sequence and description of *Alistipes senegalensis* sp. nov. *Stand. Gen. Sci.* 6:1–16.

- Miyata R, Noda N, Tamaki H, Kinjyo K, Aoyagi H, Uchiyama H and Tanaka H (2007). Influence of feed components on symbiotic bacterial community structure in the gut of the wood-feeding higher termite *Nasutitermes takasagoensis*. *Biosci. Biotechnol. Biochem.* 71:1244–1251.
- Odelson DA and Breznak JA (1985). Nutrition and growth characteristics of *Trichomitopsis termopsidis*, a cellulolytic protozoan from termites. *Appl. Environ. Microbiol.* 49:614–621.
- Ohkuma M and Kudo T (1998). Phylogenetic analysis of the symbiotic intestinal microflora of the termite *Cryptotermes domesticus*. *FEMS Microbiol. Lett.* 164:389–395.
- Ohkuma M, Noda S, and Kudo T (1999). Phylogenetic relationships of symbiotic methanogens in diverse termites. *FEMS Microbiol. Lett.* 171:14–153.
- Oksanen J, Blanchet JG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Simpson GL, Solymos P, Stevens MHH and Wagner H (2013). *vegan*: Community Ecology Package. <http://CRAN.R-project.org/package=vegan>.
- Paradis E, Claude J and Strimmer K (2004). APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* 20:289–290.
- Parks DH and Beiko RG (2013). Measures of phylogenetic differentiation provide robust and complementary insights into microbial communities. *ISME J.* 7:173–183.
- Paul K, Nonoh JO, Mikulski L and Brune A (2012). "*Methanoplasmatales*," *Thermoplasmatales*-related archaea in termite guts and other environments, are the seventh order of methanogens. *Appl. Environ. Microbiol.* 78:8245–8253.
- Pester M and Brune A (2006). Expression profiles of *fhs* (FTHFS) genes support the hypothesis that spirochaetes dominate reductive acetogenesis in the hindgut of lower termites. *Environ. Microbiol.* 8:1261–1270.
- Pester M and Brune A (2007). Hydrogen is the central free intermediate during lignocellulose degradation by termite gut symbionts. *ISME J.* 1:551–565.
- Pruesse E, Quast C, Knittel K, Fuchs B, Ludwig W, Peplies J and Glöckner FO (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res.* 35:7188–7196.
- R Development Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>.
- Rajaram S and Oono Y (2010). NeatMap – non-clustering heat map alternatives in R. *BMC Bioinform.* 11:1–9.
- Rawls JF, Mahowald MA, Ley RE and Gordon JI (2006). Reciprocal gut microbiota transplants from zebrafish and mice to germ-free recipients reveal host habitat selection. *Cell* 127:423–433.
- Roche (2009). Using multiplex identifier (MID) adaptors for the GS FLX titanium chemistry-extended MID set, Technical Bulletin: Genome Sequencer FLX System TCB no 005-2009, Roche, Branchburg, NJ.
- Schauer C, Thompson CL and Brune A (2012). The bacterial community in the gut of the cockroach *Shelfordella lateralis* reflects the close evolutionary relatedness of cockroaches and termites. *Appl. Environ. Microbiol.* 78:2758–2767.
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski, RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn, DJ and Weber CF (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75:7537–7541.
- Schmitt-Wagner D and Brune A (1999). Hydrogen profiles and localization of methanogenic activities in the highly compartmentalized hindgut of soil-feeding higher termites (*Cubitermes* spp.). *Appl. Environ. Microbiol.* 65:4490–4496.
- Shinzato N, Matsumoto T, Yamaoka I, Oshima T and Yamagishi A (1999). Phylogenetic diversity of symbiotic methanogens living in the hindgut of the lower termite *Reticulitermes speratus* analyzed by PCR and in situ hybridization. *Appl. Environ. Microbiol.* 65:837–840.
- Shinzato N, Matsumoto T, Yamaoka I, Oshima T, Yamagishi A (2001). Methanogenic symbionts and the locality of their host lower termites. *Microb. Environ.* 16:43–47.

- Sprenger WW, van Belzen MC, Rosenberg J, Hackstein JHP and Keltjens JT (2000).** *Methanomicrococcus blatticola* gen. nov., sp. nov., a methanol- and methylamine-reducing methanogen from the hindgut of the cockroach *Periplaneta americana*. *Int. J. Syst. Evol. Microbiol.* 50:1989–1999.
- Stamatakis A (2014).** RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
- Sustr V, Chronáková A, Semanová S, Tajovský K and Simek M (2014).** Methane production and methanogenic archaea in the digestive tracts of millipedes (Diplopoda). *PLoS ONE* 9:e102659.
- Tokura M, Ohkuma M and 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.
- Tsai YH and Cahill KM (1970).** Parasites of the german cockroach (*Blattella germanica* L.) in New York City. *J. Parasitol.* 56:375–377.
- Wickham H (2009).** ggplot2: elegant graphics for data analysis. Springer, New York, United States of America.
- Zellner G, Alten C, Stackebrandt E, Conway de Macario E and Winter J (1987).** Isolation and characterization of *Methanocorpusculum parvum*, gen. nov., spec. nov., a new tungsten requiring, coccoid methanogen. *Arch. Microbiol.* 147:13–20.

## Supplementary Material



**Fig. S1 | Host- and diet-specific clustering of the archaeal community of arthropods.** Unrooted BioNJ trees of the community structure displayed as cladograms based on the pairwise Soergel distances of (a) non-transformed data, (b) non-transformed dataset without lower termite samples, (c) logarithmized data set to emphasize lower abundant taxa and (d) logarithmized data set and lower termites removed. The dashed line aids in distinguishing the clusters in the data set. Color coding is the same as in Fig. 1. Numbering is the same as in Tab. 2. The symbols below the numbering indicate dietary specialization of most of the arthropod hosts and are the same as in Fig. 6.

**Fig. S2 | Phylogenetic trees of 1%-level OTUs from high-throughput sequencing of arthropod archaeal lineages.** Use the pdf bookmarks to navigate. A description of the archaeal genus-level groups is found also in the bookmarks. The numerical code of the different samples is as in Tab. 2. Please download from: [https://www.dropbox.com/s/0tnyx07wwtuqfs0/SIFig2\\_SIFig.pdf?dl=0](https://www.dropbox.com/s/0tnyx07wwtuqfs0/SIFig2_SIFig.pdf?dl=0).

Tab. S1 | Classification success of the high-throughput sequencing data.

Sample	No.	Classification success (%)				
		Phylum	Class	Order	Family	Genus
<b>Millipedes</b>						
<b><i>Spirostreptidae</i></b>						
<i>Anadenobolus monilicornis</i>	1	100.0	100.0	100.0	100.0	100.0
<b><i>Harpagophoridae</i></b>						
<i>Harpagophorida</i> sp.	2	100.0	100.0	100.0	100.0	100.0
<b>Scarab beetle larvae</b>						
<b><i>Cetoniidae</i></b>						
<i>Dicronorhina derbyana</i>	3	100.0	100.0	100.0	100.0	100.0
<i>Genyodonta lequexi</i>	4	100.0	100.0	100.0	100.0	100.0
<b><i>Scarabaeidae</i></b>						
<i>Pachnoda aemula</i>	5	100.0	100.0	100.0	100.0	100.0
<i>Pachnoda ehippiata falkei</i>	6	100.0	100.0	100.0	100.0	99.3
<i>Gnorimus tibialis</i>	7	100.0	100.0	100.0	100.0	100.0
<i>Xylotrupes gideon</i>	8	100.0	100.0	100.0	100.0	100.0
<b>Cockroaches</b>						
<b><i>Polyphagidae</i></b>						
<i>Ergaula capucina</i>	9	100.0	100.0	100.0	100.0	92.2
<b><i>Blaberidae</i></b>						
<i>Elliptorhina chopardi</i>	10	100.0	100.0	100.0	100.0	100.0
<i>Panchlora</i> sp.	11	100.0	100.0	100.0	100.0	100.0
<i>Opisthoptalia orientalis</i>	12	100.0	100.0	100.0	100.0	100.0
<i>Nauphoeta cinera</i>	13	100.0	100.0	100.0	100.0	100.0
<i>Gromphadorhina portentosa</i>	14	100.0	100.0	100.0	100.0	100.0
<i>Diploptera punctata</i>	15	100.0	100.0	100.0	100.0	100.0
<i>Panesthia angustipennis</i>	16	100.0	100.0	100.0	100.0	99.9
<i>Salganea esakii</i>	17	100.0	100.0	100.0	100.0	99.9
<i>Eublaberus posticus</i>	18	100.0	100.0	100.0	100.0	100.0
<i>Schultesia lampyridiformis</i>	19	100.0	100.0	100.0	100.0	100.0
<i>Henschoutedenia flexivitta</i>	20	100.0	100.0	100.0	100.0	100.0
<b><i>Blattidae</i></b>						
<i>Blatta orientalis</i>	21	100.0	100.0	100.0	100.0	100.0
<i>Eurycotis floridana</i>	22	100.0	100.0	100.0	100.0	100.0
<i>Shelfordella lateralis</i>	23	100.0	100.0	100.0	100.0	100.0
<b><i>Cryptocercidae</i></b>						
<i>Cryptocercus punctulatus</i>	24	100.0	100.0	100.0	100.0	99.9

Sample	No.	Classification success (%)				
		Phylum	Class	Order	Family	Genus
<b>Lower termites</b>						
<b><i>Mastotermitidae</i></b>						
<i>Mastotermes darwiniensis</i>	25	99.9	99.9	99.9	99.9	95.4
<b><i>Hodotermitidae</i></b>						
<i>Hodotermopsis sjoestedti</i>	26	100.0	100.0	100.0	100.0	93.6
<i>Hodotermes mossambicus</i>	27	100.0	100.0	100.0	100.0	100.0
<b><i>Termopsidae</i></b>						
<i>Zootermopsis nevadensis</i>	28	100.0	100.0	100.0	100.0	97.0
<b><i>Kalotermitidae</i></b>						
<i>Neotermes jouteli</i>	29	100.0	100.0	100.0	100.0	93.6
<b><i>Rhinotermitidae</i></b>						
<i>Reticulitermes santonensis</i>	30	100.0	100.0	100.0	100.0	100.0
<i>Coptotermes niger</i>	31	100.0	100.0	100.0	100.0	100.0
<b>Higher termites (<i>Termitidae</i>)</b>						
<b><i>Macrotermitinae</i></b>						
<i>Odontotermes</i> sp.	32	100.0	100.0	100.0	100.0	100.0
<i>Macrotermes</i> sp.	33	100.0	100.0	100.0	100.0	99.9
<b><i>Apicotermitinae</i></b>						
<i>Alyscotermes trestus</i>	34	100.0	100.0	100.0	100.0	100.0
<b><i>Syntermitinae</i></b>						
<i>Cornitermes</i> sp.	35	100.0	100.0	100.0	100.0	92.6
<b><i>Termitinae</i></b>						
<i>Cubitermes</i> sp.	36	100.0	100.0	100.0	100.0	100.0
<i>Cubitermes</i> sp.	37	100.0	100.0	100.0	100.0	100.0
<i>Cubitermes</i> sp.	38	100.0	100.0	100.0	100.0	100.0
<i>Ophiotermes</i> sp.	39	100.0	100.0	100.0	100.0	100.0
<i>Neocapritermes taracua</i>	40	100.0	100.0	100.0	100.0	99.4
<i>Proboscitermes</i> sp.	41	100.0	100.0	100.0	100.0	100.0
<i>Amitermes meridionalis</i>	42	100.0	100.0	100.0	100.0	100.0
<i>Microcerotermes parvus</i>	43	100.0	100.0	100.0	100.0	100.0
<b><i>Nasutitermitinae</i></b>						
<i>Trinervitermes</i> sp.	44	100.0	100.0	100.0	100.0	100.0
<i>Atlantitermes</i> sp.	45	100.0	100.0	100.0	100.0	100.0
Unidentified <i>Nasutitermitinae</i>	46	100.0	100.0	100.0	100.0	100.0
<i>Nasutitermes corniger</i>	47	100.0	100.0	100.0	100.0	94.2
<i>Nasutitermes takasagoensis</i>	48	100.0	100.0	100.0	100.0	92.4



**Tab. S2 | Bacterial genus-level groups showing the highest correlation with the archaeal arthropod gut genus-level groups.**

<b>Abbreviation</b>	<b>Family</b>	<b>Genus</b>
<b><i>Bacteroidetes</i></b>		
Bac Bac	<i>Bacteroidaceae</i>	<i>Bacteroides</i>
Mar Unc	<i>Marinilabiaceae</i>	Uncultured 1
Por Dys	<i>Porphyromonadaceae</i>	<i>Dysgonomonas</i>
Por Ter	<i>Porphyromonadaceae</i>	Termite cockroach cluster 3
Rik Ali 1	<i>Rikenellaceae</i>	<i>Alistipes</i> 1
Rik Ali 2	<i>Rikenellaceae</i>	<i>Alistipes</i> 2
Rik Ali 3	<i>Rikenellaceae</i>	<i>Alistipes</i> 3
Rik BCf	<i>Rikenellaceae</i>	BCf9-17 termite group
Rik Gut	<i>Rikenellaceae</i>	Gut Cluster C
Rik M2P	<i>Rikenellaceae</i>	M2PB4-61 termite group
<b><i>Elusimicrobia</i></b>		
End End	<i>Endomicrobiaceae</i>	<i>Endomicrobium</i>
<b><i>Firmicutes</i></b>		
Ery Tur	<i>Erysipelotrichaceae</i>	<i>Turicibacter</i>
Lac Dic	<i>Lachnospiraceae</i>	Dictyoptera cluster
Lac Gut	<i>Lachnospiraceae</i>	Gut cluster 1
Lac Ter	<i>Lachnospiraceae</i>	Termite cluster
Rum Gut	<i>Ruminococcaceae</i>	Gut cluster 2
Rum Inc 4	<i>Ruminococcaceae</i>	Incertae Sedis 4
Rum Inc 8	<i>Ruminococcaceae</i>	Incertae Sedis 8
Rum Ins	<i>Ruminococcaceae</i>	Insect guts cluster
Rum Unc 7	<i>Ruminococcaceae</i>	Uncultured 7
Rum Unc 23	<i>Ruminococcaceae</i>	Uncultured 23
Rum Unc 24	<i>Ruminococcaceae</i>	Uncultured 24
Rum Unc 29	<i>Ruminococcaceae</i>	Uncultured 29
<b><i>Fusobacteria</i></b>		
Fus Fus	<i>Fusobacteriaceae</i>	<i>Fusobacterium</i> 1
<b><i>Planctomycetes</i></b>		
Vad Ins	vadinHA49	Insect gut cluster
<b><i>Proteobacteria</i></b>		
Des Des 1	<i>Desulfovibrionaceae</i>	<i>Desulfovibrio</i> 1
Des Des 3	<i>Desulfovibrionaceae</i>	<i>Desulfovibrio</i> 3
Des TC 2	<i>Desulfovibrionaceae</i>	Termite cluster 2
Hep Hep	<i>Candidatus Hepatincola</i>	<i>Candidatus Hepatincola</i>
Rho Tha	<i>Rhodospirillales</i>	<i>Thalassospira</i>
Rsk TC 3	<i>Rs-K70 termite group</i>	Termite cluster 3
<b><i>Spirochaetes</i></b>		
Spi Tre 1a	<i>Treponema</i> 1	<i>Treponema</i> 1a
Spi Tre 1c	<i>Treponema</i> 1	<i>Treponema</i> 1c
Spi Tre 1f	<i>Treponema</i> 1	<i>Treponema</i> 1f
Spi Tre 1g	<i>Treponema</i> 1	<i>Treponema</i> 1g
Spi Tre 2	<i>Treponema</i> 2	<i>Treponema</i> 2

<b>Abbreviation</b>	<b>Family</b>	<b>Genus</b>
<b>Synergistetes</b>		
Syn TC1	<i>Synergistaceae</i>	Termite cluster 1
Syn TC2	<i>Synergistaceae</i>	Termite cluster 3
<b>TM7</b>		
Dic Dic	Dictyoptera cluster	Dictyoptera cluster
<b>TG3</b>		
Sp1 TC 1	SP1 – Termite cockroach cluster	Termite Cluster 1
Sp2 TC 1	SP2 – Termite cockroach cluster	Termite Cluster 1

# Comparative metagenomic analyses reveal bacterial functional profiles in the hindgut compartments of higher termites

Karen Rossmassler, Carsten Dietrich,  
Claire L. Thompson, Aram Mikaelyan and Andreas Brune

*Affiliations:* Max Planck Institute for Terrestrial Microbiology, 35043 Marburg, Germany- | *This manuscript is in preparation.*

*Contributions:* K.R. carried out experiments, analyzed most of the data, discussed the results, and wrote the manuscript; C.D. analyzed data and discussed the results; C.L.T. conceived the study and carried out experiments; A.M. conceived the study and carried out experiments; A.B. conceived the study, discussed results, and secured funding.

## Abstract

---

The microbial community in the highly compartmentalized gut of higher termites is integral to their nutrition. Previous metagenomic studies have explored the bacterial communities of the P3 gut compartment of wood, dung, or fungus feeding termites, but knowledge about the functional potential of soil-feeding termites, as well as the microbial communities in other compartments, is lacking. In this study we used metagenomics to compare the bacterial community structure and functional potential of six termite species representing four feeding guilds (wood, grass, wood/soil interface, and soil). The P1 and P4 gut compartments showed distinct microbial communities that nonetheless showed some patterns common to all homologous compartments. In the P1 from wood- and grass-feeding termites, abundant hemicellulases were classified as *Firmicutes*, and in interface- and soil-feeding termites peptidases such as subtilisin were classified as *Firmicutes*. In the P4, *Bacteroidetes* became the dominant community member, and the peptidases classified to *Bacteroidetes* increased accordingly. However, in the P3 gut compartment, functional profiles based on glycosyl hydrolase and peptidase abundances and phylum-level classification indicated stronger differences between feeding guilds than in the P1 and P4, especially communities from wood- and soil-feeding termites. Bacterial communities from soil-feeding termites had few genes for degrading cellulose, hemicellulose, or starch, but had a distinct suite of peptidases. These results shed light on the complexity of the structure and functional potential of the gut microbial communities found in higher termites and highlights specific communities and gut compartments for further research.

## Introduction

Termites are key ecosystem members in tropical and subtropical environments, in which they play a significant role in plant material turnover and nutrient cycling (Bignell and Eggleton, 2000). Their symbiotic relationship with their gut microbiota is vital to the termites' digestion and therefore their role in ecosystem function (Brune, 2014). However, the complex interactions between members of the gut microbiota as well as with the host termite are only just beginning to be understood. The exclusively prokaryotic microbiota of the highly compartmentalized hindgut compartments of higher termites allows different species to feed on a wide variety of organic material, including wood, grass, dung, fungus, and soil (Brune, 2014). However, little is known about the potential functional roles that microbial community members play in host nutrition, and how these roles differ between gut compartments.

To examine the functional potential of the gut microbiota from termites from different feeding guilds, previous studies have used metagenomics to investigate specific compartments. In 2007, Warnecke et al. characterized the gut community of the P3 lumen of the wood-feeding higher termite *Nasutitermes corniger* to determine the microbial role in lignocellulose degradation (Warnecke et al., 2007). They found an abundant, diverse set of genes encoding cellulase and hemicellulase, many of which could be attributed to Spirochaetes or Fibrobacteres, two major members of the microbial community (Warnecke et al., 2007). A following study compared these results to the microbial community of the P3 lumen from the dung-feeding termite *Amitermes wheeleri*, and determined that the gut microbiota of *A. wheeleri* housed and transcribed more hemicellulase genes, while that of *N. corniger* was enriched in cellulase genes (He et al., 2013). A study focused on the fungus-cultivating termite *Odontotermes yunnanensis* found the whole gut had few cellulase and hemicellulase genes, but contained abundant genes encoding debranching and oligosaccharide-degrading enzymes (Liu et al., 2013). Many of these enzymes were assigned to a major community member, Bacteroidetes, although a portion were unassigned (Liu et al., 2013). Another metagenomic study on the whole gut of fungus-cultivating termite *Macrotermes natalensis* and its fungus symbiont, *Termitomyces*, found that the fungus possessed genes for digesting complex carbohydrates, namely hemicellulose and starch, but the genes for digestion of simpler oligosaccharides were bacterial (Poulsen et al., 2014).

It is important to note that previous metagenomic studies have focused on only a portion of the microbial community in a gut compartment (i.e. the lumen from the P3, which excludes wall-associated microbes), or on the whole gut, which prevents the analysis of individual compartments. The individual hindgut compartments of higher termites show striking differences in both physicochemical parameters and microbial community structure. *Nasutitermes corniger*, perhaps the best-characterized host in terms of gut microenvironment and microbial community structure, has strong gradients in hydrogen and oxygen partial pressure, pH, and redox potential both axially and radially in the intestinal tract (Koehler et al., 2012). Accordingly, the microbial diversity and community compositions differ between gut compartments (Koehler et al., 2012). Following this pattern, previous studies on soil-feeding *Cubitermes* species also found strong gradients in pH and hydrogen partial pressure along the intestinal tract, and compartment-specific microbial community structure (Schmitt-Wagner et al., 2003). Conversely, a study focusing on the P1 microbial community from wood-, soil-, and grass-feeding termites found that it was distinct from the whole gut, and there were common bacterial groups between termite hosts (Thongaram et al., 2005). These differences in physical parameters and microbial community structure, but similarities between homologous compartments of termites from different feeding guilds leads to our hypothesis that there are distinct

functional profiles that are specific to the feeding lifestyle of the termite host and to the gut compartment.

Previous studies have highlighted the importance of the microbially-mediated digestion of organic matter in the nutrition of wood-feeding higher termites (Brune and Ohkuma, 2011, Hongoh, 2011, Ni and Tokuda, 2013). Although previous metagenomic studies have classified genes encoding enzymes for degrading cellulose, hemicellulose, and simple oligosaccharides to bacterial groups, quantitative information about the absolute abundance of such genes is lacking (Warnecke et al., 2007, He et al., 2013). Furthermore, little is known about the carbohydrate degradation potential of wood-soil interface- or soil-feeding higher termites. Previous studies on the soil-feeding *Cubitermes* highlighted a distinct microbial community that lacked key phyla found in wood-feeding termites, especially Spirochaetes and Fibrobacteres (Schmitt-Wagner et al., 2003). We hypothesized that, compared to wood-feeding termites, the gut compartments of soil-feeding and interface-feeding termites will be depleted in cellulases and hemicellulases.

One interesting aspect of nutrition in wood-feeding termites is the acquisition of nitrogen. The lignocellulose in wood is very poor in fixed nitrogen, so the host termite relies on the gut microbiota to recycle and/or fix nitrogen (Hongoh, 2011, Brune, 2014). Although peptidase activity has been detected in wood-feeding higher termites, this activity was highest in the midgut due to endogenous host peptidases (Fujita and Abe, 2002). In soil-feeding termites, peptides are consumed in the hindgut, which features high proteases activity (Ji and Brune, 2005). This led to the hypothesis that soil peptides are fermented by bacteria in the hindgut to ammonia, thus forming as a nutritional source of nitrogen for the host termite (Ji and Brune, 2006, Ngugi et al., 2011, Griffiths et al., 2013). We hypothesized that the microbial communities in the hindguts of soil-feeding termites are enriched in secreted peptidases that hydrolyze peptides to contribute fixed nitrogen to the host nutrition, but that wood- and grass-feeding termites will lack substantial amounts of secreted peptidases. Instead, we hypothesize that even between homologous compartments the functional profile of wood-feeding termites will strongly differ from soil-feeding termites, being dominated by cellulases and hemicellulases to degrade complex plant material.

In this study, we used comparative metagenomics to investigate the gut microbiota of hindgut compartments (P1, P3, and P4) of six higher termites: two wood feeders, *Nasutitermes corniger* and *Microcerotermes parvus*; a putative grass feeder, *Cornitermes* sp.; two wood/soil interface feeders, *Termes hospes* and *Neocapritermes taracua*; and a soil feeder, *Cubitermes ugandensis*. We compared absolute abundances of genes encoding carbohydrate-degrading enzymes and peptidases between feeding guilds, termite host, and homologous compartments, and determined which bacterial groups were responsible for these aspects of host nutrition.

## Materials and Methods

**Termite Collection and DNA Extraction.** *Cubitermes ugandensis* were collected from Lhiranda Hill, Kakamega, Kenya. *Nasutitermes corniger* were obtained from a laboratory-maintained colony (University of Florida). *Cornitermes* (sp.), *Microcerotermes parvus*, and *Neocapritermes taracua* were collected from Guyana. *Termes hospes* were collected from Democratic Republic of Congo.

Guts from 30-50 worker termites were dissected and separated into segments (P1, P3, and P4) as described in (Koehler et al., 2012). Gut sections were pooled in tubes containing 100  $\mu$ L phosphate buffered saline (pH 7.2) and stored at -20°C until extraction. The entire genomic content of each sample from *Cornitermes* sp., *C. ugandensis*, *M. parvus*, *N. corniger*, *N. taracua*, and *T. hospes* was extracted using the NucleoSpin Soil kit with SL2 lysis buffer and SX buffer (MN). DNA was quantified with Quant-IT dsDNA Assay on a Qubit fluorometer (Life Technologies) and purity was checked by a Nanodrop 1000 spectrophotometer (Pepqlab).

**Absolute bacterial abundance.** Copy numbers of bacterial 16S rRNA genes were determined by quantitative PCR as described in Paul et al. (Paul et al., 2012) using the primers 519fc (5'-CAGCMGCCGCGGTAA-3') and 907r (CCGTCAATTCMTTTRAGTT-3').

**16S rRNA gene sequencing.** Two amplifications strategies targeting different regions of the 16S rRNA gene were sequenced at different facilities for metagenome-independent bacterial community structure analysis. For iTAG analysis, 291 bp fragments of the V4 region of the 16S rRNA gene were amplified with the forward primer 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and the reverse primer 806R (5'-GGACTACHVGGGTWTCTAAT-3') as described in Caporaso et al. 2011 (Caporaso et al., 2011). The fragments were sequenced on an Illumina MiSeq at the Joint Genome Institute (Walnut Creek, California, USA), yielding 2x250 bp paired-end reads. For the gTAG analysis, 440 bp fragments of the V3-V4 of the 16S rRNA gene were amplified with the forward primer set 343Fmod (5'-TACGGGWGGCWGCA-3') and 784Rmod (5'-GGGTMTCTAATCCBKTT-3') as described in Koehler et al. 2012 (Koehler et al., 2012) with the following modifications: 28 cycles of amplification, extension time during cycling of 50 s, and a final extension of 5 min. gTAG amplicons were sequenced on an Illumina MiSeq at GATC Biotech (Konstanz, Germany). All samples were included in both the iTAG and gTAG analyses with the exception of the P4 from *N. corniger*, which is absent from the gTAG analysis due to insufficient DNA.

All iTAG and gTAG sequences were processed in the same way, using the mothur software suite (Schloss et al., 2009). Unpaired reads were discarded, and primer sequences were removed. Paired reads were trimmed to a minimum mean quality of 30 with a 20 bp sliding window. Reads were filtered for quality, removing any reads less than 200 bp and any reads with more than five Ns and ten nucleotides with a quality less than 15. After filtering, approximately 60,000-138,000 sequences per sample from iTAG and approximately 20,000-346,000 sequences per sample from gTAG were analyzed. The sequences were classified to the genus level at a 60% identity cutoff using a reference database, DictDB, based on the SILVA database (Quast et al., 2013) with additional sequences and curation for Dictyopteran gut microbiota (Mikaelyan et al., 2015).

**Metagenome sequencing.** Preparation of the metagenomic libraries, sequencing, quality control, and assembly was performed at the Joint Genome Institute (Walnut Creek, California, USA). 100 ng of DNA was sheared to 270 bp fragments using the Covaris E210 (Covaris) and size selected using SPRI beads (Beckman Coulter). The fragments were treated with end-repair, A-tailing, and ligation of Illumina-compatible adapters (IDT, Inc) using the KAPA-Illumina library creation kit (KAPA biosystems).

Libraries were quantified using KAPA Biosystem's next-generation sequencing library qPCR kit and run on a Roche LightCycler 480 real-time PCR instrument. The quantified libraries were then prepared for sequencing on the Illumina HiSeq sequencing platform with a TruSeq paired-end cluster kit, v3-cBot-HS, and Illumina's cBot instrument to generate clustered flowcells for sequencing. Sequencing of the flowcells was performed on the Illumina HiSeq 2000 sequencer using TruSeq SBS sequencing kits, v3-cBot-HS, following a 2x150 indexed run recipe.

The resulting sequences were screened to remove reads that match Illumina artifacts with a sliding window kmer size of 28, step size of one. Any base with was Qscore less than Q3 was trimmed from the ends of the reads. Screened, trimmed reads were screened again to remove reads less than 50 bp in length, reads with an average Qscore less than 20, and reads with 3 or more ambiguous bases.

Quality-controlled reads were assembled using SOAPdenovo v1.05 (Li et al., 2010) (<http://soap.genomics.org.cn/soapdenovo.html>) at a range of six kmers (85, 89, 93, 97, 101, 105) with the default settings. The six contig sets were dereplicated and sorted based on length. Contigs shorter than 1800 bp were assembled into longer contigs using Newbler (Life Technologies, Carlsbad, California, USA) (flags: -tr, -rip, -mi 98, -ml 80). The resulting assemblies were merged and uploaded to the Integrated Microbial Genomes (IMG/M ER) database (<https://img.jgi.doe.gov/cgi-bin/mer/main.cgi>) for gene identification and annotation by their standard pipeline (Markowitz et al., 2014). The metagenomes are publicly available on the IMG/M ER website.

**Metagenome normalization.** The normalization factor for each metagenome was determined based on the abundance of approximately 35 conserved single-copy bacterial marker genes identified by COG ID (Raes et al., 2007). To determine this, the total number of bp in each COG (the sum of the length of the reads, weighted by read depth), was divided by the consensus length of the COG. The annotations of the reads included each COG were checked manually. The median abundance of the full-length COGs was used as each metagenomes' normalization factor.

**16 rRNA sequences from metagenome.** All sequences from each metagenome annotated as 16S rRNA were downloaded from IMG/M ER. The sequences were classified to the genus level using the DictDB database as described above (Mikaelyan *et al*, unpublished results). The length of the fragments assigned to each genus was weighted by its read depth. The relative abundances of the genera were calculating by dividing the basepairs assigned to each genus by the total basepairs of 16S rRNA.

**Phylogenetic distribution of protein coding genes.** All protein coding genes were assigned to phylogenetic bins by the top BLASTp hit at a 30% identity cutoff. The genes were weighted by their read depth to estimate the phylogenetic distribution of the copies of the protein coding genes. The phylogenetic distribution of the genes was determined based on relative abundance of the protein coding genes assigned to each domain or phylum as a percentage of all protein coding genes. Any gene classified as eukaryotic was removed from further analysis to exclude the host termite's DNA from gene abundance estimations.

**Metabolic gene family abundance and classification.** Sequences encoding nutritional enzymes (glycosyl hydrolases and peptidases) were identified by pfam domains. The pfam ID(s) for each glycosyl hydrolase family was confirmed in the Carbohydrate Active Enzymes (cazymes) database (<http://www.cazy.org/>) (Lombard et al., 2014) and/or the Pfam database (<http://pfam.xfam.org/>) (Finn et al., 2014). Of the peptidase families described in the MEROPS database (<http://merops.sanger.ac.uk/>), only peptidases that found in bacteria were considered (Rawlings et al., 2014). A subset of putatively secreted bacterial peptidases was selected for analysis based on biochemical evidence for secretion for a given peptidase family (<http://pfam.xfam.org/>), or the presence of a signal peptide. Signal peptides were predicted based on representative sequences for each family (<http://www.cbs.dtu.dk/services/SignalP/>) (Petersen et al., 2011).

The sequences identified by pfam domain as encoding a glycosyl hydrolase or peptidase family were manually checked to confirm the annotation. Sequences with annotation that conflicted with the expected annotation based on the glycosyl hydrolase or peptidase family were removed. Sequence lengths were weighted by read depth to estimate the number of basepairs in the sequence and then summed to estimate the number of basepairs in each family. Full-length gene family abundances were calculated by dividing the total number of basepairs in each gene family by the approximate length of the gene in InterPro (<http://www.ebi.ac.uk/interpro/>) (Hunter et al., 2012).

The gene family abundances were normalized across the metagenomes by dividing by the normalization factor for each metagenome. The normalized gene abundance value estimates the number of full-length genes from that family per bacterial genome for direct comparison between and within metagenomes.

**Statistics.** To compare bacterial community structure between samples and between strategies (the two amplicon datasets, 16S rRNA from the metagenomes, and all coding gene copies from the metagenomes), relative abundances of taxonomic groups were visualized using non-metric multidimensional scaling (NMDS) with a Bray-Curtis similarity measure in the PAST v3 software package (Hammer et al., 2001).

Gene family abundances were assessed for significant differences between feeding guild, host species, and homologous gut compartment using the STAMP software package (Parks et al., 2014). Multiple groups were evaluated using ANOVA with the Games-Howell post-hoc test and corrected for multiple hypotheses (false discovery rate correction cutoff of 0.05) with Storey's FDR, or Benjamini-Hochberg FDR when p-values were not uniformly distributed (see references in (Parks et al., 2014)). Two groups (i.e. soil feeders vs. wood feeders) were evaluated using Welch's t-test with Welch's inverted confidence interval and corrected for multiple hypotheses as described for multiple group comparisons (Parks et al., 2014). Gene family abundance profiles were graphed using non-metric multidimensional scaling ordinations with the Bray-Curtis similarity measure with 500 bootstraps using the PAST v3 software package (Hammer et al., 2001).



## Results

**Sample Summary and Overview of Metagenomic Sequencing.** To characterize the functional potential of the gut microbiota of higher termites, we performed a metagenomic analysis on the microbial community from the P1, P3, and P4 hindgut compartments from six termite species representing four feeding guilds (Tab. 1). The eighteen resulting assemblies ranged from 712-2060 Mbp (unweighted). Accounting for read depth, the assembled sequences contained 19,067-57,891 Mbp. The vast majority of sequences could be assembled; the basepairs in assembled fraction outnumbered those in the unassembled fraction by up to two orders of magnitude. In addition, the longest contig assembled for each metagenome was as long as 628 kbp. The number of genomes captured in each metagenome, estimated by the abundance of conserved full-length single-copy bacterial genes, was highly variable between gut compartments from the same termite (for example, 799 for the P1 and 11097 for the P3 from *M. parvus*). The standard deviations for these estimates were very low, averaging 10% of the single-copy gene abundance. Furthermore, the bacterial counts per gut, estimated by QPCR of the 16S rRNA gene, followed similar relative patterns to the single-copy gene abundance from the same gut compartment. The P1 sections had the fewest copies of the bacterial 16S rRNA gene, while the P4 section from the same host had approximately twofold more 16S rRNA gene copies. The P3 sections had the most copies, up to two orders of magnitude more than the P1 from the same host.

A BLASTp analysis was used to assign the protein coding genes to the domain level at 30% identity (Fig. 1). Most sequences could be classified; 10-38% of the protein coding genes were unclassified. For all P3 gut compartments, bacterial genes were the majority of the sequences, making up 54-87% of all protein coding genes. Bacterial genes were 11-73% of protein coding genes from the P1 gut compartment, and 48-69% of protein coding genes from the P4 gut compartment. Eukaryotes comprised only 2-7% of all protein coding genes from the P3 gut compartments, but were a more significant component of sequences from the P1 gut compartment (9-64%) and the P4 (7-39%). Archaeal sequences totaled less than 1% of all protein coding genes from the P1 compartment, but they increased in relative abundance in posterior gut compartments to comprise up to 4% of protein coding genes in the P4 gut compartment.

**Community Structure.** The bacterial community structure was evaluated at the phylum level in using four strategies: the 16S rRNA genes from the metagenomes; all protein coding genes from the metagenomes; and two independent amplifications of 16S rRNA gene regions (iTAG and gTAG) (Tab. S1). Multiple strategies of evaluation were employed because of potential drawbacks associated with any one strategy: possible primer biases found in amplification-based studies; lack of appropriate references for accurately classifying protein coding genes; and limited sequence information for examining a specific gene from a metagenome. By comparing the four strategies, we determined that the community structure based on iTAG amplicons and 16S rRNA genes from the metagenomes were very similar, but the community structure determined by all protein coding genes and gTAG were different with regards to specific taxa. To explore the trends and possible causes of the discrepancies in community structures, the difference in relative abundance of the bacterial phyla was determined for iTAG and all coding sequences (Fig. 2). For the Firmicutes, the most abundant phyla in most samples, there was no consistent pattern in the difference in relative abundance between the two strategies, although individual variation was high. However, some phyla were consistently enriched in relative abundance as determined by one strategy versus the other. For example, Fibrobacteres and candidate phylum TG3 were enriched in iTAG-based community structure. TG3 was never detected by all coding sequences.

**Tab. 1** | Summary of metagenomic sequencing and sample data.

Gut section	Feeding strategy	Metagenome size (Gbp)			Longest Contig (Kbp)	SCG (10 <sup>3</sup> ) <sup>3</sup>	Bacterial Abundance (10 <sup>6</sup> gut <sup>-1</sup> )
		Assembled		Unassembled			
		Unweighted <sup>1</sup>	Weighted <sup>2</sup>				
<i>Nasutitermes corniger</i>							
P1	Wood	1.4	44.0	333	83.5	0.4 ± 0.1	0.3 ± 0.3
P3		0.6	44.3	2,464	296.0	13.2 ± 1.6	138 ± 55
P4		1.6	40.5	2,270	220.0	4.5 ± 0.3	—
<i>Microcerotermes parvus</i>							
P1	Wood	1.5	46.5	1,084	39.1	0.8 ± 0.1	2.1 ± 1.6
P3		0.7	41.1	2,135	71.1	11.1 ± 1.2	260 ± 10
P4		1.5	46.6	1,399	111.7	3.3 ± 0.3	3.6 ± 1.8
<i>Cornitermes</i> sp.							
P1	Grass/wood	1.5	42.8	2,993	249.3	8.6 ± 0.7	2.3 ± 1.6
P3		1.3	42.0	3,890	291.0	9.8 ± 0.9	7.6 ± 3.6
P4		1.3	29.9	5,944	275.8	5.9 ± 0.6	4.9 ± 3.3
<i>Termes hospes</i>							
P1	Wood/soil interface	1.5	47.9	707	84.7	0.8 ± 0.1	3.8 ± 3.0
P3		1.2	28.1	6,065	177.4	7.2 ± 0.5	11 ± 3.7
P4		1.8	36.3	3,786	110.0	7.2 ± 0.5	9.4 ± 5.6
<i>Neocapritermes taracua</i>							
P1	Wood/soil interface	1.5	40.0	3,692	177.0	3.2 ± 0.3	11 ± 4.6
P3		0.9	19.1	9,170	628.2	7.4 ± 0.5	101 ± 28
P4		1.5	34.4	4,684	232.3	6.7 ± 0.5	29 ± 14
<i>Cubitermes ugandensis</i>							
P1	Soil	2.1	57.9	10,474	175.6	8.5 ± 0.8	13 ± 5.1
P3		1.1	22.8	9,229	226.8	5.2 ± 0.5	33 ± 13.8
P4		1.3	25.4	6,425	107.5	4.7 ± 0.5	12 ± 4.7

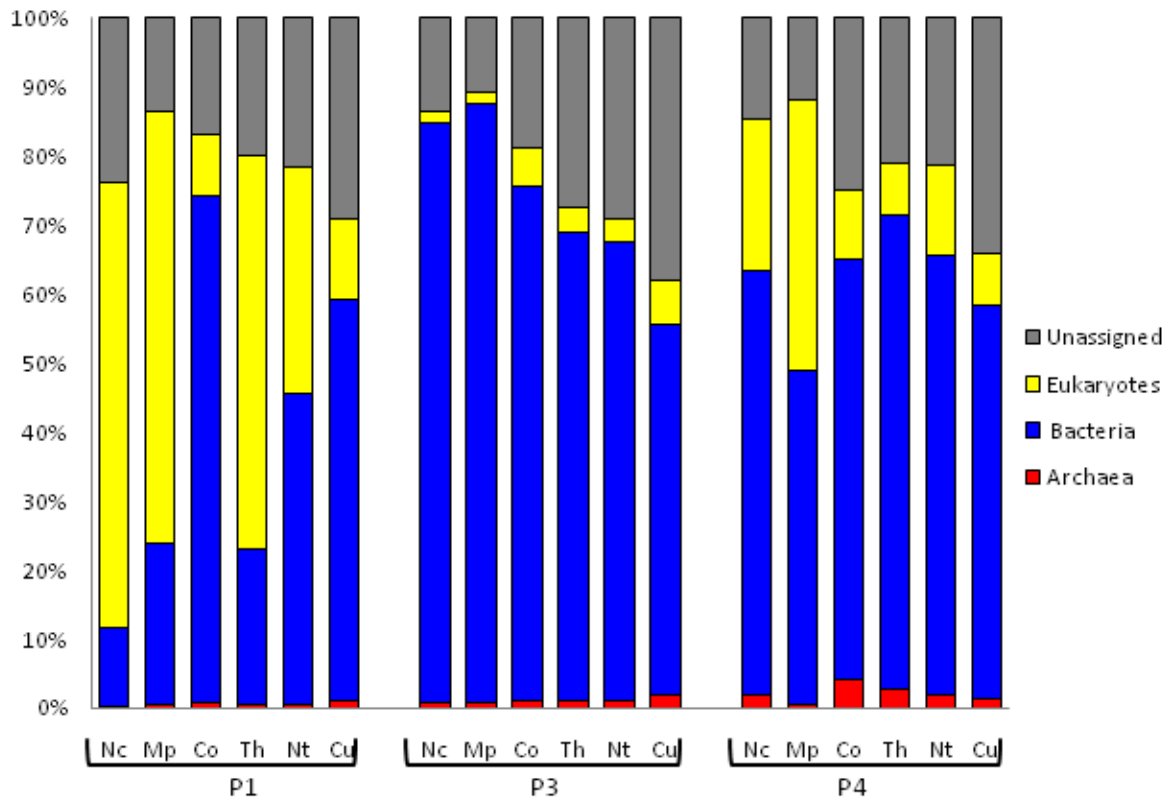
<sup>1</sup> Number of bp after assembly without accounting for depth of coverage

<sup>2</sup> Total number of bp that went into assembly, calculated by (Sequence length \* read depth)

<sup>3</sup> Median of estimated number of full-length copies of 35 conserved bacterial single-copy genes (SCG) ± standard deviation

<sup>4</sup> 16S rRNA copies per gut section bacterial

Furthermore, Proteobacteria, despite their relative rarity in iTAG-based community structures, were always found in greater relative abundances in all coding sequences. With few exceptions, Actinobacteria were also differentially enriched in all coding sequences, but Spirochaetes were differentially enriched in iTAG-based community structure. Bacteroidetes was the only phylum that showed a gut section-specific pattern; Bacteroidetes in the P1 and P3 was differentially enriched in all coding sequences, but in the P4 was differentially enriched in the iTAG-based community structure.

