**A peer-reviewed version of this preprint was published in PeerJ on 13 February 2020.**

View the peer-reviewed version (peerj.com/articles/8614), which is the preferred citable publication unless you specifically need to cite this preprint.

Hervé V, Liu P, Dietrich C, Sillam-Dussès D, Stiblik P, Šobotník J, Brune A. 2020. Phylogenomic analysis of 589 metagenome-assembled genomes encompassing all major prokaryotic lineages from the gut of higher termites. PeerJ 8:e8614 https://doi.org/10.7717/peerj.8614

PeerJ Preprints NOT PEER-REVIEWED

# Phylogenomic analysis of 589 metagenome-assembled genomes encompassing all major prokaryotic lineages from the gut of higher termites

**Vincent Hervé** [Corresp., 1], **Pengfei Liu** [1], **Carsten Dietrich** [1], **David Sillam-Dussès** [2], **Petr Stiblik** [3], **Jan Šobotník** [3], **Andreas Brune** [Corresp. 1]

[1] Research Group Insect Gut Microbiology and Symbiosis, Max Planck Institute for Terrestrial Microbiology, Marburg, Germany

[2] Laboratory of Experimental and Comparative Ethology EA 4443, Université Paris 13, Villetaneuse, France

[3] Faculty of Forestry and Wood Sciences, Czech University of Life Sciences, Prague, Czech Republic

Corresponding Authors: Vincent Hervé, Andreas Brune
Email address: vincent.herve8@gmail.com, brune@mpi-marburg.mpg.de

"Higher" termites have been able to colonize all tropical and subtropical regions because of their ability to digest lignocellulose with the aid of their prokaryotic gut microbiota. Over the last decade, numerous studies based on 16S rRNA gene amplicon libraries have largely described both the taxonomy and structure of the prokaryotic communities associated with termite guts. Host diet and microenvironmental conditions have emerged as the main factors structuring the microbial assemblages in the different gut compartments. Additionally, these molecular inventories have revealed the existence of termite-specific clusters that indicate coevolutionary processes in numerous prokaryotic lineages. However, for lack of representative isolates, the functional role of most lineages remains unclear. We reconstructed 589 metagenome-assembled genomes (MAGs) from the different gut compartments of eight higher termite species that encompass 17 prokaryotic phyla. By iteratively building genome trees for each clade, we significantly improved the initial automated assignment, frequently up to the genus level. We recovered MAGs from most of the termite-specific clusters in the radiation of, e.g., Planctomycetes, Fibrobacteres, Bacteroidetes, Euryarchaeota, Bathyarchaeota, Spirochaetes, Saccharibacteria, and Firmicutes, which to date contained only few or no representative genomes. Moreover, the MAGs included abundant members of the termite gut microbiota. This dataset represents the largest genomic resource for arthropod-associated microorganisms available to date and contributes substantially to populating the tree of life. More importantly, it provides a backbone for studying the metabolic potential of the termite gut microbiota, including the key members involved in carbon and nitrogen biogeochemical cycles, and important clues that may help cultivating representatives of these understudied clades.

PeerJ Preprints NOT PEER-REVIEWED

# Phylogenomic analysis of 589 metagenome-assembled genomes encompassing all major prokaryotic lineages from the gut of higher termites

Vincent Hervé[1]*, Pengfei Liu[1], Carsten Dietrich[1], David Sillam-Dussès[2], Petr Stiblik[3], Jan Šobotník[3], Andreas Brune[1]*

[1]Research Group Insect Gut Microbiology and Symbiosis, Max Planck Institute for Terrestrial Microbiology, 35043 Marburg, Germany

[2]Laboratory of Experimental and Comparative Ethology EA 4443, Université Paris 13, Sorbonne Paris Cité, Villetaneuse 93430, France

[3]Faculty of Forestry and Wood Sciences, Czech University of Life Sciences, Prague 6, Suchdol 16500, Czech Republic

* Corresponding authors:

Vincent Hervé[1]

Research Group Insect Gut Microbiology and Symbiosis, Max Planck Institute for Terrestrial Microbiology, Karl-von-Frisch-Strasse 10, 35043 Marburg, Germany

Email address: vincent.herve8@gmail.com

Andreas Brune[1]

Research Group Insect Gut Microbiology and Symbiosis, Max Planck Institute for Terrestrial Microbiology, Karl-von-Frisch-Strasse 10, 35043 Marburg, Germany

Email address: brune@mpi-marburg.mpg.de

Key words: metagenome-assembled genomes, gut microbiology, higher termites

ORCID:

Vincent Hervé: http://orcid.org/0000-0002-3495-561X

Andreas Brune: http://orcid.org/0000-0002-2667-4391

PeerJ Preprints NOT PEER-REVIEWED


## Abstract

"Higher" termites have been able to colonize all tropical and subtropical regions because of their ability to digest lignocellulose with the aid of their prokaryotic gut microbiota. Over the last decade, numerous studies based on 16S rRNA gene amplicon libraries have largely described both the taxonomy and structure of the prokaryotic communities associated with termite guts. Host diet and microenvironmental conditions have emerged as the main factors structuring the microbial assemblages in the different gut compartments. Additionally, these molecular inventories have revealed the existence of termite-specific clusters that indicate coevolutionary processes in numerous prokaryotic lineages. However, for lack of representative isolates, the functional role of most lineages remains unclear. We reconstructed 589 metagenome-assembled genomes (MAGs) from the different gut compartments of eight higher termite species that encompass 17 prokaryotic phyla. By iteratively building genome trees for each clade, we significantly improved the initial automated assignment, frequently up to the genus level. We recovered MAGs from most of the termite-specific clusters in the radiation of, e.g., Planctomycetes, Fibrobacteres, Bacteroidetes, Euryarchaeota, Bathyarchaeota, Spirochaetes, Saccharibacteria, and Firmicutes, which to date contained only few or no representative genomes. Moreover, the MAGs included abundant members of the termite gut microbiota. This dataset represents the largest genomic resource for arthropod-associated microorganisms available to date and contributes substantially to populating the tree of life. More importantly, it provides a backbone for studying the metabolic potential of the termite gut microbiota, including the key members involved in carbon and nitrogen biogeochemical cycles, and important clues that may help cultivating representatives of these understudied clades.


## Introduction

Termites (Blattodea: Termitoidae) are eusocial insects that have predominantly and successfully colonized tropical and subtropical areas across the world. One of the keys to this success is their rare ability to degrade lignocellulose, a very abundant but recalcitrant complex carbon substrate (Cragg et al., 2015). As major decomposers, termites play an important role in carbon cycling (Yamada et al., 2005; Dahlsjö et al., 2014; Liu et al., 2015; Griffiths et al., 2019). Lignocellulose digestion by termites is attributed to the presence of a specific microbiota colonizing the different gut compartments of the host (Brune, 2014). Even though termites produce endogenous cellulases in the labial glands and/or midgut (Tokuda et al., 2004; Fujita, Miura & Matsumoto, 2008), the digestive processes in the hindgut are the result of microbial activities.

"Lower" termites feed almost exclusively on wood, whereas "higher" termites (Termitidae family) diversified their diet and extended it from wood to plant litter, humus, and

PeerJ Preprints NOT PEER-REVIEWED

soil (Donovan, Eggleton & Bignell, 2001). Higher termites represent the most diverse and taxon-rich clade and form about 85% of the termite generic diversity (Krishna et al., 2013). Their gut morphology is more complex than that of the basal clades, and is characterized by the presence of a mixed-segment and an enlarged proctodeal segment P1. Moreover, the gut displays strong variations in pH and oxygen partial pressure along the anterior-posterior axis, which creates microenvironments within the gut (Brune, 2014).

Termites harbor a specific and complex gut microbiota (Brune & Dietrich, 2015; Bourguignon et al., 2018). Over the last decade, numerous studies targeting the 16S rRNA gene have cataloged the prokaryotic diversity of the termite gut microbiota. By analyzing the structure and composition of these microbial communities, the roles of host taxonomy (Dietrich, Kohler & Brune, 2014; Abdul Rahman et al., 2015), host diet (Mikaelyan et al., 2015a), and microenvironments found in the different gut compartments (Mikaelyan, Meuser & Brune, 2017) have emerged as the main factors shaping the termite gut microbiota. These studies have also highlighted patterns of dominant taxa associated with specific diet and/or gut compartment (Mikaelyan, Meuser & Brune, 2017). For instance, Spirochaetes tend to be the dominant phylum in the gut of wood/grass feeders, whereas their abundance is lower in litter, humus and soil feeders, in which Firmicutes are much more abundant. The accumulated 16S rRNA gene reads have revealed the existence of termite-specific clusters among both bacterial and archaeal phyla (e.g. among Fibrobacteres, Clostridia, Spirochaetes, and Euryarchaeota).

All these studies focusing on the 16S rRNA gene have helped microbiologists in answering the question “who is there?”, but the following questions “what are they doing?” and “who is doing what?” remain open. Attempts to answer the latter questions have been made, e.g., by analyzing different fractions of the gut content of *Nasutitermes* spp., which led to the identification of fiber-associated cellulolytic bacterial taxa (Mikaelyan et al., 2014), or by focusing on the diversity of individual functional marker genes, such as *nifH* (Ohkuma, Noda & Kudo, 1999) or formyl-tetrahydrofolate synthetase (Ottesen & Leadbetter, 2011). The latter approach, however, is problematic because the organismal origin of the respective genes is often obfuscated by frequent horizontal gene transfers between prokaryotes. Thus, it has been suggested that genome-centric instead of gene-centric approaches are much more relevant for elucidation of soil or gut microbiotas (Prosser, 2015). Unfortunately, the number of available isolates of termite gut microbiota and their genomes (Zheng & Brune, 2015; Yuki et al., 2018) are low compared to those from other environments. However, modern culture-independent methods, namely metagenomics and single-cell genomics have recently allowed the generation of numerous metagenome-assembled genomes (MAGs) and single-amplified genomes (SAGs), respectively, from uncultivated or difficult to cultivate organisms (Albertsen et al., 2013; Woyke, Doud & Schulz, 2017). MAGs are becoming increasingly more prominent in the literature (Bowers et al., 2017) and populate the tree of life (Parks et al., 2017). Additionally, MAGs offer the opportunity to explore the metabolic potential of these organisms and to link it with their

ecology.

To date, only a limited number of MAGs and SAGs of uncultured bacteria have been recovered from the guts of higher termites; these represent termite-specific lineages of Fibrobacteres (Abdul Rahman et al., 2016) and Cyanobacteria (Utami et al., 2018). Here, we applied a binning algorithm to 30 metagenomes from different gut compartments of eight higher termite species encompassing different feeding groups to massively recover hundreds of prokaryotic MAGs from these samples. After quality filtering, all these MAGs were taxonomically identified within a phylogenomic framework and are discussed in the context of insect gut microbiology and symbiosis.

## Materials and Methods

### Metagenomic datasets

To cover a wide range of microbial diversity, we used 30 metagenomic datasets representing the main gut compartments (crop, midgut, P1–P5 proctodeal compartments of the hindgut) and main feeding groups present in higher-termites (see Table 1). Eight species of higher termites, identified by both morphological criteria and analysis of the mitogenome, were considered: *Cornitermes* sp., *Cubitermes ugandensis*, *Microcerotermes parvus*, *Nasutitermes corniger*, *Neocapritermes taracua*, *Termes hospes* (Dietrich & Brune, 2016), *Labiotermes labralis* and *Embiratermes neotenicus* (Hervé & Brune, 2017). Field experiments were approved by the French Ministry for the Ecological and Solidarity Transition (UID: ABSCH-CNA-FR-240495-2). Processing of the termite samples and DNA extraction and purification were described previously (Rossmassler et al., 2015). Metagenomic libraries were prepared, sequenced, quality controlled, and assembled at the Joint Genome Institute (Walnut Creek, CA, USA). DNA was sequenced using Illumina HiSeq 2000 or Illumina HiSeq 2500 (Illumina Inc., San Diego, CA). Quality-controlled reads were assembled and uploaded to the Integrated Microbial Genomes (IMG/M ER) database (Markowitz et al., 2014). Accession numbers and information about these 30 metagenomes can be found in Table S1.

### Genome reconstruction

For each metagenomic dataset, both quality-controlled (QC) and assembled (contigs) reads were downloaded from IMG/M ER in August 2017. To obtain coverage profile of contigs from each metagenomic assembly, the QC reads were mapped to contigs using BWA v0.7.15 with the bwa-mem algorithm (Li & Durbin, 2009). This generated SAM files that were subsequently converted into BAM files using SAMtools v1.3 (Li et al., 2009). Combining coverage profile and tetranucleotide frequency information, genomes were reconstructed from each metagenome with

PeerJ Preprints NOT PEER-REVIEWED

MetaBAT version 2.10.2 with default parameters (Kang et al., 2019). Quality of the reconstructed genomes was estimated with CheckM v1.0.8 (Parks et al., 2015). Only MAGs that were at least 50% complete and with less than 10% contamination, were retained for subsequent analyses. These MAGs have been deposited at GenBank under the BioProject accession number PRJNA560329; genomes are available with accession numbers SRR9983610-SRR9984198 (Table S2). For each MAG, CheckM was also used to extract 16S rRNA gene sequences as well as a set of 43 phylogenetically informative marker genes consisting primarily of ribosomal proteins and RNA polymerase domains. Finally, CheckM was also used for a preliminary taxonomic classification of the MAGs by phylogenetic placement of the MAGs into the CheckM reference genome tree. When available, 16S rRNA gene sequences were classified using the k-nearest neighbor (knn) algorithm implemented in mothur v1.39.5 (Schloss et al., 2009) and the BLASTN search method with the SILVA reference database release 132 (Quast et al., 2013) and DictDb v3 (Mikaelyan et al., 2015b).

## Phylogenomic analysis

In order to improve the initial CheckM classification, genome trees were built for each clade of interest (from kingdom to family level). Using this initial CheckM classification and when available, the 16S rRNA gene classification, genomes of closely related organisms and relevant outgroups were manually selected and downloaded from NCBI and IMG/M ER. These genomes were subjected to a similar CheckM analysis to extract a set of 43 single-copy marker genes, to translate them into amino acid sequences, and to create a concatenated fasta file (6,988 positions). For each clade of interest, the amino acid sequences from the MAGs, their relatives, and outgroups were aligned with MAFFT v7.305b and the FFT-NS-2 method (Katoh & Standley, 2013), and the resulting alignment was filtered using trimAL v1.2rev59 with the gappyout method (Capella-Gutierrez et al., 2009). Smart Model Selection (Lefort, Longueville & Gascuel, 2017) was used to determine the best model of amino acid evolution of the filtered alignment based on Akaike Information Criterion. Subsequently, a maximum-likelihood phylogenetic tree was built with PhyML 3.0 (Guindon et al., 2010). Branch supports were calculated using a Chi2-based parametric approximate likelihood-ratio test (aLRT) (Anisimova & Gascuel, 2006). Finally, each tree was visualized and edited with iTOL (Letunic & Bork, 2019). Following the procedure described above, a genome tree containing only the MAGs generated in the present study was also built and visualized with GraPhlAn version 0.9.7 (Asnicar et al., 2015).

PeerJ Preprints NOT PEER-REVIEWED

### Estimation of the relative abundance of the MAGs in each metagenome

For each metagenome, raw reads were mapped against assembled contigs using BWA (Li & Durbin, 2009) with default parameters. Unmapped reads and reads mapped to more than one location were removed by using SAMtools (Li et al., 2009) with parameters: F 0x904. Reads mapped to each contigs were summarized using the "pileup.sh" script (BBmap 38.26) (Bushnell, 2014). To determine the relative abundance of each MAG in the metagenome from which it was binned, reads mapped to all contigs belonging to each MAG were calculated. The relative abundance of one MAG was estimated by dividing all reads mapped to the MAG by all reads mapped to all contigs in the metagenomes. Similarly, the MAG coverage was estimated by multiplying the mapped reads by the read length and dividing it by the MAG length.

### Statistical analyses

Statistical analyses were performed with R version 3.4.4 (R Development Core Team, 2015), and data were vizualized with the *ggplot2* (Wickham, 2016) package. Correlations between quantitative variables were investigated with Pearson's product moment correlation coefficient.