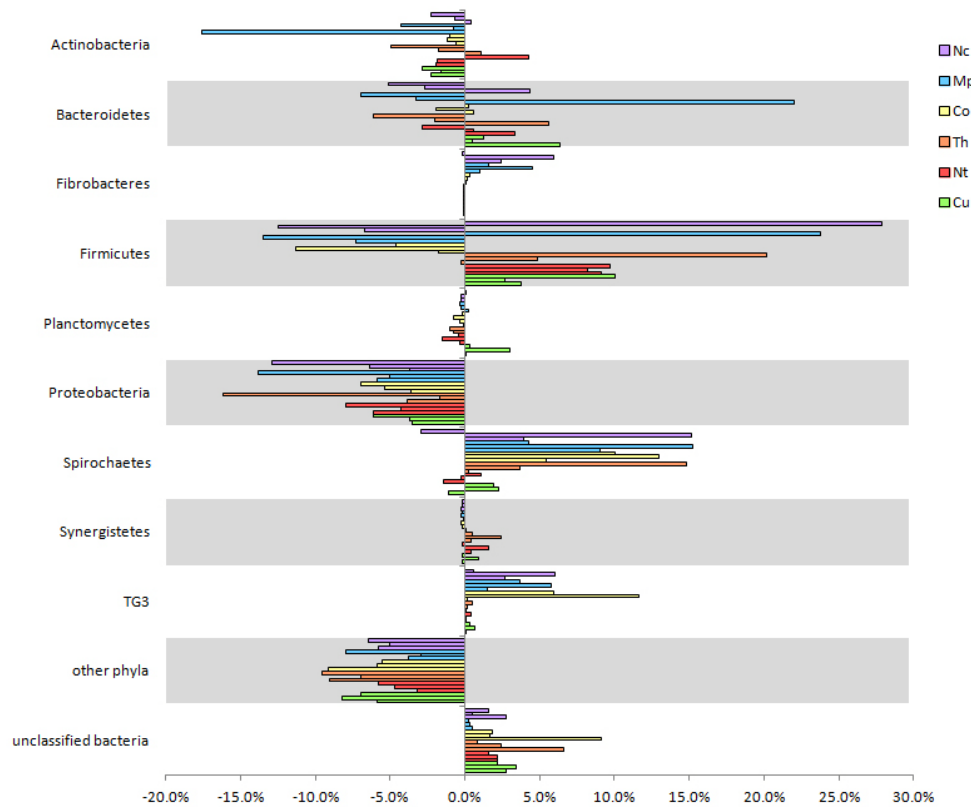


**Fig. 1 | Domain-level assignment of protein coding gene copies.** Based on best BLASTp hit. Gene copies were estimated by the read depth of each gene (read depth of one for unassembled reads). Termite hosts are abbreviated as follows: Nc, *N. corniger*; Mp, *M. parvus*; Co, *Cornitermes* sp.; Cu, *C. ugandensis*; Th, *T. hospes*; Nt, *N. taracua*.

The phylum-level community structures determined by 16S rRNA gene-based strategies (gTAG, iTAG, and 16S rRNA genes from the metagenomes) were graphed using non-metric multidimensional scaling (NMDS) (Fig. 3). Within each sample, the community structures based on iTAG and 16S rRNA sequences from the metagenome grouped very closely together. The gTAG-based community structures were less similar and did not cluster with the community structures from iTAG and 16S rRNA sequences from the same sample. At a higher phylogenetic resolution, specific groups were found to be responsible for these differences, in particular, the class Erysipelotrichi in the phylum Firmicutes. In some metagenomes this class was detected in very high abundance by iTAG and 16S rRNA genes from the metagenome, but was negligible in abundance in gTAG reads (for example, this class was 33.4% of the iTAG reads from *N. corniger* P1, and 38.4% of the 16S rRNA gene sequences from the same sample, but was 0% of the gTAG reads). This difference in abundance was due to differences binding efficiency between the primers used in iTAG and gTAG. The primers used

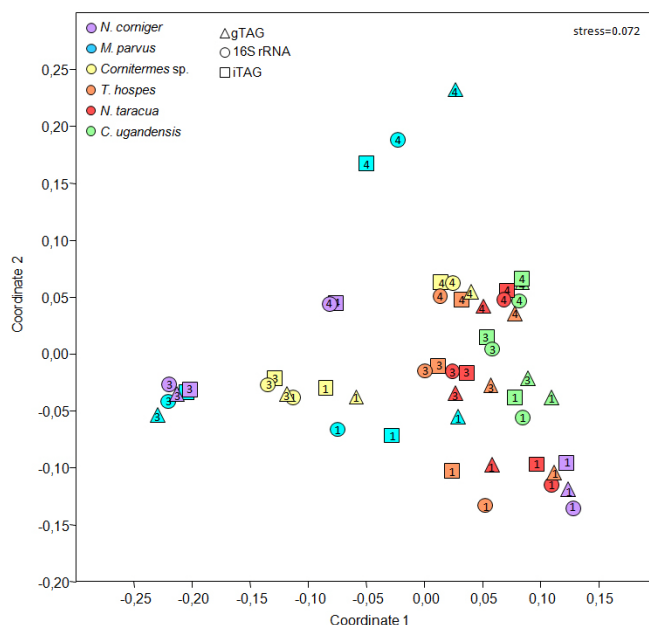
in iTAG matched with zero mismatches approximately 82% of *Erysipelotrichi* 16S rRNA gene sequences in the DictDB database (Mikaelyan et al., in prep.), but the primers used in gTAG matched only 11% of *Erysipelotrichi* 16S rRNA gene sequences with zero mismatches.

Regardless of method for determining community structure, between-sample comparisons showed groupings based on feeding strategy and gut section (Fig. 3). For example, the P3 gut section showed a strong grouping along a gradient based on feeding strategy: a tight group of the two wood feeders (*N. corniger* and *M. parvus*); the P3 from the grass-feeding *Cornitermes* sp. (along with the P1 from the same host); a group of interface feeders (*T. hospes* and *N. taracua*); and the soil feeder (*C. ugandensis*). The P1 community structures showed formed a loose group that was not based on feeding strategy, although the P1 community structures from *Cornitermes* and *C. ugandensis* grouped more closely to the P3 from their respective termite hosts. The P4 community structures displayed a slightly different grouping pattern: the grass feeder, interface feeder, and soil feeder all grouped very closely, but the P4s from the wood feeders were individually separated from the main cluster.



**Fig. 2 |** Differences in percentage points of relative abundance of phyla between iTAG and all coding sequences. Negative values represent greater relative abundance in all coding sequences; positive values represent greater relative abundance in iTAG. Termite hosts are abbreviated as follows: *N. corniger* (Nc, purple); *M. parvus* (Mp, blue); *Cornitermes* sp. (Co, yellow); *T. hospes* (Th, orange); *N. taracua* (Nt, red); *C. ugandensis* (Cu, green). For each termite host, the P1 is top bar, the P3 the middle, and the P4 the bottom.

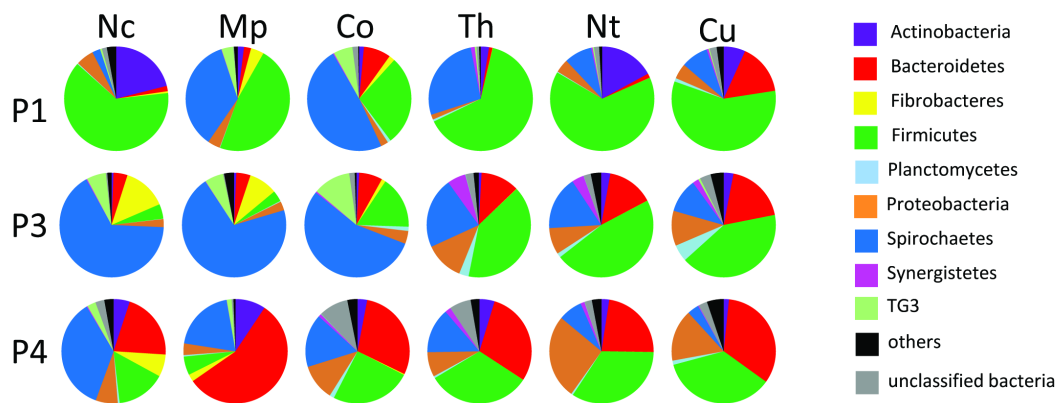
To explore which bacterial phyla were responsible for the groupings based gut section and feeding strategy we saw in the NMDS, we focused on the phylum-level community structure as determined by iTAG analysis (Fig. 4). The iTAG analysis was selected because it was most similar to the community structure of the 16S rRNA from the metagenome, but was not limited by number of sequences. The gradient according to feeding strategy in the P3 that was shown in the NMDS (Fig. 3) could be seen in the specific phyla. Spirochaetes were the overwhelming majority of bacteria in the P3 from wood feeders, were approximately have of the bacterial community in *Cornitermes* P3 and P1, and were approximately 25% or less of the bacterial community in the P3 of the interface and soil feeders. In addition, Fibrobacteres and candidate phylum TG3 were minor but significant members of the P3 from wood and grass feeders. The P1 communities from interface and soil feeders were consistently dominated by Firmicutes, but Spirochaetes were important members of the grass-feeding *Cornitermes* sp. and the wood-feeding *M. parvus*. The bacterial community from the P1 of the wood-feeding *N. corniger* was very different from that of *M. parvus*; the community was instead dominated by Firmicutes and Actinobacteria. The bacterial communities from the P4 gut sections of the grass, interface, and soil feeders were very similar and were dominated by Bacteroidetes and Firmicutes, with Proteobacteria as minor members. Bacteroidetes was the dominant phylum of the P4 from the wood feeder *M. parvus*, with Spirochaetes as a minor member. In the P4 from the other wood feeder, *N. corniger*, Spirochaetes were the dominant phylum and Bacteroidetes were a minor member.



**Fig. 3 | Non-metric multidimensional scaling (NMDS, Bray-Curtis similarity measure) of phylum-level community structure.** Data is based on iTAG (square), gTAG (triangle), and 16S rRNA from the metagenome (circle). Points are colored according to termite host: *N. corniger* (purple); *M. parvus* (blue); *Cornitermes* sp. (yellow); *T. hospes* (orange); *N. taracua* (red); *C. ugandensis* (green). The gut compartment for each sample is labeled by its number. NMDS was calculated in PAST v3 (Hammer et al., 2001).

**Glycosyl Hydrolase Abundance.** The absolute abundances of glycosyl hydrolase families (total length in bp divided by consensus gene length for each family) were calculated for each metagenome. These abundances were then normalized for comparison between metagenomes by dividing by the normalization factor for each metagenome (Tab. S2) to estimate the number of genes from each glycosyl hydrolase family per bacterial genome (Tab. S3). Each glycosyl hydrolase was evaluated using

an ANOVA (with Games-Howell post-hoc test and corrected for multiple hypotheses) in the STAMP software package (Parks et al., 2014). For most glycosyl hydrolase families, there was no significant difference in abundance between the metagenomes. However, glycosyl hydrolase families with cellulase activity (families 5, 6, 9, 44, and 45), hemicellulase activity (families 8, 10, 11, 12, and 53), and  $\alpha$ -amylase activity (families 13, 31, 57, and 77) had significant differences in abundance between metagenomes from different feeding guilds, gut sections, or termite hosts. Some glycosyl hydrolase families with cellulase activity (families 7 and 48) and hemicellulase activity (family 23) were very low in abundance and were not significantly different between metagenome groups (Tab. S3). These gene families were not included in further analyses.



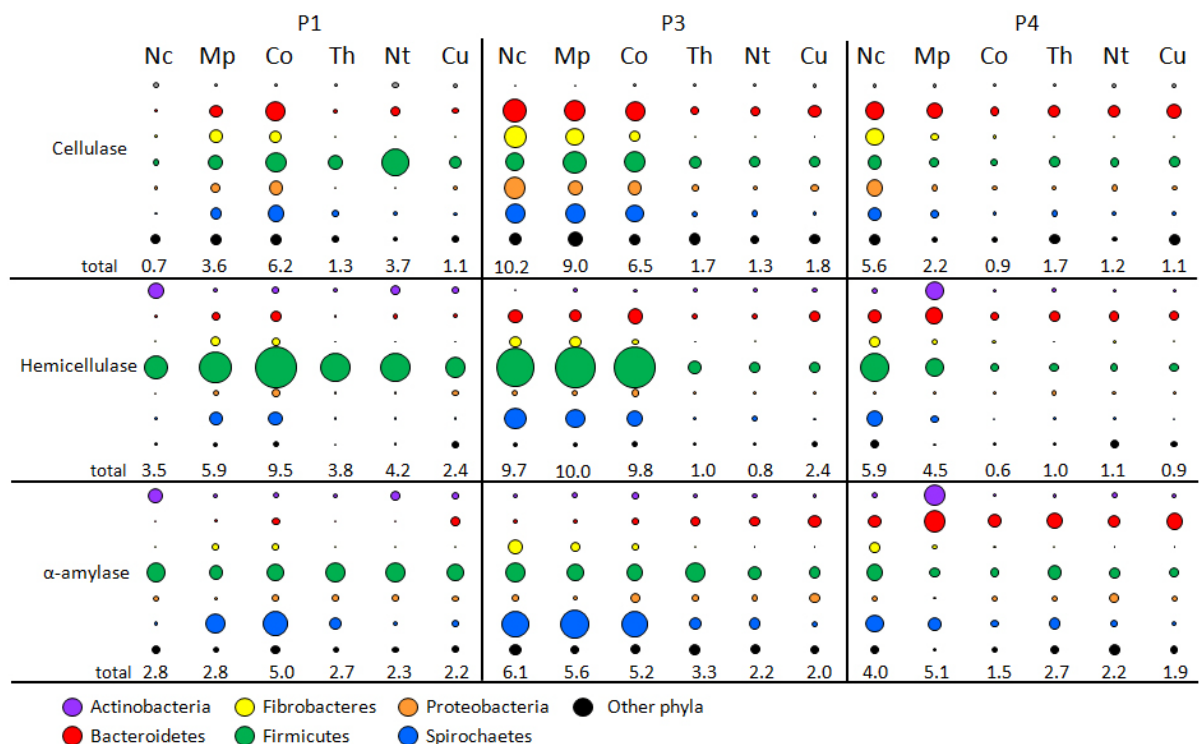
**Fig. 4 | Phylum-level community structure based on iTAG analysis.** Termite hosts are abbreviated as follows: Nc, *N. corniger*; Mp, *M. parvus*; Co, *Cornitermes* sp.; Cu, *C. ugandensis*; Th, *T. hospes*; Nt, *N. taracua*.

The genes encoding cellulases, hemicellulases, and  $\alpha$ -amylases were classified to the phylum level by their top BLASTp hit (30% identity). The classification of these genes was overlaid with the abundance data for a quantitative comparison of the bacterial community members that possessed these genes (Fig. 5). Cellulases were most abundant in the P3 of wood feeders (*N. corniger* and *M. parvus*) and grass feeders (*Cornitermes* sp.), where genes encoding these enzymes were classified nearly equally to Bacteroidetes, Fibrobacteres, Firmicutes, Proteobacteria, and Spirochaetes (Fig. 5). Similarly high abundance of cellulases and similar distribution among bacterial phyla was found in the P1 of *Cornitermes* sp. and the P4 of *N. corniger*. Cellulases were low in abundance in the P1 and P3 in interface feeders (*T. hospes* and *N. taracua*) and soil feeders (*C. ugandensis*) compared to other feeding guilds. The exceptions to this were the P1 of *N. taracua*, in which Firmicutes possessed relatively high amounts of cellulases. The total abundance of cellulases in the P1 of *N. taracua* was nearly equal to the total abundance in the homologous compartment from *M. parvus*, although the cellulases from the latter were assigned nearly equally to Bacteroidetes, Fibrobacteres, Firmicutes, Proteobacteria, and Spirochaetes. In the P1 interface and soil feeders, the cellulases were low in abundance and were mainly classified to Firmicutes. Cellulases from the P3 of interface and soil feeders, were similarly low in abundance, but were classified to Bacteroidetes and Firmicutes. In the P4 from the same termites, cellulases were also low in abundance and were mainly classified to Bacteroidetes.

Hemicellulases, although present in total abundance nearly equal to cellulases in most gut sections, were mainly classified to Firmicutes (Fig. 5). In the P1, hemicellulases were most abundant in *M. parvus* and *Cornitermes* sp., although they were also present in relatively high abundance in the

other 4 termites. In contrast, hemicellulases were in very high abundance in the P3 of wood and grass feeders, but were nearly absent in the P3 of interface and soil feeders. The hemicellulases of the P3 from wood and interface feeders were mostly classified to Firmicutes, or to Spirochaetes, Fibrobacteres, and Bacteroidetes. In the P4 gut section, hemicellulases were relatively highly abundant in the wood feeders, where they were classified mostly to Firmicutes, Spirochaetes, and Bacteroidetes for *N. corniger*, or to Actinobacteria, Firmicutes, and Bacteroidetes for *M. parvus*. The only other gut section in which hemicellulases were classified to Actinobacteria in a considerable amount was the P1 of *N. corniger*.

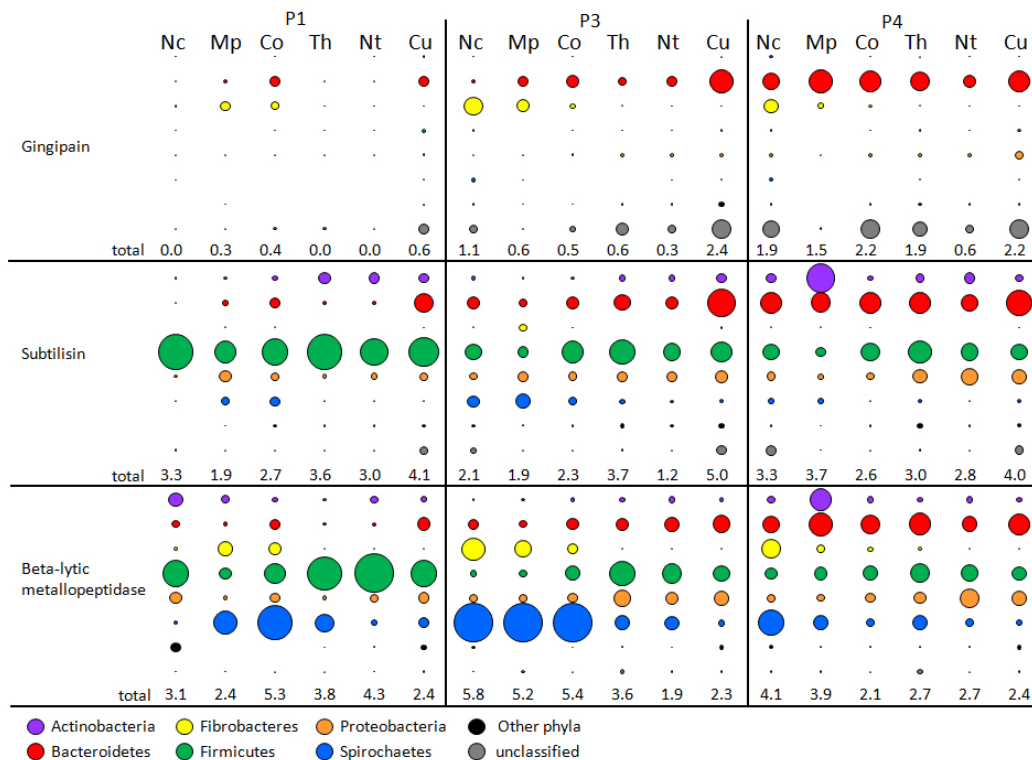
$\alpha$ -amylases were found in highest total abundance in the P3 of wood and grass feeders, where they were mostly classified to Spirochaetes, with Firmicutes and Fibrobacteres possessing minor amounts (Fig. 5). Similarly, the  $\alpha$ -amylases in the P1 of *Cornitermes* sp. and *M. parvus* were classified mainly to Spirochaetes or Firmicutes. The  $\alpha$ -amylases in the P1 of *N. corniger* were classified to Firmicutes or Actinobacteria. In the P4, the  $\alpha$ -amylases of *N. corniger* were classified to Spirochaetes, Firmicutes, Fibrobacteres, and Bacteroidetes. The  $\alpha$ -amylases in the P4 of *M. parvus* were mostly classified to Actinobacteria or Bacteroidetes. Compared to the wood and grass feeders  $\alpha$ -amylases were much lower in abundance in interface and soil feeders. However, the  $\alpha$ -amylases that were detected in these feeding guilds were mostly classified to Firmicutes in the P1; to Firmicutes, Spirochaetes, or Bacteroidetes in the P3; and Bacteroidetes or Firmicutes in the P4.



**Fig. 5 | Phylum-level classification of cellulases, hemicellulases, and  $\alpha$ -amylases. Bubbles are sized according to the estimated number of genes per genome.** Totals for each enzyme type are given as genes per genome. Termite hosts are abbreviated as follows: Nc, *N. corniger*; Mp, *M. parvus*; Co, *Cornitermes* sp.; Cu, *C. ugandensis*; Th, *T. hospes*; Nt, *N. taracua*. The major phyla included are colored as follows: Actinobacteria (purple); Bacteroidetes (red); Fibrobacteres (yellow); Firmicutes (green); Proteobacteria (orange); Spirochaetes (blue); other phyla (black).

**Peptidase Abundance.** The absolute abundances of peptidase families found in bacteria (total length in bp divided by consensus gene length for each family) were calculated for each metagenome (Tab. S4). The number of genes from each bacterial peptidase family was estimated per bacterial genome by dividing the absolute abundance by the normalization factor for each metagenome (Tab. S2). As with the glycosyl hydrolase families, all bacterial peptidase families were evaluated using an ANOVA (with Games-Howell post-hoc test and corrected for multiple hypotheses) in the STAMP software package (Parks et al., 2014). Most peptidase families were not significantly different in abundance between the metagenomes. Of the peptidase families that were significantly different, some were putatively secreted and were considered further (Tab. S4). Three families within this subset were the focus of further analysis: gingipain (family C25), beta-lytic metallopeptidase (family M23), and subtilisin (family S8).

The genes encoding gingipain, beta-lytic metallopeptidase, and subtilisin were classified to the phylum level by their top BLASTp hit (30% identity). The abundance data was overlaid onto this classification to quantitatively compare the bacterial community members that possessed these peptidases (Fig. 6). Gingipain genes were rare in the P1, and increased in abundance for all samples in the P3, peaking in abundance in the P4. For the wood feeders (*N. corniger* and *M. parvus*), the gingipain genes in the P3 were mostly classified to Fibrobacteres, with only some classified as Bacteroidetes. For grass, interface, and soil feeders, the majority of gingipain genes in both the P3 and P4 were classified as Bacteroidetes. However, approximately half of these gingipain genes could not be classified to the phylum level.



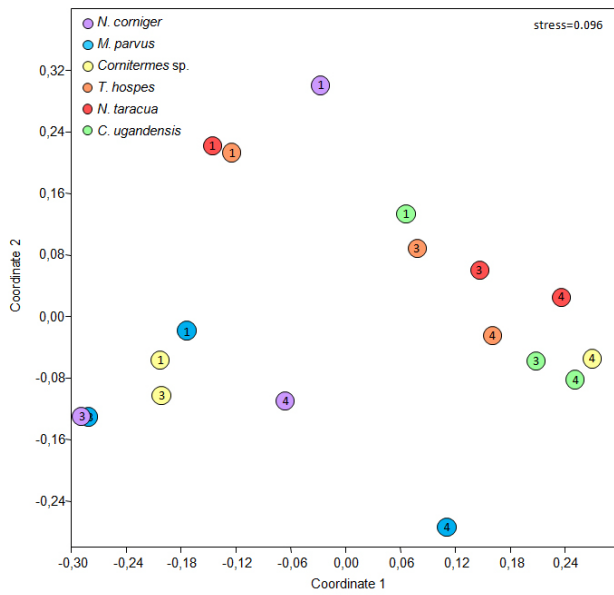
**Fig. 6 | Phylum-level classification of the peptidase families gingipain (C25), subtilisin (S8), and beta-lytic metallopeptidase (M23).** Bubbles are sized according to the estimated number of genes per genome. Totals for each peptidase family are given as genes per genome. Termite hosts are abbreviated as follows: Nc, *N. corniger*; Mp, *M. parvus*; Co, *Cornitermes* sp.; Cu, *C. ugandensis*; Th, *T. hospes*; Nt, *N. taracua*. The major phyla included are colored as follows: Actinobacteria (purple); Bacteroidetes (red); Fibrobacteres (yellow); Firmicutes (green); Proteobacteria (orange); Spirochaetes (blue); other phyla (black); unclassified (grey).



Subtilisin genes were detected in greater total abundance in interface and soil feeders than in wood and grass feeders (Fig. 6). Across all samples, the majority of subtilisin genes detected in the P1 were classified as Firmicutes. However, some subtilisin genes from all P1 gut compartments except *N. corniger* were classified as Proteobacteria, Actinobacteria, or Bacteroidetes. In the P3, the subtilisin genes could be classified nearly evenly to Firmicutes and Bacteroidetes, although Proteobacteria and Actinobacteria were minor members. Furthermore, the P3 of wood and grass feeders also contained some subtilisin genes that were classified as Spirochaetes. Subtilisin genes in the P4 were also classified as several different phyla: most were classified as Bacteroidetes, Firmicutes, or Proteobacteria. In addition, Actinobacteria possessed a minor fraction of the subtilisin genes in the P4 of all samples, with the exception of *M. parvus*, in which Actinobacteria possessed the majority of subtilisin genes.

In total, beta-lytic metallopeptidase genes were more abundant in wood and grass feeders than in interface and soil feeders. In the P1, most beta-lytic metallopeptidase genes from the wood-feeding *M. parvus* and the grass-feeding *Corniger* sp. were classified as Spirochaetes, with Firmicutes, Proteobacteria, Fibrobacteres, Proteobacteria, and Bacteroidetes as minor members. The P1 of the other wood feeder, *N. corniger*, contained beta-lytic metallopeptidase genes that were classified mainly as Firmicutes, with Actinobacteria and Proteobacteria as minor members. The beta-lytic metallopeptidase genes from the P1 of interface and soil feeders were mostly classified as Firmicutes, with Spirochaetes, Bacteroidetes, and Proteobacteria as minor members. Beta-lytic metallopeptidase genes peaked in abundance in wood and grass feeders in the P3, in which they were mostly classified as Spirochaetes, with Fibrobacteres, Bacteroidetes, Firmicutes, and Proteobacteria as minor members. In the P3 of interface and soil feeders, beta-lytic metallopeptidase genes were lower in total abundance and were nearly evenly classified to Firmicutes, Proteobacteria, Bacteroidetes, and Spirochaetes. For most samples, the total abundance of beta-lytic metallopeptidase genes decreased in the P4. In the P4 of *N. corniger*, most beta-lytic metallopeptidase genes were classified as Spirochaetes, with Fibrobacteres, Bacteroidetes, Firmicutes, Proteobacteria, and Actinobacteria as minor members. In the other wood feeder, *M. parvus*, Actinobacteria and Bacteroidetes possessed most of the beta-lytic metallopeptidase genes, with Firmicutes, Spirochaetes, Fibrobacteres, and Proteobacteria possessing minor amounts. For grass, interface, and soil feeders, the beta-lytic metallopeptidase genes found in the P4 were nearly evenly classified as Bacteroidetes and Firmicutes, with Proteobacteria, Spirochaetes, and Actinobacteria as minor members.

**Functional Profiles.** Taken together, the phylogenetic classifications and abundances of carbohydrate-active enzymes (cellulases, hemicellulases, -amylases) and peptidases (gingipain, subtilisin, and beta-lytic metallopeptidase) form a functional profile of the microbial communities. These profiles were visualized using non-metric multidimensional scaling (NMDS) to better explore similarities in functional potential between samples (Fig. 7). The ordination shows two clear clusters that correspond to feeding strategy: the profiles from wood and grass feeders in one group, and the communities from interface and soil feeders in the other. Two exceptions to this are the P4 from the grass feeder, *Cornitermes* sp., which clusters very closely with the P4 from the soil feeder, *C. ugandensis*; and the P1 from a wood feeder, *N. corniger*, which shows only low similarity to the P1 from the interface feeders, *T. hospes* and *N. taracua*. Within the interface and soil feeder group, homologous compartments cluster together into subgroups. There is some overlap between the P3 and P4 subgroups, but the P1 subgroup is more distinct. However, the cluster of profiles from wood and grass feeders shows a different pattern in homologous compartment similarity. The P3 from wood feeders are very highly similar, but the P3 from the grass feeder *Cornitermes* sp. is more similar to the P1 from the same termite and the P1 from a wood feeder, *M. parvus*. The P4 from the wood feeders show very low similarity to one another.



**Fig. 7 | Non-metric multidimensional scaling (NMDS, Bray-Curtis similarity measure) of glycosyl hydrolase and peptidase functional profiles.** Profiles are based on abundances of phylum-level classifications of gene families encoding cellulases, hemicellulases,  $\alpha$ -amylases, gingipain (C25), subtilisin (S8), and beta-lytic metalloproteinase (M23). Points are colored according to termite host: *N. corniger* (purple); *M. parvus* (blue); *Cornitermes* sp. (yellow); *T. hospes* (orange); *N. taracua* (red); *C. ugandensis* (green). The gut compartment for each sample is labeled by its number. NMDS was calculated in PAST v3 (Hammer et al., 2001).

## Discussion

The differentiated compartments in the hindgut of higher termites house complex, diverse microbial communities that are essential to the nutrition and vitality of their host (Brune, 2014). In this study, we examined the microbial communities in the homologous P1, P3, and P4 compartments from termites that feed on wood, grass, the wood/soil interface, and soil to determine the community structure and potential functional role of microbes in the digestion of plant material and peptides. By using comparative metagenomics to quantify key genes encoding glycosyl hydrolases and peptidases, we identified patterns common between homologous compartments as well as signals attributable to specific feeding guilds. Furthermore, we classified major glycosyl hydrolases and peptidases to microbial community members to give a better understanding of the role(s) that bacterial groups play in contributing to termite host's nutrition.

Even from relatively distantly related termite hosts, anterior homologous gut compartments show broadly consistent patterns. For example, the first proctodeal segment (P1) of 52 species (four subfamilies) of higher termites has higher alkalinity compared to the P3 and P4 hindgut compartments (Bignell and Eggleton, 1995). In addition, a study focusing on the P1 of four species of higher termites demonstrated that its distinct microbial community was dominated by Firmicutes (Thongaram et al., 2005). Our results are consistent with this concept in several aspects. QPCR results indicate that the P1 gut compartment had the fewest 16S rRNA gene copies compared to the P3 and P4 from the same termite host, by up to two orders of magnitude. The multivariate ordination plots based on bacterial phyla and functional profiles indicated that the microbial community structures from the P1 samples were broadly similar and separate from P3 and P4 microbial community structures, but and showed a slight influence of feeding guild in similarities in functional potential. The two interface feeders show very similar P1 functional profiles, but the P1 profiles from wood feeders are very different: the P1 functional profile of *M. parvus* is similar to the P1 functional profile from the grass feeder (*Cornitermes* sp.). However, given that the P1 from *N. corniger* contains an order of magnitude fewer bacterial cells than the P1 from *M. parvus*, it is difficult to evaluate what, if any, contribution the gut microbial from the P1 of *N. corniger* makes to host nutrition. In interface and soil feeders, hemicellulases are more abundant than cellulases, although both are less abundant than the same enzymes in wood and grass feeders. The abundances of genes encoding  $\alpha$ -amylases in the wood- and grass-feeding termites are lowest in the P1 gut compartment. This may be because most bacterial  $\alpha$ -amylases have acidic to neutral pH optima (Pandey et al., 2000, Rana et al., 2013) and may be excluded from the more alkaline environment found in the P1. However,  $\alpha$ -amylases in the P1 of grass feeders are roughly twice as abundant as  $\alpha$ -amylases in wood, interface, and soil feeders.  $\alpha$ -amylases degrade starch, a major component of grains and seeds (Huntington, 1997), but the reasons for the relatively high abundance of  $\alpha$ -amylases in a grass-feeding termite is unclear.

In contrast to the low abundance of glycosyl hydrolases, some putatively secreted peptidases were detected in high abundance in the P1. Subtilisins found in the P1 are mostly classified to Firmicutes, fitting with previous research that has found subtilisins are most active at alkaline pH values and are found in *Bacillus* species (Gupta et al., 2002). However, subtilisins are also abundant in the P3 and P4, especially in interface and soil feeders. In addition, in the posterior gut sections of these termites, subtilisins are increasingly assigned to Bacteroidetes and Actinobacteria. In the P4, Bacteroidetes are the dominant members of the microbial community, but Actinobacteria are only minor members of the bacteria communities from *N. corniger* and *M. parvus*. The digestion of peptides by subtilisin, or subtilisin-like enzymes, may be more complex and widespread than previously realized. Bacteroidetes also possess the majority of gingipain genes, which are enriched in the P4 gingipain compared to other gut compartments. The increased abundance of Bacteroidetes in

the community and Bacteroidetes-owned gingipain is consistent across all feeding guilds, suggesting the importance of the P4 gut compartment for digestion of peptidase that is distinct from the P1 and P3. Gingipains are well-studied as virulence factors of *Porphyromonas gingivalis*, which causes periodontitis and other oral diseases by fermenting a variety of host or microbial proteins for nutrition or defense (Olsen and Potempa, 2014). Although little is known about gingipains in other environments, the structural and substrate diversity of gingipains may allow them to be involved in the fermentation of microbial proteins in the termite P4 (Fitzpatrick et al., 2009, Li and Collyer, 2011).

For wood and grass feeders, the main site of digestion of complex carbohydrates is the P3; this gut section had the highest 16S rRNA gene copies and the highest abundances of genes encoding cellulases, hemicellulases, and  $\alpha$ -amylases were found in the P3 of wood- and grass-feeding termites. However, the microbial groups to which these genes were assigned differed between enzyme types. The highly abundant cellulase genes from wood- and grass-feeding termites were assigned to several different phyla, indicating a rate of cellulose digestion that requires high functional redundancy (Wohl et al., 2004, Hattenschwiler et al., 2011). In contrast, the hemicellulases were largely assigned to Firmicutes, indicating that one, albeit very large and diverse, bacterial phylum is filling this ecological role. However, the taxonomic assignment of the majority of hemicellulase genes to Firmicutes is suspect because there is recent evidence for horizontal gene transfer of carbohydrate metabolism genes from Clostridia to a Spirochaete (Caro-Quintero et al., 2012). Given that the microbial community of the P3 in wood- and grass- feeders is strongly and consistently dominated by Spirochaetes, with Firmicutes a minor community member only in grass-feeding termites, it is likely that a large number of hemicellulase genes are incorrectly assigned to Firmicutes. Furthermore, Fibrobacteres and candidate phylum TG3 are numerically minor members of the gut community of wood- and grass-feeding termites, but a recent study on the fiber-associated community from the P3 lumen of *N. corniger* determined they contribute significantly to the digestion of cellulose (Mikaelyan et al., 2014). However, genome information on these phyla is limited, resulting in their poor representation in reference databases (Suen et al., 2011, Sorokin et al., 2014). Current work using compositional binning that uses phylogenetic markers instead of whole genome sequence classification is exploring the functional role of Fibrobacteres and candidate phylum TG3 in the digestion of cellulose in wood- and grass-feeding higher termites (Dietrich, et al. in prep.).

Despite the similarities between the homologous P1 and P4 compartments, the P3 shows stronger patterns related to feeding guild. For wood feeders, the lower abundances of cellulases and hemicellulases in the P1 than in the P3, along with the huge number of bacterial cells, indicate that the P3 is the primary site for digestion of complex carbohydrates. In the grass feeder, *Cornitermes* sp., cellulase and hemicellulase abundances were nearly identical between its P1 and its P3, a pattern that is mirrored in its 16S rRNA gene copies. This indicates that for this termite, carbohydrate digestion is not restricted to one gut compartment. The P3 of wood and grass feeders also contains the highest abundance of beta-lytic metallopeptidases, which were classified to Spirochaetes. These enzymes are involved in cell lysis and may play a role in bacterial nutrition or defense, but they have also been studied as virulence factors in pathogenic *Treponema* species (Seshadri et al., 2004, Bamford et al., 2010, Wu and Chen, 2011). The wood, grass, interface, and soil-feeding termites clearly have distinct bacterial peptidase profiles that they may be using in different ways. The subtilisins in interface and soil feeders may be fermenting small extracellular peptides, while the beta-lytic metallopeptidases in wood and grass feeders may be lysing other bacteria, especially in the cell-rich P3. The digestion of peptides has not been the focus of previous work on wood and grass feeders, but instead was postulated to be a major contributor to host nutrition in soil feeders (Ji and Brune, 2005, Ji and Brune, 2006, Ngugi et al., 2011). This work demonstrates that the gut microbiota of wood feeders are not only digesting complex carbohydrates and the gut microbiota of soil feeders are not only digesting

peptides, but both feeding strategies employ a distinct suite of glycosyl hydrolases and peptidases for the extraction of nutrients.

The microbial community structure and functional profiles of the gut microbiota of higher termites shows complicated patterns related to homologous compartment and feeding guild. Although there are some similarities in bacterial community structure in the P1 and P4 homologous compartments, the P3 shows more distinctions in bacterial community structure and functional profiles that are distinct for wood-, grass-, interface-, and soil-feeding termites. However, these functional profiles only reveal the presence of enzymes that could be used for the digestion of complex carbohydrates and proteins. A metatranscriptomic analysis of the bacterial communities from wood- and soil-feeding termites is currently in progress to further explore the differences in bacterial activity and gene expression related to feeding strategy.

## **Acknowledgements**

This study was supported by the Max Planck Society. We thank David Sillam-Dussès and James Nonoh for providing termites. We thank Kristina Lang for the QPCR analysis. The work conducted by the U.S. Department of Energy Joint Genome Institute is supported by the Office of Science, Biological and Environmental Research Program of the U. S. Department of Energy, and by the University of California, Lawrence Berkeley National Laboratory under Contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under Contract No. DE-AC02-06NA25396.