## Results and Discussion

### Metagenomes and MAGs overview

Metagenomic reads were generated from the P1, P3 and P4 proctodeal compartments of the gut of the two termite species *Embiratermes neotenicus* and *Labiotermes labralis*. These six metagenomes were combined with 24 previously published metagenomes from the gut of higher termites (Rossmassler et al., 2015) in order to obtain data encompassing different gut compartments from eight species of higher termites feeding on different lignocellulosic substrates ranging from wood to soil (Table 1). Metagenomic binning of these 30 termite gut metagenomes yielded 1732 bins in total (Table S1). For further analysis, we selected only those bins that represented high-quality (135 bins, >90% complete and <5% contamination) and medium-quality (454 bins, >50% complete and <10% contamination) MAGs (Table 1, Table S1). The present study focused on these 589 MAGs, which showed on average a 38.6-fold coverage (Table S2).

The number of MAGs recovered from the different metagenomes did not show a Gaussian distribution. Instead, we found a significant and positive relationship between the number of metagenome-assembled reads and the number of MAGs recovered ($r = 0.85$, $p < 0.0001$), indicating that assembly success and sequencing depth were important predictors of genome reconstruction success (Figure 1). This is in agreement with benchmarking reports on metagenomic datasets (Sczyrba et al., 2017) and underscore that a good quality assembly is a

prerequisite for high binning recovery, which is important to consider when designing a metagenomic project for the purpose of binning. A significantly higher number of assembled reads and of MAGs recovered was observed in the current dataset compared to the Rossmassler et al., 2015 dataset (Wilcoxon test, $p < 0.005$), highlighting the importance of this new dataset (Figure 1).

## MAGs taxonomy and abundance

We investigated the phylogenomic context of the 589 MAGs. An initial automated classification of the MAGs using CheckM and when available, the taxonomic assignment of the 16S rRNA gene, identified representatives of 15 prokaryotic phyla (Table S3). Initially, 142 MAGs (24% of the dataset) remained unclassified at the phylum level, and key taxa of the termite microbiota, such as Fibrobacteres and *Treponema*, were absent or only poorly represented. This is partly explained by the lack of representative genomes for certain taxa in the reference genome tree provided in the current version of CheckM (e.g., only one Fibrobacteres genome and one Elusimicrobia genome, and an absence of Bathyarchaeota and Kiritimatiellaeota genomes). New tools incorporating larger databases, such as GTDB-Tk (Parks et al., 2018), will probably resolve such issues.

We improved the taxonomic resolution of the classification by iteratively constructing genome trees for each clade of interest that included all recently published reference genomes. This approach allowed the successful classification of all 589 MAGs at least at the phylum level and in some cases down to the genus level (Table S2). Thirty-eight MAGs were from the archaeal domain, and 551 MAGs were from the bacterial domain, which together represented a total of 17 prokaryotic phyla (Figure 2). Obvious patterns in the taxonomic distribution of the MAGs according to the sample origin were not apparent, which reflects the lack of effects of the gut compartments and/or of the diet of the host on the genome taxonomy (Figures S1 and S2). Among the most abundant phyla, genomes were recovered from different gut compartments and diets, which indicated a good coverage of the diversity among gut compartments and host diets.

We computed the relative abundance of each MAG. These abundances ranged from 0.005% to 4.63% (Table S2), with a mean value of 0.23%, which can be considered as abundant (Delmont et al., 2018). The average mean value indicated that the present dataset includes abundant members of the termite gut microbiota, which was confirmed when we looked at the taxonomic distribution of the MAGs (Figure 3), in particular when we linked it to the host diet. Indeed, similarities were observed when we compared taxonomic patterns of the MAG relative abundance with previously published 16S rRNA gene amplicon-based surveys (Abdul Rahman et al., 2015; Mikaelyan, Meuser & Brune, 2017). For instance, Spirochaetes were the most abundant phylum within the wood-feeding termite *Nasutitermes corniger,* and their proportion decreases along the humification gradient, being less abundant in the gut of humus feeders and litter feeders and even less abundant in soil feeders, in the favor of other phyla such as

Firmicutes. Fibrobacteres were preferentially abundant within wood- and litter-feeder samples. Interestingly, a significant and negative relationship between the number of metagenome-assembled reads and the MAG relative abundance ($r = -0.34$, $p < 0.0001$) was observed. This could be partly explained by the fact that increasing sequencing depth would increase the number of metagenome-assembled reads and thus allow the binning of sequences from less abundant organisms. However, since quantity of metagenome-assembled reads and relative abundance are not independent variables, it also implies that MAG relative abundances can not be directly quantitatively compared between samples but only within a single sample. Thus, proportions of taxa within a sample using relative abundance can be used to describe such sample.

## Archaea

The archaeal domain was represented by members of the phyla Euryarchaeota and Bathyarchaeota (Figure 4, Figure S3). Euryarchaeota were represented by 23 MAGs that were classified as members of the genera *Methanobrevibacter* (family *Methanobacteriaceae*; 3 MAGs) and, *Methanimicrococcus* (family *Methanosarcinaceae*; 3 MAGs), and members of the family *Methanomassiliicoccaceae* (16 MAGs), one of them in the genus *Candidatus* Methanoplasma. MAGs assigned as Euryarchaeota encompassed three (*Methanobacteriales*, *Methanosarcinales,* and *Methanomassiliicoccales*) of the four orders of methanogens found in termite guts (Brune, 2018); *Methanomicrobiales* were absent from the present dataset. This genomic resource will be extremely valuable for a better understanding of the genomic basis of methanogenesis in the termite gut and more generally for investigating the functional role of archaea in arthropod guts. Indeed, Euryarchaota have been found to be present in virtually all termite species investigated (Brune, 2018), and a global 16S rRNA gene survey has revealed that this phylum is the most abundant archaeal clade in arthropod-associated microbiota (Schloss et al., 2016). Bathyarchaeota were represented by 15 MAGs, which formed a termite-specific cluster, with Bathyarchaeota reconstructed from sediments of the White Oak River (WOR) estuary (North Carolina, USA) as next relatives (Lazar et al., 2016) (Figure 4). Bathyarchaeota is a lineage formerly known as Miscellaneous Crenarchaeota Group (MCG), which has been reported to occur in the gut of soil-feeding termites (Friedrich et al., 2001). To date, MAGs of Bathyarchaeota have been mostly derived from aquatic environments (Zhou et al., 2018). Here, we identify the members of this lineage as Bathyarchaeota and provide the first genomes from this environment. Interestingly, Bathyarchaeota MAGs were particularly abundant in humus-, litter- and soil-feeding termites (Figure 3); a genomic characterization, combined with analyses of their abundance and localization, should shed light on the metabolic potential of these organisms and their functional role in termite guts.

## Firmicutes

Firmicutes was by far the most abundantly represented phylum. The 237 MAGs (40% of the total dataset) represented three classes (*Bacilli*, *Clostridia* and *Erysipelotrichia*) and ten families,

including four members of *Streptococcaceae* (*Bacilli*) and three members of *Turicibacteraceae* (*Erysipelotrichia*). *Clostridia* was the most diverse and rich class (191 MAGs), in which *Ruminococcaceae* (91 MAGs), *Defluviitaleaceae* (64 MAGs), *Lachnospiraceae* (4 MAGs), *Peptococcaceae* (4 MAGs), *Eubacteriaceae* (2 MAGs), *Symbiobacteriaceae* (2 MAGs), Family XIII *incertae sedis* (2 MAGs) and Clostridiales *incertae sedis* (2 MAGs) families were identified. Interestingly, among the *Defluviitaleaceae*, the genomes were mainly recovered from the P1 compartment (51 MAGs, i.e., 80% of the family members) whereas *Ruminococcaceae* were predominantly recovered from the P3 compartment (58 MAGs, i.e., 64% of the family members). Further studies should investigate the potential metabolic specialization of these two families in relation to the gut physicochemical properties. A fourth class-level lineage could not be further classified for lack of reference genomes. In a recent global 16S rRNA gene-based survey, it has been suggested that many novel lineages of Firmicutes in insect-associated metagenomes are hidden (Schulz et al., 2017). Our present study confirms this idea but our genome trees also provide evidence of new lineages. Here we report the first genomes of uncultured termite-specific lineages that were already detected in previous 16S rRNA gene-based surveys (Bourguignon et al., 2018). For example, the phylogenomic tree of the most abundant family *Ruminococcaceae* (Figure S4) showed various termite-specific clusters, including a cluster of 18 MAGs closely related to *Sporobacter termitidis* isolated from *Nasutitermes lujae* (Grech-Mora et al., 1996). *Lachnospiraceae*, *Ruminococcaceae*, *Turicibacteraceae* (previously classified as *Erysipelotrichaceae*), and *Defluviitaleaceae* (previously classified as *Lachnospiraceae*) have been reported among the dominant taxa in termite guts (Mikaelyan, Meuser & Brune, 2017), but most of them remain uncultivated and/or with few representative genomes. As such, many questions regarding their ecology and metabolism remain open. With 237 Firmicutes MAGs recovered from different gut compartments and from hosts with different diets, the present study provides the material for further genomic exploration of the role of these bacteria in plant polysaccharide degradation, based for instance on CAZyme distribution (Lombard et al., 2014). Since diet has been shown to be the main factor shaping gut community composition in higher termites (Mikaelyan et al., 2015a), one might hypothesize the existence of different arsenals of lignocellulolytic enzymes, potentially reflecting the host diet specificity (balance between cellulose, lignin, and hemicelluloses). More generally, Firmicutes and especially *Ruminococcaceae* are also abundant and metabolically important in rumen systems (Svartström et al., 2017; Söllinger et al., 2018; Stewart et al., 2018). At a broader scale, our dataset will allow comparative studies between intestinal tract microbiota of ruminants and phytophagous or xylophagous invertebrates, which would allow a better understanding of plant polysaccharide degradation across the tree of life.

## Actinobacteria

Actinobacteria was the second most abundant phylum with 71 MAGs, including members of the classes *Acidimicrobiia*, *Actinobacteria*, *Coriobacteriia* and *Thermoleophilia* (Figure S5). Eight

families were represented, namely *Propionibacteriaceae* (11 MAGs), *Promicromonosporaceae* (4 MAGs), *Eggerthellaceae* (4 MAGs), *Microbacteriaceae* (2 MAGs), *Nocardioidaceae* (2 MAGs), *Acidimicrobiaceae* (1 MAG), *Nocardiaceae* (1 MAG) and *Conexibacteraceae* (1 MAG). Among these 71 MAGs, 36 were recovered from humus feeders, 33 from soil feeders but only 2 from wood feeders, which suggests a higher prevalence in termites with a more humified diet. This phylum is known to be present and of significant abundance in both the nest (Sujada, Sungthong & Lumyong, 2014) and gut of termites (Le Roes-Hill, Rohland & Burton, 2011), but to be more abundant in the nest (Moreira et al., 2018). This was for instance the case for the families *Acidimicrobiaceae*, *Nocardiaceae*, *Promicromonosporaceae*, *Microbacteriaceae*, *Nocardioidaceae,* and *Propionibacteriaceae*, which were more abundant in the nest than in the gut of workers or soldiers of *Procornitermes araujoi* (Moreira et al., 2018). Therefore, one of the key questions regarding this phylum concerns their effective role in lignocellulose degradation in the termite guts. Are they just present in the surrounding environment of the termite and thus sometimes transit from the gut or are they actively involved in food digestion? The MAGs obtained in the present study will allow to address such questions by evaluating gene expression of these organisms using metatranscriptomic data from higher termites (He et al., 2013; Marynowska et al., 2017).

## Spirochaetes

The phylum Spirochaetes was represented by 68 MAGs from wood-, soil-, litter- and humus-feeding termites. It has long been known that Spirochaetes are a diverse and important lineage in termite gut (Paster et al., 1996; Lilburn, Schmidt & Breznak, 1999), especially because of their involvement in reductive acetogenesis (Leadbetter et al., 1999; Ohkuma et al., 2015) and in hemicellulose degradation (Tokuda et al., 2018). In terms of abundance, Spirochaetes are among the dominant phyla in termite guts and may represent more than half of the bacterial relative abundance in some species (Diouf et al., 2018a). Three Spirochaetes orders, namely *Brevinematales* (1 MAG), *Leptospirales* (4 MAGs) and *Spirochaetales* (59 MAGs), were identified (Figure 5, Figure S6). In the latter order, 54 MAGs recovered from the P1, P3 and P4 compartments of wood-, litter-, humus-, and soil-feeding hosts were assigned to the termite-specific cluster *Treponema* I (Ohkuma, Iida & Kudo, 1999; Lilburn, Schmidt & Breznak, 1999) and represent the first genomes of this cluster from higher termites. Indeed, to date only two *Treponema* I genomes are available, and both were recovered from isolates, namely *T. azotonutricium* and *T. primitia*, from the hindgut of the lower termite *Zootermopsis angusticollis* (Graber, Leadbetter & Breznak, 2004). Thus, our dataset significantly expands the genomic resources for this taxonomic group. Subclusters of this clade have been identified on a dedicated genome tree (Figure 5). The genome tree topology is in agreement with a previous phylogenomic Spirochaetes study (Gupta, Mahmood & Adeolu, 2013). Regarding Spirochaetes classification, our tree topology suggests that the genus *Treponema* could be elevated at least to the family rank due to the presence of distinct *Treponema* clusters (Figure 5). This observation is

PeerJ Preprints NOT PEER-REVIEWED

also in agreement with the recent Genome Taxonomy Database, which now proposes a *Treponemataceae* family and a *Treponematales* order (Parks et al., 2018).

## Fibrobacteres

Thirteen members of the Fibrobacteres phylum were recovered from the P1, P3, and P4 compartments of wood-, litter-, humus-, and soil-feeding termites. These genomes encompass the 3 classes of this phylum, namely *Chitinispirillia* (5 MAGs), *Chitinivibrionia* (previously known as TG3 candidate phylum; 2 MAGs), and *Fibrobacteria* (6 MAGs) (Figure 6, Figure S7). Phylogenomic analysis of this phylum suggests the existence of a termite-specific cluster among *Fibromonadales* (Figure 6). This is in agreement with a previous 16S rRNA gene-based phylogenetic analysis and a phylogenomic analysis that identify the family *Fibromonadaceae* as a termite-specific cluster within this order (Abdul Rahman et al., 2016). Members of the phylum Fibrobacteres are abundant in the hindgut of wood-feeding higher termites (Hongoh et al., 2006), where they have been identified as fiber-associated cellulolytic bacteria (Mikaelyan et al., 2014). To date, available Fibrobacteres genomes from termite guts belong to *Chitinivibrionia* (previously classified as TG3 phylum) and *Fibrobacteria* (Abdul Rahman et al., 2016). Here we also added five members of *Chitinispirillia* to the list of termite-associated Fibrobacteres genomes. Interestingly, we did not recover MAGs within the *Fibrobacterales,* which harbors the *Fibrobacter* genus, a clade that was also absent from 16S rRNA gene-based surveys of termite gut microbiota (Hongoh et al., 2006; Mikaelyan et al., 2015b; Bourguignon et al., 2018). Members of this genus have been isolated from the gastrointestinal tracts of mammals and bird herbivores (Neumann, McCormick & Suen, 2017), where they are potentially involved in cellulose and hemicellulose degradation (Neumann & Suen, 2018). This suggests co-evolution patterns among different Fibrobacteres clades within animal hosts with a lignocellulose-based diet.

## Proteobacteria and Bacteroidetes

Sixty-seven MAGs of Proteobacteria belonging to *Alphaproteobacteria* (21 MAGs), *Betaproteobacteria* (15 MAGs), and *Deltaproteobacteria* (21 MAGs) were recovered from all hindgut compartments of litter-, humus-, and soil-feeding termites. Among the *Deltaproteobacteria*, six orders were identified, namely *Desulfobacterales* (3 MAGs, all assigned to *Desulfobulbus*), *Desulfovibrionales* (5 MAGs), *Desulfuromonadales* (1 MAG), *Myxococcales* (4 *Cystobacterineae* and 4 *Polyangiaceae*), Rs-K70 group (1 MAG), and *Syntrophobacterales* (1 *Syntrophaceae* and 2 *Syntrophorhabdaceae*). *Desulfovibrionaceae* (*Desulfovibrionales*) members of gut and termite-gut clusters have been found to be highly prevalent in termite guts (Bourguignon et al., 2018). Similarly, we identified 3 *Desulfovibrionaceae* MAGs that form a monophyletic clade and 2 *Desulfovibrionaceae* MAGs that fall into a cluster of gut-associated genomes (Figure S8). This family, among others, is composed of various sulfate-reducing bacteria; this functional group has already been identified

in different termite species (Kuhnigk et al., 1996). Thus, these MAGs could provide new genomic resources to further investigate this metabolism in the termite gut.