## References

- Bamford CV, Francescutti T, Cameron CE, Jenkinson HF and Dymock D (2010).** Characterization of a novel family of fibronectin-binding proteins with M23 peptidase domains from *Treponema denticola*. *Mol. Oral Microbiol.* 25:369–383.
- Bignell DE and Eggleton P (1995).** On the elevated intestinal pH of higher termites (Isoptera, Termitidae). *Insect. Soc.* 42:57–69.
- Bignell DE and Eggleton P (2000).** Termites in ecosystems, pp. 363–387. *In* Abe T, Bignell DE and Higashi M. (eds.), *Termites: Evolution, Sociality, Symbioses, Ecology*, vol. 1. Kluwer Academic Publisher, Dordrecht.
- Brune A (2014) Symbiotic digestion of lignocellulose in termite guts. *Nat. Rev. Microbiol.* 12: 168–180.
- Brune A and Ohkuma M (2011).** Role of the termite gut microbiota in symbiotic digestion, pp. 439–475. *In* Bignell DE, Roisin Y and Lo N (eds), *Biology of Termites: A Modern Synthesis*, Springer, Dordrecht.
- Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Lozupone CA, Turnbaugh PJ, Fierer N and Knight R (2011)** Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci. USA* 108: 4516–4522.
- Caro-Quintero A, Ritalahti KM, Cusick KD, Loeffler FE and Konstantinidis KT (2012).** The Chimeric Genome of *Sphaerochaeta*: Nonspiral Spirochetes That Break with the Prevalent Dogma in Spirochete Biology. *Mbio* 3:e00025–12
- Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J and Punta M. (2014).** Pfam: the protein families database. *Nucl. Res.* 42: D222–D230.
- Fitzpatrick RE, Wijeyewickrema LC and Pike RN (2009).** The gingipains: scissors and glue of the periodontal pathogen, *Porphyromonas gingivalis*. *Fut. Microbiol.* 4: 471–487.
- Fujita A and Abe T (2002).** Amino acid concentration and distribution of lysozyme and protease activities in the guts of higher termites. *Physiol. Entom.* 27: 76–78.
- Griffiths BS, Bracewell JM, Robertson GW and Bignell DE (2013).** Pyrolysis-mass spectrometry confirms enrichment of lignin in the faeces of a wood-feeding termite, *Zootermopsis nevadensis* and depletion of peptides in a soil-feeder, *Cubitermes ugandensis*. *Soil Biology and Biochemistry* 57: 957–959.
- Gupta R, Beg QK and Lorenz P (2002).** Bacterial alkaline proteases: molecular approaches and industrial applications. *Appl. Microbiol. Biotechnol.* 59: 15–32.
- Hammer O, Harper DAT, Ryan PD (2001).** PAST: paleontological statistics software package for education and data analysis. *Paleontologica Electronica* 4:9.
- He S, Ivanova N, Kirton E, Allgaier M, Bergin C, Scheffrahn RH, Kyrpides NC, Warnecke F, Tringe SG and Hugenholtz P (2013).** comparative metagenomic and metatranscriptomic analysis of hindgut paunch microbiota in wood- and dung-feeding higher termites. *Plos One* 8:e61126
- Hongoh Y (2011).** Toward the functional analysis of uncultivable, symbiotic microorganisms in the termite gut. *Cellular and Molecular Life Sciences* 68: 1311–1325.
- Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, Bernard T, Binns D, Bork P, Burge S, de Castro E, Coggill P, Corbett M, Das U, Daugherty L, Duquenne L, Finn RD, Fraser M, Gough J, Haft D, Hulo N, Kahn D, Kelly E, Letunic I, Lonsdale D, Lopez R, Madera M, Maslen J, McAnulla C, McDowall J, McMenamin C, Mi H, Mutowo-Mueller P, Mulder N, Natale D, Orengo C, Pesseat S, Punta M, Quinn AF, Rivoire C, Sangrador-Vegas A, Selengut JD, Sigrist CJ, Scheremetjew M, Tate J, Thimmajananthan M, Thomas PD, Wu CH, Yeats C and Yong SY (2012).** InterPro in 2011: new developments in the family and domain prediction database. *Nucl. Res.* 40: D306–D312.
- Huntington GB (1997).** Starch utilization by ruminants: From basics to the bunk. *J. Anim. Sci.* 75: 852–867.
- Ji R and Brune A (2005).** Digestion of peptidic residues in humic substances by an alkali-stable and humic-acid-tolerant proteolytic activity in the gut of soil-feeding termites. *Soil Biol. Biochem.* 37: 1648–1655.

- Ji R and Brune A (2006).** Nitrogen mineralization, ammonia accumulation, and emission of gaseous NH<sub>3</sub> by soil-feeding termites. *Biogeochem.* 78:267–283.
- Koehler T, Dietrich C, Scheffrahn RH and Brune A (2012).** High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation of the gut in wood-feeding higher termites (*Nasutitermes* spp.). *Appl. Environ. Microbiol.* 78:4691–4701.
- Li N and Collyer CA (2011).** Gingipains from *Porphyromonas gingivalis* - Complex domain structures confer diverse functions. *Eur. J. Microbiol. Immunol.* 1: 41–58.
- Li R, Zhu H, Ruan J, et al. (2010).** De novo assembly of human genomes with massively parallel short read sequencing. *Gen. Res.* 20: 265–272.
- Liu N, Zhang L, Zhou H, Zhang M, Yan X, Wang Q, Long Y, Xie L, Wang S, Huang Y, Zhou Z (2013).** Metagenomic Insights into Metabolic Capacities of the Gut Microbiota in a Fungus-Cultivating Termite (*Odontotermes yunnanensis*). *Plos One* 8:e69184.
- Lombard V, Ramulu HG, Drula E, Coutinho PM and Henrissat B (2014).** The carbohydrate-active enzymes database (CAZy) in 2013. *Nucl. Res.* 42: D490–D495.
- Markowitz VM, Chen IM, Chu K, Szeto E, Palaniappan K, Pillay M, Ratner A, Huang J, Pagani I, Tringe S, Huntemann M, Billis K, Varghese N, Tennessen K, Mavromatis K, Pati A, Ivanova NN and Kyrpides NC (2014).** IMG/M 4 version of the integrated metagenome comparative analysis system. *Nucl. Res.* 42: D568–D573.
- Mikaelyan A, Strassert JFH, Tokuda G and Brune A (2014).** The fibre-associated cellulolytic bacterial community in the hindgut of wood-feeding higher termites (*Nasutitermes* spp.). *Environ. Microbiol.* 16: 2711–2722.
- Ngugi DK, Ji R and Brune A (2011).** Nitrogen mineralization, denitrification, and nitrate ammonification by soil-feeding termites: a <sup>15</sup>N-based approach. *Biogeochem.* 103: 355–369.
- Ni J and Tokuda G (2013).** Lignocellulose-degrading enzymes from termites and their symbiotic microbiota. *Biotech. Adv.* 31: 838–850.
- Olsen I and Potempa J (2014).** Strategies for the inhibition of gingipains for the potential treatment of periodontitis and associated systemic diseases. *J. Oral Microbiol.* 6: doi:http://dx.doi.org/10.3402/jom.v6.24800.
- Pandey A, Nigam P, Soccol CR, Soccol VT, Singh D and Mohan R (2000).** Advances in microbial amylases. *Biotech. Appl. Biochem.* 31: 135–152.
- Parks DH, Tyson GW, Hugenholtz P and Beiko RG (2014).** STAMP: statistical analysis of taxonomic and functional profiles. *Bioinform.* 30: 3123–3124.
- Paul K, Nonoh JO, Mikulski L and Brune A (2012).** "Methanoplasmatales," Thermoplasmatales-Related Archaea in Termite Guts and Other Environments, Are the Seventh Order of Methanogens. *Appl. Environ. Microbiol.* 78: 8245–8253.
- Petersen TN, Brunak S, von Heijne G and Nielsen H (2011).** SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat. Meth.* 8: 785–786.
- Poulsen M, Hu H, Li C, Chen Z, Xu L, Otani S, Nygaard S, Nobre T, Klaubauf S, Schindler PM, Hauser F, Pan H, Yang Z, Sonnenberg A9, de Beer ZW, Zhang Y, Wingfield MJ, Grimmelikhuijzen CJ, de Vries RP, Korb J, Aanen DK, Wang J, Boomsma JJ, Zhang G (2014).** Complementary symbiont contributions to plant decomposition in a fungus-farming termite. *Proc. Natl. Acad. Sci. USA* 111: 14500–14505.
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J and Gloeckner FO (2013).** The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucl. Res.* 41: D590–D596.
- Raes J, Korb JO, Lercher MJ, von Mering C and Bork P (2007).** Prediction of effective genome size in metagenomic samples. *Gen. Biol.* 8: R10–R10.
- Rana N, Walia A and Gaur A (2013).** alpha-Amylases from Microbial Sources and Its Potential Applications in Various Industries. *Nat. Acad. Sci. Lett.* 36: 9–17.
- Rawlings ND, Waller M, Barrett AJ and Bateman A (2014).** MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucl. Res.* 42: D503–D509.

- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ and Weber CF (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75: 7537–7541.
- Schmitt-Wagner D, Friedrich MW, Wagner B and Brune A (2003). Phylogenetic diversity, abundance, and axial distribution of bacteria in the intestinal tract of two soil-feeding termites (*Cubitermes* spp.). *Appl. Environ. Microbiol.* 69: 6007–6017.
- Seshadri R, Myers GS, Tettelin H, Eisen JA, Heidelberg JF, Dodson RJ, Davidsen TM, DeBoy RT, Fouts DE, Haft DH, Selengut J, Ren Q, Brinkac LM, Madupu R, Kolonay J, Durkin SA, Daugherty SC, Shetty J, Shvartsbeyn A, Gebregeorgis E, Geer K, Tsegaye G, Malek J, Ayodeji B, Shatsman S, McLeod MP, Smajs D, Howell JK, Pal S, Amin A, Vashisth P, McNeill TZ, Xiang Q, Sodergren E, Baca E, Weinstock GM, Norris SJ, Fraser CM and Paulsen IT (2004). Comparison of the genome of the oral pathogen *Treponema denticola* with other spirochete genomes. *Proc. Natl. Acad. Sci. USA* 101: 5646–5651.
- Sorokin DY, Gumerov VM, Rakitin AL, Beletsky AV, Damste JSS, Muyzer G, Mardanov AV and Ravin NV (2014). Genome analysis of *Chitinivibrio alkaliphilus* gen. nov., sp. nov., a novel extremely haloalkaliphilic anaerobic chitinolytic bacterium from the candidate phylum Termite Group 3. *Environ. Microbiol.* 16: 1549–1565.
- Suen G, Weimer PJ, Stevenson DM, Aylward FO, Boyum J, Deneke J, Drinkwater C, Ivanova NN, Mikhailova N, Chertkov O, Goodwin LA, Currie CR, Mead D and Brumm PJ (2011). The Complete Genome Sequence of *Fibrobacter succinogenes* S85 Reveals a Cellulolytic and Metabolic Specialist. *Plos One* 6:e18814
- Thongaram T, Hongoh Y, Kosono S, Ohkuma M, Trakulnaleamsai S, Noparatnaraporn N and Kudo T (2005). Comparison of bacterial communities in the alkaline gut segment among various species of higher termites. *Extrem.* 9: 229–238.
- Warnecke F, Luginbühl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, Cayouette M, McHardy AC, Djordjevic G, Aboushadi N, Sorek R, Tringe SG, Podar M, Martin HG, Kunin V, Dalevi D, Madejska J, Kirton E, Platt D, Szeto E, Salamov A, Barry K, Mikhailova N, Kyrpides NC, Matson EG, Ottesen EA, Zhang X, Hernández M, Murillo C, Acosta LG, Rigoutsos I, Tamayo G, Green BD, Chang C, Rubin EM, Mathur EJ, Robertson DE, Hugenholtz P and Leadbetter JR (2007). Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nat.* 450:560–565.
- Wohl DL, Arora S and Gladstone JR (2004). Functional redundancy supports biodiversity and ecosystem function in a closed and constant environment. *Ecol.* 85: 1534–1540.
- Wu JW and Chen XL (2011). Extracellular metalloproteases from bacteria. *Appl. Microbiol. Biotech.* 92: 253–262.



## Supplementary Material

**Tab. S1 | Phylum-level classification of the bacterial community. Analysis strategies used: gTAG (amplicon of V3-V4), iTAG (amplicon of V4), 16S (16S rRNA genes from metagenomes), and All\_CDS (all protein coding sequences from metagenomes).** Termite hosts are abbreviated as follows: Nc, *N. corniger*; Mp, *M. parvus*; Co, *Cornitermes* sp.; Cu, *C. ugandensis*; Th, *T. hospes*; Nt, *N. taracua*. Missing values from All\_CDS occur when the phyla lacks a representative with a fully sequenced genome. Nc\_P4 lacked sufficient DNA to perform a gTAG analysis.

**Please download from:** [https://www.dropbox.com/s/ki99mmtflnm1s2i/Chpater\\_7\\_TAB\\_S1.xlsx?dl=0](https://www.dropbox.com/s/ki99mmtflnm1s2i/Chpater_7_TAB_S1.xlsx?dl=0).

**Tab. S2 | SCG based normalization factors for metagenomes.** Determined by mean value of full-length single-copy conserved bacterial genes. Genes were identified by COG ID. Copy numbers were calculated by dividing the bp in the assembled and unassembled fractions by the consensus length of the gene, then summing to get the total per metagenome.

**Please download from:** [https://www.dropbox.com/s/l5d5a5pfdabv0g3/Chpater\\_7\\_TAB\\_S2.xlsx?dl=0](https://www.dropbox.com/s/l5d5a5pfdabv0g3/Chpater_7_TAB_S2.xlsx?dl=0).

**Tab. S3 | Glycosyl hydrolase genes per average bacterial genome.** Estimated by dividing gene copy numbers of a glycosyl hydrolase family by each metagenome's normalization factor. Gene copy numbers were calculated by summing the bp of each gene family from the assembled and unassembled fractions, then dividing by the consensus length.

**Please download from:** [https://www.dropbox.com/s/2loecbcb7lz3im2/Chpater\\_7\\_TAB\\_S3.xlsx?dl=0](https://www.dropbox.com/s/2loecbcb7lz3im2/Chpater_7_TAB_S3.xlsx?dl=0).

**Tab. S4 | Peptidase genes per bacterial average genome.** Estimated by dividing gene copy numbers of a peptidase family by each metagenome's normalization factor. Gene copy numbers were calculated by summing the bp of each gene family from the assembled and unassembled fractions, then dividing by the consensus length. Only families that were significantly different between feeding guilds, gut sections, or termite hosts were checked for potential secretion. Families were considered to be putatively secreted based on previous research or the presence of a signal peptide.

**Please download from:** [https://www.dropbox.com/s/lbyye36c0bkrnpm/Chpater\\_7\\_TAB\\_S4.xlsx?dl=0](https://www.dropbox.com/s/lbyye36c0bkrnpm/Chpater_7_TAB_S4.xlsx?dl=0).



## Genome sequences of *Fibrobacteres* and TG3 obtained by marker-assisted compositional binning of multiple metagenomes

Carsten Dietrich, Karen Rossmassler and Andreas Brune

*Affiliations:* Max Planck Institute for Terrestrial Microbiology, 35043 Marburg, Germany | *This Manuscript is in preparation.* | *Contributions:* C.D. conceived the study, carried out all experiments, analyzed data, and wrote the manuscript. K.R. contributed data and discussed the results; A.B. discussed the results and secured funding.

### Abstract

---

The termite gut is a complex ecosystem densely populated with a microbiota consisting of exclusive and deep-branching lineages. The taxonomy of these mostly uncultured symbionts has been revealed using culture-independent methods based on the 16S rRNA gene, and their ecophysiology is often predicted from properties of their closest cultured relatives, which is dangerous since genomes already substantially differ at the species level. Recent studies employing metagenomics to estimate total ecosystem functions and to link functions to microorganisms perform well when most microbial members are well represented in the reference database, but the termite gut contains two poorly studied but highly abundant phylum-level groups: *Fibrobacteres*, with one available genome sequence; and *candidate division* TG3, with only a draft genome and therefore excluded from metagenomic classification databases. Here, we present a reference-independent but phylogenetic-marker-assisted binning method to obtain population genomes of termite gut members of the *Fibrobacteres* and *cand. div.* TG3 with an average contamination of only 2%. Genomic analyses predicted that both phylum-level groups act complementarily in wood degradation but employ different strategies to ferment the products of wood hydrolysis. Furthermore, population genomes from this study harbor genes encoding an electron transport chain with high affinity for oxygen and a full pathway for nitrogen fixation — two adaptations to the unique termite gut environment. This binning method will be beneficial in unraveling the function of uncultured microorganisms in many ecosystems.

## Introduction

Termites harbor a dense and diverse microbiota in their guts (Brune, 2014 and reference therein), up to 95% of which are exclusively found in this ecosystem. The symbionts seem to coevolve with their termite hosts (Hongoh et al., 2005). Such an evolutionary interplay indicates the reciprocal dependence of each partner on the other; the host provides a stable environment and continuous substrate supply, and the microorganisms in return provide essential compounds or ecosystem services (Brune and Dietrich, in press). Although most functions of the microbiota are reported to be conserved, the microbial players change depending on diet and/or host phylogeny (Dietrich et al., 2014).

In higher wood-feeding termites, members of a putative new bacterial division, *candidate division* termite group 3 (TG3), have been detected in high abundance (Hongoh et al., 2005; Hongoh et al., 2006; Köhler et al., 2012; Dietrich et al., 2014 – chapter 3 of this thesis) and appear to reside in the hindgut (Hongoh et al., 2005; Köhler et al., 2012; Mikaelyan et al., 2014). Mikaelyan et al. (2014) showed that in *Nasutitermes corniger*, members of *cand. div.* TG3 are associated with wood fibers, along with members of the *Fibrobacteres* and *Treponema* cluster. The members of these bacterial guilds are likely involved in wood hydrolysis as high cellulose hydrolysis activity has been measured around the wood fibers.

The results of Mikaelyan et al. (2014) provided the first insights into the functional role of the fiber-associated bacteria. Further ecophysiological characterization of these bacteria cannot be obtained from isolates as culturing of termite gut bacteria has been mostly unsuccessful. A common approach around this problem is to predict ecophysiological properties based on those of the closest cultured relatives, but this is dangerous because prokaryotic lineages in the termite gut are usually deep branching and only far related to standard organisms. The physiological meaning of deep branching is not clarified but it might indicate a high specialization to the termite gut environment. Such adaptations are likely due to horizontal gene transfer that occurs especially in densely populated microbial ecosystems, such as guts, and results in species with genomes that very different from closely related organisms (Cordero and Polz, 2014).

Culture-independent metagenomic analyses have enhanced our understanding of bacterial diversity. The ground-breaking, first termite gut metagenome study (Warnecke et al., 2007) implicated both *Spirochaetes* and *Fibrobacteres* in wood degradation. More recent studies mostly focused on the differential abundance of glycosyl hydrolases (Liu et al., 2013; He et al., 2013; Poulsen et al., 2014). Such approaches identify the gene but possibly not the identity of the microorganism carrying the gene if this group is not well represented in public databases. This issue became clear in studies of the first genomes of the two phylum-level groups *Fibrobacteres* (Suen et al., 2011) and *cand. div.* TG3 (Sorokin et al., 2014) in which *blastp*-based analysis of the protein-coding genes yielded top hits within different phyla. This problem was first enumerated for the termite gut by Rossmassler et al. (Chapter 7 of this thesis), who found a huge discrepancy between the estimated abundance of the 16S rRNA gene of *Fibrobacteres* and *cand. div.* TG3 and that of their respective protein-coding genes. *Fibrobacter succinogenesis* from the rumen (Suen et al., 2011) is included in the genome section of the RefSeq database, which is commonly used for *blastp*-based metagenomic analysis. *Chitinivibrio alkaliphilus*, the first isolate of *cand. div.* TG3 from a hypersaline soda lake (Sorokin et al., 2014), is excluded from common metagenomic classification databases because of its draft genome status, but even the inclusion of this genome would not compensate for the lack of representation of the genomic variation in this phylum-level group.

Progress in overcoming the problems of metagenomic analyses has been made recently by binning metagenomic data into population genomes, but most of the automated binning tools still struggle with highly complex metagenomes (e.g., Bragg and Tyson, 2014). Many improved binning methods rely either on composition (Laczny et al., 2014; Wu et al., 2014; Kang et al., 2014) or on both composition and (differential) coverage (Albertsen et al., 2013; Wu et al., 2014; Kang et al., 2014). Some of the software used is completely automated; however, automation also assumes theoretical models that might not be generally applicable since usually only the culturable fraction is available in these databases used for the formulating of theoretical models — a major drawback with metagenomes of high complexity consisting of hitherto non-sequenced microbial guilds. Laczny et al. (2014) introduced semiautomatic and human-augmented binning of metagenomic fragments using centered log-ratio-transformed k-mer frequencies, followed by two-dimensional embedding using Barnes-Hut stochastic neighbor embedding (BH-SNE) as described in (van der Maaten, 2014). This semiautomatic approach is promising, especially because of its independence from reference databases and low number of theoretical assumptions. Recently, we demonstrated that the reconstruction of complete termite mitochondrial genomes from whole metagenomes of termite gut sections is feasible (Dietrich and Brune, 2015; Appendix A). Here, we employ phylogenetic marker assisted BH-SNE based compositional binning of metagenomic sequences to retrieve 33 high-quality population genomes of the poorly studied phylum-level groups *Fibrobacteres* and *cand. div.* TG3 from the termite gut with an average rate of contamination of only 2%.

## Materials and Methods

**Data sets.** The data sets from Rossmassler et al. (Chapter 7 of this thesis) were downloaded from the Joint Genome Institute (JGI) as tarball files (for accession numbers, see Tab. 1). We used the xml tree of the respective folders to download the archives using the command line tool *curl*. The downloaded archives contained all necessary files, e.g., nucleotide sequences of the contigs, their open reading frames (ORFs), and coverage. Basic annotation files were also included: general feature format (GFF) files, gene product (GP) names, clusters of orthologous groups (COGs), KEGG orthology (KO), and phylogenetic distribution (pd) files.

**Genome classification.** To screen for *Fibrobacteres* and *cand. div.* TG3 genomes present in the different metagenomes from Rossmassler et al. (Chapter 7 of this thesis), we used the *AMPHORA2* pipeline (Wu and Scott, 2012). Initially, we analyzed the *Chitinivibrio alkaliphilus* draft genome (Sorokin et al., 2014), which resulted in the identification of all 31 marker genes of Wu and Scott (2012). Subsequent phylogenetic analysis yielded a poor classification success (see Results section). Therefore, we created maximum-likelihood phylogenetic trees (Wu and Scott, 2012) based on the data supplied by the *AMPHORA2* pipeline and the additional 31 identified marker genes of the *Chitinivibrio alkaliphilus* draft genome by keeping the original parameters of *AMPHORA2* tree building. Briefly, we employed *RAxML* v8.1.3 (Stamatakis, 2014) with the protein- $\Gamma$  model and the Whelan and Goldman (WAG) substitution matrix. The resulting database consisting of the updated alignments, and their respective filters as well as the updated trees and hidden Markov models were used to screen the metagenomes of Rossmassler et al. (chapter 7 of this thesis) with the *AMPHORA2* pipeline. First, the marker genes were identified in a hidden Markov model search against the translated ORFs of the metagenomes using the *hmmsearch* binary of *hmmer* (Eddy, 2011) with the recommended *E-value* threshold of  $10^{-7}$ . Next, positively identified sequences were aligned against the seed alignment of the respective hidden Markov models. Sequences were then filtered with the supplied probabilistic filters (Wu et al., 2012) and trimmed with the reference sequences from the respective marker protein tree. The resulting sequences of the 31 markers identified in the metagenomes were placed under maximum likelihood into the phylogenetic trees using the evolutionary placement algorithm (Berger et al., 2011) implemented in *RAxML*. Subsequently, the resulting jplace files were evaluated by the *AMPHORA2* phylotyping script, which involves a step determining the least-common ancestor. Placements with less than 90% bootstrap support were considered as “unclassified”.

**Classification of the 16S rRNA genes.** For analysis of 16S rRNA genes, we used the GFF file of the metagenomes to identify the contigs carrying a 16S rRNA gene (originally identified by *RNAmmer*; Lagesen et al., 2007). The coordinates and sense of these genes on the contigs were used to extract the pure 16S rRNA gene sequences with *subseq* and *reverseComplement* commands of *R* package *Biostrings* (Pages et al., 2010). The extracted sequences were classified with the Naïve Bayesian Classifier implemented in *mothur* (Schloss et al., 2009) in conjunction with *DictDB*, a reference database optimized for the gut microbiota of *Dictyoptera* (Dietrich et al., 2014; Chapter 3 of this thesis). The resulting taxonomy was evaluated on all levels. Only taxonomic levels showing a minimal bootstrap support of 90% were considered as classified.

**Tab. 1 | Statistics of binning-relevant parameters of metagenomes of termites with a considerable number of genomes of the phyla *Fibrobacteres* and *can.* div. TG3.**

Termite gut section	IMG ID <sup>a</sup>	No contig length cut-off				2 kbp contig length cut-off				Perpl. <sup>d</sup>	Contigs binned
		Size (10 <sup>9</sup> bp)		Contigs (10 <sup>3</sup> )		Size (10 <sup>9</sup> bp)		Contigs (10 <sup>3</sup> )			
		w <sup>b</sup>	u <sup>c</sup>	w <sup>b</sup>	u <sup>c</sup>	w <sup>b</sup>	u <sup>c</sup>	w <sup>b</sup>	u <sup>c</sup>		
<b><i>Cornitermes</i> sp.</b>											
P1	2552	42.8	1.5	48,296	2,690	17.5	0.25	2,767	47	41	4,426
P3	2450	42.0	1.3	46,379	2,528	19.3	0.23	2,944	44	34	2,987
P4	2834	29.9	1.3	64,193	2,921	3.9	0.05	692	13	9	1,105
<b><i>Microcerotermes parvus</i></b>											
P1	2507	46.5	1.5	98,485	3,202	4.8	0.14	1,604	46	3	490
P3	2449	41.1	0.7	38,902	1,016	18.7	0.22	4,471	54	28	4,259
P4	2509	46.7	1.5	88,446	2,999	8.6	0.17	1,979	52	8	576
<b><i>Nasutitermes corniger</i></b>											
P1	2238	44.4	1.4	78,205	2,546	7.3	0.21	2,233	66	4	0
P3	2119	44.3	0.6	39,254	895	21.8	0.21	4,724	47	34	6,521
P4	2308	40.6	1.6	73,228	3,061	7.7	0.20	1,938	61	11	3,078

<sup>a</sup> IMG ID, taxon object identifier of the raw read data in the Integrated Microbial Genomes database (330000xxxx; <http://img.jgi.doe.gov/>).

<sup>b</sup> Weighted analysis by considering contig coverage levels.

<sup>c</sup> Unweighted analysis by not considering contig coverage levels.

<sup>d</sup> Perplexity estimate derived from a single-copy-based estimation of Shannon entropy. Shannon entropy can be used estimate perplexity in the Barnes-Hut stochastic neighbor embedding (van der Maaten, 2014)

**Binning of contigs into population genomes.** We binned contigs based on composition. The best results were obtained using the semi-automated approach of Laczny et al. (2014). Contigs with a minimum size of 2 kbp were selected since this cut-off provides an optimal resolution for k-mer-based binning with a low k value (Laczny et al., 2014; Kang et al., 2014). Tetramer frequencies were obtained with the *oligonucleotideFrequency* function of the R package *Biostrings* (Pages et al., 2010). Reverse complements (a correction was made for palindromes) were summed up, which resulted in 136 distinct dimensions. If applicable, we subtracted symmetric tetramers counts originating from 16S rRNA genes because it is often difficult to bin contigs containing these genes (Laczny et al., 2014). Tetramer profiles were normalized by the total sum method and were then incremented by a pseudo-count smaller than the smallest value (here tenfold) in the normalized tetramer frequency data set, as recommended for log-transformation of sparse matrices (Costea et al., 2014). The addition of the pseudo-count is necessary because it circumvents  $\log(0)$  in the subsequent centered log-ratio (clr) transformation. This procedure is a slight modification of the original protocol, which recommends the addition of 1 to the k-mer profiles prior to normalization. However, the addition of 1 did not result in a symmetric embedding with many clusters, and was outperformed by the addition of the pseudo-count, which is based on the normalized data set (see Results).

Each of the clr-transformed data sets was reduced to 30 dimensions in a principal component analysis. These lower-dimensional data were embedded into two dimensions using Barnes-Hut stochastic neighbor embedding, which is a rapid implementation of t-distributed stochastic neighbor embedding (van der Maaten, 2014). All data were embedded using  $\theta$  of 0.5. The perplexity  $p$  was optimized for each metagenome separately. Perplexity can be estimated by  $2^H$ , where  $H$  is the Shannon entropy. The term  $2^H$  is highly similar to the ecological parameter “effective number of species”. Therefore, we estimated  $H$  from the number of single-copy genes. We conducted a search of 111 single-copy genes against the translated ORFs of the metagenomes. Briefly, we downloaded the hidden Markov models of the single-copy genes proposed by Dupont et al. (2012) and McLean et al. (2013) from the Craig Venter Institute webpage (<http://www.jcvi.org/cgi-bin/tigrfams/index.cgi>) and searched against the metagenome with an approach based on hidden Markov models using the trusted cut-off flag. Positively identified genes were weighted by the coverage level to obtain the Shannon entropy for each gene separately. The median number was taken as final Shannon entropy for the estimation of perplexity for each metagenome (see Tab. 1).

Two-dimensional embeddings were visualized with R software (R Development Core Team, 2008). Contigs containing the phylogenetic markers were also visualized as a guide for extraction of *Fibrobacteres* and *cand. div.* TG3 clusters. If contigs contained multiple markers, a majority vote was carried out on these contigs. Taxonomic affiliation was given when at least 51% of the identified marker genes were classified as concordant taxonomy; in all other cases, the contig was label as “unclassified”. Next, contigs were interactively extracted using in-house R functions that work with R base functions (e.g., *identify*) and the functions of the *alphahull* package (Beatriz and Rodriguez-Casal, 2010); this approach is similar but faster than that of Albertsen et al. (2013). A separate publication on the binning R package is planned.

**Phylogenetic analysis and estimation of contamination.** After binning metagenomic contigs into population genomes, the amount of contamination was estimated. A recent publication suggests that the only means of estimating contamination of genomes are searching for the presence of indicative gene sets or phylogenetically analyzing phylogenetic markers (Parks et al., 2014). Phylogenetic analysis of metagenomes is becoming increasingly popular; in taxonomic profiling, it even substitutes for *blast*-based analysis (Wu and Scott, 2012; Darling et al., 2014). When dealing with new bacterial



divisions, the major problem with reference-tree-based estimations of contamination is the lack of reference sequences. For *Fibrobacteres*, there is only a singleton, and for *cand. div. TG3*, the situation is even worse since the draft genome of *Chitinivibrio alkaliphilus* is not part of common metagenome classification databases. We therefore decided to use the 31 marker genes proposed by Wu and Scott (2012). and trees were built as described above (Genome classification), except that when a marker was identified multiple times in a population genome, sequences were clustered at 90% identity by keeping the longer sequence. Sequences that were too short (i.e., start later than the latest starting reference sequence and/or end before the earliest ending reference sequence) were added under maximum likelihood with the evolutionary placement algorithm (Berger et al., 2011) using the same parameters that were used for the backbone phylogenetic tree. Each tree was evaluated separately. Genes that did not form a monophyletic cluster with either *Fibrobacter succinogenes* or *Chitinivibrio alkaliphilus* were considered as contamination, as were sequences from *Fibrobacteres* population genomes in the *cand. div. TG3* cluster and vice versa.

Finally, high-quality target sequences were concatenated into a super alignment using an in-house *R* script. Only genomes and population genomes showing all 31 genes were used. A tree was inferred from this super alignment using *RAxML* v8.1.3 (Stamatakis, 2014) with the protein- $\Gamma$  model and the Whelan and Goldman (WAG) substitution matrix. Population genomes that did not contain all 31 markers were added using the evolutionary placement algorithm. Final trees were visualized and annotated in the *R* package *APE* (Paradis et al., 2004).

**Genomes from public databases.** We also analyzed seven genomes from other studies: genomes of the bacterial divisions *Fibrobacteres* (*Fibrobacter succinogenes* S85 from rumen; CP00179; Suen et al., 2011) and *cand. div. TG3* (*Chitinivibrio alkaliphilus* ACht1 from a hypersaline soda lake; NZ\_ASJR00000000, draft; Sorokin et al., 2014); population genomes of the cellulolytic *Clostridium cellulolyticum* H10 (NC\_011898), *Clostridium thermocellum* DSM 1313 (NC\_017304), and *Clostridium phytofermentans* ISDg (NC\_010001); and genomes of two non-cellulolytic but hydrogenotrophic acetogenic termite gut *Spirochaetes* as a control: *Treponema primitia* ZAS-2 (NC\_015578) and *Treponema azotonutricium* ZAS-9 (NC\_015577).

**Metabolic potential.** For each population genome, we estimated completeness on the basis of the 111 single-copy genes given in Dupont et al. (2012) and McLean et al. (2013). We also (re)valuated the completeness of *Fibrobacter succinogenes* and *Chitinivibrio alkaliphilus*. Both contained 100 single-copy genes; therefore, we used this count as a basis for an estimation of completeness. We then evaluated single-copy genes occurring multiple times. The total number of identified single-copy genes in each population genome divided by the number of unique single-copy genes was used for an estimation of the number of strains.

For analysis of glycosyl hydrolases, we used the hidden Markov models of *dbCAN* version 3 (Yin et al., 2012). After identification of relevant genes, we tested whether the genes occurred as full-length or partial sequences using the supplied consensus length database file of the *dbCAN*. An *E-value* cut-off of  $10^{-3}$  or  $10^{-5}$  was applied for full-length and partial sequences, respectively, as recommended (Yin et al., 2012). Glycosyl hydrolase families were classified into groups of similar function (e.g., cellulases) according to (Allgaier et al., 2010). Additionally, the class “lignin-degrading’ enzymes” was included according to Lévassieur et al. (2013). Positively identified genes were normalized by the number strains and estimation of completeness.

For the analysis of the metabolic potential, we used the existing KEGG and COG annotations. An initial overview was obtained using the iPath explorer (Yamada et al., 2011). Pathway-

specific analyses were carried out using the KEGG mapping function of the KEGG database (<http://www.genome.jp/kegg/>). If a particular gene in a pathway was missing, we checked whether the gene was present in the COG annotation. If the gene was not present, we downloaded the respective TIGRFAM hidden Markov model from the Craig Venter Institute webpage (<http://www.jcvi.org/cgi-bin/tigrfams/index.cgi>) and carried out a hidden Markov model search with the respective model using the trusted cut-off flag.

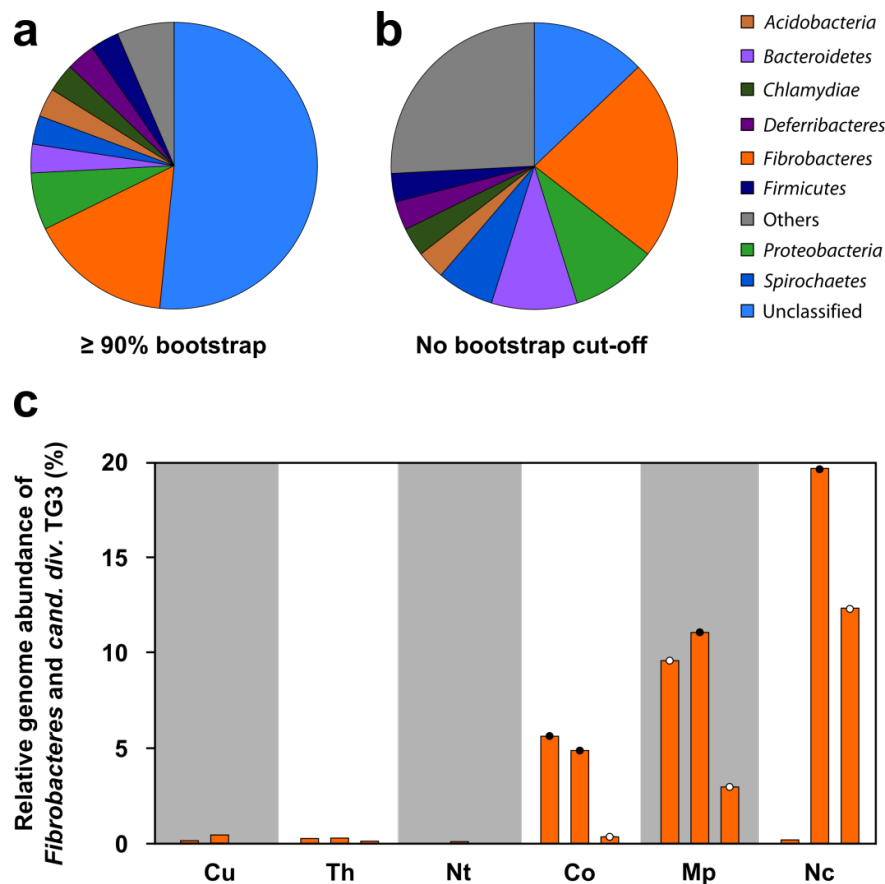
## Results

**Identification of suitable metagenomes.** To date, the only publicly available genomes for the phylum *Fibrobacteres* and *cand. div.* TG3 are *Fibrobacter succinogenes* (Suen et al., 2011) and *Chitinivibrio alkaliphilus* (draft; Sorokin et al., 2014). Despite the high value of both genomes, the genetic diversity of both phylum-level groups is far from covered. Taxonomic assignments based on reference databases in metagenomic analyses of microbial communities containing both of these groups would suffer greatly with such deficiencies. Rossmassler et al. (Chapter 7 of this thesis) showed that the estimation of abundance of especially these two phylum-level groups significantly differ depending on whether the estimation is based on 16S rRNA genes or protein-coding genes. Therefore, we decided to first analyze 31 marker genes of (Wu and Scott, 2012) of *Chitinivibrio alkaliphilus* with a minimal bootstrap support of 90%. Over 50% of these genes could not be classified (Fig. 1a). Most of the genes were assigned as phylum *Fibrobacteres*, which reflects that *Fibrobacter succinogenes* is the closest relative with a genome sequence available in public databases but is very distantly related. When less-stringent conditions were used (no bootstrap cut-off), still the largest fraction of sequences were classified as *Fibrobacteres* at phylum level (Fig. 1b). However, majority of genes were assigned to diverse phyla, which supported the phylum-level position of *cand. div.* TG3.

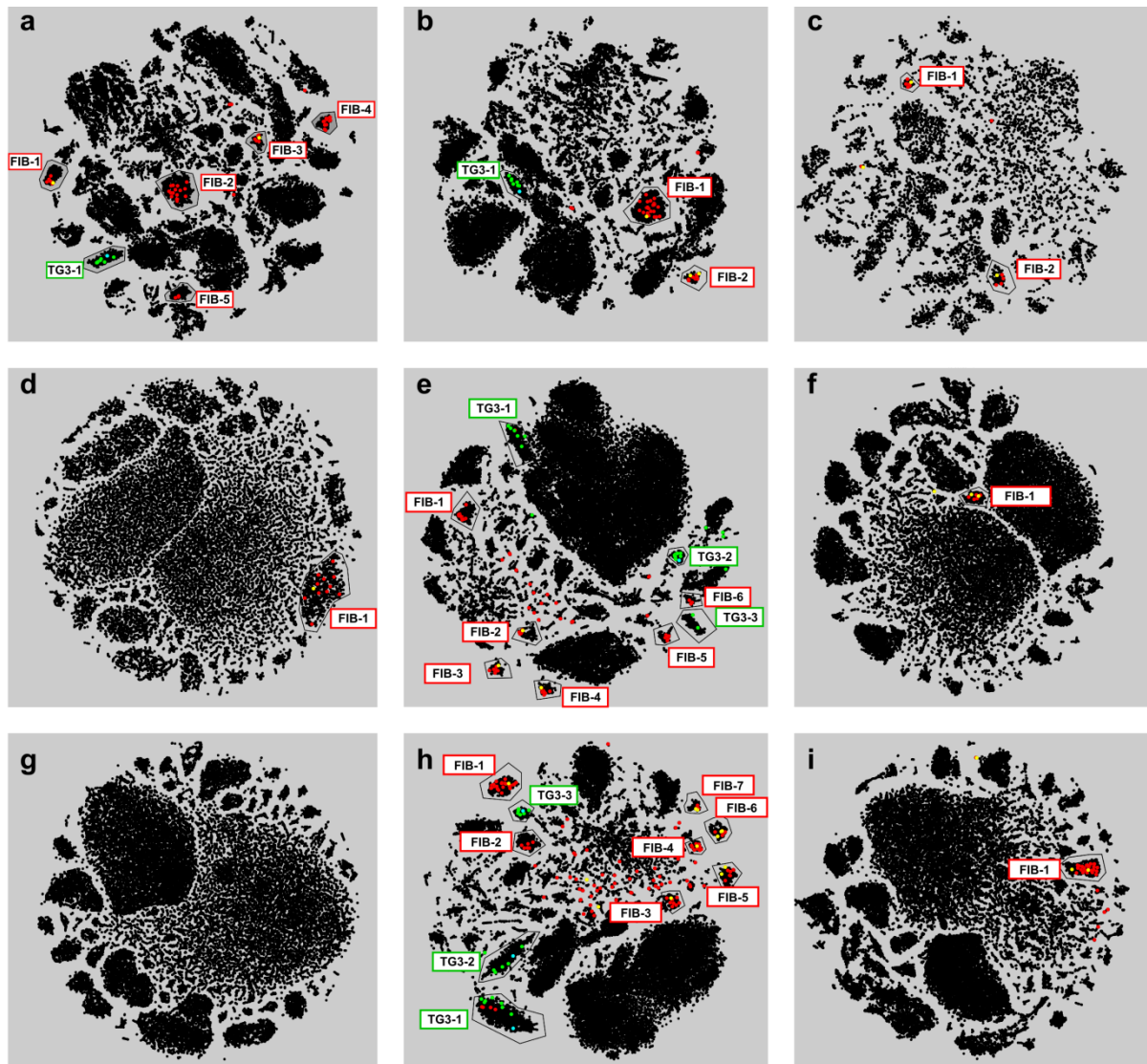
To compensate for the lack of TG3 reference genomes in the metagenomic analysis, we slightly modified the *AMPHORA2* database by adding the respective sequences of *Chitinivibrio alkaliphilus* to the alignment, hidden Markov models, and trees. We added the entry *cand. div.* TG3 at the phylum level to the binary taxonomy database files. We analyzed the metagenomic data sets from Rossmassler et al. (Chapter 7 of this thesis) using the *AMPHORA2* pipeline and this slight modification. Since the marker genes of this pipeline usually occur as single-copy genes, the results can be directly used as a proxy for genome abundance of the phylum *Fibrobacteres* and *cand. div.* TG3. It was evident that both groups are present in high abundance only in the different gut section samples of the wood-feeding or interface-feeding termites *Cornitermes* sp., *Nasutitermes corniger* and *Microcerotermes parvus* (Fig. 1c). The highest abundance of both groups was detected in the P3 compartment of *Nasutitermes corniger*, where approximately 20% of the genomes are classified as *Fibrobacteres* or *cand. div.* TG3. This result contradicts the *blastp*-based protein-coding-gene estimation of Rossmassler et al. (Chapter 7 of this thesis), who found only a very small fraction of genes related to *Fibrobacteres* (8% based on estimation of protein-coding genes, but 25% based on estimation of 16S rRNA genes; for TG3, no protein-coding genes were assigned). In contrast, only less than 1% of the genomes in the gut sections of soil-feeding termites could be classified to the target groups. Therefore, we decided to work only with the gut sections of wood-feeding and interface-feeding termites.