Our dataset comprises 33 MAGs of Bacteroidetes (Figure S9), including members of the families *Candidatus* Azobacteroides (4 MAGs), *Lentimicrobiaceae* (5 MAGs), *Paludibacteraceae* (2 MAGs, both assigned to the *Paludibacter* genus), *Rikenellaceae* (2 MAGs), *Marinilabiliaceae* (1 MAG), and *Prolixibacteraceae* (1 MAG). These Bacteroidetes were found in the P1, P3, and P4 compartments and in wood-, litter-, humus-, and soil-feeding termites. In Blattodea guts, different clusters of *Alistipes* (Bacteroidetes) have been found in a 16S rRNA gene survey (Mikaelyan et al., 2015b). Two MAGs from *Labiotermes labralis* belonging to the *Rikenellaceae* family and closely related to *Alistipes* have been identified. Additionally, among Bacteroidetes, four MAGs, all originating from P4 compartments, fall into the *Candidatus* Azobacteroides family that contains symbionts of flagellates from guts of lower termites (Hongoh et al., 2008b; Yuki et al., 2015). We also recovered two MAGs assigned to *Paludibacter*; *Paludibacter propionicigenes* and *Paludibacter jiangxiensis* are both strictly anaerobic, propionate-producing bacteria isolated from rice paddy field (Ueki et al., 2006; Qiu et al., 2014). Propionate is produced by fermenting bacteria in the gut of termites (Odelson & Breznak, 1983); these bacteria utilize glucose generated by cellulose degradation to form succinate and propionate (Tokuda et al., 2014). *P. propionicigenes* might be involved in nitrogen fixation, as *nifH* transcripts assigned to this species are the most abundant in the gut of the wood-feeding beetle *Odontotaenius disjunctus* (Ceja-Navarro et al., 2014).

## Saccharibacteria, Synergistetes and Planctomycetes

Fifteen MAGs of *Candidatus* Saccharibacteria (also known as candidate division TM7) were reconstructed from the P1, P3, and P4 compartments of wood-, litter-, humus- and soil-feeding termites (Figure S10). Most of them originated from humus feeders (11 MAGs), especially from the P3 compartment (8 out of these 11 MAGs). Similarly, six MAGs of Synergistetes, all belonging to the *Synergistaceae* family that contains a termite/cockroach cluster (Mikaelyan et al., 2015b), were recovered from the P3 and P4 compartments of humus- and soil-feeding termites (Figure S11). Both Saccharibacteria and Synergistetes were recently highlighted as numerically important clades of the termite gut microbiota, with some OTUs being present in the gut of the majority of 94 termite species collected across four continents (Bourguignon et al., 2018). Genomic analysis of these MAGs should help in understanding the roles of these bacteria in termite gut and also provide clues for designing successful isolation media to study their physiology.

Twelve MAGs were assigned to the phylum Planctomycetes, including 4 to the class *Phycisphaerae* (and among them 2 classified as *Phycisphaerales*) and 7 to the class *Planctomycetia* (all classified as *Planctomycetaceae*) (Figure S12). These MAGs were recovered from the P3, P4, and P5 compartments and were restricted to humus- and soil-feeding termites.

The recovery of Planctomycetes was expected, especially from the *Planctomycetaceae* family, which also contains a termite / cockroach cluster (Mikaelyan et al., 2015b). Interestingly, we found three MAGs from the P4 and P5 compartments of *Cubitermes ugandensis*, with one 16S rRNA gene sequence assigned to the termite/cockroach cluster 2 (according to DictDb v3.0 classification), described in a previous study investigating the gut microbiota of the same termite species (Köhler et al., 2008). When such 16S rRNA gene information is available, it will allow the direct linkage between prokaryotic taxonomy and potential metabolisms.

## Other phyla

Nine members of the phylum Elusimicrobia were identified, including members of the class *Endomicrobia* (7 members) and *Elusimicrobia* (1 member) (Figure S13). These were found in all hindgut compartments and were restricted to humus- and soil-feeding termites. Currently, only three complete genomes of Elusimicrobia from insect guts are available: *Elusimicrobium minutum* from the gut of a humivorous scarab beetle larva (Herlemann et al., 2009), and *Endomicrobium proavitum* (Zheng & Brune, 2015) and *Candidatus* Endomicrobium trichonymphae (Hongoh et al., 2008a) from the termite gut. Here, we provided 9 additional genomes from the guts of humus- and soil-feeding termites.

The Chloroflexi phylum was represented by eight MAGs, including seven belonging to the class *Dehalococcoidia*, found exclusively in the P3 and P4 compartments of humus- and soil-feeding termites (Figure S10). Their function in termite gut remains unclear, but Chloroflexi, including *Dehalococcoidia*, were found to be enriched in lignin-amended tropical forest soil (DeAngelis et al., 2011), where oxygen concentration and redox potential are highly variable, as in the termite gut (Brune, 2014). Therefore, their ability to use oxygen as final electron acceptor and their potential involvement in lignin degradation could be investigated by comparative genomics.

Minor phyla were also present in our dataset. Two MAGs assigned as Cloacimonetes (Figure S14) and five MAGs assigned as Kiritimatiellaeota were recovered from the P3 compartment of the two humus-feeding termites *Neocapritermes taracua* and *Termes hospes* (Figure S15). Kiritimatiellaeota have been reported to be present in the digestive tract of various animals (Spring et al., 2016). The few clones obtained from termite guts, which had been tentatively classified as uncultured Verrucomicrobia, were mostly obtained with planctomycete-specific primers (Köhler et al., 2008), underscoring the potential biases in amplicon-based studies toward certain taxa. Similarly, one MAG of Microgenomates (also known as candidate division OP11), which probably represents a lineage of Pacebacteria that was discovered only in a recent amplicon-based analysis but occurs in the majority of termites investigated (Bourguignon et al., 2018), was reconstructed from the P3 compartment of *Termes hospes* (Figure S10).

Finally, four MAGs classified as Acidobacteria were reconstructed from either the P3 or

P4 compartments of humus- and soil-feeding termites (Figure S16), which show a moderately alkaline or circumneutral pH in comparison to the highly alkaline P1. Of these four genomes, two were assigned to the family *Holophagaceae* and one to the *Acidobacteriaceae*. Acidobacteria can represent a significant fraction of the termite gut microbiota, especially in wood-feeding termites (Hongoh et al., 2005; Wang et al., 2016; Bourguignon et al., 2018). Moreover, *Holophagaceae* and *Acidobacteriaceae* have been reported to be present in moderately acidic lignocellulosic substrates, such as peatland soil (Schmidt et al., 2015) and decaying wood (Hervé et al., 2014). Genomic analysis should help us in identifying the metabolic potential of these MAGs for lignocellulose degradation.

### Phyla not represented by MAGs

Several bacterial phyla and one archaeal phylum containing prominent taxa that have been identified in previous 16S rRNA gene surveys of termite guts were not represented among the MAGs recovered in the present study. They include Cyanobacteria (class *Melainabacteria*; Utami et al., 2018), Lentisphaerae (Köhler et al., 2012; Sabree & Moran, 2014), Verrucomicrobia (Wertz et al., 2012), and Thaumarchaeota (Friedrich et al., 2001; Shi et al., 2015). Also intracellular symbionts of termite tissues, such as *Wolbachia* (Proteobacteria) (Diouf et al., 2018b) were not recovered. Possible reasons are a low relative abundance and/or a high phylogenetic diversity of the respective lineages. Although larger metagenomes should improve the chances of their recovery in the medium- and high-quality bins, targeted single-cell based approaches have proven to be quite effective in recovering these genomes (Ohkuma et al., 2015; Yuki et al., 2015; Utami et al., 2019).

## Conclusions

The 589 MAGs reported here represent the largest genomic resource for arthropod-associated microorganisms available to date. Moreover, almost all major prokaryotic lineages previously identified in 16S rRNA gene amplicon-based surveys of the gut of higher termites were recovered from our 30 metagenomes. This provides the foundations for studying the prokaryotic metabolism of the termite gut microbiota, including the key members involved in carbon and nitrogen biogeochemical cycles, and important clues that may help in cultivating representatives of these understudied clades.

## Acknowledgments

The authors thank all JGI staff, particularly their project manager Tijana Glavina del Rio, for their excellent service. The technical assistance of Katja Meuser is highly appreciated.

PeerJ Preprints NOT PEER-REVIEWED

## Tables

**Table 1. Recovery of metagenome-assembled genomes (MAGs) from the 30 termite gut metagenomes analyzed in this study.** The host termite, its mitochondrial genome accession number, dietary preference, and the originating gut compartments are indicated. *C* crop (foregut), *M* midgut, *P1–P5* proctodeal compartments (hindgut). The sample codes used for the metagenomes are the combination of host ID and gut compartment.

## Figure legends

**Figure 1: Relationship between the number of MAGs recovered and the number of assembled reads in the respective metagenomes.** The linear regression line and the Pearson correlation coefficient (r) are shown for the entire dataset.

**Figure 2: Distribution of the 589 MAGs among bacterial and archaeal phyla.** This maximum-likelihood tree was inferred from a concatenated alignment (amino acids) of 43 protein-coding genes using the LG+G+I model of evolution.

**Figure 3: Relative abundance of the MAGs from different phyla among the respective metagenomes.** Circle size indicates the relative abundance of the MAGs among the respective metagenome sample; color indicates host diet.

**Figure 4: Phylogenomic tree of the archaeal domain.** This maximum-likelihood tree was inferred from a concatenated alignment of 43 proteins using the LG+G+I+F model of amino-acid evolution. Branch supports were calculated using a Chi2-based parametric approximate likelihood-ratio test. Names in bold included MAGs recovered in the present study. Clusters shaded in brown consist exclusively of MAGs from termite guts and clusters shaded in gray contain genomes from termite guts. The Asgard group was used as outgroup.

**Figure 5: Phylogenomic tree of the Spirochaetes phylum.** This maximum-likelihood tree was inferred from a concatenated alignment of 43 proteins using the LG+G+I+F model of amino-acid evolution. Branch supports were calculated using a Chi2-based parametric approximate likelihood-ratio test. Names in bold included MAGs recovered in the present study. Clusters shaded in brown consist exclusively of genomes from termite guts and clusters shaded in gray contain genomes from termite guts. Elusimicrobia and Cyanobacteria were used as outgroup.

**Figure 6: Phylogenomic tree of the Fibrobacteres phylum.** This maximum-likelihood tree was inferred from a concatenated alignment of 43 proteins using the LG+G+I+F model of amino-acid evolution. Branch supports were calculated using a Chi2-based parametric approximate likelihood-ratio test. Names in bold included MAGs recovered in the present study. Clusters

shaded in brown consist exclusively of genomes from termite guts and clusters shaded in gray contain genomes from termite guts. Bacteroidetes were used as outgroup.

## Supplementary information

**Supplementary Table S1:** Metagenome characteristics.

**Supplementary Table S2:** Final taxonomic assignment and characteristics of the MAGs

**Supplementary Table S3:** Initial taxonomic assignment of the MAGs

**Figure S1: Phylogenomic distribution of the MAGs according to the host diet.** The outer rings show the occurrence of MAGs in termites with different diets. The maximum likelihood tree was inferred from a concatenated alignment of 43 proteins using the LG+G+I model of amino-acid evolution.

**Figure S2: Phylogenomic distribution of the MAGs according to the gut compartment of the host.** The outer rings show the occurrence of MAGs in the different termite gut compartments: *C* crop (foregut), *M* midgut, *P1–P5* proctodeal compartments (hindgut). The maximum-likelihood tree was inferred from a concatenated alignment of 43 proteins using the LG+G+I model of amino-acid evolution.

**Figure S3: Phylogenomic tree of the Archaea.** This maximum-likelihood tree was inferred from a concatenated alignment of 43 proteins using the LG+G+I+F model of amino-acid evolution. Branch supports were calculated using a Chi2-based parametric approximate likelihood-ratio test. Asgard group was used as outgroup. Names in bold included MAGs recovered in the present study.

**Figure S4: Phylogenomic tree of the *Ruminococcaceae* family (Firmicutes).** This maximum-likelihood tree was inferred from a concatenated alignment of 43 proteins using the LG+G+I model of amino-acid evolution. Branch supports were calculated using a Chi2-based parametric approximate likelihood-ratio test. *Dorea* and *Butyrivibrio* (*Lachnospiraceae*) species were used as outgroup. Names in bold included MAGs recovered in the present study.

**Figure S5: Phylogenomic tree of the Actinobacteria.** This maximum-likelihood tree was inferred from a concatenated alignment of 43 proteins using the LG+G+I+F model of amino-acid evolution. Branch supports were calculated using a Chi2-based parametric approximate likelihood-ratio test. Chloroflexi species were used as outgroup. Names in bold included MAGs recovered in the present study.

**Figure S6: Phylogenomic tree of the Spirochaetes.** This maximum-likelihood tree was inferred from a concatenated alignment of 43 proteins using the LG+G+I+F model of amino-acid

evolution. Branch supports were calculated using a Chi2-based parametric approximate likelihood-ratio test. Elusimicrobia and Cyanobacteria were used as outgroup. Names in bold included MAGs recovered in the present study.

**Figure S7: Phylogenomic tree of the Fibrobacteres.** This maximum-likelihood tree was inferred from a concatenated alignment of 43 proteins using the LG+G+I+F model of amino-acid evolution. Branch supports were calculated using a Chi2-based parametric approximate likelihood-ratio test. Bacteroidetes were used as outgroup. Names in bold included MAGs recovered in the present study.

**Figure S8: Phylogenomic tree of the *Desulfovibrionaceae* family (*Deltaproteobacteria*).** This maximum-likelihood tree was inferred from a concatenated alignment of 43 proteins using the LG+G+I model of amino-acid evolution. Branch supports were calculated using a Chi2-based parametric approximate likelihood-ratio test. *Desulfonatronum* species were used as outgroup. Names in bold included MAGs recovered in the present study.

**Figure S9: Phylogenomic tree of the Bacteroidetes.** This maximum-likelihood tree was inferred from a concatenated alignment of 43 proteins using the LG+G+I model of amino-acid evolution. Branch supports were calculated using a Chi2-based parametric approximate likelihood-ratio test. Chlorobi species were used as outgroup. Names in bold included MAGs recovered in the present study.

**Figure S10: Phylogenomic tree of the Chloroflexi, Saccharibacteria and Microgenomates.** This maximum-likelihood tree was inferred from a concatenated alignment of 43 proteins using the LG+G+I+F model of amino-acid evolution. Branch supports were calculated using a Chi2-based parametric approximate likelihood-ratio test. Actinobacteria species were used as outgroup. Names in bold included MAGs recovered in the present study.

**Figure S11: Phylogenomic tree of the Synergistetes.** This maximum-likelihood tree was inferred from a concatenated alignment of 43 proteins using the LG+G+I+F model of amino-acid evolution. Branch supports were calculated using a Chi2-based parametric approximate likelihood-ratio test. Elusimicrobia species were used as outgroup. Names in bold included MAGs recovered in the present study.

**Figure S12: Phylogenomic tree of the Planctomycetes.** This maximum-likelihood tree was inferred from a concatenated alignment of 43 proteins using the LG+G+I+F model of amino-acid evolution. Branch supports were calculated using a Chi2-based parametric approximate likelihood-ratio test. Verrucomicrobia species were used as outgroup. Names in bold included MAGs recovered in the present study.

**Figure S13: Phylogenomic tree of the Elusimicrobia.** This maximum-likelihood tree was inferred from a concatenated alignment of 43 proteins using the LG+G+I+F model of amino-acid evolution. Branch supports were calculated using a Chi2-based parametric approximate

likelihood-ratio test. Spirochaetes species were used as outgroup. Names in bold included MAGs recovered in the present study.