**Optimization of binning.** When we used the default conditions proposed by Laczny et al. (2014; *clr* transformation of pentamer frequencies and a standard perplexity of 30), only a small number of clusters were found in the two-dimensional embedding of the metagenomic contigs. For the sample with the highest estimate of *Fibrobacteres* and *cand. div.* TG3 genomes, i.e., the *Nasutitermes corniger* gut section P3, only two vague clusters containing markers of either *Fibrobacteres* or *cand. div.* TG3 were detected (data not shown). Although we carried out the centered log-ratio transformation, an overall asymmetry of the two-dimensional embedding was obvious. This is an atypical feature of Barnes-Hut stochastic neighbor embeddings (van der Maaten, 2014). Asymmetry and the lack of clustering are indications of either a bad choice of perplexity or insufficient data (e.g., the lack of different features). Since all metagenomic samples contain taxa from diverse phyla, we considered the original data as suitable for binning. We found that perplexity can be approximated using the

Shannon entropy of single-copy genes. Remarkably, also the choice of the right pseudo-count was critical for principal component analysis and Barnes-Hut stochastic neighbor embedding, which is in agreement with Costea et al. (2014). The default pseudo-count is 1, which represents the smallest unit in k-mer counting for contigs. The addition is primarily made to circumvent  $\log(0)$ , a procedure necessary for the clr transformation. But the addition of this pseudo-count is asymmetric. Therefore, we decided to first normalize the total sum and then to add a pseudo-count that is smaller than the smallest value in the data set. This procedure has been proposed by Costea et al. (2014) and strongly enhances clustering on a rational basis. Although this modification appears minor, it had a great impact on the performance of clustering and overall symmetry (Fig. 2).



**Fig. 1 | Phylogenetic analysis of 31 marker genes.** (a, b) Phylogenetic placement at phylum-level of *Chitinivibrio alkaliphilus* with (a) stringent criteria (bootstrap  $\geq 90\%$ ) and (b) loose criteria (no bootstrap cut-off). (c) Combined relative genome abundance of the phylum *Fibrobacteres* and *cand. div. TG3* in metagenomes analyzed in this study. Circles indicate whether 16S rRNA gene sequences classified to the different divisions were also detected (no circle, no *Fibrobacteres* and no *cand. div. TG3* 16S rRNA genes detected; ○ 16S rRNA genes of only the phylum *Fibrobacteres* were detected; ● 16S rRNA genes of both phylum-level groups, *Fibrobacteres* and *cand. div. TG3*, were detected). Cu, *Cubitermes ugandensis*; Th, *Termes hospes*; Nt, *Neocarpitermes taracua*; Co, *Cornitermes* sp.; Mp, *Microcerotermes parvus*; Nc, *Nasutitermes corniger*.



**Fig. 2 | Two-dimensional embeddings of clr-transformed tetranucleotide frequencies of metagenomes that contained considerable amounts of genomes of the phyla *Fibrobacteres* and *cand. div. TG3*.** (a–c) *Cornitermes* sp. gut sections P1, P3, and P4, respectively; (d–f) *Microcerotermes parvus* gut sections P1, P3, and P4, respectively; (g–i) *Nasutitermes corniger* gut sections P1, P3, and P4, respectively. Clusters were chosen by contigs that were positively classified as either *Fibrobacterales* or *Chitinivibrionia* by the evolutionary placement of 31 phylogenetic marker genes. Cluster annotation: first three letters represent the division (FIB, *Fibrobacteres*; TG3, *cand. div. TG3*); number is that assigned to the population in the respective compartment.

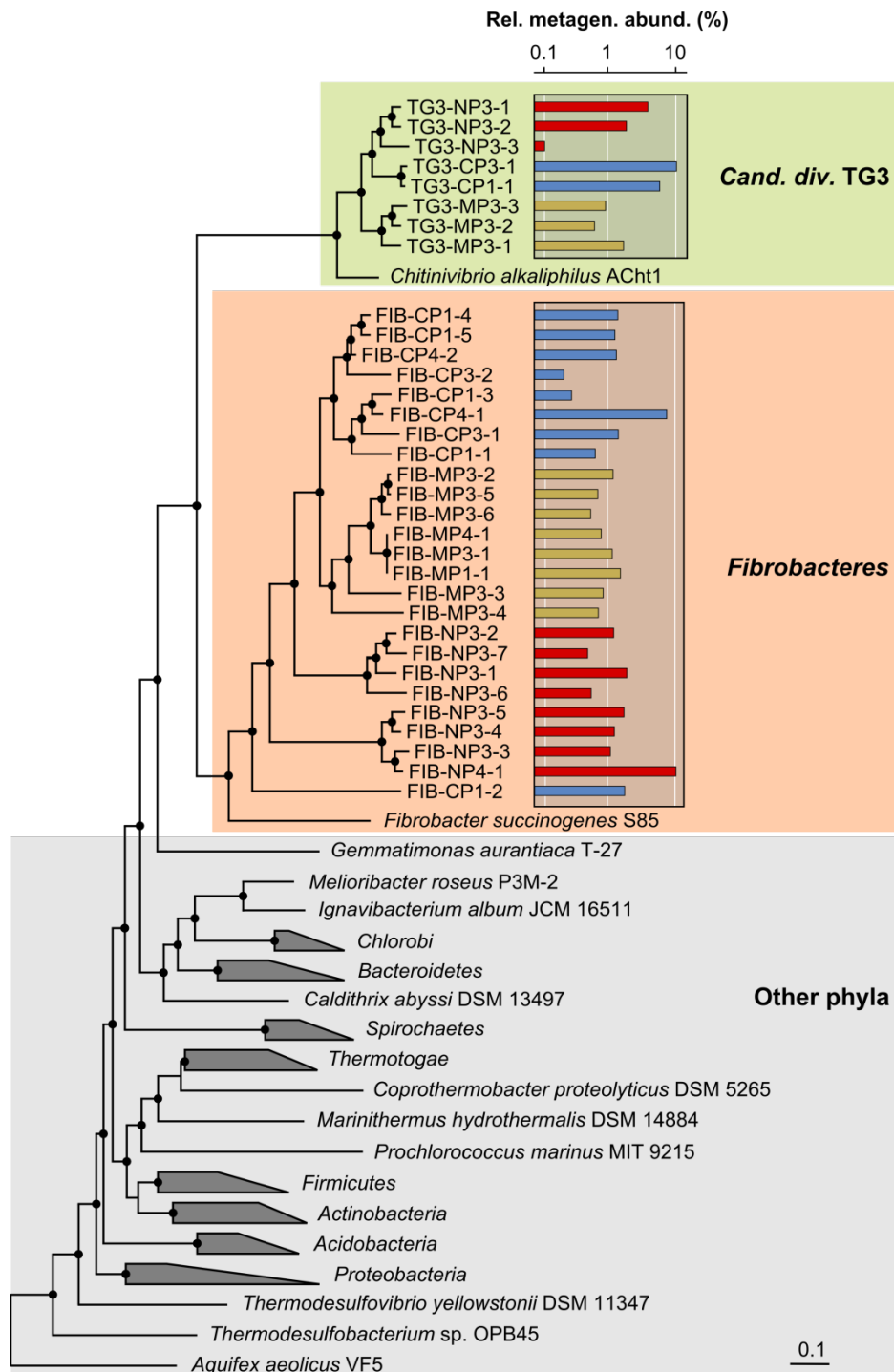
**Binning into population genomes.** After we applied the necessary 2 kbp cut-off, only a very small fraction of contigs remained (0.5–5.4%; Tab. 1), but the absolute number of contigs was high (13,339–66,010). However, 10–50% of the respective metagenome size could be retained since the remaining contigs were long. The fraction of remaining bases was always highest in the P3 compartment of the respective termite host, which indicated that host contamination might play an important role in successful assembly of long contigs and underlines that the number of contigs alone is not a good estimate of metagenome quality. We carried out Barnes-Hut stochastic neighbor embedding on clr-transformed tetramer frequencies of the metagenomic contigs, each with different perplexities (Tab. 1)

and an individual choice of the pseudo-count, which resulted in nine highly symmetric embeddings with many distinct clusters (Fig. 2). When contigs related to *Fibrobacteres* or *cand. div. TG3* were visualized (based on the 16S rRNA gene and 31 elite phylogenetic markers), one to ten clusters containing these contigs could be determined (except for the P1 section of *Nasutitermes corniger*, where no cluster could be identified). We obtained a total of 25 *Fibrobacteres* and 8 *cand. div. TG3* population genomes (Tab. S1). Many of these were more than 90% complete, and some were even 100% complete. However, an estimation of the number of strains in these population genomes showed that these genomes are a composite of one to three strains of each guild. Based on these data, we estimated the genome size of *Fibrobacteres* and *cand. div. TG3* in the termite gut to be  $4.2 \pm 0.9$  Mbp and  $3.7 \pm 0.5$  Mbp, respectively, which indicated that the genome of the two phylum-level groups are similar in size. Many populations showed different relative genome abundances, which might indicate that bins represent different strains or species (Tab. S1; Fig. 3).

**Phylogenetic placement of the population genomes.** After we extracted the population genomes, we asked whether these bins contained also non-target sequences. Since the phylum *Fibrobacteres* and *cand. div. TG3* are each only represented by a single genome, we decided to estimate contamination using a *de novo* phylogenetic analysis of the 31 markers of Wu and Scott (2012). Each marker tree contained 0 to 2 sequences (the sum of sequences from all population genomes) that either could not end up within monophyletic clusters with either *Chitinivibrio alkaliphilus* or *Fibrobacter succinogenes* or could end up in the wrong cluster of these two. The latter scenario did not occur. Using this phylogenetic method, we found 23 of 1,075 non-target sequences from all markers. Therefore, we estimated that the number of non-target sequences in the population genomes of this study were 0–3% (average 2.1%).

Next, we wanted to test the phylogeny of the population genomes. Therefore, we again used the 31 marker genes from genomes obtained in this study and published genomes to produce a concatenated sequence alignment of all filtered 31 marker sequences and subjected it to maximum-likelihood analysis. The resulting phylogenetic tree clearly documents that both phylum-level groups *Fibrobacteres* and *cand. div. TG3* are deep-branching lineages (Fig. 3). Sorokin et al. (2014) were not able to judge whether their isolate represents a new phylum or a new class within the phylum *Fibrobacteres*. Here, we were able to show that they are different clusters. Given the diversity within other established phyla of the bacterial domain (e.g., *Firmicutes* and *Bacteroidetes*), the phylogenetic analysis proved that the two groups are different divisions since the internal phylogenetic variance of *Fibrobacteres* and *cand. div. TG3* is higher than that of all other phyla. Another argument for the phylum-level assignment is the split between *Fibrobacteres* and *cand. div. TG3*. In the literature, *Melioribacter roseus* is considered to be its own phylum (Podosokorskaya et al., 2013). If we set the branching point of this organism as the definition of the phylum level, *Fibrobacteres* and *cand. div. TG3* definitely are two different phyla.

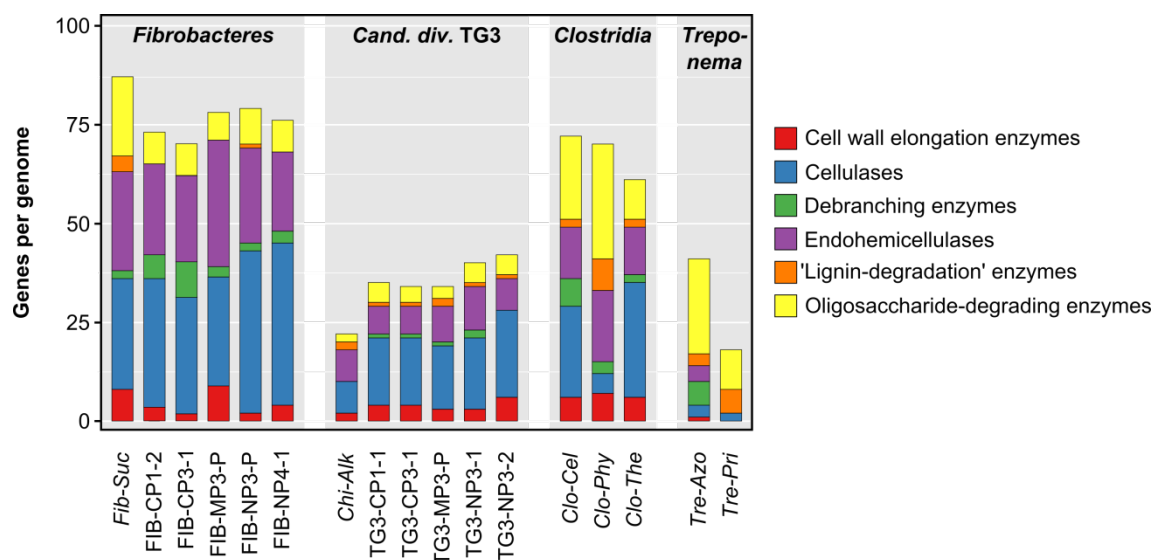
Within the clusters that exclusively consisted of the population genomes of this study, a termite-host-specific clustering was evident, which supports the coevolution findings of Hongoh et al. (2005). The only exception was the population genome FIB-CP1-2, which could be a very basal lineage of termite gut *Fibrobacteres*.



**Fig. 3 | Phylogenetic tree of important bacterial phyla and population genomes from this study inferred by concatenation of 31 phylogenetic markers.** Shadings allow readers to focus on the groups *Fibrobacteres* and *cand. div. TG3*. The relative metagenome abundance is plotted on logarithmic scale. Colors of the bars indicate the termite host species from which the population genome was obtained (red, *Nasutitermes corniger*; blue, *Cornitermes* sp.; yellow, *Microcerotermes parvus*).

**Genes encoding glycosyl hydrolases.** For analysis of genes encoding glycosyl hydrolases, we decided to use only population genomes with a minimum completeness of 80%. The application of this cut-off resulted in five population genomes for each of the phylum-level groups *Fibrobacteres* and *cand. div.* TG3. In certain cases, we used the pooled population genomes from a termite gut section.

When we carried out a search driven by hidden Markov models for genes belonging to glycosyl hydrolase families (GHFs) relevant for cellulose degradation, the number of these genes in the two phylum-level groups differed dramatically (Fig. 4). Population genomes identified as belonging to the phylum *Fibrobacteres* contained 70 to 80 of these genes per genome, whereas bins belonging to *cand. div.* TG3 contained less than the half of this number, with no exceptions. When we compared the *Fibrobacteres* population genomes to the only genome of the phylum *Fibrobacteres*, *Fibrobacter succinogenes*, we found that the type strain genome had a similar number of such genes (87 genes). In contrast, *Chitinivibrio alkaliphilus*, the draft genome and only genome of *cand. div.* TG3, had only 23 GHF-encoding genes related to cellulose degradation. This was expected since the primary substrate of *C. alkaliphilus* is chitin (Sorokin et al., 2014). Population genomes from *cand. div.* TG3 contained about 25% more cellulolysis-relevant genes than *C. alkaliphilus*, but still only about half of the number from *Fibrobacteres*.



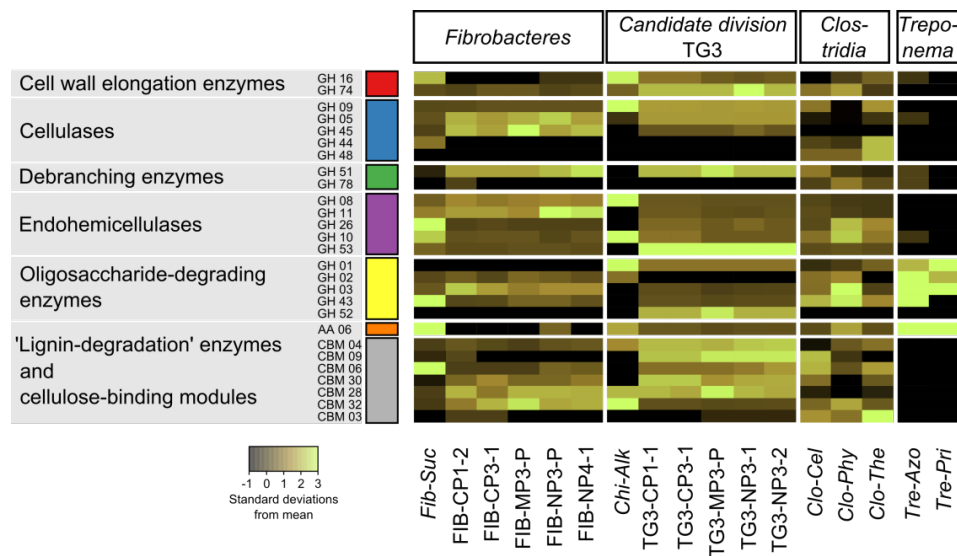
**Fig. 4 | Differences in the content of genes encoding various major classes of glycosyl hydrolases.** Population genomes were normalized by the genome completeness estimation and estimation of strains. Other, cultured, relevant bacteria were included for comparison (*Fib-Suc*, *Fibrobacter succinogenes*; *Chi-Alk*, *Chitinivibrio alkaliphilus*; *Clo-Cel*, *Clostridium cellulolyticum*, *Clo-Phy*, *Clostridium phytofermentans*; *Clo-The*, *Clostridium thermocellum*; *Tre-Azo*, *Treponema azotonutricium*; *Tre-Pri*, *Treponema primitia*).

Compared to other cellulose-degrading bacteria, e.g., *Clostridium cellulolyticum*, members of the *Fibrobacteres* surpassed the cellulolytic potential of these strains. In contrast, *cand. div.* TG3 seemed to be on the lower end of the potential of cellulose degradation. When we classified each GHF into groups of similar functions, a general trend was visible. The genes of the following GHFs were almost exclusively most abundant (in decreasing order): cellulases, hemicellulases, oligosaccharide-degrading enzymes, cell-wall-elongation enzymes, debranching enzymes, and “lignin degradation” enzymes.



After classification, it was obvious that the termite gut population genomes contained only a small number of genes encoding oligosaccharide-degrading enzymes (Fig. 4).

At the level of single GHF-encoding genes, we identified a set of GHF-encoding genes shared between *Fibrobacteres* and *cand. div.* TG3 but also a set of division-specific GHF-encoding genes (Fig. 5). Genes encoding GH 05, 08, 09, 10, 11, 45, and 56 are shared by *Fibrobacteres* and *cand. div.* TG3, but each is highly abundant only in one of the phylum-level groups. For example, genes encoding GH 09, 51, and 59 were more abundant in *cand. div.* TG3, whereas genes encoding GH 03, 05, and 45 were abundant only in *Fibrobacteres*. Furthermore, we found that genes encoding GH 01 and 52 seem to be exclusively encoded by *cand. div.* TG3, which indicated that the two phylum-level groups in termite guts might not be redundant but synergistic in cellulose degradation. Interestingly, all *cand. div.* TG3 bins contained genes that were classified as being involved in 'lignin degradation'. Proteins encoded by all such genes fall into the AA6 family, a family of enzymes with auxiliary activities recently added to the CAZY database (Levasseur et al., 2013); enzymes in this family have at least 1,4-benzoquinone reductase activity. Such activities have recently also been shown in the termite gut bacterium *Treponema primitia*, and proteins encoded by genes of the AA6 family are also active against a variety of other aromatic compounds (Lucey and Leadbetter, 2014). Together, our results and those of Lucey and Leadbetter (2014) indicate that *cand. div.* TG3 might be involved in the degradation or modification of lignin.



**Fig. 5 | Detailed analysis of glycosyl hydrolases encoded by the population genomes obtained in this study and selected relevant bacteria.** Data were normalized by the z.score transformation according to the glycosyl hydrolase families to allow families to be compared. *Fib-Suc*, *Fibrobacter succinogenes*; *Chi-Alk*, *Chitinivibrio alkaliphilus*; *Clo-Cel*, *Clostridium cellulolyticum*, *Clo-Phy*, *Clostridium phytofermentans*; *Clo-The*, *Clostridium thermocellum*; *Tre-Azo*, *Treponema azotonutricium*; *Tre-Pri*, *Treponema primitia*.

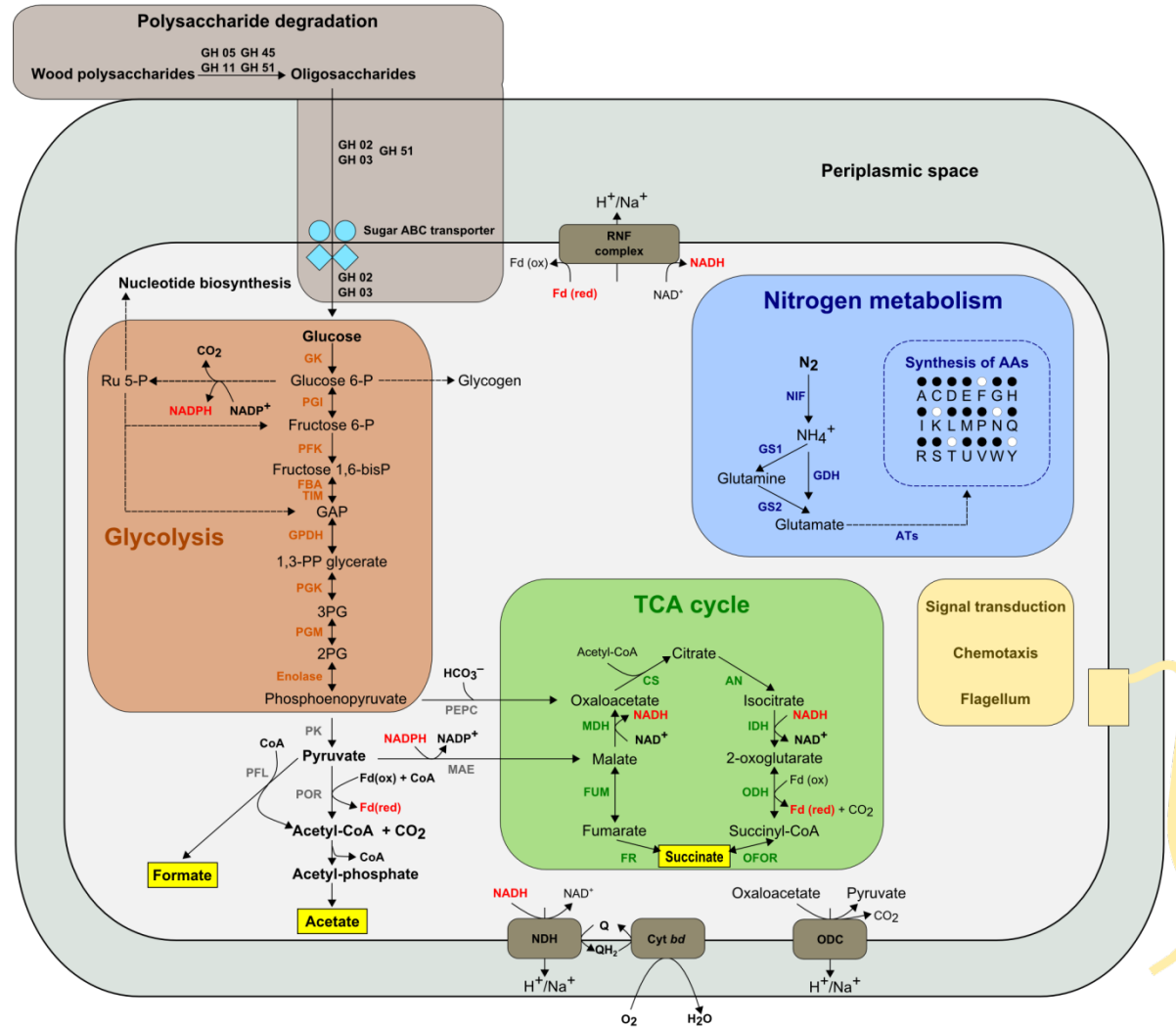
**Potential metabolism of *Fibrobacteres* and *can. div.* TG3.** When we treated the population genomes from this study as regular genomes and annotated them properly using the KEGG and COG systems, we were able to construct the potential metabolism of phylum *Fibrobacteres* and *can. div.* TG3. We focused on pathways involved in the degradation of cellulose and the subsequent pyruvate metabolism as well as on other pathways relevant for ecosystem functioning.

*Fibrobacteres.* Our analysis of *Fibrobacteres* from *Nasutitermes corniger* revealed that these bacteria are most likely motile, with genes for chemotaxis and a complete flagellum with motor proteins. Their potential metabolism is as follows (Fig. 6). The primary substrate, wood, is first degraded extracellularly with enzymes classified as GHFs 05, 11, 45, and 51, which results in wood-derived oligosaccharides. These are then degraded either extracellularly, periplasmatically, or cytoplasmically, mainly catalyzed by GHs 02, 03, and 51. The important sugars xylose and glucose are taken up by ABC transporters. Xylose can enter glycolysis via the pentose phosphate pathway at the level of either fructose 6-phosphate or glyceraldehyde 3-phosphate. Glucose and/or the intermediates of xylose are then converted to pyruvate by standard glycolysis. The two major fermentation products are acetate and formate. Pyruvate is oxidized by pyruvate:ferredoxin oxidoreductase to form acetyl-CoA, which is subsequently converted to acetyl-phosphate by acetyl-CoA:phosphate acetyltransferase, which in turn is converted to acetate by acylphosphatase. Pyruvate formate lyase converts pyruvate and CoA to acetyl-CoA and formate.

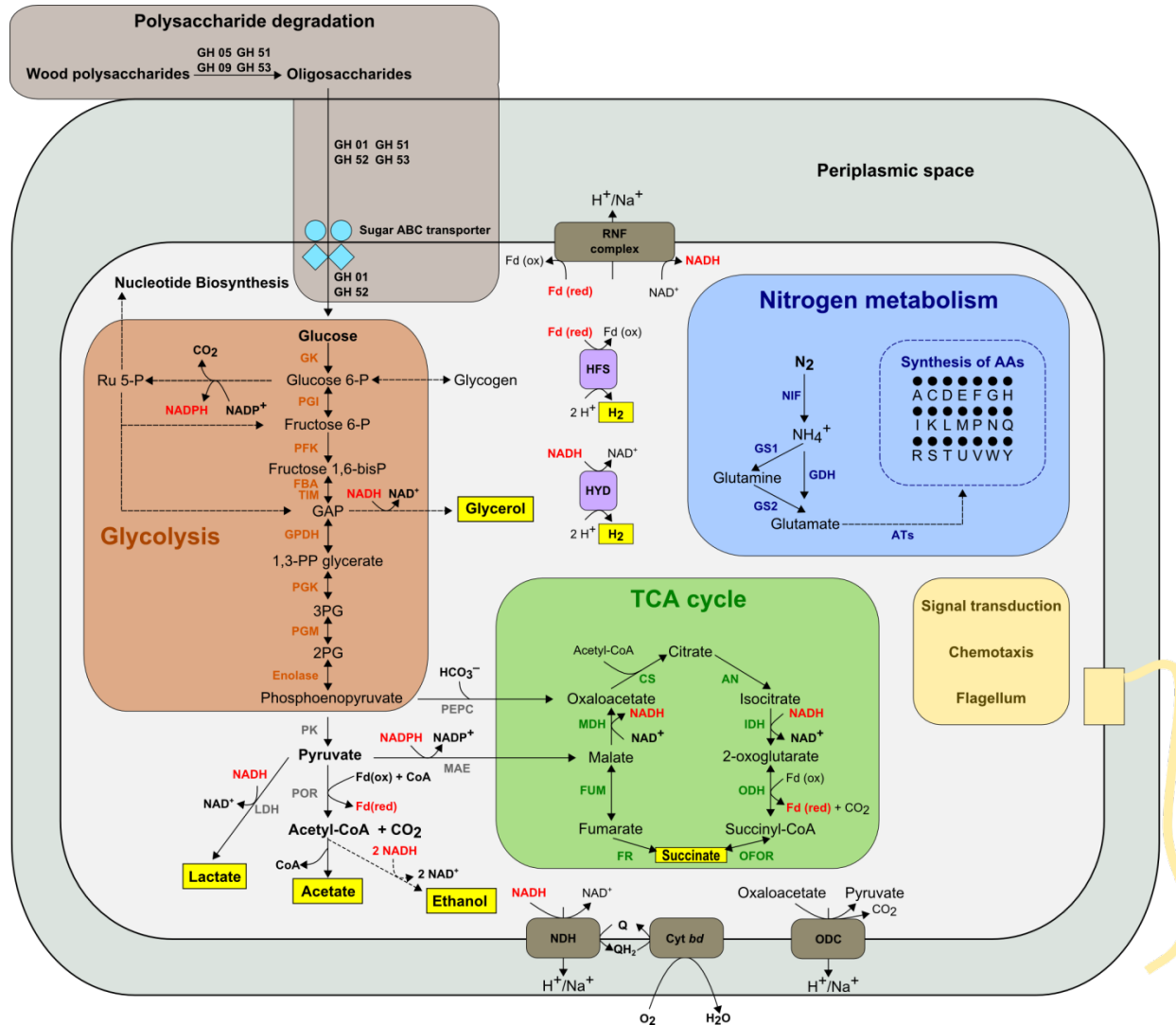
We could not identify any hydrogenases in the population genomes from *Fibrobacteres* — a feature shared with *Fibrobacter succinogenes* (Suen et al., 2011). However, we found that reduced ferredoxin could be oxidized by a ferredoxin:NAD<sup>+</sup> oxidoreductase complex by simultaneous reduction of NAD<sup>+</sup> to NADH. However, we could not identify an additional sink for NADH, which means that NADH could not be utilized under anoxic conditions. We are currently searching for other genes involved, e.g., in glycerol, ethanol, or lactate production. We found a decarboxylating malate dehydrogenase that may run in reverse direction and may allow *Fibrobacteres* to produce malate. *Fibrobacteres* from *Nasutitermes corniger* possess genes for the complete TCA cycle, and this cycle is entered either at the level of oxaloacetate or malate. It is possible that oxaloacetate can be converted to pyruvate via a translocating oxaloacetate decarboxylase.

The presence of a complete TCA (when running in classical mode and not only replenishment of intermediates) would result in even more NADH, which needs to be oxidized. When we searched the population genomes for suitable candidates, we did not consider the possibility of an external terminal electron acceptor. Surprisingly, we found all subunits of a proton translocating NADH dehydrogenase complex. Electrons would then be transferred by reduced quinol to a non-translocating *bd*-type cytochrome oxidase. Such an enzyme reduces oxygen without the translocation of protons (Borisov et al., 2011) and instead reoxidizes quinol to form quinone.

*Fibrobacteres* population genomes encode aminotransferases for the synthesis of 16 different amino acids, and the bacteria are most likely auxotrophic for phenylalanine, lysine, asparagine, threonine, and tyrosine. The ammonium and subsequent glutamate required for amino acid synthesis could be obtained by nitrogen fixation as the genomes carry the whole gene set necessary for nitrogen fixation and subsequent glutamate synthesis from 2-oxoglutarate and ammonium via glutamate dehydrogenase. Alternatively, glutamate could also be formed by glutamine synthetase and ferredoxin-dependent glutamate synthase.



**Fig. 6 | Potential metabolism deduced from the *Fibrobacteres* population genomes in the *Nasutitermes corniger* hindgut.** Yellow shading indicates potential metabolic end products.



**Fig. 7 | Potential metabolism deduced from the *can. div.* TG3 population genomes in the *Nasutitermes corniger* hindgut.** Yellow shading indicates potential metabolic end products.

*Cand. div.* TG3. Members of the *cand. div.* TG3 are likely motile as they possess the full gene set for a flagellum and its motor proteins and have a variety of genes involved in chemotaxis (Fig. 7).

The potential metabolism of members of *cand. div.* TG3 in the termite gut is as follows. Wood is primarily degraded extracellularly to oligosaccharide by GHs 05, 09, 51, and 53. Oligosaccharides are converted to monomeric sugars either in the periplasm or in the cytoplasm, depending on the size of the oligosaccharides, by the GHs 01 and 52. ABC transporters import these monomeric sugars. As in the *Fibrobacteres* from the termite gut, xylose can be metabolized by the pentose phosphate pathway. Intermediates of xylose metabolism can enter glycolysis at the level of fructose 6-phosphate or glyceraldehyde 3-phosphate. Glucose is processed by standard glycolysis. Glycerol is produced atypically from glyceraldehyde in an NADH-dependent reaction.

Pyruvate can be either transformed to lactate via reduction by lactate dehydrogenase or by oxidation to acetyl-CoA. The latter can be directly converted to acetate by acetyl-CoA hydrolase. The subsequent acetate could be further reduced to ethanol by alcohol dehydrogenase. However, genes encoding for aldehyde dehydrogenase could not be identified.

Reduced ferredoxin and NADH could be recycled by enzymes involved in hydrogen production. We identified NADH-dependent hydrogenase and reduced-ferredoxin-dependent hydrogenase. Like members of the *Fibrobacteres*, members of *cand. div.* TG3 possess a putative translocating ferredoxin:NAD<sup>+</sup> oxidoreductase complex that oxidizes ferredoxin and reduces NAD<sup>+</sup> simultaneously. Such an enzyme would enable *cand. div.* TG3 to gain additional energy from hydrogen production by the creation of a membrane potential and NADH for NADH-dependent hydrogenase.

Surprisingly, *cand. div.* TG3 possesses the full TCA cycle and an electron transport chain containing a non-translocating *bd*-type cytochrome oxidase, similar to that of *Fibrobacteres* from the termite gut (see above).

Also like *Fibrobacteres* in the termite gut, *cand. div.* TG3 carries all genes necessary for nitrogen fixation. Subsequently, glutamate can also be synthesized directly from 2-oxoglutarate and ammonium by glutamate dehydrogenase, but also by way of glutamine via glutamine synthetase and ferredoxin-dependent glutamate synthase. The presence of all genes encoding all aminotransferases allows *cand. div.* TG3 to synthesize all amino acids.

## Discussion

The termite gut is a small but complex ecosystem that consists of many distinct microhabitats colonized by microorganisms highly specific to termites (Brune and Dietrich, 2015). These deep-branching lineages still do not exist in culture; hence, their physiology and potential lie entirely in the dark. Although metagenomic analyses have become increasingly popular, emphasis usually lies on particular gene sets. Metagenomic analyses of the termite gut have mostly searched for the presence of glycosyl hydrolases and rarely for functional marker genes that are important for ecosystem function, even though the ultimate goal is to link functions to organisms. And even though a *blastp* classification is often employed, many genes, especially from taxa underrepresented or missing in the reference databases, are misassigned (e.g. Chapter 7)

The termite gut contains a plethora of microbial taxa whose functions are not yet known – for some microorganisms even at the phylum level. The phylum *Fibrobacteres* and *cand. div.* TG3 are, together with *Spirochaetes*, the taxa with the largest discrepancy between the estimations of abundance based on 16S rRNA genes and protein-coding genes in data sets of Rossmassler et al. (Chapter 7 of this thesis). Here we show that binning the sequences into population genomes via Barnes-Hut stochastic neighbor embedding based on symmetric tetramer frequencies provides insights into how members of these dominant but poorly studied phyla adapt to the termite gut environment.

**Fulfilling the promise of metagenomics.** In this study, we were able to bin genomic fragments of *Fibrobacteres* and *cand. div.* TG3 into population genomes using human-augmented compositional binning (Laczny et al., 2014). For successful binning, we had to optimize the parameters to obtain an optimal binning result, which was monitored using the 31 elite phylogenetic markers of Wu and Scott (2012). We found that the right combination of the pseudo-count for the clr transformation and the choice of perplexity for Barnes-Hut stochastic neighbor embedding is critical. Both parameters should be optimized for each metagenome separately. Although perplexity seems to be a relatively elastic parameter (van der Maaten, 2014), an estimation of the Shannon entropy using single-copy genes enhances the global clustering. However, this would not be valid if too few features were present (e.g., a diverse community of extremely close relatives) and would result in a biased perplexity estimate. In agreement with Costea et al. (2014), we found that the addition of a suitable pseudo-count in log-transformation (here clr) is critical. We also found that this small adjustment had a huge impact, especially on local clustering. Through this optimization, we were able to bin the metagenomic fragments into 33 population genomes of the phylum *Fibrobacteres* and *cand. div.* TG3. When we analyzed contamination, we found an average population genome contamination of 2.1%. A recent study (Parks et al., 2014) estimated that most public genomes are contaminated by up to 5%, which documents that the population genomes from our study are highly accurate. In an inspection of the completeness of the recovered population genomes, most of the genomes from our study were at least 50% complete, but we even recovered 100% complete genomes, which is very rare for metagenomic studies or single-cell genomes (Parks et al., 2014). We obtained at least one high-quality population genome from each of the suitable gut metagenomes from the termite species *Cornitermes* sp., *Microcerotermes parvus*, and *Nasutitermes corniger*.

**Phylogenetic status of candidate division TG3.** Past culture-independent studies have revealed that most termite gut microbial lineages form exclusive phylogenetic clusters (e.g., Hongoh et al., 2003; Schmitt-Wagner et al., 2003; Hongoh et al., 2005; Stingl et al., 2005). Even new bacterial divisions have been discovered (e.g., Hongoh et al., 2003; Hongoh et al., 2005; Stingl et al., 2005). One of these phylum-level groups, *cand. div.* TG3, has been studied at phylogenetic and morphological levels (Hongoh et al., 2005; Hongoh et al., 2006). Recently, Sorokin et al. (2014) described the first isolate of *cand. div.* TG3, *Chitinivibrio alkaliphilus*, and tried to infer whether *cand. div.* TG3 is a separate phylum. This analysis is necessary because the proposed phylum status is based on 16S rRNA gene phylogeny and would have to be revised with a more highly resolved phylogenetic analysis, perhaps even one independent of 16S rRNA, ideally by incorporating many conserved genes. Non-horizontally transferred genes can show contradicting tree topologies, which underlines the need for multigene phylogenies. Although Sorokin et al. (2014) used such a multigene approach, they still could not conclude whether *cand. div.* TG3 is a valid candidate division. Their tree indicated that *Fibrobacter succinogenes* and *Chitinivibrio alkaliphilus* are the closest relatives. However, since no other genomes are available for *Fibrobacteres* and *cand. div.* TG3, this comparison does not allow one to conclude whether the two divisions can be considered separate.

Here, we obtained 33 population genomes of members of the *Fibrobacteres* and *cand. div.* TG3 present in different abundances in the termite gut metagenomes (Tab. 2; Fig. 3). Phylogenetic analysis of all population genomes based on the elite marker genes of Wu and Scott (2012) revealed that *Fibrobacteres* as well as *cand. div.* TG3 are each a deep-branching monophyletic group (Fig. 3). This is agreement with earlier phylogenetic analysis based on the 16S rRNA gene (Hongoh et al., 2003; Hongoh et al., 2005; Mikaelyan et al., unpubl); therefore, we consider *cand. div.* TG3 a bacterial division. Another argument for the phylum-level status is the huge dissimilarity within a potential comprises *Fibrobacteres* and TG3. Compared to most established phyla, this variation surpasses those of the other phyla, which indicates that a split into two phyla is reasonable.

**Synergism in cellulolysis between *Fibrobacteres* and candidate division TG3.** The fiber fraction of *Nasutitermes corniger* is a diversely populated microhabitat with a high activity of cellulases (Mikaelyan et al., 2014). Besides *Treponema* cluster Ic, *Fibrobacteres* termite cluster II (formerly *Fibrobacteres* subphylum I; Hongoh et al., 2006) and TG3 termite cluster I are the only highly abundant members in the fiber fraction. Since all these groups would share the same substrate, competition would be a feasible scenario. In our analyses, we found that members of the *Fibrobacteres* from termite guts contain a similar number of glycosyl hydrolase genes per genome as the type strain *Fibrobacter succinogenes* (Suen et al., 2011). *cand. div.* TG3 genome populations contained substantially fewer genes (about 50% fewer) encoding glycosyl hydrolases relevant for cellulose degradation – a feature they share with *Chitinivibrio alkaliphilus* (Fig. 4). The fact that *C. alkaliphilus* carries few genes involved in cellulose degradation is not surprising because it is a chitinolytic organism. However, *C. alkaliphilus* also possesses genes involved in the degradation of wood polysaccharides (Sorokin et al., 2014). Members of *cand. div.* TG3 from termite gut carry 25% more cellulose degradation genes on average than *C. alkaliphilus*, and this might be an adaptation to the termite gut.

When we analyzed glycosyl hydrolase families in more detail, we found that although *Fibrobacteres* seem to be the more potent cellulose degraders, *cand. div.* TG3 population genomes carry GHF-encoding genes that do not occur or only occur in only low abundance in the population genomes of *Fibrobacteres* from the termite gut. This suggests that *Fibrobacteres* and *cand. div.* TG3 bacteria

cooperatively degrade wood. A dual cellulolytic system involved in the cooperative digestion of wood has also been proposed for the termite host and its gut microbiota (see references in Brune and Ohkuma, 2011). A recent study (Poulsen et al., 2014) even revealed a complementary set of glycosyl hydrolases in the fungus-feeding *Macrotermitinae*, with partners here being termite host, farmed fungus, and gut microbiota. Our study here showed that even bacterial lineages within the same termite gut possess complementary sets of glycosyl hydrolase families.

**Strategies to prevent wash out.** The metabolism of the complex termite gut ecosystem is highly efficient. This efficiency is attained by the concerted effort of many microorganisms and their termite host. Consequently, the termite gut can be regarded as a bioreactor (Brune, 1998), and the concepts and rules of a (bio)chemical reactor are also valid. One main property of a reactor is its retention time. If a bacterium grows more slowly than the retention time, the bacterium would be washed out. To circumvent this wash out, microorganisms have evolved different strategies. The most efficient strategy in terms of energy economy is the attachment to surfaces. Members of the *Fibrobacteres* and *cand. div. TG3* are both associated with wood fibers in *Nasutitermes corniger* (Mikaelyan et al., 2014). To support this finding, we included an analysis of genes encoding for various cellulose-binding modules in Fig. 5. The presence and use of these binding modules might not be necessary to circumvent wash out since the same retention applies to wood as well. We think that binding to cellulose might be an adaptation that is necessary to stay in proximity of the hydrolytic products released during wood degradation. For example, two included termite *Treponema* species from the termite gut (Fig. 4) have a high potential to degrade oligosaccharides. To compete with these bacteria that metabolize wood hydrolysates, members of *Fibrobacteres* and *cand. div. TG3* could bind to wood fibers, thereby also minimizing the effect of Brownian movement, which would result in dispersion of the bacteria. Another adaptation of both *Fibrobacteres* and *cand. div. TG3* to avoid wash out in the termite gut could be motility and chemotaxis. *Chitinivibrio alkaliphilus* from a hypersaline soda lake encodes flagellum proteins (Sorokin et al., 2014), but *Fibrobacter succinogenes* from the rumen is reported to have only twitching motility (Suen et al., 2011).

**Adaptations of the energy metabolism to the termite gut ecosystem.** The termite gut displays a remarkable spatial heterogeneity with a constant substrate supply (Brune et al., 1995; Köhler et al., 2012). One of the most spatially variable physicochemical gut parameter is oxygen. In the guts of ruminants, with their large volume, the oxygen influx per unit volume is tiny. In contrast, the termite gut surface-to-volume ratio is so large that the influx of oxygen is significant (see Brune, 1998 and references therein). The steady influx of oxygen is consumed near the gut wall but might not always be used for energy conservation. The presence and isolation of microaerobic microorganisms from the termite gut (e.g., Wertz and Breznak, 2007; Wertz et al., 2012) documents that also facultative reductive energy metabolism is common. Although *Fibrobacter succinogenes* and *Chitinivibrio alkaliphilus* apparently do not harbor genes encoding respiratory enzymes (Suen et al., 2011; Sorokin et al., 2014), we found genes in the population genomes of *Fibrobacteres* and *cand. div. TG3* encoding an electron transport chain that reduces oxygen with a *bd*-type cytochrome oxidase coupled via quinone/quinol to a proton translocating NADH dehydrogenase. Such an electron transport chain is found in many bacteria from diverse phyla, including many pathogens. The habitat of most pathogens contains only little oxygen, and under oxygen limitation, many bacteria express this type of electron transport chain (Ingledeew and Poole, 1984; Poole and Cook, 2000). The *bd*-type cytochrome oxidase in this chain is not homologous to other respiratory oxygen reductases and also does not translocate protons for chemiosmotic energy conservation. This may appear to be wasteful, but it results in a very



high affinity for oxygen (Borisov et al., 2011) – an important feature that allows an oxygen-dependent respiratory or competitive oxygen-dependent lifestyle in ecosystems with low oxygen tension, such as in the termite gut.

Whether members of the *Fibrobacteres* and *cand. div.* TG3 use their electron transport chain primarily for detoxification or for energy conservation is not known. But since the ATP yield of their electron transport chain is probably comparable to that of other bacteria, we hypothesize that it is an alternative pathway to fermentation. Population genomes from both *Fibrobacteres* and *cand. div.* TG3 also harbor genes for reduction of endogenous electron acceptors. *Fibrobacteres* population genomes carry genes with the potential for producing acetate, formate, and possibly malate. We found no hydrogenase genes in *Fibrobacteres* population genomes, none were detected in the genome of *Fibrobacter succinogenes* (Suen et al., 2011), and physiological data indicates that *F. succinogenes* does not produce any hydrogen (Fonty et al., 1995; Chaucheyras-Durand et al., 2010). The population genomes of *cand. div.* TG3, in contrast, carry genes for producing a wider spectrum of products, i.e., acetate, ethanol, lactate, and possibly malate. These *cand. div.* TG3 bacteria might actively produce hydrogen in the termite gut, as they harbor genes for one ferredoxin-dependent hydrogenase and one NADH-dependent hydrogenase. Since *C. alkaliphilus* also encodes these two types of hydrogenases, they might be a feature shared by members of this phylum.

Both *Fibrobacteres* and *cand. div.* TG3 population genomes harbor genes encoding the complete ferredoxin:NAD<sup>+</sup> oxidoreductase complex; this membrane-bound electron transport complex is found in many bacteria. It was first discovered in *Rhodobacter capsulatus* (Schmehl et al., 1993), where it transports electrons from NADH to ferredoxin and then to nitrogenase; it has been questioned whether the reaction can run in the opposite direction as a ferredoxin oxidoreductase, thereby creating a membrane potential (Biegel and Müller, 2010). If this reverse reaction occurred in the termite gut in members of *cand. div.* TG3, who have the potential to produce hydrogen by the oxidation of ferredoxin or NADH, a membrane potential could be created. In such a scenario, NADH would serve as the primary substrate for hydrogen production instead of ferredoxin. However, with the high hydrogen partial pressure in termite guts (Brune et al., 1995; Köhler et al., 2012), it is questionable whether an NADH-dependent hydrogen production is feasible.

**Adaptions to the low nitrogen content of lignocelluloses.** Termites generally feed on recalcitrant diets. For wood-feeding termites, the energy conserved in wood polysaccharides equals the energy of the resulting monomers minus the energy expended for hydrolysis. Although wood contains plenty of energy, it has a very low nitrogen content. Therefore, termites rely on their gut microbiota to acquire nitrogenous compounds (Brune and Ohkuma, 2011 and references therein). The uptake and hydrolysis of nitrogenous compounds in termites is mostly localized to the midgut (Fujita and Abe, 2002). In contrast to foregut fermenters, such as ruminants, termites need to transfer the hindgut microbiota to the midgut via proctodeal trophallaxis before the microbial proteins can be digested. Microorganisms carrying out nitrogen fixation in higher termites have not yet been identified. Nitrogen fixation *nifH* genes have been identified in the termite gut, but since the *nifH* gene was frequently horizontally transferred, it is almost impossible to link these genes to microbial taxonomy (Ohkuma et al., 1999; Yamada et al., 2007). We found that the population genomes of both *Fibrobacteres* and *cand. div.* TG3 harbor all genes required for nitrogen fixation. In addition, we identified genes encoding enzymes for the production of glutamate and other amino acids, i.e., for all amino acids in *cand. div.* TG3 and for 16 amino acids in *Fibrobacteres*. Since each of the cultured representatives of *Fibrobacteres* and *cand. div.* TG3 do not harbor genes for nitrogen fixation (Suen et

al., 2011; Sorokin et al., 2014), we postulate that the acquisition of nitrogen fixation (perhaps via horizontal gene transfer) is an adaptation of their relatives in the termite gut to the low nitrogen content of their host's diet.

**Application of this binning method to other ecosystems and its potential for microbiology.** In this study we used a phylogenetic marker assisted binning method to create population genomes of deep-branching high-level taxa that are poorly represented in public databases. Albertsen et al. (2013) already showed that microorganisms sensitive to short-term physical and chemical treatments (e.g. high temperature treatment in phenol) can be binned by a differential coverage approach. However, not all microorganisms have such an exclusive property (when all organisms are altered in the same fashion, no differential signal to bin would be present).

The method employed in this study allows to bin metagenomic fragments by compositional features with minimal theoretical assumptions. By the use of 31 phylogenetic marker genes of Wu and Scott (2012) it is possible to optimize the binning conditions of each individual metagenome. It was even possible to bin contigs containing a 16S rRNA gene — usually a major problem in compositional binning. The use of 31 phylogenetic markers and the 16S rRNA genes allows an accurate determination of the taxonomy of the binned population genome. The optimized and customizable method of this study will be of great value for all metagenomic studies that focus on microbial assemblages that contain lineages that are only distantly related to cultured strains. The resulting genomes would allow to reveal the true genetic diversity of microorganisms.

## Acknowledgements

This study was funded by a grant of the Deutsche Forschungsgemeinschaft (DFG) in the Collaborative Research Center SFB 987 and the LOEWE program of the state of Hessen (Synmikro). Carsten Dietrich received a scholarship from the International Max Planck Research School for Environmental, Cellular and Molecular Microbiology (IMPRS-MIC). We thank Karen Brune for correcting the manuscript.

## References

- Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW and Nielsen PH (2013). Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nat. Biotechnol.* 31:533–538.
- Allgaier M, Reddy A, Park JI, Ivanova N, D'Haeseleer P, Lowry S, Sapra R, Hazen TC, Simmons BA, VanderGheynst JS and Hugenholtz P (2010). Targeted discovery of glycoside hydrolases from a switchgrass-adapted compost community. *PLoS One* 5:e8812.
- Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, ZIJaz U, Lahti L, Loman NJ, Andersson AF and Quince C (2014). Binning metagenomic contigs by coverage and composition. *Nat. Meth.* 11:1144–1146.
- Beatriz PL and Rodriguez-Casal R (2010). Generalizing the convex hull of a sample. The R package alphahull.. *J. Stat. Soft.* 34:1–28.
- Berger SA, Krompass D and Stamatakis A (2011). Performance, accuracy, and web server for evolutionary placement of Short Sequence Reads under Maximum Likelihood. *Syst. Biol.* 60:291–302
- Biegel E and Müller V (2010). Bacterial Na<sup>+</sup>-translocating ferredoxin:NAD<sup>+</sup> oxidoreductase. *Proc. Natl. Acad. Sci. USA* 107: 18138–18142.
- Borisov VB, Gennis RB, Hemp J and Verkhovsky MI (2011) The cytochrome bd respiratory oxygen reductases. *Biochim. Biophys. Acta.* 1807(11):1398–413.
- Bragg L and Tyson GW (2014). Metagenomics Using Next-Generation Sequencing. pp. 183–201 *In* Paulsen IT and Holmes AJ (eds.), *Environmental Microbiology*. Humana Press, New York
- Brune A (1998). Termite guts: the world's smallest bioreactors. *Trends Biotechnol.* 16:16–21.
- Brune A (2014). Symbiotic digestion of lignocellulose in termite guts. *Nat. Rev. Microbiol.* 12:168–180.
- Brune A and Dietrich C (2015). The termite gut microbiota: Digesting the diversity in the light of ecology and evolution. *Ann. Rev. Microbiol.* (in Press).
- Brune A and Ohkuma M (2011). Role of the termite gut microbiota in symbiotic digestion, pp. 439–475. *In* Bignell DE, Roisin Y, Lo N (eds.), *Biology of Termites: A Modern Synthesis*, Springer, Dordrecht.
- Brune A, Emerson D and Breznak JA (1995). The termite gut microflora as an oxygen sink: microelectrode determination of oxygen and pH gradients in guts of lower and higher termites. *Appl. Environ. Microbiol.* 61:2681–2687.
- Chaucheyras-Durand F, Masségla S, Font G and Forano E (2010). Influence of the Composition of the Cellulolytic Flora on the Development of Hydrogenotrophic Microorganisms, Hydrogen Utilization, and Methane Production in the Rumens of Gnotobiotically Reared Lambs. *Appl. Environ. Microbiol.* 76:7931–7937.
- Cordero OX and Polz MF (2014). Explaining microbial genomic diversity in light of evolutionary ecology. *Nat. Rev. Microbiol.* 12:263–273.
- Costea PL, Zeller G, Sunagawa S, and Bork P (2014). A fair comparison. *Nat. Meth.* 11:359
- Darling AE, Jospin G, Lowe E, Matsen FA, Bik HM and Eisen JA (2014). PhyloSift: phylogenetic analysis of genomes and metagenomes. *PeerJ* 2:e243.
- Dietrich C and Brune A (2015). The complete mitogenomes of six higher termite species reconstructed from metagenomic datasets (*Cornitermes* sp., *Cubitermes ugandensis*, *Microcerotermes parvus*, *Nasutitermes corniger*, *Neocapritermes taracua*, and *Termes hospes*). *Mitochondr. DNA* (in Press).
- Dietrich C, Köhler T and Brune A (2014). The cockroach origin of the termite gut microbiota: patterns in bacterial community structure reflect major evolutionary events. *Appl. Environ. Microbiol.* 80:2261–2269.
- Dupont CL, Rusch DB, Yooshef S, Lombardo M-J, Richter RA, Valas R, Novotny M, Yee-Greenbaum J, Selengut JD, Haft DH, Halpern AL, Lasken RS, Nealson K, Friedman R and Venter JC (2012). Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. *ISME J.* 6:1186–1199.
- Eddy SR (2011). Accelerated profile HMM searches. *PLoS Comp. Biol.* 7:e1002195.

- Fonty G, Jouany JP, Forano E and Gouet P (1995). L'écosystème microbien du réticulo-rumen, p. 299–347. In R. Jarrige (ed.), Nutrition des ruminants domestiques. INRA, Paris, France.
- Fujita A and Abe T (2002). Amino acid concentration and distribution of lysozyme and protease activities in the guts of higher termites. *Physiol. Entomol.* 27:76–78.
- He S, Ivanova N, Kirton E, Allgaier M, Bergin C, Scheffrahn RH, Kyrpides NC, Warnecke F, Tringe SG and Hugenholtz P (2013). Comparative metagenomic and metatranscriptomic analysis of hindgut paunch microbiota in wood- and dung-feeding higher termites. *PLoS ONE* 8:e61126.
- Hongoh Y, Deevong P, Hattori S, Inoue T, Noda S, Noparatnaraporn N, Kudo T and Ohkuma M (2006). Phylogenetic diversity, localization and cell morphologies of the candidate phylum TG3 and a subphylum in the phylum *Fibrobacteres*, recently found bacterial groups dominant in termite guts. *Appl. Environ. Microbiol.* 72:6780–6788.
- Hongoh Y, Deevong P, Inoue T, Moriya S, Trakulnaleamsai S, Ohkuma M, Vongkaluang C, Noparatnaraporn N and Kudo T (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.
- Hongoh Y, Ohkuma M and Kudo T (2003). Molecular analysis of bacterial microbiota in the gut of the termite *Reticulitermes speratus* (Isoptera; Rhinotermitidae). *FEMS Microbiol. Ecol.* 44:231–242.
- IngledeW WJ and Poole RK (1984). The respiratory chains of *Escherichia coli*. *Microbiol Rev.* 48:222–271.
- Kang DD, Froula J, Egan R and Wang Z (2014). A robust statistical framework for reconstructing genomes from metagenomic data. *bioRxiv* 1:1–14.
- Köhler T, Dietrich C, Scheffrahn RH and Brune A (2012). High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation of the gut in wood-feeding higher termites (*Nasutitermes* spp.). *Appl. Environ. Microbiol.* 78:4691–4701.
- Laczny CC, Pinel N, Vlassis N and Wilmes P (2014). Alignment-free visualization of metagenomic data by nonlinear dimension reduction. *Sci. Rep.* 4:4516.
- Lagesen K, Hallin P, Rødland EA, Stærfeldt, HH, Rognes T and Ussery DW (2007). RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucl. Res.* 35(9):3100–3108.
- Levasseur A, Drula E, Lombard V, Coutinho PM and Henrissat B (2013). Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol Biofuels* 6:41.
- Liu N, Zhang L, Zhou H, Zhang M, Yan X, Wang Q, Long Y, Xie L, Wang S, Huang Y and Zhou Z (2013). Metagenomic Insights into Metabolic Capacities of the Gut Microbiota in a Fungus-Cultivating Termite (*Odontotermes yunnanensis*). *PLoS ONE* 8:e69184.
- Lucey KS and Leadbetter JR (2014). Catechol 2,3-dioxygenase and other meta-cleavage catabolic pathway genes in the 'anaerobic' termite gut spirochete *Treponema primitia*. *Mol. Ecol.* 6:1531–43.
- McLean JS, Lombardo M-J, Badger JH, Edlund A, Novotny M, Yee-Greenbaum J, Vyahhi N, Hall AP, Yang Y, Dupont CL, Ziegler MG, Chitsaz H, Allen AE, Yooseph S, Tesler G, Pevzner PA, Friedman RM, Neilson KH, Venter JC and Lasken RS (2013). Candidate phylum TM6 genome recovered from a hospital sink biofilm provides genomic insights into this uncultivated phylum. *Proc. Natl. Acad. Sci. USA* 110:E2390–E2399.
- Mikaelyan A, Lampert N, Köhler T, Rohland J, Boga H, Meuser K and Brune A (2015). DictDb: an expanded reference database for the highly resolved classification of the bacterial gut microbiota of termites and cockroaches. Submitted Manuscript.
- Mikaelyan A, Strassert JFH, Tokuda G and Brune A (2014). The fiber-associated cellulolytic bacterial community in the hindgut of wood-feeding higher termites (*Nasutitermes* spp.). *Environ. Microbiol.* 16:2711–2722.
- Ngugi DK and Brune A (2012). Nitrate reduction, nitrous oxide formation, and anaerobic ammonia oxidation to nitrite in the gut of soil-feeding termites (*Cubitermes* and *Ophiotermes* spp.). *Environ. Microbiol.* 14:860–871.
- Ohkuma M, Noda S, Kudo T (1999). Phylogenetic diversity of nitrogen fixation genes in the symbiotic microbial community in the gut of diverse termites. *Appl. Environ. Microbiol.* 65:4926–4934.

- Pages H, Aboyou P, Gentleman R and DebRoy S (2010).** Biostrings: string objects representing biological sequences, and matching algorithms. R package version 2.16.9. R Foundation for Statistical Computing, Vienna, VA. <http://www.R-project.org>.
- Paradis E, Claude J and Strimmer K (2004).** APE: analyses of phylogenetics and evolution in R language. *Bioinform.* 20:289–290.
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P and Tyson GW (2014).** CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *PeerJ PrePrints* 2:e554v1.
- Pester M, Tholen A, Friedrich MW and Brune A (2007).** Methane oxidation in termite hindguts: absence of evidence and evidence of absence. *Appl. Environ. Microbiol.* 73:2024–2028.
- Podosokorskaya OA, Kadnikov VV, Gavrilov SN, Mardanov AV, Merkel AY, Karnachuk OV, Ravin, NV, Bonch-Osmolovskaya EA and Kublanov IV (2013).** Characterization of *Melioribacter roseus* gen. nov., sp. nov., a novel facultatively anaerobic thermophilic cellulolytic bacterium from the class Ignavibacteria, and a proposal of a novel bacterial phylum *Ignavibacteriae*. *Environ. Microbiol.*, 15: 1759–1771.
- Poole RK and Cook GM (2000).** Redundancy of aerobic respiratory chains in bacteria? Routes, reasons and regulation. *Adv Microb Physiol.* 43:165–224.
- Poulsen M, Hu H, Li C, Chen Z, Xu L, Otani S, Nygaard S, Nobre T, Klaubauf S, Schindler PM, Hauser F, Pan H, Yang Z, Sonnenberg ASM, de Beer AW, Zhang Y, Wingfield MJ, Grimmelikhuijzen CJP, de Vries RP, Korb J, Aanen DK, Wang J, Boomsma JJ and Zhang G (2014).** Complementary symbiont contributions to plant decomposition in a fungus-farming termite. *Proc. Natl. Acad. Sci. USA* 111:14500–14505.
- R Development Core Team (2008).** R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ and Weber CF (2009).** Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* 75:7537–7541.
- Schmehl M, Jahn A, zu Vilsendorf AM, Hennecke S, Masepohl B, Schuppler M, Marxer M, Oelze J and Klipp W (1993).** Identification of a new class of nitrogen fixation genes in *Rhodobacter capsulatus*: a putative membrane complex involved in electron transport to nitrogenase. *Mol. Gen. Genet.* 241: 602–615.
- Schmitt-Wagner D, Friedrich MW, Wagner B and Brune A (2003).** Phylogenetic diversity, abundance, and axial distribution of bacteria in the intestinal tract of two soil-feeding termites (*Cubitermes* spp.). *Appl. Environ. Microbiol.* 69:6007–6017.
- Sorokin DY, Gumerov VM, Rakitin AL, Beletsky AV, Damsté JSS, Muyzer G, Mardanov AV and Ravin NV (2014).** Genome analysis of *Chitinivibrio alkaliphilus* gen. nov., sp. nov., a novel extremely haloalkaliphilic anaerobic chitinolytic bacterium from the candidate phylum TG3. *Environ. Microbiol.* 16:1549-1565.
- Stamatakis A (2014).** RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinf.* 30:1312–1313.
- Stingl U, Radek R, Yang H and Brune A (2005).** 'Endomicrobia': Cytoplasmic symbionts of termite gut protozoa form a separate phylum of prokaryotes. *Appl. Environ. Microbiol.* 71:1473–1479.
- Suen G, Weimer PJ, Stevenson DM, Aylward FO, Boyum J, Deneke J, Drinkwater C, Ivanova NN, Mikhailova N, Chertkov O, Goodwin LA, Currie CR, Mead D and Brumm PJ (2011).** The complete genome sequence of *Fibrobacter succinogenes* S85 reveals a cellulolytic and metabolic specialist. *PLoS One* 6(4):e18814.
- van der Maaten L (2014).** Accelerating t-SNE using Tree-Based Algorithms. *J. Mach. Learn. Res.* 15:3221–3245.

- Warnecke F, Luginbühl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, Cayouette M, McHardy AC, Djordjevic G, Aboushadi N, Sorek R, Tringe SG, Podar M, Martin HG, Kunin V, Dalevi D, Madejska J, Kirton E, Platt D, Szeto E, Salamov A, Barry K, Mikhailova N, Kyrpides NC, Matson EG, Ottesen EA, Zhang X, Hernández M, Murillo C, Acosta LG, Rigoutsos I, Tamayo G, Green BD, Chang C, Rubin EM, Mathur EJ, Robertson DE, Hugenholtz P and Leadbetter JR (2007). Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nat.* 450:560–565.
- Wertz JT and Breznak JA (2007). Physiological ecology of *Stenoxybacter acetivorans*, an obligate microaerophile in termite guts. *Appl. Environ. Microbiol.* 73:6829–6841.
- Wertz JT, Kim E, Breznak JA, Schmidt TM and Rodrigues JLM (2012). Genomic and physiological characterization of the *Verrucomicrobia* isolate *Diplosphaera colotermitum* gen. nov., sp. nov. reveals microaerophily and nitrogen fixation genes. *Appl. Environ. Microbiol.* 78:1544–1555.
- Wu M and Scott AJ (2012). Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. *Bioinf.* 28(7):1033–1034.
- Wu M, Chatterji S and Eisen JA (2012). Accounting for alignment uncertainty in phylogenomics. *PLoS One* (1):e30288.
- Wu YW, Tang YH, Tringe SG, Simmons BA and Singer SW (2014). MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbi.* 2:26.
- Yamada A, Inoue T, Noda Y, Hongoh H and Ohkuma M (2007). Evolutionary trend of phylogenetic diversity of nitrogen fixation genes in the gut community of wood-feeding termites. *Mol. Ecol.* 16:3768–3777.
- Yamada T, Letunic I, Okuda S, Kanehisa M and Bork P (2011). iPath2.0: interactive pathway explorer. *Nucl. Res.* 39:W412–415.
- Yin Y, Mao X, Yang JC, Chen X, Mao F and Xu Y (2012). dbCAN: a web resource for automated carbohydrate-active enzyme annotation. *Nucl. Res.* 40:445–451.

## Supplementary Material

Tab. S1 | Primary statistics of the genome populations obtained in this study.

Bin	Size (Mb)	Meta-genome abundance (%)	Contigs	N50	Max. contig length (kbp)	Completeness (%)	Estimated number of strains
<b>Phylum <i>Fibrobacteres</i></b>							
FIB-CP1-1	3.0	0.8	417	4,263	56.9	44	1.2
FIB-CP1-2	8.2	2.1	2,296	2,613	41.0	81	2.0
FIB-CP1-3	1.0	0.3	231	2,846	23.1	23	1.2
FIB-CP1-4	3.0	1.7	312	7,552	50.2	60	1.3
FIB-CP1-5	3.7	1.5	635	3,245	89.1	56	1.5
<b>FIB-CP1-P</b>	<b>18.9</b>	<b>6.4</b>	<b>3,891</b>	<b>2,933</b>	<b>89.1</b>	<b>100</b>	<b>4.0</b>
FIB-CP3-1	8.1	1.7	2,205	2,673	61.2	98	2.5
FIB-CP3-2	1.5	0.3	255	4,496	27.2	32	1.3
<b>FIB-CP3-P</b>	<b>9.6</b>	<b>2.0</b>	<b>2,460</b>	<b>2,775</b>	<b>61.2</b>	<b>100</b>	<b>2.8</b>
FIB-CP4-1	5.6	8.9	976	3,389	46.8	90	1.6
FIB-FCP4-2	1.5	1.6	129	8,630	72.6	38	1.0
<b>FIB-FCP4-P</b>	<b>7.1</b>	<b>10.5</b>	<b>1,105</b>	<b>3,782</b>	<b>72.6</b>	<b>100</b>	<b>1.7</b>
FIB-MP1-1	1.8	1.8	490	3,017	24.9	54	1.0
<b>FIB-MP1 -P</b>	<b>1.8</b>	<b>1.8</b>	<b>490</b>	<b>3,017</b>	<b>24.9</b>	<b>54</b>	<b>1.0</b>
FIB-MP3-1	3.1	1.4	567	3,693	40.5	43	1.3
FIB-MP3-2	2.6	1.4	643	2,897	41.2	36	1.6
FIB-MP3-3	1.6	1.0	327	3,627	32.8	46	1.3
FIB-MP3-4	1.7	0.9	286	4,454	39.0	55	1.2
FIB-MP3-5	1.3	0.8	305	2,897	40.0	42	1.2
FIB-MP3-6	1.1	0.7	205	4,127	34.4	25	1.2
<b>FIB-MP3 -P</b>	<b>11.4</b>	<b>6.2</b>	<b>2,333</b>	<b>3,464</b>	<b>41.2</b>	<b>97</b>	<b>3.3</b>
FIB-MP4-1	1.9	11.9	576	2,904	10.6	54	1.0
<b>FIB-MP4 -P</b>	<b>1.9</b>	<b>11.9</b>	<b>576</b>	<b>2,904</b>	<b>10.6</b>	<b>54</b>	<b>1.0</b>
FIB-NP3-1	3.6	2.3	685	4,361	23.6	54	1.9
FIB-NP3-2	3.4	1.5	623	4,163	31.3	43	1.8
FIB-NP3-3	2.2	1.3	439	3,861	22.6	34	1.3
FIB-NP3-4	2.8	1.5	638	3,273	33.8	37	1.5
FIB-NP3-5	2.5	2.1	446	4,564	26.5	30	1.4
FIB-NP3-6	2.0	0.7	373	3,965	28.7	35	1.4
FIB-NP3-7	1.2	0.6	215	4,132	23.1	16	1.0
<b>FIB-NP3 -P</b>	<b>17.7</b>	<b>10.0</b>	<b>3,419</b>	<b>3,951</b>	<b>33.8</b>	<b>92</b>	<b>4.2</b>
FIB-NP4-1	13.4	11.9	3,078	3,303	37.7	78	2.9
<b>FIB-NP4 -P</b>	<b>13.4</b>	<b>11.9</b>	<b>3,078</b>	<b>3,303</b>	<b>37.7</b>	<b>78</b>	<b>2.9</b>
<b>Candidate division TG3</b>							
TG3-CP1-1	6.9	7.0	535	8,195	117.0	100	2.1
<b>TG3-CP1-P</b>	<b>6.9</b>	<b>7.0</b>	<b>535</b>	<b>8,195</b>	<b>117.0</b>	<b>100</b>	<b>2.1</b>
TG3-CP3-1	6.7	12.4	527	8,298	89.5	100	2.1
<b>TG3-CP3-P</b>	<b>6.7</b>	<b>12.4</b>	<b>527</b>	<b>8,298</b>	<b>89.5</b>	<b>100</b>	<b>2.1</b>

Bin	Size (Mb)	Meta-genome abundance (%)	Contigs	N50	Max. contig length (kbp)	Completeness (%)	Estimated number of strains
TG3-MP3-1	3.1	2.0	833	3,072	23.6	43	1.4
TG3-MP3-2	3.4	0.7	483	3,426	63.1	100	1.2
TG3-MP3-3	1.8	1.1	610	2,529	13.5	29	1.1
<b>TG3-MP3-P</b>	<b>8.3</b>	<b>3.8</b>	<b>1,926</b>	<b>2,896</b>	<b>63.1</b>	<b>100</b>	<b>2.1</b>
CNP3-1	7.8	4.7	1,647	3,509	49.9	91	2.1
CNP3-2	6.9	2.3	1,139	4,331	78.1	92	1.7
CNP3-3	1.0	0.1	316	2,612	16.0	50	1.0
<b>TG3-NP3-P</b>	<b>15.7</b>	<b>7.1</b>	<b>3,102</b>	<b>3,612</b>	<b>78.1</b>	<b>100</b>	<b>3.7</b>



## Chapter 9 General discussion

Carsten Dietrich and Andreas Brune

*Affiliations:* Max Planck Institute for Terrestrial Microbiology, 35043 Marburg, Germany. | *This chapter is based on the manuscript:* Brune A and Dietrich C (2015). The termite gut microbiota: Digesting the diversity in the light of ecology and evolution. *Ann. Rev. Microbiol.* 69. | *Contributions:* C.D. contributed to the manuscript and designed the figures. A.B. contributed to the manuscript.

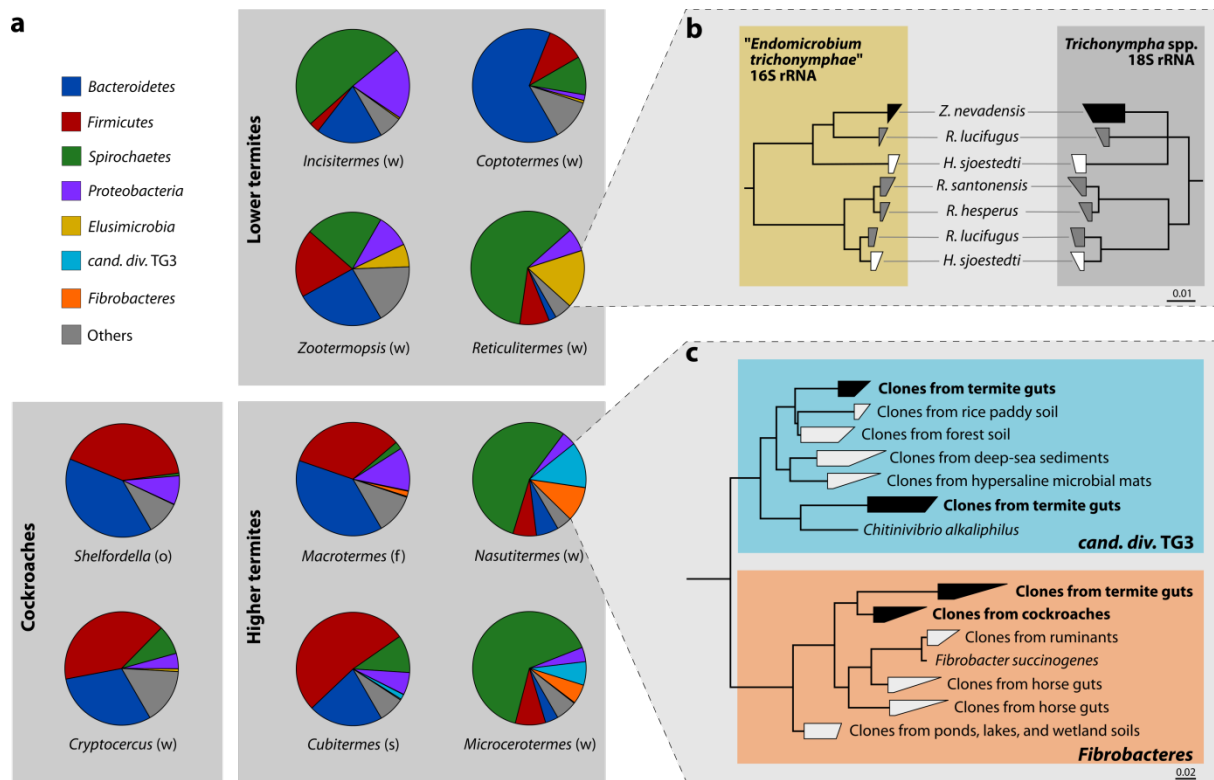
## Diversity of the termite gut microbiota

With the presence of Bacteria, Archaea, and Eukarya (flagellate protists in lower termites), the gut microbiota of termite encompasses all three domains of life (Ohkuma and Brune, 2011; Brune, 2014). While the flagellates can be identified already on the basis of their morphological features, the characterization of the bacterial and archaeal communities requires molecular tools. The isolates obtained from termite guts are typically not very abundant (Brune, 2006), and particularly the termite-specific lineages remain mostly uncultured.

**Termite gut flagellates.** Most termite gut flagellates belong to the phylum *Parabasalia* (Ohkuma and Brune, 2011). Three of the six classes of parabasalids (i.e., the traditional hypermastigids) are composed of species that are unique to the guts of lower termites (Noda et al., 2012). Their large cells with multiple flagella allow the phagocytosis of wood particles and high motility to prevent washout – probably adaptations to diet and gut environment. The ancestral *Trichomonadea*, which are found also in other habitats, are generally much smaller and feed on bacterial cells or dissolved nutrients. The larger cell size in some cellulolytic lineages is probably a response to the same evolutionary pressure (Noda et al., 2012). Many, but not all species of lower termites harbor oxymonadid flagellates (phylum *Preaxostyla*). Some lineages developed special holdfasts that attach to the hindgut cuticle, and the cells can be so small that they disappear within the bacterial biofilm (Tamschick and Radek, 2013).

Molecular studies revealed that the diversity of termite gut flagellates is larger than expected (e.g., Gile et al., 2013; Tai et al., 2014), and many species await a detailed phylogenetic and ultrastructural characterization. New lineages are still being discovered (e.g. Radek et al., 2014), and even seemingly identical morphospecies turn out to comprise different phylotypes (e.g., *Trichonympha* species: James et al., 2013; Tai et al., 2013; Appendix B), adding to the notion that each termite species has unique symbionts. First attempts to assess diversity and community structure of termite gut flagellates by amplicon sequencing indicated the need for an improved phylogenetic framework and universal primer sets (Tai et al., 2014; Rahman et al., 2015).

**Bacteria.** The bacterial gut microbiota of termites comprises only a few dominant phyla, with distinct differences between the major host groups (Fig. 2a). Over the past two decades, clone libraries of 16S rRNA genes have provided a wealth of information on bacterial diversity in a variety of termite guts (see Hongoh and Ohkuma, 2010; Ohkuma and Brune, 2011). Many of the libraries, particularly from older studies, were relatively small and the diversity of the gut communities was severely undersampled. The application of next-generation sequencing technologies resolved these issues and also allowed the study of differences in community structure across a wide range of termite species (Chapters 3, 4 and 6), among individuals of the same species obtained from geographically separated colonies or subjected to different dietary regimens (Boucias et al., 2013; Reid et al., 2014), or between different gut compartments or luminal fractions (Chapters 2 and 7; Mikaelyan et al., 2014).



**Fig. 2 | Diversity of the gut microbiota of termites and cockroaches.** (a) Phylum-level differences between representatives of major host groups (data from Chapter 3 of this thesis). Abbreviations: f, fungus-feeding; o, omnivorous; s, soil-feeding; w, wood-feeding). (b) Comparison of the phylogenetic trees (small-subunit rRNA) of *Candidatus Endomicrobium trichonymphae* and their *Trichonympha* hosts from various lower termites, illustrating that the strict cospeciation between the flagellates and their symbionts does not extend to the termite host (data from Ikeda-Ohtsubo and Brune, 2009). (c) Phylogenetic tree (16S rRNA) of the *Fibrobacteres* and the *candidate division TG3*, illustrating the presence of termite-specific clusters and their relationship to clones from other environments (modified from Mikaelyan et al., 2015).

*Spirochaetes* are characteristic members of all termite gut communities (Breznak and Leadbetter, 2006; Hongoh, 2011). They are phylogenetically highly diverse and comprise various monophyletic groups of termite-specific lineages (e.g., Hongoh et al., 2005; Mikaelyan et al., 2015). Individual lineages differ in abundance between host groups (Chapters 3, 4, 5), occur either as free-swimming cells or associated with the surface of flagellates (Breznak and Leadbetter, 2006; Inoue et al., 2008) or the fiber fraction (Mikaelyan et al., 2014), and may comprise different functional guilds (Brune, 2014). The highest proportion of spirochetes is found in wood-feeding termites, but there are striking differences between termite genera also among fungus-cultivating termites (Chapters 3 and 4; Otani et al., 2014).

*Bacteroidetes* are highly abundant in fungus-cultivating termites (e.g., Chapters 3 and 4; Hongoh et al., 2006) and contribute to the similarity of their gut microbiota with that of cockroaches (Chapter 3). Many of the predominant taxa (e.g., *Alistipes*, *Dysgonomonas*, *Paludibacter*, and *Parabacteroides*) with a general preference for intestinal habitats that are readily isolated from termite guts (Sakamoto and Ohkuma, 2013; Yang et al., 2014) but encountered also in the guts of mammals. However, there are also many family-level clades (e.g., *Bacteroidales* cluster V) that consist exclusively of representatives encountered in termites and cockroaches (Noda et al., 2009; Schauer et al., 2012; Chapter 3). The situation is similar for the *Firmicutes*, which are represented by common gut bacteria (mostly *Lachnospiraceae* and *Ruminococcaceae*) and highly specific lineages associated with the hindgut cuticle

of arthropods (e.g., *Candidatus* Arthromitus [Thompson et al., 2012]) or the alkaline gut compartments of higher termites (Chapter 2 and 7; Schmitt-Wagner et al., 2003; Thongaram et al., 2003). Also Proteobacteria are more abundant in cockroaches and *Macrotermitinae* than in other termite groups. Representatives encountered in all host groups are various *Desulfovibrio*-related lineages and another, deep-branching clade of *Deltaproteobacteria* (Rs-K70 Cluster), both of which comprise strains associated with flagellates (Strassert et al., 2012; Rosenthal et al., 2013).