**Figure S14: Phylogenomic tree of the Cloacimonetes.** This maximum-likelihood tree was inferred from a concatenated alignment of 43 proteins using the LG+G+I+F model of amino-acid evolution. Branch supports were calculated using a Chi2-based parametric approximate likelihood-ratio test. Fibrobacteres species were used as outgroup. Names in bold included MAGs recovered in the present study.

**Figure S15: Phylogenomic tree of the Kiritimatiellaeota.** This maximum-likelihood tree was inferred from a concatenated alignment of 43 proteins using the LG+G+I model of amino-acid evolution. Branch supports were calculated using a Chi2-based parametric approximate likelihood-ratio test. Chlamydiae species were used as outgroup. Names in bold included MAGs recovered in the present study.

**Figure S16: Phylogenomic tree of the Acidobacteria.** This maximum-likelihood tree was inferred from a concatenated alignment of 43 proteins using the LG+G+I+F model of amino-acid evolution. Branch supports were calculated using a Chi2-based parametric approximate likelihood-ratio test. Proteobacteria species were used as outgroup. Names in bold included MAGs recovered in the present study.

## References

Abdul Rahman N, Parks DH, Vanwonterghem I, Morrison M, Tyson GW, Hugenholtz P. 2016. A phylogenomic analysis of the bacterial phylum Fibrobacteres. *Frontiers in Microbiology* 6. DOI: 10.3389/fmicb.2015.01469.

Abdul Rahman N, Parks DH, Willner DL, Engelbrektson AL, Goffredi SK, Warnecke F, Scheffrahn RH, 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* 3:5. DOI: 10.1186/s40168-015-0067-8.

Albertsen M, Hugenholtz P, Skarshewski A, Nielsen KL, Tyson GW, Nielsen PH. 2013. Genome sequences of rare, uncultured bacteria obtained by differential coverage binning of multiple metagenomes. *Nature Biotechnology* 31:533–538. DOI: 10.1038/nbt.2579.

Anisimova M, Gascuel O. 2006. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Systematic biology* 55:539–52. DOI: 10.1080/10635150600755453.

Asnicar F, Weingart G, Tickle TL, Huttenhower C, Segata N. 2015. Compact graphical representation of phylogenetic data and metadata with GraPhlAn. *PeerJ* 3:e1029. DOI:

10.7717/peerj.1029.

Bourguignon T, Lo N, Dietrich C, Šobotník J, Sidek S, Roisin Y, Brune A, Evans TA. 2018. Rampant host switching shaped the termite gut microbiome. *Current Biology* 28:649-654.e2. DOI: 10.1016/j.cub.2018.01.035.

Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA, Yooseph S, Sutton G, Glöckner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Ettema TJG, Tighe S, Konstantinidis KT, Liu W-T, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Kyrpides NC, Schriml L, Garrity GM, Hugenholtz P, Sutton G, Yilmaz P, Meyer F, Glöckner FO, Gilbert JA, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Lapidus A, Meyer F, Yilmaz P, Parks DH, Eren AM, Schriml L, Banfield JF, Hugenholtz P, Woyke T. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nature Biotechnology* 35:725–731. DOI: 10.1038/nbt.3893.

Brune A. 2014. Symbiotic digestion of lignocellulose in termite guts. *Nature Reviews Microbiology* 12:168–180. DOI: 10.1038/nrmicro3182.

Brune A. 2018. *Methanogenesis in the digestive tracts of insects and other arthropods*. Berlin, Heidelberg, Heidelberg: Springer Berlin Heidelberg. DOI: 10.1007/978-3-540-77587-4.

Brune A, Dietrich C. 2015. The gut microbiota of termites: Digesting the diversity in the light of ecology and evolution. *Annual Review of Microbiology* 69:145–166. DOI: 10.1146/annurev-micro-092412-155715.

Bushnell B. 2014. *BBMap: a fast, accurate, splice-aware aligner*. DOI: 10.1186/1471-2105-13-238.

Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T, Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973. DOI: 10.1093/bioinformatics/btp348.

Ceja-Navarro JA, Nguyen NH, Karaoz U, Gross SR, Herman DJ, Andersen GL, Bruns TD, Pett-Ridge J, Blackwell M, Brodie EL. 2014. Compartmentalized microbial composition, oxygen gradients and nitrogen fixation in the gut of *Odontotaenius disjunctus*. *The ISME journal* 8:6–18. DOI: 10.1038/ismej.2013.134.

Cragg SM, Beckham GT, Bruce NC, Bugg TD, Distel DL, Dupree P, Etxabe AG, Goodell BS, Jellison J, McGeehan JE, McQueen-Mason SJ, Schnorr K, Walton PH, Watts JE, Zimmer M. 2015. Lignocellulose degradation mechanisms across the Tree of Life. *Current Opinion*

*in Chemical Biology* 29:108–119. DOI: 10.1016/j.cbpa.2015.10.018.

Dahlsjö CAL, Parr CL, Malhi Y, Meir P, Chevarria OVC, Eggleton P. 2014. Termites promote soil carbon and nitrogen depletion: Results from an *in situ* macrofauna exclusion experiment, Peru. *Soil Biology and Biochemistry* 77:109–111. DOI: 10.1016/j.soilbio.2014.05.033.

DeAngelis KM, Allgaier M, Chavarria Y, Fortney JL, Hugenholtz P, Simmons B, Sublette K, Silver WL, Hazen TC. 2011. Characterization of trapped lignin-degrading microbes in tropical forest soil. *PLoS One* 6:e19306. DOI: 10.1371/journal.pone.0019306.

Delmont TO, Quince C, Shaiber A, Esen ÖC, Lee ST, Rappé MS, McLellan SL, Lücker S, Eren AM. 2018. Nitrogen-fixing populations of Planctomycetes and Proteobacteria are abundant in surface ocean metagenomes. *Nature Microbiology* 3:804–813. DOI: 10.1038/s41564-018-0176-9.

Dietrich C, Brune A. 2016. 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*). *Mitochondrial DNA* 27:3903–3904. DOI: 10.3109/19401736.2014.987257.

Dietrich C, Kohler T, Brune A. 2014. The cockroach origin of the termite gut microbiota: Patterns in bacterial community structure reflect major evolutionary events. *Applied and Environmental Microbiology* 80:2261–2269. DOI: 10.1128/AEM.04206-13.

Diouf M, Hervé V, Mora P, Robert A, Frechault S, Rouland-Lefèvre C, Miambi E. 2018a. Evidence from the gut microbiota of swarming alates of a vertical transmission of the bacterial symbionts in *Nasutitermes arborum* (Termitidae, Nasutitermitinae). *Antonie van Leeuwenhoek* 111:573–587. DOI: 10.1007/s10482-017-0978-4.

Diouf M, Miambi E, Mora P, Frechault S, Robert A, Rouland-Lefèvre C, Hervé V. 2018b. Variations in the relative abundance of *Wolbachia* in the gut of *Nasutitermes arborum* across life stages and castes. *FEMS Microbiology Letters* 365. DOI: 10.1093/femsle/fny046.

Donovan SE, Eggleton P, Bignell DE. 2001. Gut content analysis and a new feeding group classification of termites. *Ecological Entomology* 26:356–366. DOI: 10.1046/j.1365-2311.2001.00342.x.

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*. *Applied and Environmental Microbiology* 67:4880–4890. DOI: 10.1128/AEM.67.10.4880-4890.2001.

Fujita A, Miura T, Matsumoto T. 2008. Differences in cellulose digestive systems among castes in two termite lineages. *Physiological Entomology* 33:73–82. DOI: 10.1111/j.1365-

3032.2007.00606.x.

Graber JR, Leadbetter JR, Breznak JA. 2004. Description of *Treponema azotonutricium* sp. nov. and *Treponema primitia* sp. nov., the first spirochetes isolated from termite guts. *Applied and Environmental Microbiology* 70:1315–1320. DOI: 10.1128/AEM.70.3.1315-1320.2004.

Grech-Mora I, Fardeau M-L, Patel BKC, Ollivier B, Rimbault A, Prensier G, Garcia J-L, Garnier-Sillam E. 1996. Isolation and characterization of *Sporobacter termitidis* gen. nov., sp. nov., from the digestive tract of the wood-feeding termite *Nasutitermes lujae*. *International Journal of Systematic Bacteriology* 46:512–518. DOI: 10.1099/00207713-46-2-512.