While the above-mentioned phyla are consistently represented in all termites, others may be absent or of low abundance in some groups. Members of the *Elusimicrobia* make up a large proportion of the bacterial community in many lower termites (Chapter 3) and have been identified as endosymbionts of certain flagellates (*Candidatus* Endomicrobium [Brune, 2012]). The fiber-associated members of *Fibrobacteres* and the *cand. div.* TG3 are abundant in wood-feeding higher termites but have been detected also in other lineages (Chapters 3 and 4; Hongoh et al., 2006). *Planctomycetes* form large populations only in the posterior hindgut compartments of soil-feeding *Termitinae* but occur also in other groups (Chapter 3; Rahman et al., 2015). *Verrucomicrobia* related to *Candidatus* Nucleococcus, an intranuclear symbiont of termite gut flagellates (Sato et al., 2014), are abundant in several lower termites but found also in hosts that lack flagellates (Chapter 3).

**Archaea.** There are four major lineages of *Euryarchaeota* in termite guts: *Methanosarcinales*, *Methanomicrobiales*, *Methanobacteriales*, and a deep-branching clade distantly related to the non-methanogenic *Thermoplasmatales* (see Brune, 2010; Hongoh and Ohkuma, 2010). The latter were identified as a new order of methanogens, which was initially referred to as “*Methanoplasmatales*” but is now called *Methanomassiliicoccales* after the first isolate of the order. Comparative genome analysis of *Candidatus* Methanoplasma termitum, a highly enriched culture from the gut of a higher termite (Paul et al., 2012), with strains from the human gut indicated a new mode of energy metabolism in all members of this lineage (Lang et al., 2015).

Highest diversity of Archaea is found in the *Termitinae* (Chapter 6), particularly in soil-feeding lineages, where each hindgut compartment harbors a distinct archaeal community of presumably hydrogenotrophic and methylotrophic populations (see Brune, 2010; Paul et al., 2012). Archaeal communities in lower termites are dominated by *Methanobrevibacter* species, but Chapter 6 revealed that their diversity is larger than indicated by the earlier, clone-based studies. This includes an uncultured lineage of *Thaumarchaeota* that had been detected in soil-feeding termites (Chapter 6; Friedrich et al., 2001).

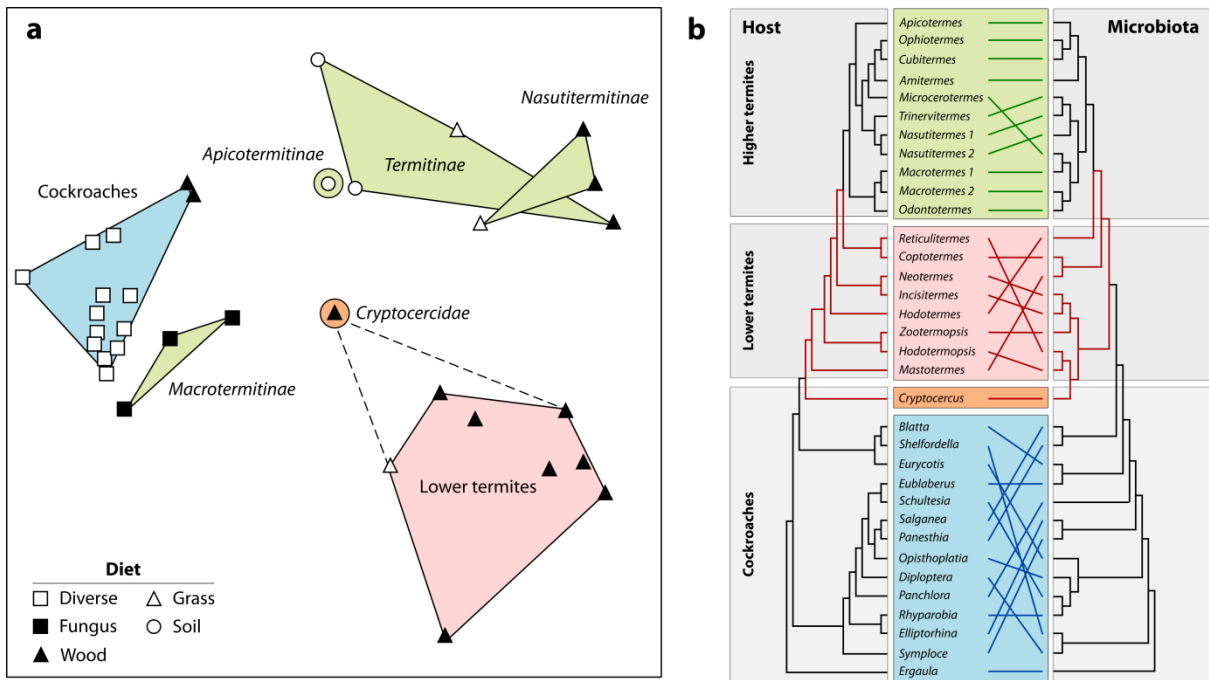
## Coevolutionary patterns

A coevolution between termites and members of their gut microbiota had been postulated already in the early studies of bacterial diversity in termite guts, which had observed clusters of phylogenetically related bacteria in closely related but geographically isolated hosts (e.g., Hongoh et al., 2005, Noda et al., 2009). The most recent high-throughput sequencing approaches, which allow a much larger taxon sampling, confirmed the presence of distinct coevolutionary patterns in the microbiota across the entire host range (Chapters 3, 4, 5 and 6; Tai et al., 2014; Rahman et al., 2015). While the distribution of certain microbial lineages is in agreement with a cospeciation of symbiont and host, others seem to reflect changes in microhabitats and functional niches that occurred during host evolution.

**Evidence for cospeciation.** The high similarity of the gut microbiota within and among colonies of the same termite species (e.g., Reid et al., 2014; Rahman et al., 2015) suggests that proctodeal trophallaxis stabilizes the microbial community structure within a colony and ensures the faithful transfer of symbionts across generations, which should ultimately lead to cospeciation. However, cladogenesis with the termite host has so far been firmly established only for the symbiosis of *Candidatus* *Azobacteroides* *pseudotriconymphae*, an endosymbiont of *Pseudotriconympha* flagellates, which again cospeciate with termites of the family *Rhinotermitidae*, giving rise to a termite-specific clade of bacterial symbionts (Noda et al., 2007). In the cases of the endosymbiotic *Candidatus* *Endomicrobium* *triconymphae* and the ectosymbiotic *Candidatus* *Armantifilum* *devescovinae* (Ikeda-Ohtsubo and Brune, 2009; Desai et al., 2010), the trees of the entire bacterial clades are perfectly congruent with that of their flagellate hosts (Fig. 2b), but lateral transfers of flagellates between different termite lineages limits periods of cospeciation between termites and bacterial symbionts.

**Habitat-specific lineages.** However, not all flagellate symbionts are cospeciating with their respective host. An example is *Candidatus* *Desulfovibrio* *triconymphae*, which is found in many but not all species of *Triconympha* and seems to be acquired in a non-hereditary manner (Sato et al., 2009). There are numerous other cases among *Elusimicrobia*, *Spirochaetes*, and *Bacteroidetes* (e.g., Desai et al., 2010; Inoue et al., 2008), where closely related bacteria colonize the surface or cytoplasm of distantly related flagellates. By contrast, the same habitats may be colonized also by entirely unrelated symbionts of presumably similar function, such as nitrogen fixation (Desai and Brune, 2012; Hongoh et al., 2008b), the provision of amino acids and vitamins (Hongoh et al., 2008a; Strassert et al., 2012), and the propelling of the host cell (Wenzel et al., 2003), which strongly suggests a selective effect of microenvironment and/or functional niche.

There seems to be a predisposition in certain bacterial lineages to colonize a particular habitat whenever the opportunity arises. Lineages of recognized flagellate symbionts are usually part of termite-specific clusters (Ikeda-Ohtsubo et al., 2010; Noda et al., 2009), which suggests that the gut serves as a reservoir of bacteria that are capable of colonizing suitable microhabitats. Members of the Bacteroidales are regularly encountered at the gut wall (e.g., Nakajima et al., 2006; Chapter 2), whereas certain lineages of *Clostridiales* show a clear preference for alkaline gut compartments (Schmitt-Wagner et al., 2003; Chapters 2 and 7). Many of the taxa among *Bacteroidetes* and *Firmicutes* that are encountered also in the guts of other insects and mammals seem to have a general preference for intestinal habitats.



**Fig. 3 | Phylogenetic patterns in the bacterial communities in the guts of termites and cockroaches.** (a) Similarity-based analysis of community structure shows a strong clustering according to host groups but also weak dietary signals (from Chapter 3). (b) Comparison of host phylogeny (data from Bourguignon et al., 2015) and phylogenetic cluster analysis of the "core microbiota" (data from Chapter 3), showing a clear separation of the gut microbiota according to the major hosts groups but a strong phylogenetic signal only in higher termites.

Microhabitats and niches are not constant factors but change during host evolution. The loss of flagellates in higher termites and the new opportunities arising for cellulolytic bacteria explain the disappearance of flagellate symbionts and the eventual appearance of presumably fiber-associated lineages in the wood-feeding groups (Chapters 2, 3, 4, 7 and 8; Mikaelyan et al., 2014). The termite-specific clades in the *Fibrobacteres* and *cand. div. TG3* (Fig. 2c) are composed of lineages that are apparently specific for certain termite genera but are not necessarily cospeciating across the entire host range (Chapter 3 and 4; Hongoh et al., 2006). The sister-group position of clones from leaf-feeding cockroaches suggests niche selection as the determinant (Mikaelyan et al., 2015).

**Host patterns and "core microbiota".** There is a strong phylogenetic signal in the overall structure of the termite gut microbiota (Fig. 3a). Changes in the abundance of particular lineages coincide with major events in host evolution (Chapters 3; Tai et al., 2014; Rahman et al., 2015) and are reflected also in inventories of functional genes (Yamada et al., 2007; Zhang and Leadbetter, 2012). However, the overall similarity of the microbiota of congeneric termites (Chapter 2; Rahman et al., 2015) contrasts with the simultaneous presence of phylotypes that are in obvious contradiction to cospeciation across the entire host range (Chapter 3, 4 and 6). A relationship between host phylogeny and community structure has been observed also in comparative analyses of the gut microbiota of other insects (Colman et al., 2012) and mammals (Ley et al., 2008; Ochman et al., 2010), but also here, the importance of dietary factors as determinants of community structure seems to increase with phylogenetic distance.

The phylogenetic analysis of an entire community will always result in a mixed signal of both hereditary and environmentally acquired lineages, some of which may be cospeciating with certain host clades, whereas others are only occasionally encountered. These problems are at least partially resolved if the analysis is restricted to a "core microbiota" of similar phylotypes that are represented in the majority of the host species and are selected using a tree-based definition (Chapters 2–8), i.e., classification against a curated reference database (e.g. Chapter 2 and 6; Mikaelyan et al., 2015). The size of the core is scale dependent because the chance that a lineage is represented across the entire host range decreases with increasing evolutionary distance (Chapter 3; Otani et al., 2014; Rahman et al., 2015).

A cluster analysis of the bacterial "core microbiota" across a wide range of termites and cockroaches resolved the general phylogenetic relationship among the major host groups (Fig. 3b). Within the host groups, however, a strong phylogenetic signal was present only among the higher termites (family *Termitidae*), where the trees of host and microbiota were almost congruent. Among the lower termites, which were represented mostly by members of different families, the trees were highly divergent and the sister-group position of *Cryptocercus* was lost, indicating that signals of cospeciation are weak or entirely absent at least at the family level. The entire absence of any phylogenetic signal among the cockroaches underlines that the high similarity in community structure between the phylogenetically highly divergent host lineages is not a product of coevolution.

## Drivers of community structure

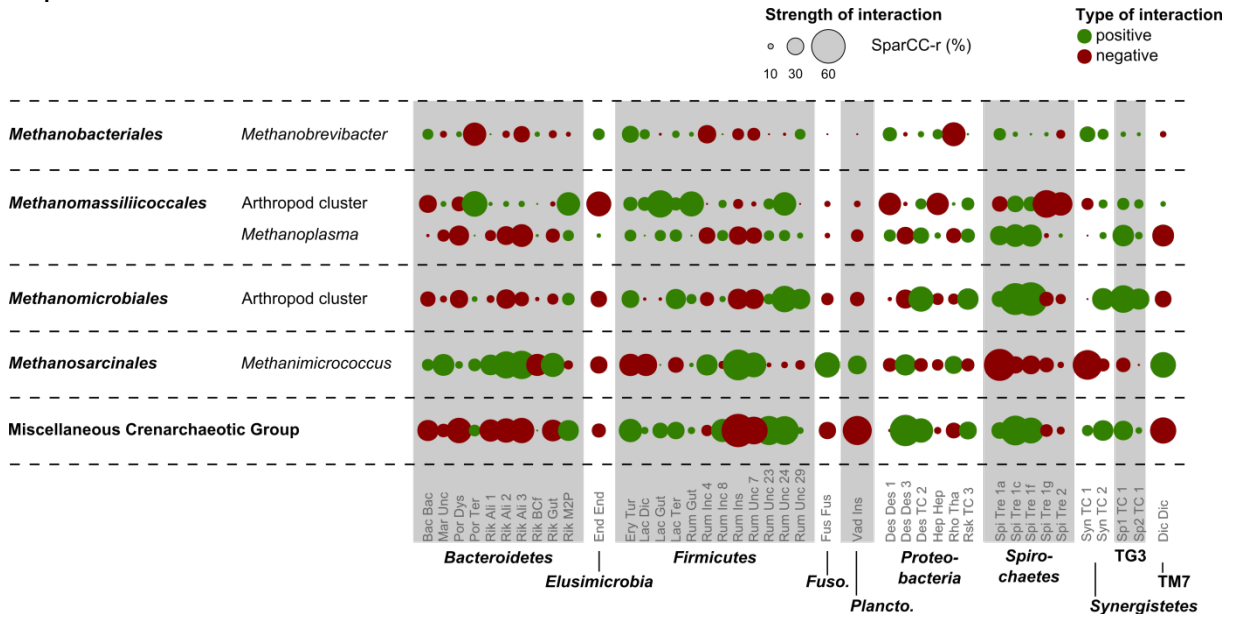
In contrast to the hereditary symbioses of insects and their intracellular symbionts, a digestive symbiosis is an open association, and both deterministic and stochastic processes should contribute to the assembly of the gut microbial community (Douglas, 2015). The guts of cockroaches are colonized by the ingestion of bacteria acquired from the environment, either together with the food source or by coprophagy (Klass et al., 2008, Nalepa, 2011), which introduces a strong stochastic element into habitat selection and would explain the highly similar yet individually variable community structure of cockroaches (Chapter 3; Schauer et al., 2014; Bauer et al., 2015). In termites, however, proctodeal trophallaxis adds another deterministic component, which attenuates the stochastic element by ensuring a reliable transfer of symbionts across generations and allows reciprocal adaptations that create host specificity and may eventually lead to cospeciation.

The obvious drivers of community structure in the termite gut ecosystem are differences in the microhabitats and functional niches, which are often difficult to distinguish. Examples are the microorganisms colonizing the gut wall, which contribute to oxygen reduction but at the same time must be equipped to deal with the toxic effects of reactive oxygen species, and the bacterial lineages colonizing the woofcud particles in higher termites, which must be able to attach to the fiber but also fulfill an important function in digestion (e.g. Chapter 8 and 9). The resulting patterns of biodiversity depend on the level of resolution (i.e., they differ between compartments, luminal fluid, gut wall, and other surfaces [Chapter 2, 5, 7 and 8; Nakajima et al., 2006; Mikaelyan et al., 2014]). Since a habitat always entails biotic factors, cooperative metabolic interactions between species have to be considered (Chapter 6; Rosenthal et al., 2011). In Chapter 6 potential interactions between the archaeal and the bacterial microbiota were discussed (Fig. 4).

Presently, direct experimental evidence for habitat selection is available only for vertebrate guts, where the inoculation of germ-free hosts with the microbiota of unrelated donors results in communities that closely resemble that of conventional individuals (Rawls et al., 2006; Seedorf et al., 2014). Also changes in community structure provoked by artificial diet shifts may help to identify cases of niche selection (e.g. Miyata et al., 2007). However, the strong discrepancies between the results of similar studies indicate that conclusions have to be regarded with caution, particularly if the underlying mechanisms are not investigated. In lower termites, changes in bacterial community structure or function (Boucias et al., 2013; Huang et al., 2013) may merely reflect shifts in the flagellate community in response to diet (Ikeda-Ohtsubo et al., 2010; Tanaka et al., 2006). Also the strong stochastic element in the gut microbiota of cockroaches may mask a diet response (Schauer et al., 2014).

It is reasonable to expect the presence of molecular mechanisms that allow the host to actively control the composition of its gut microbiota, either by favoring beneficial bacteria or by excluding undesired competitors and harmful pathogens (Douglas, 2015; Engel and Moran, 2013). Antimicrobial peptides and reactive oxygen species are important components of the innate immune system of insects and seem to regulate the abundance and composition of the microbiota in *Drosophila* (Douglas, 2014). Like other insects, termites produce small antifungal peptides and Gram-negative bacteria binding proteins with chitinase activity are secreted with the saliva (e.g., Bulmer et al. 2012). It is possible that host-specific members of the gut microbiota have developed mechanisms to evade such host defenses.





**Fig. 4 | Potential dependencies between the archaeal and bacterial members of the gut microbiota of arthropods shown as a correlogram of the SparCC correlation results.** Figure taken from Chapter 6 (Fig. 10). Data were filtered by applying a minimum threshold of  $|r|= 0.4$  for each archaeal and bacterial genus-level group. Colors of the points indicate directionality of the possible interaction: red, negative; green, positive. Point size indicates the strength of interaction as measured by SparCC-r. Bacterial genus-level groups are abbreviated according to the first three letters of the family and the genus-level group abbreviation (see Chapter 6 Tab. S2).

## Differences in functional potential between termite-specific groups and relatives from other environments

The termite gut is a small but complex ecosystem that consists of many distinct microhabitats colonized by microorganisms highly specific to termites (e.g. Hongoh et al., 2005; Chapters 3, 4 and 6). A plausible explanation of the appearance of such groups is that the reliable transfer of symbionts across termite generations allowed reciprocal adaptations that ultimately create host specificity.

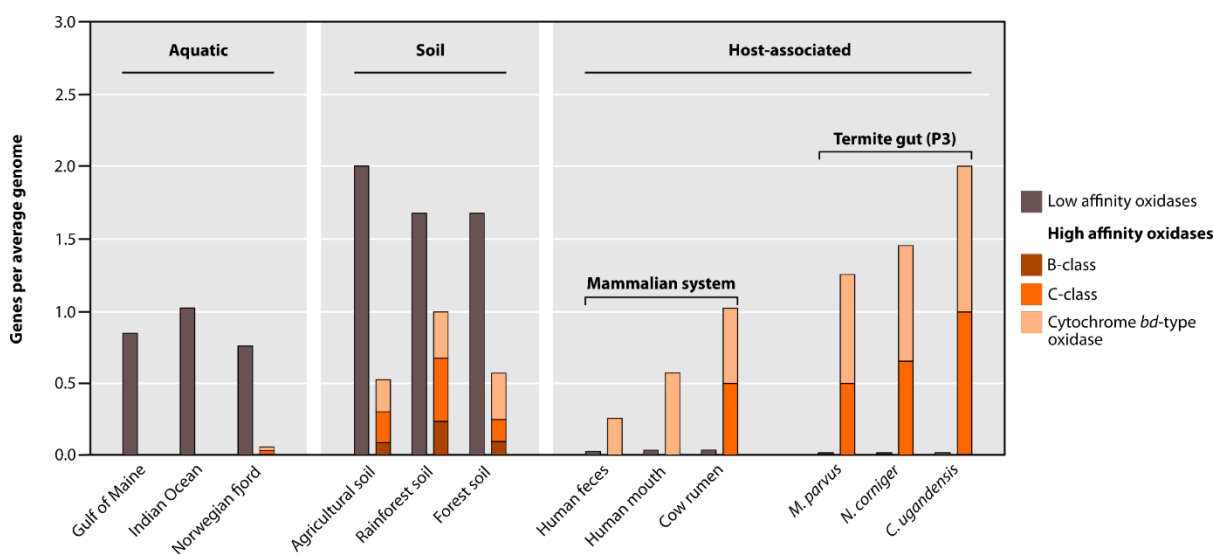
**Flagellate symbionts.** For intracellular organisms, i.e. endosymbionts, genome decay is the most evident genome modification (Thomson et al., 2003), which has been documented for the cockroach fat-body endosymbiont *Blattabacterium cuenoti* (e.g. Sabree, et al. 2012). Symbionts of termite gut flagellates are no exception to this rule (see the following example). When *Elusimicrobia* genomes are compared there is evidence that genomes of flagellate symbionts “decay”. While the genome size of the free-living *Elusimicrobium minutum* from the humivorous scarab beetle *Pachnoda ephippiata* is about 1.6 Mb (Herlemann et al., 2009), the flagellate endosymbiont Rs-D17 has a more than 30% smaller genome size (1.1 Mb; Hongoh et al. 2008a). Despite such streamlining adaptations, the latter was found to contain an interesting pseudo gene in the genome: *dnaA*. The loss of the DnaA protein is common for endosymbionts and might imply that replication is controlled by the host cell to prevent over-enthusiastic propagation of the bacterium within the cytosol (Thomson et al., 2003) — but it might also simply be the result of streamlining adaptations by using a recombination-dependent replication of the chromosome as a compensating mechanism (Hongoh et al., 2009). As a result, Rs-D17 appears to have lost the ability to survive in a host-independent manner which could explain the host-specificity and cospeciation between *Candidatus Endomicrobium trichonymphae* and its *Trychonympha* spp. hosts reported by Ikeda-Ohtsubo and Brune (2009) and Appendix B.

**Symbionts with no intracellular lifestyle.** Whether processes similar to those suggested for intracellular symbionts also apply to free-living symbionts in the termite gut is still unclear. However, adaptations of microorganisms might be needed to remain in the termite gut as a constant member. A major selective pressure for the microorganisms is to overcome outflow due to the low retention time of the total termite gut system (e.g. Brune 2014). These adaptations have probably compelled many organisms to associate with surface microhabitats.

Mikaelyan et al. (2014) showed that several spirochete lineages and members of the phylum *Fibrobacteres* and *cand. div. TG3* are associated with wood particles in wood-feeding higher termites, in Chapter 8 many carbohydrate binding modules could be identified that potentially support the fiber associated lifestyle. When the population genomes of termite-specific TG3 bacteria were compared to the draft genome of the only TG3 isolate, *Chitinivibrio alkaliphilus*, it was found that termite lineages contain 25% more cellulose degrading enzymes (Chapter 8), which might be an adaptation to the wood-feeding lifestyle of the termite host.

A striking feature of termite-specific members of the TG3 and *Fibrobacteres* is the presence of all genes required for nitrogen fixation (Chapter 8). In addition, genes encoding enzymes for the production of glutamate and other amino acids were identified in *cand. div. TG3* and *Fibrobacteres* bacteria. Since each of the cultured and sequenced representatives of *Fibrobacteres* and *cand. div. TG3* do not harbor genes for nitrogen (*Fibrobacter succinogenes* from Suen et al., 2011; *Chitinivibrio alkaliphilus* from Sorokin et al., 2014), acquisition of nitrogen fixation should be an adaptation of their relatives in the termite gut to the low nitrogen content in their host’s diet. Because *nifH* genes were frequently subject to horizontal genes transfer (Ohkuma et al., 1999; Yamada et al., 2007), their

taxonomic annotation by mere sequence comparison can be cumbersome. A preliminary phylogenetic analysis of the *nifH* genes from TG3 from Chapter 8 indicates that these bacteria likely acquired these genes from *Spirochaetes* (preliminary results from an ongoing collaborative project with Hao Zheng). If these sequences represent functional genes, these bacteria might be also responsible for the nitrogen fixation observed in wood-feeding higher termites (e.g. Slaytor and Chappell, 1994). Regulation of nitrogenase activity in bacteria and archaea is complex, and occurs also at the post-translational level (Leigh and Dodsworth, 2007). Therefore, without fully understanding the expression of these genes in the termite gut, the results from the metagenomes will have to be interpreted with some caution.

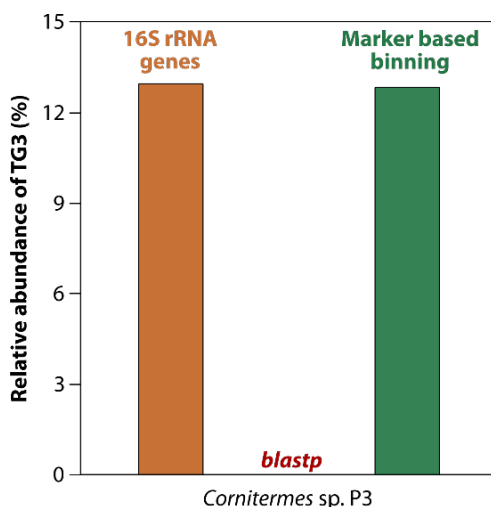


**Fig. 5 | Abundance of low and high affinity oxidases in metagenomes of microbial communities from different environments.** All data except the termite gut information was taken from Morris and Schmidt (2013). Oxidase genes in termite gut metagenomes (raw data from Chapter 7) were searched by a hidden Markov model search (Eddy, 2011) with hidden Markov models downloaded from the Craig Venter Institute webpage (<http://www.jcvi.org/cgi-bin/tigrfams/index.cgi>).

Interestingly, our analysis of the population genomes of *Fibrobacteres* and *cand. div.* TG3 also showed the presence of genes related to respiration under hypoxic conditions (chapter 8), a feature that these members do not share with their closest cultured relatives from the same phyla (*Fibrobacter succinogenes* from Suen et al., 2011; *Chitinivibrio alkaliphilus* from Sorokin et al., 2014). A subsequent analysis of the oxidase(s) in question was compared to metagenomes from other environments (Fig. 5). While datasets from soil or open water report a high abundance of low affinity oxidase genes those from host-associated microbial communities characteristically lack them (Morris and Schmidt, 2013). On the other hand, host-associated environments show high occurrences of high affinity oxidase genes. The termite gut microbiota of higher termites shows even the highest content in this comparison (raw data taken from Chapter 7). The high abundance of high affinity oxidase genes might be correlated with the considerable influx of oxygen into the termite hindgut, resulting from the large surface-to-volume ratio of termite guts (Brune, 1998). In contrast to the cow rumen where the oxygen influx seems negligible, in termite guts it results in a microoxic periphery which is estimated to cover 60% of the total P3 gut compartment volume (Breznak, 2000). This explains why many isolates

of termite microorganisms are capable of oxygen reduction (e.g. Wertz and Breznak, 2007; Wertz et al., 2012). Apart from being used for respiration, oxygen also is used as electron sink by fermenting bacteria, as indicated by the shift from the production of propionate to acetate during the metabolism of lactate (Tholen et al., 2000). Even strict anaerobes, such as methanogens colonizing the hindgut wall, are capable of oxygen reduction using  $H_2$  as reductant (Tholen et al., 2007).

**The need for reference-independent methods for functional analysis of termite-specific groups.** Metagenomic studies on the termite gut have offered several critical insights into the ecological processes governed by microbial symbionts. However, despite a rise in their popularity in recent years (see Brune, 2014), emphasis is typically laid on particular gene sets, most commonly glycosyl hydrolases (Warnecke et al., 2007; He et al., 2013; Poulsen et al. 2014). Most investigators annotate functions in their metagenomes using *blastp* (e.g. He et al., 2013; Poulsen et al. 2014; Chapter 7) or hidden Markov model searches (e.g. Prestat et al., 2014). Comparatively less effort, however, is placed on linking important ecological processes to particular members of the communities, despite one of the key goals of metagenomics being to ultimately provide a link between community structure and function (Riesenfeld et al., 2004).



**Fig. 6 | Relative abundance of TG3 in the metagenome of *Cornitermes* sp. determined by different methods.** Based on 16S rRNA gene abundance, TG3 bacteria in the P3 compartment of *Cornitermes* sp. constitute almost 13% of the bacterial community. In contrast, *blastp* based relative metagenome abundance by protein coding genes is zero due to the lack of genomes of related microorganisms in the reference database. When metagenomic fragments are binned into population genomes (Chapter 8), the relative metagenome abundance of TG3 bacteria is within the same range as the 16S rRNA gene based relative abundance. Data taken from Chapter 7 and 8.

Taxonomic annotation for metagenomic sequences is commonly based on top hit(s) in a reference database (*blastp*), an approach that is often incapable of accurate assignments due to the widespread horizontal gene transfer between organisms and due to the cultivation bias of microbial genomes present in databases (Hugenholtz, 2002). In chapter 8 population genomes of *Fibrobacteres* and *cand. div.* TG3 were obtained by marker-assisted compositional binning of multiple metagenomes. We optimized the parameters to obtain an optimal binning result for each termite gut microbiota metagenome. This was achieved by monitoring the distribution of 31 elite phylogenetic markers in the metagenomes of Wu and Scott (2012). We observed this method to be superior to *blastp* based annotations, especially for the accurate binning of sequences from lineages that are only distantly related to cultured strains (For comparison see Fig. 6; data from Chapters 7 and 8). The extreme example here being *cand. div.* TG3, from which the only available genome is that of its sole cultured representative *Chitinivibrio alkaliphilus*. Apart from being only a distant relative of TG3 bacteria in

the termite gut, because its genome is only a draft, it is excluded from commonly used metagenome classification databases.

For some processes in termite guts, especially in higher soil-feeding termites, many interesting processes were detected such as the anaerobic ammonium oxidation in the P4 compartment of *Cubitermes ugandensis* (Ngugi and Brune, 2012) — the responsible organisms, however, remain unknown. Good guesses would include highly abundant organisms whose relatives have been previously reported to catalyze these or related reactions. For the example of anaerobic ammonium oxidation, such guesses would include members of the *Planctomycetes* (Kuenen, 2008) or members of the *Thaumarchaeota* (related reaction; e.g. Pester et al., 2011) — both lineages occur in *Cubitermes* spp. (Chapters 3, 4 and 6) but are highly abundant in the P4 compartment (Friedrich et al. 2001; Köhler et al. 2008). The reconstruction of population genomes from metagenomes (e.g. from the data of Chapter 7) of those organisms, would allow to further elaborate hypotheses and to design new experiments to explore this unique reaction.

## References

- Bauer E, Lampert N, Mikaelyan A, Köhler T, Maekawa K and Brune A (2015).** Physicochemical conditions, metabolites, and community structure of the bacterial microbiota in the gut of wood-feeding cockroaches (*Blaberidae: Panesthiinae*). *FEMS Microbiol. Ecol.* (in Press).
- Boucias DG, Cai Y, Sun Y, Lietze V-U, Sen R, Raychoudhury R and Scharf ME (2013).** The hindgut lumen prokaryotic microbiota of the termite *Reticulitermes flavipes* and its responses to dietary lignocellulose composition. *Mol. Ecol.* 22:1836–1853.
- Bourguignon T, Lo N, Cameron SL, Šobotnik J, Hayashi Y, Shigenobu S, Watanabe D, Roisin Y, Miura T and Evans TA (2015).** The evolutionary history of termites as inferred from 66 mitochondrial genomes. *Mol. Biol. Evol.* 32:406–421.
- Breznak JA (2000).** Ecology of prokaryotic microbes in the guts of wood- and litter-feeding termites, pp. 209–231. *In* Abe T, Bignell DE and Higashi M (eds.). *Termites: Evolution, Sociality, Symbiosis, Ecology*, vol. 1. Kluwer Academic Publishers, Dordrecht.
- Breznak JA and Leadbetter JR (2006).** Termite gut spirochetes, pp. 318–329. *In* Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E (eds.), *The Prokaryotes*, 3rd ed., Volume 7: *Proteobacteria: Delta and Epsilon Subclasses. Deeply Rooting Bacteria*, vol. 7. Springer, New York.
- Brune A (1998).** Termite guts: the world's smallest bioreactors. *Trends Biotechnol.* 16:16–21.
- Brune A (2006).** Symbiotic associations between termites and prokaryotes, pp. 439–474. *In* Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E (eds.), *The Prokaryotes*, 3rd ed., Volume 1: *Symbiotic associations, Biotechnology, Applied Microbiology*, vol. 1. Springer, New York.
- Brune A (2010).** Methanogens in the digestive tract of termites, pp. 81–100. *In* Hackstein JHP (ed.), *(Endo)symbiotic Methanogenic Archaea*. Springer, Heidelberg.
- Brune A (2012).** Endomicrobia: intracellular symbionts of termite gut flagellates. *J. Endocytobiosis Cell Res.* 23:11–15.
- Brune A (2014).** Symbiotic digestion of lignocellulose in termite guts. *Nat. Rev. Microbiol.* 12:168–180.
- Bulmer MS, Denier D, Velenovsky J, Hamilton C (2012).** A common antifungal defense strategy in *Cryptocercus* woodroaches and termites. *Insect. Soc.* 59:469–478.
- Colman DR, Toolson EC and Takacs-Vesbach CD (2012).** Do diet and taxonomy influence insect gut bacterial communities? *Mol. Ecol.* 21:5124–5137.
- Desai MS and Brune A (2012).** *Bacteroidales* ectosymbionts of gut flagellates shape the nitrogen-fixing community in dry-wood termites. *ISME J.* 6:1302–1313.
- Desai MS, Strassert JFH, Meuser K, Hertel H, Ikeda-Ohtsubo W, Radek R and Brune A (2010).** Strict cospeciation of devescoviniid flagellates and *Bacteroidales* ectosymbionts in the gut of dry-wood termites (*Kalotermitidae*). *Environ. Microbiol.* 12:2120–2132.
- Douglas AE (2014).** The molecular basis of bacterial-insect symbiosis. *J. Mol. Biol.* 426:3830–3837.
- Douglas AE (2015).** Multiorganismal insects: Diversity and function of resident microorganisms. *Annu. Rev. Entomol.* 60:17–34.
- Eddy SR (2011).** Accelerated profile HMM searches. *PLoS Comp. Biol.* 7:e1002195.
- Engel P and Moran NA (2013).** The gut microbiota of insects - diversity in structure and function. *FEMS Microbiol. Rev.* 37:699–735.
- Friedrich MW, Schmitt-Wagner D, Lueders T and Brune A (2001).** Axial differences in community structure of *Crenarchaeota* and *Euryarchaeota* in the highly compartmentalized gut of the soil-feeding termite *Cubitermes orthognathus*. *Appl. Environ. Microbiol.* 67:4880–4890.
- Gile GH, Carpenter KJ, James ER, Scheffrahn RH and Keeling PJ (2013).** Morphology and Molecular Phylogeny of *Staurojoenina mulleri* sp. nov. (Trichonymphida, Parabasalia) from the Hindgut of the Kalotermitid *Neotermes jouteli*. *J. Euk. Microbiol.* 60:203–213.

- He S, Ivanova N, Kirton E, Allgaier M, Bergin C, Scheffrahn RH, Kyrpidis NC, Warnecke F, Tringe SG and Hugenholtz P (2013). Comparative metagenomic and metatranscriptomic analysis of hindgut paunch microbiota in wood- and dung-feeding higher termites. *PLoS ONE* 8:e61126.
- Herlemann DPR, Geissinger O, Ikeda-Ohtsubo W, Kunin V, Sun H, Lapidus A, Hugenholtz P and Brune A (2009). Genomic analysis of *Elusimicrobium minutum* the first cultivated representative of the phylum *Elusimicrobia* (formerly Termite Group 1). *Appl. Environ. Microbiol.* 75:2841–2849.
- Hongoh Y (2011). Toward the functional analysis of uncultivable, symbiotic microorganisms in the termite gut. *Cell. Mol. Life Sci.* 68:1311–1325.
- Hongoh Y and Ohkuma M (2010). Termite gut flagellates and their methanogenic and eubacterial symbionts, pp. 55–79. *In* Hackstein JHP (ed.), (Endo)symbiotic Methanogenic Archaea, Springer, Heidelberg.
- Hongoh Y, Deevong P, Inoue T, Moriya S, Trakulnaleamsai S, Ohkuma M, Vongkaluang C, Noparatnaraporn N and Kudo T (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.
- Hongoh Y, Ekpornprasit L, Inoue T, Moriya S, Trakulnaleamsai S, Ohkuma M, Noparatnaraporn N and Kudo T (2006). Intracolony variation of bacterial gut microbiota among castes and ages in the fungus-growing termite *Macrotermes gilvus*. *Mol. Ecol.* 15:505–516.
- Hongoh Y, Sharma VK, Prakash T, Noda S, Taylor TD, Kudo T, Sakaki Y, Toyoda A, Hattori M and Ohkuma M (2008a). Complete genome of the uncultured Termite Group 1 bacteria in a single host protist cell. *Proc. Natl. Acad. Sci. USA* 105:5555–5560.
- Hongoh Y, Sharma VK, Prakash T, Noda S, Toh H, Taylor TD, Kudo T, Sakaki Y, Toyoda A, Hattori M and Ohkuma M (2008b). Genome of an endosymbiont coupling N<sub>2</sub> fixation to cellulolysis within protist cells in termite gut. *Science* 322:1108–1109.
- Huang X-F, Bakker MG, Judd TM, Reardon KF and Vivanco JM (2013). Variations in diversity and richness of gut bacterial communities of termites (*Reticulitermes flavipes*) fed with grassy and woody plant substrates. *Microb. Ecol.* 65:531–536.
- Hugenholtz P (2002). Exploring prokaryotic diversity in the genomic era. *Genome Biol.* 3:S0003.
- Ikeda-Ohtsubo W, Faivre N and Brune A (2010). Putatively free-living 'Endomicrobia' – ancestors of the intracellular symbionts of termite gut flagellates? *Environ. Microbiol. Rep.* 2:554–559.
- Ikeda-Ohtsubo W und Brune A (2009). Cospeciation of termite gut flagellates and their bacterial endosymbionts: *Trichonympha* species and 'Candidatus Endomicrobium trichonymphae'. *Mol. Ecol.* 18:332–342.
- Inoue J, Noda S, Hongoh Y and Ohkuma M (2008). Identification of endosymbiotic methanogen and ectosymbiotic spirochetes of gut protists of the termite *Coptotermes formosanus*. *Microb. Environ.* 23:94–97.
- James ER, Tai V, Scheffrahn RH and Keeling PJ (2013). *Trichonympha burlesquei* from *Reticulitermes virginicus* and evidence against a cosmopolitan distribution of *Trichonympha agilis* in many termite hosts. *Int. J. Syst. Evol. Microbiol.* 63:3873–3876.
- Klass K-D, Nalepa CA and Lo N (2008). Wood-feeding cockroaches as models for termite evolution (*Insecta: Dictyoptera*): *Cryptocercus* vs. *Parasphaeria boleiriana*. *Mol. Phylogenet. Evol.* 46:809–817.
- Köhler T, Dietrich C, Scheffrahn RH and Brune A (2012). High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation of the gut in wood-feeding higher termites (*Nasutitermes* spp.). *Appl. Environ. Microbiol.* 78:4691–4701.
- Köhler T, Stingl U, Meuser K and Brune A (2008). Novel lineages of *Planctomycetes* densely colonize the alkaline gut of soil-feeding termites (*Cubitermes* spp.). *Environ. Microbiol.* 10:1260–1270.
- Kuennen JG (2008). Anammox bacteria: from discovery to application. *Nat. Rev. Microbiol.* 6:320–326.
- Lang K, Schuldes J, Klingl A, Poehlein A, Daniel R and Brune A (2015). New mode of energy metabolism in the seventh order of methanogens as indicated by comparative genome analysis of "Candidatus Methanoplasma termitum". *Appl. Environ. Microbiol.* 81:1338–52.
- Leigh JA and Dodsworth JA (2007). Nitrogen regulation in bacteria and archaea. *Annu. Rev. Microbiol.* 61:349–377.

- Ley RE, Hamady M, Lozupone C, Turnbaugh PJ, Ramey RR, Bircher JS, Schlegel ML, Tucker TA, Schrenzel MD, Knight R and Gordon JI (2008). Evolution of Mammals and Their Gut Microbes. *Science* 320:1647–1651.
- Mikaelyan A, Lampert N, Köhler T, Rohland J, Boga H, Meuser K and Brune A (2015). DictDb: an expanded reference database for the highly resolved classification of the bacterial gut microbiota of termites and cockroaches. Submitted manuscript.
- Mikaelyan A, Strassert JFH, Tokuda G and Brune A (2014). The fiber-associated cellulolytic bacterial community in the hindgut of wood-feeding higher termites (*Nasutitermes* spp.). *Environ. Microbiol.* 16:2711–2722.
- Miyata R, Noda N, Tamaki H, Kinjyo K, Aoyagi H, Uchiyama H and Tanaka H (2007). Influence of feed components on symbiotic bacterial community structure in the gut of the wood-feeding higher termite *Nasutitermes takasagoensis*. *Biosci. Biotechnol. Biochem.* 71:1244–1251.
- Morris RL and Schmidt TM (2013). Shallow breathing: bacterial life at low O<sub>2</sub>. *Nat. Rev. Microbiol.* 11:205–212.
- Nakajima H, Hongoh Y, Noda S, Yoshida Y, Usami R, Kudo T and Ohkuma M (2006). Phylogenetic and morphological diversity of *Bacteroidales* members associated with the gut wall of termites. *Biosci. Biotechnol. Biochem.* 70:211–218.
- Nalepa CA (2011). Altricial development in wood-feeding cockroaches: the key antecedent of termite eusociality, pp. 69–95. In Bignell DE, Roisin Y, Lo N (eds.), *Biology of Termites: A Modern Synthesis*. Springer, Dordrecht.
- Ngugi DK and Brune A (2012). Nitrate reduction, nitrous oxide formation, and anaerobic ammonia oxidation to nitrite in the gut of soil-feeding termites (*Cubitermes* and *Ophiotermes* spp.). *Environ. Microbiol.* 14:860–871.
- Noda S, Hongoh Y, Sato T and Ohkuma M (2009). Complex coevolutionary history of symbiotic *Bacteroidales* bacteria of various protists in the gut of termites. *BMC Evol. Biol.* 9:158.
- Noda S, Kitade O, Inoue T, Kawai M, Kanuka M, Hiroshima K, Hongoh Y, Constantino R, Uys V, Zhong J, Kudo T and Ohkuma M (2007). Cospeciation in the triplex symbiosis of termite gut protists (*Pseudotriconympha* spp.), their hosts, and their bacterial endosymbionts. *Mol. Ecol.* 16:1257–1266.
- Noda S, Mantini C, Meloni D, Inoue J-I, Kitade O, Viscogliosi E, Ohkuma M (2012). Molecular phylogeny and evolution of parabasalids with improved taxon sampling and new protein markers of actin and elongation factor-1. *PLoS ONE* 7:e29938.
- Ochman H, Worobey M, Kuo C-H, Ndjanga J-BN, Peeters M, Hahn B, Hugenholtz P (2010). Evolutionary relationships of wild hominids recapitulated by gut microbial communities. *PLoS Biol.* 8:e1000546.
- Ohkuma M and Brune A (2011). Diversity, structure, and evolution of the termite gut microbial community, pp. 413–438. In Bignell DE, Roisin Y, Lo N (eds.), *Biology of Termites: A Modern Synthesis*. Springer, Dordrecht.
- Ohkuma M, Noda S, Kudo T (1999). Phylogenetic diversity of nitrogen fixation genes in the symbiotic microbial community in the gut of diverse termites. *Appl. Environ. Microbiol.* 65:4926–4934.
- Otani S, Mikaelyan A, Nobre T, Hansen LH, Koné NA, Sørensen SJ, Aanen DK, Boomsma JJ, Brune A and Poulsen M (2014). Identifying the core microbial community in the gut of fungus-growing termites. *Mol. Ecol.* 23:4631–4644.
- Paul, K, Nonoh JO, Mikulski L and Brune A (2012). "Methanoplasmatales," *Thermoplasmatales*-related archaea in termite guts and other environments, are the seventh order of methanogens. *Appl. Environ. Microbiol.* 78:8245–8253.
- Pester M, Schleper C and Wagner M (2011). The *Thaumarchaeota*: an emerging view of their phylogeny and ecophysiology. *Curr. Opin. Microbiol.* 14:300–306.
- Poulsen M, Hu H, Li C, Chen Z, Xu L, Otani S, Nygaard S, Nobre T, Klaubauf S, Schindler PM, Hauser F, Pan H, Yang Z, Sonnenberg ASM, de Beer AW, Zhang Y, Wingfield MJ, Grimmelikhuijzen CJP, de Vries RP, Korb J, Aanen DK, Wang J, Boomsma JJ and Zhang G (2014). Complementary symbiont contributions to plant decomposition in a fungus-farming termite. *Proc. Natl. Acad. Sci. USA* 111:14500–14505.
- Prestat E, David MM, Hultman J, Tǿ N, Lamendella R, Dvornik J, Mackelprang R, Myrold DD, Jumpponen A, Tringe SG, Holman E, Mavromatis K, Jansson JK (2014). FOAM (Functional Ontology Assignments for Metagenomes): a Hidden Markov Model (HMM) database with environmental focus. *Nucleic Acids Res.* 2014 42:e145.
- Radek R, Strassert JFH, Krüger J, Meuser K, Scheffrahn RH and Brune A (2014). Phylogeny and ultrastructure of *Oxymonas jouteli*, a rostellum-free species, and *Opisthomitus longiflagellatus* sp. nov., oxymonadid flagellates from the gut of *Neotermes jouteli*. *Protist* 165:384–399



- Rahman NA, Parks DH, Willner DL, Engelbrekton AL, Goffredi SK, Warncke F, Scheffrahn H and Hugenholtz P (2015). A molecular survey of Australian and North American termite genera indicates that vertical inheritance is the primary force shaping termite gut microbiomes. *Microbiome* (in press)
- Rawls JF, Mahowald MA, Ley RE and Gordon JI (2006). Reciprocal Gut Microbiota Transplants from Zebrafish and Mice to Germ-free Recipients Reveal Host Habitat Selection. *Cell* 127:423–433.
- Reid NM, Addison SL, West MA, Lloyd-Jones G (2014). The bacterial microbiota of *Stoloterms ruficeps* (Stolotermitidae), a phylogenetically basal termite endemic to New Zealand. *FEMS Microbiol. Ecol.* 90:678–688.
- Riesenfeld CS, Schloss PD and Handelsman J (2004). Metagenomics: genomic analysis of microbial communities. *Ann. Rev. Genet.* 38:525–552.
- Rosenthal AZ, Matson EG, Eldar A and Leadbetter JR (2011). RNA-seq reveals cooperative metabolic interactions between two termite-gut spirochete species in co-culture. *ISME J.* 5:1133–1142.
- Rosenthal AZ, Zhang X, Lucey KS, Ottesen EA, Trivedi V, Choi HMT, Pierce NA and Leadbetter JR (2013). Localizing transcripts to single cells suggests an important role of uncultured *Deltaproteobacteria* in the termite gut hydrogen economy. *Proc. Natl. Acad. Sci. USA* 110:16163–16168
- Sabree ZL, Huang CY, Arakawa G, Tokuda G, Lo N, Watanabe H and Moran NA (2012). Genome shrinkage and loss of nutrient-providing potential in the obligate symbiont of the primitive termite *Mastotermes darwiniensis*. *Appl. Environ. Microbiol.* 78:204–210.
- Sakamoto M and Ohkuma M (2013). *Bacteroides reticulotermidis* sp. nov., isolated from the gut of the subterranean termite (*Reticulitermes speratus*). *Int. J. Syst. Evol. Microbiol.* 63:691–695.
- Sato T, Hongoh Y, Noda S, Hattori S, Ui S and Ohkuma M (2009). *Candidatus* Desulfovibrio trichonymphae, a novel intracellular symbiont of the flagellate *Trichonympha agilis* in termite gut. *Environ. Microbiol.* 11:1007–1015.
- Sato T, Kuwahara H, Fujita K, Noda S, Kihara K, Yamada A, Ohkuma M and Hongoh Y (2014). Intranuclear verrucomicrobial symbionts and evidence of lateral gene transfer to the host protist in the termite gut. *ISME J.* 8:1008–1019.
- Schauer C, Thompson CL and Brune A (2012). The bacterial community in the gut of the cockroach *Shelfordella lateralis* reflects the close evolutionary relatedness of cockroaches and termites. *Appl. Environ. Microbiol.* 78:2758–2767.
- Schauer C, Thompson CL and Brune A (2014). Pyrotag sequencing of the gut microbiota of the cockroach *Shelfordella lateralis* reveals a highly dynamic core but only limited effects of diet on community structure. *PLoS ONE* 9:e85861.
- Schmitt-Wagner D, Friedrich MW, Wagner B and Brune A (2003). Phylogenetic diversity, abundance, and axial distribution of bacteria in the intestinal tract of two soil-feeding termites (*Cubitermes* spp.). *Appl. Environ. Microbiol.* 69:6007–6017.
- Seedorf H, Griffin NW, Ridaura VK, Reyes A, Cheng J, Rey FE, Smith MI, Simon GM, Scheffrahn RH, Woebken D, Spormann AM, Van Treuren W, Ursell LK, Pirrung M, Robbins-Pianka A, Cantarel BL, Lombard V, Henrissat B, Knight R and Gordon JI (2014). Bacteria from diverse habitats colonize and compete in the mouse gut. *Cell* 159:254–266
- Slaytor M and Chappell DJ (1994). Nitrogen metabolism in termites. *Comp. Biochem. Physiol.* 107:1–10.
- Sorokin DY, Gumerov VM, Rakitin AL, Beletsky AV, Damsté JSS, Muyzer G, Mardanov AV and Ravin NV (2014). Genome analysis of *Chitinivibrio alkaliphilus* gen. nov., sp. nov., a novel extremely haloalkaliphilic anaerobic chitinolytic bacterium from the candidate phylum TG3. *Environ. Microbiol.* 16:1549–1565.
- Strassert JFH, Köhler T, Wienemann THG, Ikeda-Ohtsubo W, Faivre N, Franckenberg S, Plarre R, Radek R and Brune A (2012). '*Candidatus* Ancillula trichonymphae', a novel lineage of endosymbiotic *Actinobacteria* in termite gut flagellates of the genus *Trichonympha*. *Environ. Microbiol.* 14:3259–3270.
- Suen G, Weimer PJ, Stevenson DM, Aylward FO, Boyum J, Deneke J, Drinkwater C, Ivanova NN, Mikhailova N, Chertkov O, Goodwin LA, Currie CR, Mead D and Brumm PJ (2011). The complete genome sequence of *Fibrobacter succinogenes* S85 reveals a cellulolytic and metabolic specialist. *PLoS One* 6(4):e18814.
- Tai V, James ER, Nalepa CA, Scheffrahn RH, Perlman SJ and Keeling PJ (2014). The role of host phylogeny varies in shaping microbial diversity in the hindguts of lower termites. *Appl. Environ. Microbiol.* 81:1059–70

- Tai V, James ER, Perlman SJ and Keeling PJ (2013).** Single-cell DNA barcoding using sequences from the small subunit rRNA and internal transcribed spacer region identifies new species of *Trichonympha* and *Trichomitopsis* from the hindgut of the Termite *Zootermopsis angusticollis*. PLoS ONE 8:e58728.
- Tamschick S and Radek R (2013).** Colonization of termite hindgut walls by oxymonad flagellates and prokaryotes in *Incisitermes tabogae*, *I. marginipennis* and *Reticulitermes flavipes*. Eur. J. Protistol. 49:1–14.
- Tanaka H, Aoyagi H, Shina S, Dodo Y, Yoshimura T, Nakamura R, Uchiyama H (2006).** Influence of the diet components on the symbiotic microorganisms community in hindgut of *Coptotermes formosanus* Shiraki. Appl. Microbiol. Biotechnol. 71:907–917.
- Tholen A and Brune A (2000).** Impact of oxygen on metabolic fluxes and *in situ* rates of reductive acetogenesis in the hindgut of the wood-feeding termite *Reticulitermes flavipes*. Environ. Microbiol. 2:436–449.
- Tholen A, Pester M and Brune A (2007).** Simultaneous methanogenesis and oxygen reduction by *Methanobrevibacter cuticularis* at low oxygen fluxes. FEMS Microbiol. Ecol. 2:303–312.
- Thompson CL, Vier R, Mikaelyan A, Wienemann T and Brune A (2012).** 'Candidatus Arthromitus' revised: Segmented filamentous bacteria in arthropod guts are members of *Lachnospiraceae*. Environ. Microbiol. 14:1454–1465.
- Thomson N, Bentley S, Holden M and Parkhill J (2003).** Fitting the niche by genomic adaptation. Nat. Rev. Microbiol. 1:92–93.
- Thongaram T, Kosono S, Ohkuma M, Hongoh Y, Kitada M, Yoshinaka T, Trakulnaleamsai S, Noparatnaraporn N and Kudo T (2003).** Gut of higher termites as a niche for alkaliphiles as shown by culture-based and culture-independent studies. Microb. Environ. 18:152–159.
- Warnecke F, Luginbühl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, Cayouette M, McHardy AC, Djordjevic G, Aboushadi N, Sorek R, Tringe SG, Podar M, Martin HG, Kunin V, Dalevi D, Madejska J, Kirton E, Platt D, Szeto E, Salamov A, Barry K, Mikhailova N, Kyrpides NC, Matson EG, Ottesen EA, Zhang X, Hernández M, Murillo C, Acosta LG, Rigoutsos I, Tamayo G, Green BD, Chang C, Rubin EM, Mathur EJ, Robertson DE, Hugenholtz P and Leadbetter JR (2007).** Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. Nat. 450:560–565.
- Wenzel M, Radek R, Brugerolle G and König H (2003).** Identification of the ectosymbiotic bacteria of *Mixotrichia paradoxa* involved in movement symbiosis. Eur. J. Protistol. 39:11–24.
- Wertz JT and Breznak JA (2007).** *Stenoxybacter acetivorans* gen. nov., sp. nov., an acetate-oxidizing obligate microaerophile among diverse O<sub>2</sub>-consuming bacteria from termite guts. Appl. Environ. Microbiol. 73:6819–6828.
- Wertz JT, Kim E, Breznak JA, Schmidt TM and Rodrigues JLM (2012).** Genomic and physiological characterization of the *Verrucomicrobia* isolate *Diplosphaera colotermitum* gen. nov., sp. nov. reveals microaerophily and nitrogen fixation genes. Appl. Environ. Microbiol. 78:1544–1555.
- Wu M and Scott AJ (2012).** Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. Bioinf. 28:1033–1034.
- Yamada A, Inoue T, Noda Y, Hongoh Y and Ohkuma M (2007).** Evolutionary trend of phylogenetic diversity of nitrogen fixation genes in the gut community of wood-feeding termites. Mol. Ecol. 16:3768–3777.
- Yang YJ, Zhang N, Ji SQ, Lan X, Zhang KD, Shen YL, Li FL, Ni JF (2014).** *Dysgonomonas macrotermidis* sp. nov., isolated from the hindgut of a fungus-growing termite. Int. J. Syst. Evol. Microbiol. 64:2956–2961.
- Zhang X and Leadbetter JR (2012).** Evidence for cascades of perturbation and adaptation in the metabolic genes of higher termite gut symbionts. mBio 3:e00223–12.

## The complete mitogenomes of six higher termite species reconstructed from metagenomic datasets

Carsten Dietrich and Andreas Brune

*Affiliations:* Max Planck Institute for Terrestrial Microbiology, 35043 Marburg, Germany | *This chapter is published in:* Dietrich C and Brune A (2015). The complete mitogenomes of six higher termite species reconstructed from metagenomic datasets (*Cornitermes* sp., *Cubitermes ugandensis*, *Microcerotermes parvus*, *Nasutitermes corniger*, *Neocapritermes taracua*, and *Termes hospes*). *Mitochondr. DNA Dec* 4:1-2. | *Contributions:* C.D. Conceived the study, analyzed data, and wrote the manuscript. A.B. conceived the study and contributed to the manuscript.

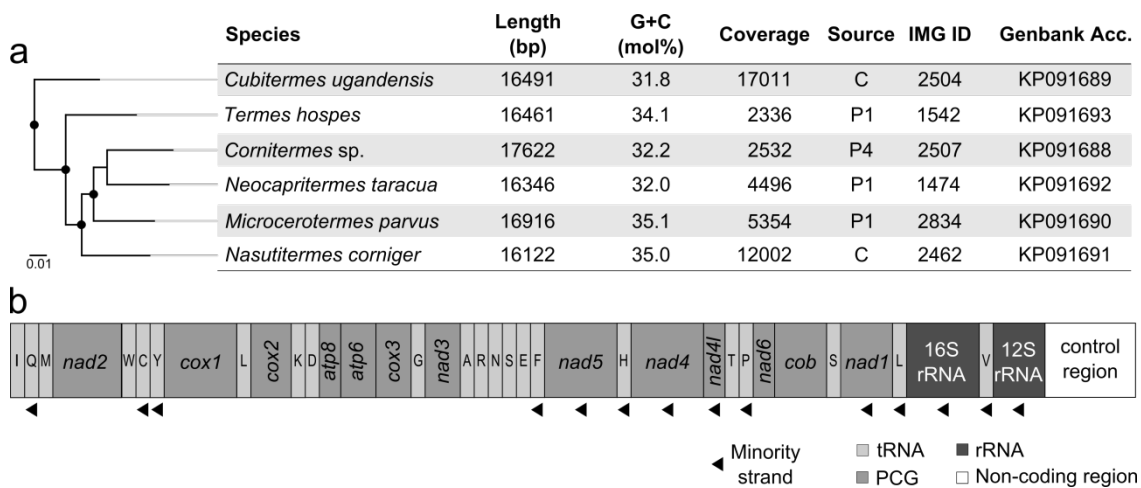
### Abstract

---

We reconstructed the complete mitochondrial genomes of six higher termite species from metagenomic datasets of their isolated hindgut compartments. The reads were retrieved and assembled with the mitochondrial-baiting and iterative-mapping algorithm (*MITObim*), which yielded closed mitogenomes without additional finishing efforts (average coverage ranging from 2,300- to 17,000-fold). The genomes ranged from 16.1 to 17.6 kbp in size and had GC contents between 32 and 35 mol%; each contained the same 37 genes present also in the mitochondria of other termite species. Our study substantially increases the number of termite mitogenomes available for phylogenetic studies and offers a facile strategy for identifying host species in metagenomic studies of their associated microbiota.

## Announcement

Termites are a diverse and ecological important group of insects that originated from an ancestral cockroach (Inward et al., 2007). The detailed analysis of the evolutionary relationships among termites requires multi-gene phylogenies, which can be obtained from entire mitochondrial genomes (Cameron et al., 2013). However, despite the scientific interest in termites and their role on lignocellulose digestion (Brune, 2014), only few mitochondrial genomes have been sequenced. While most previous studies used primer walking as the sequencing strategy (Cameron and Whiting, 2007; Tokuda et al., 2012; Cameron et al., 2013), Qian (2014) assembled the mitogenome of *Zootermopsis nevadensis* from the data of the recently published first termite genome (Terrapon et al., 2014). Here, we show that this strategy can be employed to reconstruct complete mitochondrial genomes from the host information contained in metagenomic datasets of their associated microbiota.



**Fig. 1 | Characteristics of the mitogenomes of six higher termite species.** (a) Maximum likelihood tree based on the amino acid sequences of all protein-coding genes (PCG) of the mitogenomes (●, > 90% bootstrap support) and a brief description of their associated metadata. Source, dataset from the gut compartment with the highest frequency of host reads; IMG ID, taxon object identifier of the raw read data in the Integrated Microbial Genomes database (330000xxxx; <http://img.jgi.doe.gov/>). (b) Linearized Schematic gene map (identical for all species); genes encoding tRNAs and rRNAs are labeled according their product for simplicity.

Using the quality-trimmed reads obtained from the metagenomes of six higher termite species, we assembled the mitogenomes using the mitochondrial baiting and iterative mapping algorithm (MITObim; Hahn et al., 2013) with default parameters and the mitogenome of *Nasutitermes triodiae* (accession number JX144940) as initial template. The assemblies yielded high-quality mitogenomes with coverages ranging from 2300- to 17000-fold (Fig. 1a). The genome length ranged from 16.0 to 17.6 kbp, and the G+C contents varied between 32 and 35 mol%. The nucleotide compositions of all mitogenomes were asymmetric (A: 41.4–43.6; C: 19.8–22.4; G: 11.6–12.9; T: 23.5–26.1); the values were in the same range as that of the four other mitogenomes of higher termites sequenced to date. The sequences were analyzed using the Improved de novo Metazoan Mitochondrial Genome Annotation (MITOS) webserver (Bernt et al. 2013; <http://mitos.bioinf.uni-leipzig.de/>), which provides

a convenient and accurate means of gene prediction and annotation of mitochondrial genomes, including the necessary quality control.

All mitogenomes contained the same 37 genes that are present in all termite mitogenomes sequenced to date; the genes varied slightly in length: 13 protein-coding genes (155–1682 bp), a 12S rRNA (805–817 bp) gene, a 16S rRNA (1357–1388 bp) gene, and 22 tRNA genes (62–75 bp). Gene order and directionality were identical in all genomes; as in other termites, the rRNAs and several subunits of respiratory complex I were encoded on the minority strand (Fig. 1b). Also the intergenic regions (1–94 bp), intergenic overlapping regions (1–41 bp), and A+T-rich control region contributed to the length heterogeneity of the genomes.

In addition to illustrating that mitochondrial genomes can be reconstructed from metagenomic datasets, our study substantially increases the number of available mitogenomes of higher termites, which will be important for future phylogenetic studies of this important insect group. Furthermore, the study contributes novel sequences to the sets of molecular markers, such as cytochrome oxidase and 16S rRNA genes, that are frequently used for identifying host species via DNA barcoding methods in large-scale surveys of microbial symbionts (Dietrich et al., 2014).

## References

- Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsche G, Pütz J, Middendorf M and Stadler F (2013).** *MITOS: Improved de novo Metazoan Mitochondrial Genome Annotation*. *Mol. Phylogenet. Evol.* 69(2):313-319
- Brune A (2014).** Symbiotic digestion of lignocellulose in termite guts. *Nat. Rev. Microbiol.* 12:168-180.
- Cameron SL, Lo N, Bourguignon T, Svenson GJ and Evans TA (2012).** A mitochondrial genome phylogeny of termites (Blattodea: Termitoidea): robust support for interfamilial relationships and molecular synapomorphies define major clades. *Mol. Phylogenet. Evol.* 65(1):163-73.
- Cameron SL and Whiting MF (2007).** Mitochondrial genomic comparisons of the subterranean termites from the Genus *Reticulitermes* (Insecta: Isoptera: Rhinotermitidae). *Genome* 50(2): 188-202.
- Dietrich C, Köhler T and Brune A (2014).** The cockroach origin of the termite gut microbiota: patterns in bacterial community structure reflect major evolutionary events. *Appl. Environ. Microbiol.* 80:2261-2269.
- Hahn C, Bachmann L and Chevreux B (2013).** Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. *Nucl. Acids Res.* 41:e129.
- Inward D, Beccaloni G and Eggleton P (2007).** Death of an order: a comprehensive molecular phylogenetic study confirms that termites are eusocial cockroaches. *Biol. Lett.* 3:331-335.
- Qian ZQ (2014).** The complete mitogenome of the dampwood termite *Zootermopsis nevadensis* (Insecta: Isoptera: Termopsidae). doi:10.3109/19401736.2014.936419.
- Terrapon N, Li C, Robertson HM, Ji L, Meng X, Booth W, Chen Z, et al. (2014).** Molecular traces of alternative social organization in a termite genome. *Nat. Commun.* 5:3636
- Tokuda G, Isagawa H and Sugio K (2012).** The complete mitogenome of the Formosan termite, *Coptotermes formosanus* Shiraki. *Insect. Soc.* 59:17-24.

## Population structure of Endomicrobia in single host cells of termite gut flagellates (*Trichonympha* spp.)

Hao Zheng<sup>1</sup>, Carsten Dietrich<sup>1</sup>,  
Claire L. Thompson<sup>2</sup>, Katja Meuser<sup>1</sup>, Andreas Brune<sup>1,2</sup>

*Affiliations:*<sup>1</sup> Max Planck Institute for Terrestrial Microbiology, 35043 Marburg, Germany; <sup>2</sup> LOEWE Center for Synthetic Microbiology, SYNMIKRO, Philipps-Universität Marburg, 35043 Marburg, Germany. | *This chapter is published in:* Zheng H, Dietrich C, Thompson CL, Meuser K und Brune A (2015). Population structure of Endomicrobia in single host cells of termite gut flagellates (*Trichonympha* spp.). *Microb. Environ.* 30:92-98. | *Contributors:* H.Z. planned and carried out most experiments, analyzed data, discussed results, and wrote the manuscript. C.D. analyzed data, discussed results, and contributed to the manuscript. C.L.T. conceived the study and carried out experiments. K.M. carried out experiments. A.B. conceived the study, discussed results, wrote the manuscript, and secured funding.

### Abstract

---

The gut microbiota of many phylogenetically lower termites is dominated by cellulolytic flagellates of the genus *Trichonympha*, which are consistently associated with bacterial symbionts. In the case of “Endomicrobia”, an unusual lineage of endosymbionts of the *Elusimicrobia* phylum that is present also in other gut flagellates, previous studies have documented strict host specificity, leading to cospeciation of “*Candidatus* Endomicrobium trichonymphae” with their respective flagellate hosts. However, it remained unclear if one *Trichonympha* species can harbor more than one Endomicrobia phylotype. In this study, we picked single *Trichonympha* cells from the gut of *Zootermopsis nevadensis* and *Reticulitermes santonensis* and characterized their Endomicrobia populations based on the internal transcribed spacer (ITS) region sequences. We found that each host cell harbors a homogeneous population of symbionts that were specific for their respective host species but phylogenetically distinct between each host lineage, corroborating that cospeciation is caused by vertical inheritance. The experimental design of our study also allowed identification of an unexpectedly large amount of tag-switching between samples, which indicates that any high-resolution analysis of microbial community structure using the pyrosequencing technique has to be interpreted with great caution.

## Introduction

The termite hindgut is colonized by dense assemblages of prokaryotes (bacteria and archaea) and eukaryotic protists that play essential roles in both digestion and host nutrition (Brune and Ohkuma, 2011; Hongoh, 2011). The gut microbiota of the phylogenetically lower termites is dominated by cellulolytic flagellates that are unique to termites and comprise diverse lineages from the phyla *Parabasalia* and *Preaxostyla* (order *Oxymonadida*) (Brune, 2014). In most cases, the flagellates are associated with large numbers of prokaryotic symbionts that colonize both the surface and cytoplasm (Hongoh and Ohkuma, 2010; Ohkuma and Brune, 2011) and sometimes even the nucleus of their hosts (Sato et al., 2014). There is strong evidence that the symbionts complement the nitrogen metabolism of the flagellates and have an important nutritional role in the hindgut microecosystem (Hongoh and Ohkuma 2010; Brune, 2011; Hongoh, 2011; Brune 2014).

The predominant flagellates in the hindgut of several termite lineages are hypermastigid flagellates of the genus *Trichonympha*. Originally discovered in *Reticulitermes* spp. (*Rhinotermitidae*) (Leidy, 1881; Grassi and Sandias, 1897; Koidzumi, 1921), members of the genus have been described also in several other termite families (*Termopsidae*, *Kalotermitidae*) and in their sister group, the wood-feeding cockroaches of the genus *Cryptocercus* (for references please see Yamin, 1979). While some termite species contain only a single species of *Trichonympha*, others harbor several representatives of this genus.

Termites of the genus *Zootermopsis* harbor at least four *Trichonympha* species. *Trichonympha campanula* (Kofoid and Swezi, 1919a), *Trichonympha sphaerica* (Kofoid and Swezi, 1919b), and *Trichonympha collaris* (Kirby, 1932) have long been recognized based on differences in morphology; the fourth species was discovered using molecular approaches. The first evidence of the existence of this fourth species was the presence of an additional phylotype of small subunit (SSU) rRNA gene sequences in a picked cell suspension of *T. campanula* from *Zootermopsis nevadensis* that was differentiated from the cells of *T. campanula* by fluorescence in situ hybridization (Ikeda-Ohtsubo, 2007). A few years later, this species, *Trichonympha postcylindrica*, was described based on specimens from the gut of *Zootermopsis angusticollis* (Tai et al., 2013).

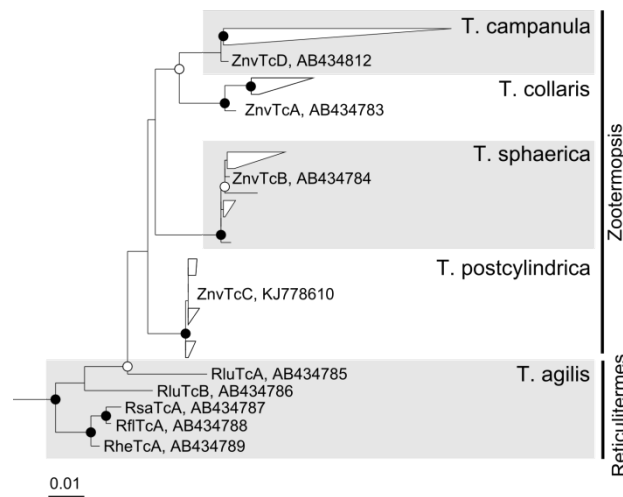
Although the same four *Trichonympha* species seem to be present in both *Z. nevadensis* and *Z. angusticollis* (Kirby, 1932; Ikeda-Ohtsubo, 2007 ; Tai et al., 2013), their relative abundance may differ even among individuals from the same colony (Kirby, 1932). Moreover, there is a considerable range of sequence divergence within the populations of each *Trichonympha* species in *Z. angusticollis* (Tai et al., 2013). The current knowledge about the diversity of *Trichonympha* species in the two *Zootermopsis* species, including the hitherto unpublished sequence of *T. postcylindrica* from *Z. nevadensis* (Ikeda-Ohtsubo, 2007), is summarized in Fig. 1.

Members of the genus *Trichonympha* fall into three distinct phylogenetic clusters (Carpenter et al., 2009; Ikeda-Ohtsubo and Brune, 2009; Ohkuma et al., 2009), and all are associated with bacterial symbionts, quite often more than one bacterial species per flagellate host (Sato et al., 2009; Sato et al., 2014; Strassert et al., 2012). All members of *Trichonympha* Cluster I harbor large populations of a specific intracellular symbiont, “*Candidatus Endomicrobium trichonymphae*” (Stingl et al., 2005; Ikeda-Ohtsubo et al., 2007; Ohkuma et al., 2007), which belongs to a distinct lineage of uncultured, insect-associated bacteria (“Endomicrobia”) in the *Elusimicrobia* phylum (Brune, 2012). The presence of homogeneous bacterial symbiont populations in a single flagellate host cell (Hongoh et al., 2008) and an almost perfect congruence of the phylogenies of the bacterial symbionts and their host flagellates (Ikeda-Ohtsubo and Brune, 2009) suggest that “*Ca. Endomicrobium trichonymphae*” is propagated exclusively by vertical transmission (cytoplasmic inheritance) in the *Trichonympha*



lineage. In this case, one would expect the same number of Endomicrobia species as *Trichonympha* species in a given termite. This is in apparent contradiction with the composition of a metagenomic library constructed from DNA prepared from the enrichment of “*Ca. Endomicrobium trichonymphae*” in the gut of *Z. nevadensis*, which contains a larger number of 16S rRNA phylotypes of Endomicrobia than host species (IMG Project ID: Gi01566) (Ikeda-Ohtsubo, 2007). It remains unclear whether all individuals of each host species harbor the same symbionts and whether a single *Trichonympha* cell carries more than one Endomicrobia phylotype.

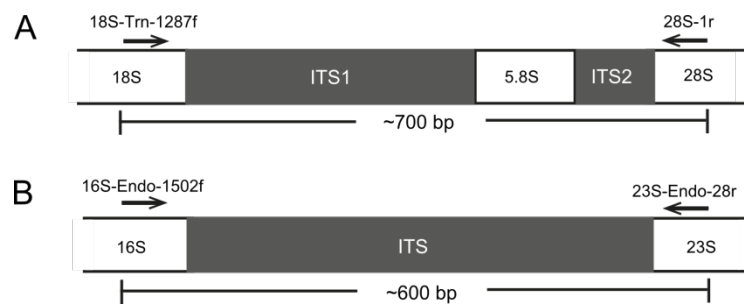
To clarify the situation, we isolated single *Trichonympha* cells from the gut of *Z. nevadensis* and *R. santonensis* and characterized their Endomicrobia populations by high-throughput sequencing, using the internal transcribed spacer (ITS) region between their rRNA genes to resolve even low levels of strain variation within both symbionts and hosts.



**Fig. 1 | Diversity of the 18S rRNA gene sequences of *Trichonympha* flagellates.** Depicted is phylogeny of *Trichonympha* from *Zootermopsis nevadensis* (Ikeda-Ohtsubo, 2007), *Zootermopsis angusticollis* (grouped nodes) (Tai et al., 2013), and different *Reticulitermes* species (Rlu, *Reticulitermes lucifugus*; Rsa, *Reticulitermes santonensis*; Rfl, *Reticulitermes flavipes*; Rhe, *Reticulitermes hesperus*) (Ikeda-Ohtsubo and Brune, 2009). The maximum-likelihood tree was rooted with *Trichonympha* species of Clusters II and III. Circles indicate node support (○, > 80%; ●, > 95% bootstrap confidence).

## Materials and Methods

**Termites.** *Zootermopsis nevadensis* was collected near the Chilao Flats Campground, Angeles National Forest, California, USA in 2006. *Reticulitermes santonensis*, which is synonymous with *Reticulitermes flavipes* (Austin et al., 2005), was collected near La Gautrelle, Ile d'Oléron, France in 2010. Since then, both species have been maintained in the laboratory on a diet of pine wood and water. Worker termites (pseudergates) were used for all experiments.



**Fig. 2 | Strategy used for the PCR amplification of the ITS regions.** Depicted are the ITS regions of (A) *Trichonympha* flagellates and (B) their Endomicrobia symbionts. The primers, their target positions in the proximal regions of the flanking rRNA genes, and the resulting amplicon length are indicated.

**Micromanipulation and whole-genome amplification.** Termites were dissected and the entire hindgut content was carefully diluted in Solution U (Trager, 1934) in Eppendorf tubes. Aliquots of this suspension (10  $\mu$ L) were transferred to the wells of a ThermoFisher -coated microscope slide and inspected with an inverted microscope with phase-contrast optics (50-fold magnification). Individual *Trichonympha* cells were identified by their morphology and captured using a micropipette attached to a micromanipulator as described before (Thompson et al., 2012), and transferred to a PCR tube with 50  $\mu$ L Solution U. The tubes were incubated at 95  $^{\circ}$ C for 10 min to lyse the flagellates, cooled on ice for 2 min, and centrifuged at slow speed (50  $\times$  g) for 1 min to remove cell debris. The supernatant (including the endosymbionts) was used as template for multiple displacement amplification (MDA) with the REPLI-g UltraFast Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions, except that the reaction time was increased to 4 h.

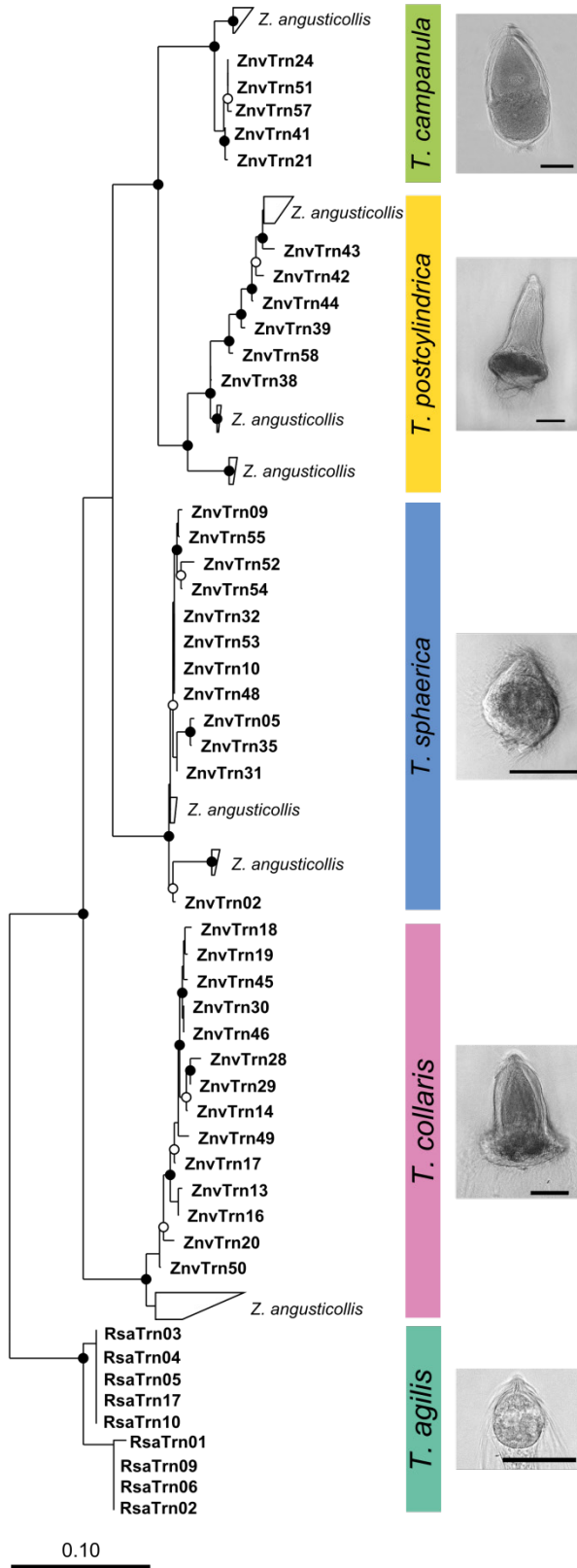
**Library preparation and sequencing.** The ITS regions of flagellates and Endomicrobia were amplified by PCR using the MDA products (25-fold diluted) as template and specific primer pairs for the proximal regions of the flanking rRNA genes (Fig. 2). Combination of the *Trichonympha*-specific forward primer 18S-Tri-1287f (AAGATTCACGTAGCTGGG; this study) and the universal reverse primer 28S-1r (ATGCTTAAATTCAGCGGGT) (Moreira et al, 2007) yielded PCR products of ~700 bp (30 cycles of amplification: 30 s at 94  $^{\circ}$ C, 30 s at 54  $^{\circ}$ C, and 60 s at 72  $^{\circ}$ C), which were directly sequenced as previously described (Thompson et al., 2012).

The Endomicrobia-specific primers 16S-Endo-1502f (AAGGTAGCCGTACGAGA) and 23S-Endo-28r (ACAGTCTTAGCCAAGGCA) were designed on the basis of all Endomicrobia sequences represented in public databases. Both primers were barcoded as described previously (Köhler et al., 2012). Endomicrobia sequences were also amplified directly from undiluted DNA extracted from whole-gut homogenates of *Z. nevadensis* and *R. santonensis* (35 cycles of amplification: 30 s at 94  $^{\circ}$ C, 30 s at 56  $^{\circ}$ C, and 60 s at 72  $^{\circ}$ C). All samples were commercially sequenced in a single sequencing run (454 GS FLX 64 with Titanium technology; GATC Biotech, Konstanz, Germany).

**Bioinformatics.** The ITS sequences of picked flagellates were aligned de novo with MAFFT version 7 (Kato and Standley, 2013). After manual curation of the alignment, a neighbor-joining tree was constructed using the ARB software suite (Ludwig et al., 2004).

The ITS sequences of Endomicrobia were processed as previously described (Dietrich et al., 2014). Briefly, pyrotag reads with a minimum length of 250 bp and a maximum expected error of 0.5 were selected and demultiplexed using their barcode sequences (no mismatch allowed). After removal of barcodes and primer sequences, the sequences were clustered at the 99% similarity level with UPARSE (Edgar, 2013). Sequences were dereplicated, and representative phlotypes (most abundant sequence in the respective cluster) were aligned de novo using the MAFFT aligner in L-INS-I mode with 100 iterations (Kato and Standley, 2013). If necessary, the alignment was manually refined so that all sequences were unambiguously aligned. Maximum-likelihood trees were constructed using RAxML version 8.1.3 (Stamatakis, 2014) with the 16-state GTR- $\Gamma$  model and 1,000 bootstraps. The heatmap was generated using the R software with the package heatmap.plus (Day, 2012).

**Sequence accession numbers.** The ITS region sequences of flagellates have been deposited in GenBank under accession numbers KJ778566–KJ778610. Representative sequences of all Endomicrobia phlotypes obtained from flagellate samples and whole-gut homogenates have been deposited in GenBank under accession numbers KP058245–KP058309.

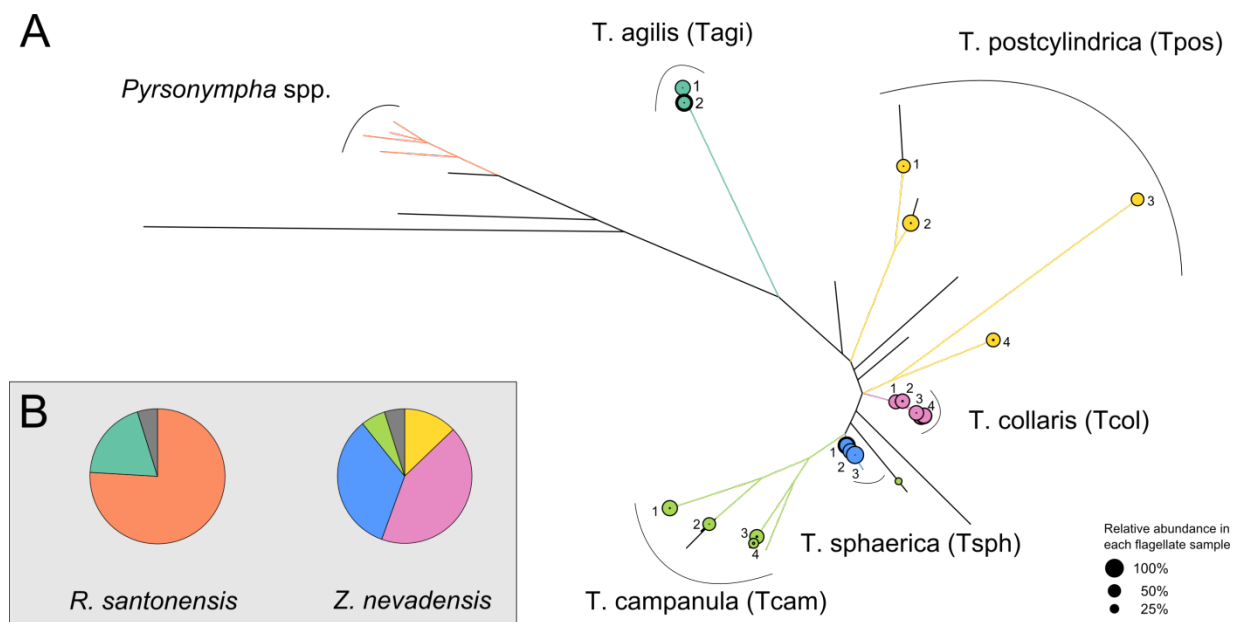


**Fig. 3 | Neighbor-joining tree of the ITS sequences.** The tree contains sequences obtained from picked flagellates, illustrating the phylogenetic diversity of the *Trichonympha* species from *Zootermopsis nevadensis* (this study) and their relationship to those from *Z. angusticollis* (sequences are grouped) (Tai et al., 2013). The tree is based on an alignment of 450 base positions and was rooted with the ITS sequences of *T. agilis* from *R. santonensis*. Phase-contrast photomicrographs illustrate the morphology of the respective species (scale bars = 50 μm). Circles indicate node support (○, > 80%; ●, >95% bootstrap confidence).

## Results

**Isolation and identification of single *Trichonympha* cells.** We successfully amplified the ITS regions of host flagellates (Fig. 2A) and Endomicrobia symbionts (Fig. 2B) from the whole-genome amplification products obtained from picked *Trichonympha* cells from *Z. nevadensis* (34 of 60 cells) and *R. santonensis* (9 of 20 cells). Direct sequencing of the PCR products obtained with flagellate-specific primers yielded clean signals for all species except for *T. campanula*, where multiple bases at several positions of the trace file indicated sequence polymorphism in the ITS region (Fig. S1).

Comparative sequence analysis showed that each of the sequences obtained from the morphotypes of *T. sphaerica*, *T. postcylindrica*, *T. collaris*, and *T. campanula* in *Z. nevadensis* clustered with their respective relatives from *Z. angusticollis* (Fig. 3), confirming the morphological assignment of the picked flagellates. The nine sequences from the morphotype of *T. agilis* formed a sister group of the *Trichonympha* species from *Z. nevadensis*.

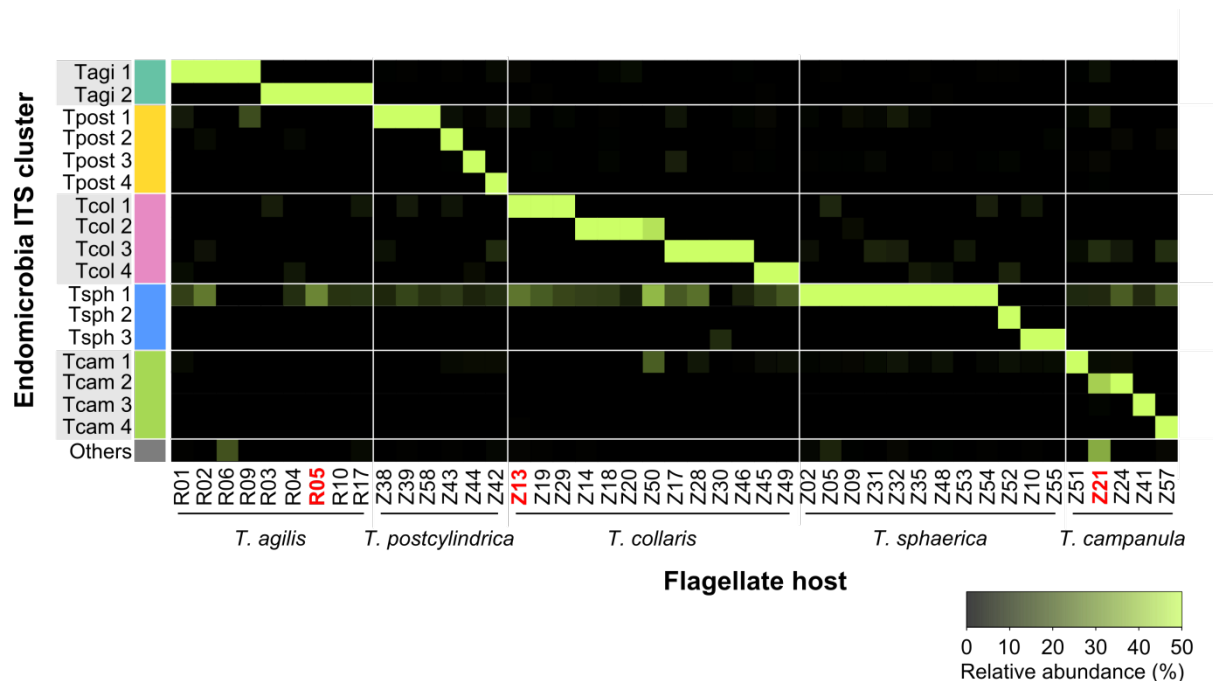


**Fig. 4 | Endomicrobia phylotypes.** (A) Maximum-likelihood tree of all Endomicrobia phylotypes (99% sequence similarity) detected in the whole-genome amplification (MDA) products of all single flagellate cells. Branches shared by phylotypes from the same host species are color coded; the area of the circles indicates the relative abundance of the major phylotype in each library. Black dots indicate the position of phylotypes that originated from whole-gut samples of *Z. nevadensis* or *R. santonensis*. (B) Pie charts indicate the relative abundance of the different Endomicrobia clusters in the whole gut samples. The colors represent the major phylotypes from picked *Trichonympha* cells; phylotypes from whole-gut samples are shown in grey. Phylotypes in the *Pyrrsonympha* cluster were identified based on the sequences retrieved from picked flagellates from *R. santonensis* (unpublished results).

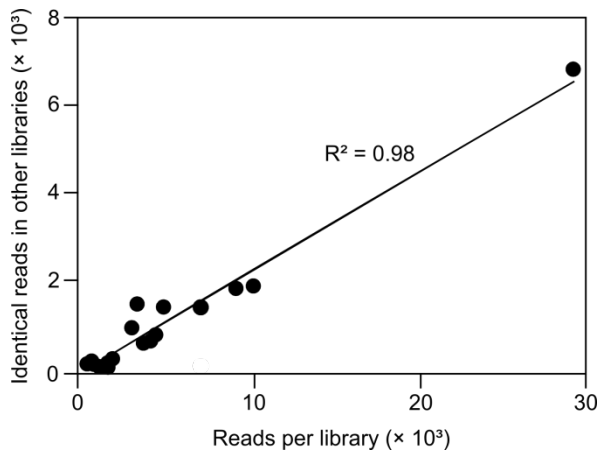
**Endomicrobia populations in single *Trichonympha* cells.** Pyrosequencing analysis of PCR products obtained with Endomicrobia-specific primers yielded variable read numbers per flagellate sample (after quality control, see Fig. S2). Although the number of Endomicrobia phylotypes (99% sequence identity) obtained from all samples (17 phylotypes from 45 single host cells) was much larger than the number of flagellate species investigated (5 species), they all clustered more or less closely according to their respective hosts (Fig. 4A).

While the Endomicrobia obtained from nine different cells of *T. agilis* comprised only two closely related phylotypes, the symbionts of the other flagellates, particularly *T. postcylindrica* and *T. campanula*, were far more diverse (Fig. 4A). Phylogenetically distinct flagellates carried different phylotypes of symbionts, and even the Endomicrobia in host cells with identical ITS sequences sometimes differed slightly in their respective phylotype (e.g., *T. sphaerica* ZnvTrn10 and ZnvTrn48).

When we analyzed the diversity of Endomicrobia sequences obtained from whole-gut DNA of *Z. nevadensis*, we found several additional phylotypes (represented by black branches in Fig. 4A) within the radiation of the Endomicrobia sequences retrieved from picked *Trichonympha* cells, which indicated that the *Trichonympha* populations in this termite were not exhaustively sampled. Nonetheless, the vast majority of the Endomicrobia in *Z. nevadensis* consisted of phylotypes retrieved from *T. collaris* (41%) and *T. sphaerica* (33%), while the diverse populations of Endomicrobia associated with *T. postcylindrica* (13%) and *T. campanula* (6%) represented only a small part of the community (Fig. 4B). In the case of *R. santonensis*, the two phylotypes of Endomicrobia from *T. agilis* accounted for 23% of the phylotypes in the gut homogenate, and the majority of the reads clustered with ITS sequences of picked *Pyrrsonympha* flagellates (unpublished results), which form a distinct branch in the phylogeny of Endomicrobia (Stingl et al., 2005).



**Fig. 5 | Heatmap of the relative abundance of different phylotypes of Endomicrobia within the pyrotag libraries of all single flagellate cells.** The samples were ordered according to the clusters indicated in Fig. 2. The samples that were checked for purity by cloning and Sanger sequencing are in red.



**Fig. 6 | Correlation of the number of reads for the major phylotype in each pyrotag library of individual flagellates to the number of identical reads recovered from the libraries of other flagellate species.** Each dot represents the results obtained for an individual sample.

**Pyrosequencing artifacts.** Although each of the pyrotag libraries of Endomicrobia from single flagellates consistently contained only one major phylotype (Fig. 5), they always comprised a smaller fraction of reads that were identical with the dominant phylotypes in other libraries. The shared presence of the same Endomicrobia phylotype in different flagellate samples of *Z. nevadensis* may be caused by the interspecific transfer of endosymbionts within the gut, but the presence of identical phylotypes in *Trichonympha* species of different termites was quite surprising, because it can be explained only by multiple recent transfers of endosymbionts between flagellates of different termite species. Therefore, we suspected that the shared phylotypes are artifacts that either resulted from the picking process (*e.g.*, the symbionts stem from other, lysed flagellate cells present in the same gut) or were generated during the pyrosequencing process.

To clarify this, we reanalyzed three of the Endomicrobia samples obtained from individual flagellates (R05, Z13 and Z21), which were selected because they had particularly large proportions of suspicious sequences (see Fig. 5). To exclude the possibility that the minor phylotypes in the corresponding pyrotag libraries stem from a contamination during the picking process or originated during whole-genome amplification or library preparation (*e.g.*, by contaminated tags), we used aliquots of the samples that had been preserved directly before pyrosequencing (*i.e.*, after the PCR step that introduced the sample-specific tags). The amplicons were ligated into a plasmid vector, and 30 clones of each sample were sequenced. In each case, all sequences in the library were identical to the major phylotype obtained from the respective flagellate, which clearly identifies the minor phylotypes in the corresponding pyrotag libraries as artifacts generated during the pyrosequencing process.

This interpretation is supported by the observation that the reads of Endomicrobia phylotype Tsph1, which was present in most *T. sphaerica* samples and overrepresented in the pyrosequencing run, were recovered in varying abundance from all other flagellate samples (Fig. 5; Fig. S2). To test this quantitatively, we compared the number of reads for the major phylotype in each flagellate sample to the number of identical reads recovered from samples of other flagellate species (Fig. 6). The strong linear correlation corroborates that the minor phylotypes in each sample are not real but artifacts caused by tag switching.

## Discussion

**Host specificity of Endomicrobia.** In this study, we showed that each of the *Trichonympha* cells from both *Z. nevadensis* and *R. santonensis* harbor a single phylotype of “*Candidatus Endomicrobium trichonymphae*”. The symbiont populations in each host lineage are phylogenetically distinct even if multiple host species are present in the same gut, corroborating the assumption that the apparent cospeciation between the partners in this symbiosis is caused by vertical (cytoplasmic) inheritance of the endosymbionts (Ikeda-Ohtsubo and Brune, 2009).

Although Ikeda-Ohtsubo and Brune (2009) were able to document a congruent topology of the SSU rRNA trees of the symbiotic pairs, the relatively small number of sequences in their clone libraries (10–20 clones) did not allow us to exclude the presence of minor phlotypes of Endomicrobia within each *Trichonympha* cell. Moreover, since the libraries were obtained from flagellate suspensions containing numerous host cells (100–200 cells per sample), it remained possible that individual flagellates harbored a phylotype entirely different from that of the majority of the cells in the suspension.

Prompted by the observation that a large Endomicrobia clone library of *Z. nevadensis* (353 clones) contained a much larger number of 16S rRNA phlotypes than *Trichonympha* species described for this flagellate (Ikeda-Ohtsubo, 2007), we reinvestigated the subject. Since the resolution of 16S rRNA analysis was not sufficient due to the high similarity of the sequences, we used the much more variable ITS region, harnessing the advantages of whole-genome amplification and next-generation sequencing technology to achieve a highly resolved analysis of their community structure based on the genomic DNA of the symbiont populations in individual host cells. Beyond the artifacts inherent to the pyrotag method (see below), it is now safe to conclude that (i) each *Trichonympha* cell harbors a homogeneous population of Endomicrobia symbionts, (ii) *Trichonympha* cells of different species always harbor a phylogenetically distinct symbiont population, and (iii) the phylogenetic diversity of Endomicrobia in the two termites investigated in this study is caused by a microdiversity of *Trichonympha* populations that extends beyond the framework of the currently described species.

It cannot be denied that also the present study reaches its limits when it comes to excluding the possibility of an exchange of symbionts between individuals of the same (or closely related) *Trichonympha* populations present within the same gut. While the *Trichonympha* species in the wood-feeding cockroach *Cryptocercus punctulatus* still reproduce sexually during each molt of the host, this trait has been lost in the *Trichonympha* lineages of termites at an uncertain evolutionary time point (Cleveland, 1949). However, artificial feeding of *Z. angusticollis* with the molting hormone 20-hydroxyecdysone still seems to trigger a sexual cycle of their *Trichonympha* flagellates (Messer and Lee, 1989), and it is possible that such copulation events still allow a mixing of Endomicrobia phlotypes among host cells of the same species on rare occasions. Nevertheless, the strong and ongoing genome reduction demonstrated in “*Ca. Endomicrobium trichonymphae*” strain Rs-D17 (Hongoh, 2008) indicates the absence of genetic exchange and a frequent population bottleneck during the transmission (Wernegreen, 2002), a well-documented phenomenon in the intracellular symbionts in the bacteriocytes of aphids (Mira and Moran, 2002; Moran et al., 2008).

**Endomicrobia populations in Zootermopsis and Reticulitermes.** Assuming that all Endomicrobia species have the same copy number of rRNA genes, the Endomicrobia community in the gut of *Z. nevadensis* is dominated by phlotypes from *T. collaris* (43%) and *T. sphaerica* (31%). These outnumber the diverse populations of “*Ca. Endomicrobium trichonymphae*” associated with *T. postcylindrica* and *T. campanula*, which account for another 19% of the community (Fig. 4B).



Considering that the colonization density of Endomicrobia differs among the four *Trichonympha* phylotypes in *Z. nevadensis* (Ikeda-Ohtsubo, 2007), this is in reasonable agreement with the relative abundance of *Trichonympha* species in this termite reported by Kirby (1932), who found that the three species (he did not recognize *T. postcylindrica* as separate from *T. campanula*), were usually present in comparable numbers. The phylogenetic position of the remaining phylotypes (5% of the reads in the library) recovered only from the gut of *Z. nevadensis* but not from picked flagellates (Fig. 4A) suggests that they represent Endomicrobia populations of *Trichonympha* lineages that are not included among the picked cells, but whose presence is indicated by the microdiversity of the ITS sequences obtained for each morphotype (Fig. 3).

Also in the case of the *T. agilis* cells picked from the gut of *R. santonensis*, the microdiversity of Endomicrobia (Fig. 4A) reflects that of the host flagellates (Fig. 3), which indicates a strict host specificity of the endosymbionts. Kirby (1932) reports that with the exception of *Reticulitermes lucifugus*, all *Reticulitermes* species he investigated harbor only a single species of *Trichonympha*, which is in agreement with the number of SSU rRNA genes of *Trichonympha* species obtained from different *Reticulitermes* species (Fig. 1). However, the ITS sequences of picked flagellates reveal the presence of two closely related phylotypes in *R. santonensis* (Fig. 3), which is in agreement with the two phylotypes of “*Ca. Endomicrobium trichonymphae*” obtained from these samples (Fig. 4).

The reads obtained for the amplified ITS regions of individual flagellates of *T. campanula* often contained traces of sequence polymorphism, suggesting that the rRNA gene clusters have multiple and slightly divergent copies. So far, there is no genome information for any termite gut flagellate, but the draft genome of the distantly related parabasalid *Trichomonas vaginalis* contains 254 copies of the 18S rRNA gene (Carlton et al., 2007). ITS regions experience low selective constraint and are considered to evolve rapidly by large insertions and/or deletions (Sawada et al., 1997; García-Martínez et al., 1999).

**Methodological pitfalls of pyrotag sequencing.** In amplicon pyrosequencing, the addition of sequence tags to the amplified target sequences allows a parallel processing of numerous samples in the same sequencing run (Hamady et al., 2008). The tags are usually added to both ends of the amplicons because this allows identification of cases of tag switching (Binladen et al., 2007), which result in impossible tags combinations at each terminus and may affect a large proportion of the reads in pyrotag libraries (van Orsouw et al., 2007; Carlsen et al., 2012).

However, tag switching can be recognized and removed during quality control only if both termini of the amplicon are present in the same read. If the length of the amplicon exceeds the read length, a common problem with most sequencing platforms, it is not possible to identify mistagged sequences without any prior information. In this case, they either go entirely unnoticed or are purged from the dataset by omitting all reads that do not pass a certain frequency threshold (e.g., Tai et al., 2014).

Also in this study, the necessities of primer design dictated an amplicon length of about 600 bp (ITS region of Endomicrobia; Fig. 2B) that exceeded even the read length of the 454 Titanium technology (up to 400 bp) (Glenn, 2011). However, the low diversity of the Endomicrobia community and the highly improbable presence of identical phylotypes in host flagellates of different termites allowed us to identify reads with suspicious tags. Their frequency amounted to 15% of the reads in the entire sequencing run (Fig. S2), which is at the upper end of the proportion of mistagged reads reported in previous studies (van Orsouw et al., 2007). The complete absence of these reads from the processed samples prior to the sequencing run excludes that they result from contaminated tags and indicated that they originated during the pyrosequencing step, probably from tag switching during emulsion PCR, as already suggested by other authors (Carlsen et al., 2012). Since the number of

mistagged reads in each sample will depend on the number of reads carrying a particular tag (Fig. 6; Fig. S2), mistagged reads cannot be removed simply by applying a frequency threshold. Particularly in cases that are highly sensitive to tag switching, *e.g.*, the identification of core communities shared across multiple samples, such artifacts have to be excluded using an appropriate experimental design. Otherwise, the simultaneous presence of OTUs in different samples of a pyrotag run has to be interpreted with the necessary caution (Dietrich et al., 2014).

## **Acknowledgements**

This study was funded by the Max Planck Society. Hao Zheng is a doctoral student in the International Max Planck Research School for Environmental, Cellular and Molecular Microbiology (IMPRS-MIC), Marburg. Claire Thompson was funded by a postdoctoral fellowship of the LOEWE Center for Synthetic Microbiology, Marburg. We thank Jared Leadbetter (Caltech) for providing termites.

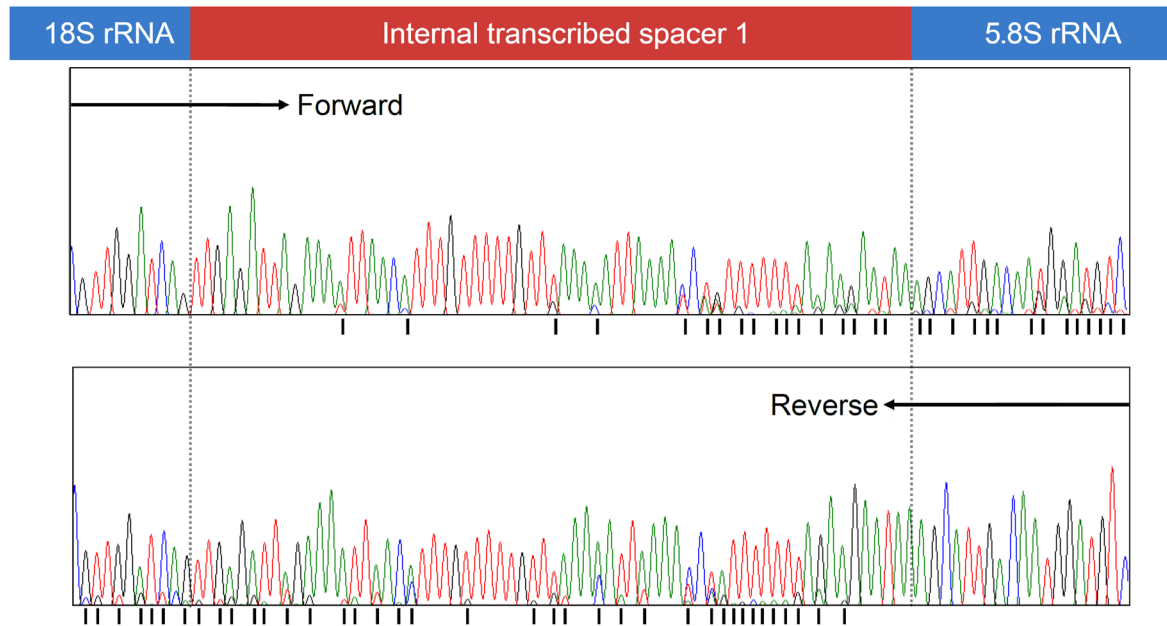
## References

- Austin JW, Szalanski AL, Scheffrahn RH, Messenger MT, Dronnet S and Bagnères AG (2005).** Genetic evidence for the synonymy of two *Reticulitermes* species: *Reticulitermes flavipes* and *Reticulitermes santonensis*. *Ann. Entomol. Soc. Am.* 98: 395-401.
- Binladen J, Gilbert MTP, Bollback JP, Panitz F, Bendixen C, Nielsen R and Willerslev E (2007).** The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS One* 2: e197.
- Brune A (2011).** Microbial symbioses in the digestive tract of lower termites, p. 3-25. In E. Rosenberg, and U. Gophna (ed.), *Beneficial Microorganisms in Multicellular Life Forms*. Springer, Heidelberg.
- Brune A (2012).** Endomicrobia: intracellular symbionts of termite gut flagellates. *J. Endocytobiosis Cell Res.* 23: 11-15.
- Brune A (2014).** Symbiotic digestion of lignocellulose in termite guts. *Nat. Rev. Microbiol.* 12: 168-180.
- Brune A and Ohkuma M (2011).** Role of the termite gut microbiota in symbiotic digestion, p. 439-475. In Bignell DE, Roisin Y, and Lo N (eds.), *Biology of Termites: A Modern Synthesis*. Springer, Dordrecht.
- Carlsen T, Aas AB, Lindner D, Vralstad T, Schumacher T and Kausrud H (2012).** Don't make a mista(g)ke: is tag switching an overlooked source of error in amplicon pyrosequencing studies? *Fungal Ecol.* 5: 747-749.
- Carlton JM, Hirt RP, Silva JC, et al. (2007).** Draft genome sequence of the sexually transmitted pathogen *Trichomonas vaginalis*. *Science* 315: 207-212.
- Carpenter KJ, Chow L and Keeling PJ (2009).** Morphology, phylogeny, and diversity of Trichonympha (Parabasalia: Hypermastigida) of the wood-feeding cockroach *Cryptocercus punctulatus*. *J. Eukaryotic Microbiol.* 56: 305-313.
- Cleveland LR (1949).** Hormone-induced sexual cycles of flagellates. I. gametogenesis, fertilization, and meiosis in *Trichonympha*. *J. Morphol.* 85: 197-295.
- Day A (2012).** heatmap.plus – heatmap with more sensible behavior. R package version 1.3. Available at <http://cran.r-project.org/web/packages/heatmap.plus/index.html>.
- Dietrich C, Köhler T and Brune A (2014).** The cockroach origin of the termite gut microbiota: patterns in bacterial community structure reflect major evolutionary events. *Appl. Environ. Microbiol.* 80: 2261-2269.
- Edgar RC (2013).** UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods* 10: 996-998.
- García-Martínez J, Acinas SG, Antón AI and Rodríguez-Valera F (1999).** Use of the 16S-23S ribosomal genes spacer region in studies of prokaryotic diversity. *J. Microbiol. Methods* 36: 55-64.
- Glenn TC (2011).** Field guide to next-generation DNA sequencers. *Mol. Ecol. Resour.* 11: 759-769.
- Grassi B and Sandias A (1897).** The constitution and development of the society of termites: observations on their habits; with appendices on the parasitic protozoa of *Termitidae*, and on the Embiidae. *Quart. J. Microsc. Sci.* 39: 245-332.
- Hamady M, Walker JJ, Harris JK, Gold NJ and Knight R (2008).** Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nat. Methods* 5: 235-237.
- Hongoh Y (2011).** Toward the functional analysis of uncultivable, symbiotic microorganisms in the termite gut. *Cell. Mol. Life Sci.* 68: 1311-1325.
- Hongoh Y and Ohkuma M (2010).** Termite gut flagellates and their methanogenic and eubacterial symbionts, p. 55-79. In J. H. P. Hackstein (ed.), *(Endo)Symbiotic Methanogenic Archaea*. Springer, Heidelberg.
- Hongoh Y, Sharma VK, Prakash T et al (2008).** Complete genome of the uncultured Termite Group 1 bacteria in a single host protist cell. *Proc. Natl. Acad. Sci. USA* 105: 5555-5560.
- Ikeda-Ohtsubo W (2007).** Endomicrobia in termite guts: symbionts within a symbiont. Doctoral thesis. Philipps-Universität Marburg, Marburg, Germany.
- Ikeda-Ohtsubo W and Brune A (2009).** Cospeciation of termite gut flagellates and their bacterial endosymbionts: *Trichonympha* species and 'Candidatus Endomicrobium trichonymphae'. *Mol. Ecol.* 18: 332-342.

- Ikeda-Ohtsubo W, Desai M, Stingl U and Brune A (2007).** Phylogenetic diversity of 'Endomicrobia' and their specific affiliation with termite gut flagellates. *Microbiology* 153: 3458-3465.
- Katoh K and Standley DM (2013).** MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30: 772-780.
- Kirby H (1932).** Flagellates of the genus *Trichonympha* in termites. *Univ. Calif. Publ. Zool.* 37: 349-476.
- Kofoed CA and Swezy (1919a).** Studies on the parasites of the termites III. On *Trichonympha campanula* sp. nov. *Univ. Calif. Publ. Zool.* 20: 41-98.
- Kofoed CA and Swezy O (1919b).** Studies on the parasites of the termites IV. On *Leidyopsis sphaerica* gen. nov., sp. nov. *Univ. Calif. Publ. Zool.* 20: 99-116.
- Köhler T, Dietrich C, Scheffrahn RH and Brune A (2012).** High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation of the gut in wood-feeding higher termites (*Nasutitermes* spp.). *Appl. Environ. Microbiol.* 78: 4691-4701
- Koidzumi M (1921).** Studies on the intestinal protozoa found in the termites of Japan. *Parasitology* 13: 235-305.
- Leidy J (1881).** The parasites of the termites. *J. Acad. Nat. Sci. Philadelphia.* 2nd Ser. 8: 425-477.
- Ludwig W, Strunk O, Westram R, et al. (2004).** ARB: a software environment for sequence data. *Nucleic Acids Res.* 32: 1363-1371.
- Messer AC and Lee MJ (1989).** Effect of chemical treatments on methane emission by the hindgut microbiota in the termite *Zootermopsis angusticollis*. *Microb. Ecol.* 18: 275-284.
- Mira A and Moran NA (2002).** Estimating population size and transmission bottlenecks in maternally transmitted endosymbiotic bacteria. *Microb. Ecol.* 44: 137-143.
- Moran NA, McCutcheon JP and Nakabachi A (2008).** Genomics and evolution of heritable bacterial symbionts. *Annu. Rev. Genet.* 42: 165-190.
- Moreira D, von der Heyden S, Bass D, López-García P, Chao E and Cavalier-Smith T (2007).** Global eukaryote phylogeny: Combined small- and large-subunit ribosomal DNA trees support monophyly of Rhizaria, Retaria and Excavata. *Mol. Phylogenet. Evol.* 44: 255-266.
- Ohkuma M and Brune A (2011).** Diversity, structure, and evolution of the termite gut microbial community, p. 413-438. In D. E. Bignell, Y. Roisin and N. Lo (ed.), *Biology of Termites: A Modern Synthesis*. Springer, Dordrecht.
- Ohkuma M, Noda S, Hongoh Y, Nalepa CA and Inoue T (2009).** Inheritance and diversification of symbiotic trichonymphid flagellates from a common ancestor of termites and the cockroach *Cryptocercus*. *Proc. R. Soc. B* 276: 239-245.
- Ohkuma M, Sato T, Noda S, Ui S, Kudo T and Hongoh Y (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.
- Sato T, Hongoh Y, Noda S, Hattori S, Ui S and Ohkuma M (2009).** *Candidatus* Desulfovibrio trichonymphae, a novel intracellular symbiont of the flagellate *Trichonympha agilis* in termite gut. *Environ. Microbiol.* 11: 1007-1015.
- Sato T, Kuwahara H, Fujita K, Noda S, Kihara K, Yamada A, Ohkuma M and Hongoh Y (2014).** Intranuclear verrucomicrobial symbionts and evidence of lateral gene transfer to the host protist in the termite gut. *ISME J.* 8: 1008-1019.
- Sawada H, Takeuchi T and Matsuda I (1997).** Comparative analysis of *Pseudomonas syringae* pv. *actinidiae* and pv. *phaseolicola* based on phaseolotoxin-resistant ornithine carbamoyltransferase gene (*argK*) and 16S-23S rRNA intergenic spacer sequences. *Appl. Environ. Microbiol.* 63: 282-288.
- Stamatakis A (2014).** RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312-1313.
- Stingl U, Radek R, Yang H and Brune A (2005).** "Endomicrobia": cytoplasmic symbionts of termite gut protozoa form a separate phylum of prokaryotes. *Appl. Environ. Microbiol.* 71: 1473-1479.

- Strassert JFH, Köhler T, Wienemann THG, Ikeda-Ohtsubo W, Faivre N, Franckenberg S, Plarre R, Radek R and Brune A (2012). '*Candidatus Ancillula trichonymphae*', a novel lineage of endosymbiotic *Actinobacteria* in termite gut flagellates of the genus *Trichonympha*. *Environ. Microbiol.* 14: 3259-3270.
- Tai V, James ER, Nalepa CA, Scheffrahn RH, Perlman SJ and Keeling PJ (2014).** The role of host phylogeny varies in shaping microbial diversity in the hindguts of lower termites. *Appl. Environ. Microbiol.* doi: 10.1128/AEM.02945-14.
- Tai V, James ER, Perlman SJ and Keeling PJ (2013).** Single-cell DNA barcoding using sequences from the small subunit rRNA and internal transcribed spacer region identifies new species of *Trichonympha* and *Trichomitopsis* from the hindgut of the termite *Zootermopsis angusticollis*. *PLoS One* 8: e58728.
- Thompson CL, Vier R, Mikaelyan A, Wienemann T and Brune A (2012).** '*Candidatus Arthromitus*' revised: segmented filamentous bacteria in arthropod guts are members of *Lachnospiraceae*. *Environ. Microbiol.* 14: 1454-1465.
- Trager W (1934).** The cultivation of a cellulose-digesting flagellate, *Trichomonas termopsidis*, and of certain other termite protozoa. *Biol. Bull.* 66: 182-190.
- van Orsouw NJ, Hogers RCJ, Janssen A et al. (2007).** Complexity reduction of polymorphic sequences (CRoPSTM): A novel approach for large-scale polymorphism discovery in complex genomes. *PLoS One* 2: e1172.
- Wernegreen JJ (2002).** Genome evolution in bacterial endosymbionts of insects. *Nat. Rev. Genet.* 3: 850-861.
- Yamin MA (1979).** Flagellates of the orders *Trichomonadida* Kirby, *Oxymonadida* Grassé, and *Hypermastigida* Grassi & Foà reported from lower termites (*Isoptera* families *Mastotermitidae*, *Kalotermitidae*, *Hodotermitidae*, *Termopsidae*, *Rhinotermitidae*, and *Serritermitidae*) and from the wood-feeding roach *Cryptocercus* (*Dictyoptera: Cryptocercidae*). *Sociobiology* 4: 1-119.

## Supplementary Material



**Fig. S1 | Sequence polymorphism in the ITS region of *Trichonympha campanula*.** Illustrated is a trace file of the direct Sanger sequencing of the PCR product obtained from a singleflagellate MDA sample (ZnvTrn21). The positions of multiple peaks (indicated by bars) are affected by the sequencing direction but always start in the ITS region, suggesting the presence of both substitutions and indels. For phylogenetic analysis, we used the consensus of the sequences obtained forward and reverse primers, always selecting the stronger signal at the polymorphic positions.



**Fig. S2 | Composition of Endomicrobia ITS libraries obtained by pyrotag analysis of the whole-genome amplification products of individual *Trichonympha* cells.** The color code indicates the host species of the major phylotypes. Other phylotypes are shown in grey. The area of each circle represents the number of reads in each library. The names of the samples that were checked for purity by cloning and Sanger sequencing are in red.





---

# Curriculum vitae

## Basic information

Name	Carsten Dietrich
Date of Birth	6th of July, 1985
Place of Birth	Erfurt, Germany
Marital Status	Single
Nationality	German

## Higher education

Diploma in Environmental engineering, THM Giessen	2011
Doctorate in Microbiology, University of Marburg	2015

## Awards and Honors

Among the best 10% of his diploma year	2011
IMPRS—MIC scholarship for doctoral research	2011–2014

## Professional experience

Research Assistant at the THM Giessen, Germany	2008–2009
Quality insurance at the Pharmaserv GmbH & Co. KG, Giessen	2009–2010
Research Assistant at the Max Planck Institute Marburg, Germany	2010–2011
PhD Student at the Max Planck Institute Marburg, Germany	2011–2015

## Publications

- Brune A und **Dietrich C** (2015). The termite gut microbiota: Digesting the diversity in the light of ecology and evolution. *Ann. Rev. Microbiol.* 69 (release in october 2015).
- Dietrich C** und Brune A (2014). The complete mitogenomes of six higher termite species reconstructed from metagenomic datasets (*Cornitermes* sp., *Cubitermes ugandensis*, *Microcerotermes parvus*, *Nasutitermes corniger*, *Neocapritermes taracua*, and *Termes hospes*). *Mitochondr. DNA.* Dec 4:1-2.
- Dietrich C\***, Köhler T\* und Brune A (2014). The cockroach origin of the termite gut microbiota: patterns in bacterial community structure reflect major evolutionary events. *Appl. Environ. Microbiol.* 80:2261-2269.
- Dietrich C\***, Nonoh J\*, Lang K, Mikulski L, Meuser K, Köhler T, Boga HI, Ngugi DK, Sillam-Dussès S und Brune A (submitted). Habitat selection and vertical inheritance drive archaeal community structure in arthropod guts.
- Guichard P, Desfosses A, Maheshwari A, Hachet V, **Dietrich C** und Brune A, Ishikawa T, Sachse, C und Gönczy P (2012). Cartwheel architecture of *Trichonympha* basal body. *Science* 337:553–554.
- Köhler T, **Dietrich C**, Scheffrahn RH und Brune A (2012). High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation of the gut in wood-feeding higher termites (*Nasutitermes* spp.). *Appl. Environ. Microbiol.* 78:4691–4701.
- Li H\*, **Dietrich C\***, Zhu N, Mikaelyan A, Ma B, Pi R, Liu Y, Yang M, Brune A und Mo J (submitted). Age polyethism drives community structure of the bacterial gut microbiota in the fungus-cultivating termite *Odontotermes formosanus*.
- Meyer BH, Peyfoon E, **Dietrich C**, Hitchen P, Panico M, Morris HR, Dell A und Albers SV (2013). Agl16, a thermophilic glycosyltransferase mediating the last step of N-glycan biosynthesis in the thermoacidophilic crenarchaeon *Sulfolobus acidocaldarius*. *J. Bacteriol.* 195:2177–2186.
- Mikaelyan A, **Dietrich C**, Köhler T, Meuser K, Sillam-Dussès D und Brune A (submitted). Dietary and phylogenetic determinants of bacterial gut community structure in higher termites.
- Zheng H, **Dietrich C**, Radek R und Brune A (submitted). *Endomicrobium proavitum*, the first isolate of Endomicrobia class. nov. (phylum *Elusimicrobia*) – an ultramicrobacterium with an unusual cell cycle that fixes nitrogen with a Group IV nitrogenase.
- Zheng H, **Dietrich C**, Thompson CL, Meuser K und Brune A (2015). Population structure of Endomicrobia in single host cells of termite gut flagellates (*Trichonympha* spp.). *Microb. Environ.* 30: 92–98.

\* These authors contributed equally.



---

## Contributions

**Chapter 1** | C.D. contributed to the manuscript and designed the figures. A.B. contributed to the manuscript.

**Chapter 2** | T.K. designed and carried out experiments, analyzed data and wrote the manuscript. C.D. designed and carried out experiments, analyzed data and contributed to the manuscript. R.H.F. provided termites and Figure 1. A.B. conceived the study, wrote the manuscript, and secured funding.

**Chapter 3** | C.D. designed and carried out experiments, analyzed data and wrote the manuscript. T.K. designed and carried out experiments, analyzed data and contributed to the manuscript. A.B. conceived the study, wrote the manuscript, and secured funding.

**Chapter 4** | A.M. designed and carried out experiments, analyzed data and wrote the manuscript. C.D. designed and carried out experiments, analyzed data and contributed to the manuscript. T.K. designed experiments. K.M. carried out experiments. A.B. conceived the study, wrote the manuscript, and secured funding.

**Chapter 5** | H.L. conceived and designed the study, carried out experiments, analyzed data, discussed the results and wrote the manuscript. C.D. analyzed data, discussed the results and contributed to the manuscript. N.Z. carried out experiments. A.M. analyzed data and discussed the results. B.M. carried out experiments. R.P. carried out experiments. Y.L. carried out experiments. M.Y. carried out experiments. A.B. discussed the results and contributed to the manuscript. J.M. contributed to the manuscript.

**Chapter 6** | C.D. designed and conducted the high-throughput sequencing analysis; created a reference database including all previously published data; conducted the final phylogenetic analysis, all bioinformatics, and statistical analyses; and wrote the manuscript. J.O.N. created the majority of the clone libraries, and contributed to the phylogenetic analysis and the evaluation of the results. K.L. created clone libraries and contributed to the phylogenetic analysis. L.M. created clone libraries. K.M. prepared libraries for the high-throughput sequencing analysis. T.K. improved the primer set for the high-throughput sequencing analysis. H.I.B. collected termites and created clone libraries. D.K.N. collected termites and prepared DNA. D.S.D. collected termites. A.B. conceived the study, discussed the results, and secured funding.

**Chapter 7** | K.R. carried out experiments, analyzed most of the data, discussed the results, and wrote the manuscript; C.D. analyzed data and discussed the results; C.L.T. conceived the study and carried out experiments; A.M. conceived the study and carried out experiments; A.B. conceived the study, discussed results, and secured funding.

**Chapter 8** | C.D. conceived the study, carried out all experiments, analyzed data, and wrote the manuscript. K.R. contributed data and discussed the results; A.B. discussed the results and secured funding.

**Chapter 9** | C.D. contributed to the manuscript and designed the figures. A.B. contributed to the manuscript.

**Appendix A** | C.D. conceived the study, analyzed data, and wrote the manuscript. A.B. conceived the study and contributed to the manuscript.

**Appendix B** | H.Z. planned and carried out most experiments, analyzed data, discussed results, and wrote the manuscript. C.D. analyzed data, discussed results, and contributed to the manuscript. C.L.T. conceived the study and carried out experiments. K.M. carried out experiments. A.B. conceived the study, discussed results, wrote the manuscript, and secured funding.



---

## **Erklärung der Eigenständigkeit**

Ich versichere, dass ich die Dissertation „Ecological and evolutionary drivers of microbial community structure in termite guts“ selbstständig und ohne unerlaubte Hilfe angefertigt habe und mich keiner als der von mir ausdrücklich bezeichneten Quellen und Hilfen bedient habe. Diese Dissertation wurde in der jetzigen oder einer ähnlichen Form noch bei keiner anderen Hochschule eingereicht und hat noch keinen sonstigen Prüfungszwecken gedient.

Marburg, Februar 2015

---