Griffiths HM, Ashton LA, Evans TA, Parr CL, Eggleton P. 2019. Termites can decompose more than half of deadwood in tropical rainforest. *Current Biology* 29:R118–R119. DOI: 10.1016/j.cub.2019.01.012.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Systematic Biology* 59:307–321. DOI: 10.1093/sysbio/syq010.

Gupta RS, Mahmood S, Adeolu M. 2013. A phylogenomic and molecular signature based approach for characterization of the phylum Spirochaetes and its major clades: proposal for a taxonomic revision of the phylum. *Frontiers in microbiology* 4:217. DOI: 10.3389/fmicb.2013.00217.

He S, Ivanova N, Kirton E, Allgaier M, Bergin C, Scheffrahn RH, Kyrpides NC, Warnecke F, Tringe SG, Hugenholtz P. 2013. Comparative metagenomic and metatranscriptomic analysis of hindgut paunch microbiota in wood- and dung-feeding higher termites. *PloS one* 8:e61126. DOI: 10.1371/journal.pone.0061126.

Herlemann DPR, Geissinger O, Ikeda-Ohtsubo W, Kunin V, Sun H, Lapidus A, Hugenholtz P, Brune A. 2009. Genomic analysis of "*Elusimicrobium minutum*" the first cultivated representative of the phylum "Elusimicrobia" (formerly termite group 1). *Applied and Environmental Microbiology* 75:2841–2849. DOI: 10.1128/AEM.02698-08.

Hervé V, Brune A. 2017. The complete mitochondrial genomes of the higher termites *Labiotermes labralis* and *Embiratermes neotenicus* (Termitidae: Syntermitinae). *Mitochondrial DNA Part B* 2:109–110. DOI: 10.1080/23802359.2017.1289349.

Hervé V, Le Roux X, Uroz S, Gelhaye E, Frey-Klett P. 2014. Diversity and structure of bacterial communities associated with *Phanerochaete chrysosporium* during wood decay. *Environmental Microbiology* 16:2238–2252. DOI: 10.1111/1462-2920.12347.

Hongoh Y, Deevong P, Hattori S, Inoue T, Noda S, Noparatnaraporn N, Kudo T, 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. *Applied and Environmental Microbiology* 72:6780–8. DOI: 10.1128/AEM.00891-06.

Hongoh Y, Deevong P, Inoue T, Moriya S, Trakulnaleamsai S, Ohkuma M, Vongkaluang C, Noparatnaraporn N, Kudo T. 2005. Intra- and interspecific comparisons of bacterial diversity and community structure support coevolution of gut microbiota and termite host. *Applied and environmental microbiology* 71:6590–9. DOI: 10.1128/AEM.71.11.6590-6599.2005.

Hongoh Y, Sharma VK, Prakash T, Noda S, Taylor TD, Kudo T, Sakaki Y, Toyoda A, Hattori M, Ohkuma M. 2008a. Complete genome of the uncultured Termite Group 1 bacteria in a single host protist cell. *Proceedings of the National Academy of Sciences* 105:5555–5560. DOI: 10.1073/pnas.0801389105.

Hongoh Y, Sharma VK, Prakash T, Noda S, Toh H, Taylor TD, Kudo T, Sakaki Y, Toyoda A, Hattori M, Ohkuma M. 2008b. Genome of an endosymbiont coupling N2 fixation to cellulolysis within protist cells in termite gut. *Science (New York, N.Y.)* 322:1108–9. DOI: 10.1126/science.1165578.

Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 7:e7359. DOI: 10.7717/peerj.7359.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* 30:772–80. DOI: 10.1093/molbev/mst010.

Köhler T, Dietrich C, Scheffrahn RH, 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.). *Applied and environmental microbiology* 78:4691–701. DOI: 10.1128/AEM.00683-12.

Köhler T, Stingl U, Meuser K, Brune A. 2008. Novel lineages of Planctomycetes densely colonize the alkaline gut of soil-feeding termites (*Cubitermes* spp.). *Environmental Microbiology* 10:1260–1270. DOI: 10.1111/j.1462-2920.2007.01540.x.

Krishna K, Grimaldi DA, Krishna V, Engel MS. 2013. Treatise on the Isoptera of the World. *Bulletin of the American Museum of Natural History* 377:2433–2705. DOI: 10.1206/377.7.

Kuhnigk T, Branke J, Krekeler D, Cypionka H, König H. 1996. A feasible role of sulfate-reducing bacteria in the termite gut. *Systematic and Applied Microbiology* 19:139–149. DOI: 10.1016/S0723-2020(96)80039-7.

783 Lazar CS, Baker BJ, Seitz K, Hyde AS, Dick GJ, Hinrichs K-U, Teske AP. 2016. Genomic
784 evidence for distinct carbon substrate preferences and ecological niches of Bathyarchaeota
785 in estuarine sediments. *Environmental Microbiology* 18:1200–1211. DOI: 10.1111/1462-
786 2920.13142.

787 Leadbetter JR, Schmidt TM, Graber JR, Breznak JA. 1999. Acetogenesis from H2 plus CO2 by
788 spirochetes from termite guts. *Science (New York, N.Y.)* 283:686–9.

789 Lefort V, Longueville J-E, Gascuel O. 2017. SMS: Smart Model Selection in PhyML. *Molecular*
790 *Biology and Evolution* 6:461–464. DOI: 10.1093/molbev/msx149.

791 Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new
792 developments. *Nucleic Acids Research*. DOI: 10.1093/nar/gkz239.

793 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform.
794 *Bioinformatics (Oxford, England)* 25:1754–60. DOI: 10.1093/bioinformatics/btp324.

795 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R.
796 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.
797 DOI: 10.1093/bioinformatics/btp352.

798 Lilburn, Schmidt, Breznak. 1999. Phylogenetic diversity of termite gut spirochaetes.
799 *Environmental Microbiology* 1:331–345. DOI: 10.1046/j.1462-2920.1999.00043.x.

800 Liu G, Cornwell WK, Cao K, Hu Y, Van Logtestijn RSP, Yang S, Xie X, Zhang Y, Ye D, Pan
801 X, Ye X, Huang Z, Dong M, Cornelissen JHC. 2015. Termites amplify the effects of wood
802 traits on decomposition rates among multiple bamboo and dicot woody species. *Journal of*
803 *Ecology* 103:1214–1223. DOI: 10.1111/1365-2745.12427.

804 Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-
805 active enzymes database (CAZy) in 2013. *Nucleic Acids Research* 42:D490–D495. DOI:
806 10.1093/nar/gkt1178.

807 Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Pillay M, Ratner A, Huang J,
808 Woyke T, Huntemann M, Anderson I, Billis K, Varghese N, Mavromatis K, Pati A, Ivanova
809 NN, Kyrpides NC. 2014. IMG 4 version of the integrated microbial genomes comparative
810 analysis system. *Nucleic Acids Research* 42:D560–D567. DOI: 10.1093/nar/gkt963.

811 Marynowska M, Goux X, Sillam-Dussès D, Rouland-Lefèvre C, Roisin Y, Delfosse P,
812 Calusinska M. 2017. Optimization of a metatranscriptomic approach to study the
813 lignocellulolytic potential of the higher termite gut microbiome. *BMC Genomics* 18:681.
814 DOI: 10.1186/s12864-017-4076-9.

815 Mikaelyan A, Dietrich C, Köhler T, Poulsen M, Sillam-Dussès D, Brune A. 2015a. Diet is the
816 primary determinant of bacterial community structure in the guts of higher termites.

*Molecular Ecology* 24:5284–5295. DOI: 10.1111/mec.13376.

Mikaelyan A, Köhler T, Lampert N, Rohland J, Boga H, Meuser K, Brune A. 2015b. Classifying the bacterial gut microbiota of termites and cockroaches: A curated phylogenetic reference database (DictDb). *Systematic and Applied Microbiology* 38:472–482. DOI: 10.1016/j.syapm.2015.07.004.

Mikaelyan A, Meuser K, Brune A. 2017. Microenvironmental heterogeneity of gut compartments drives bacterial community structure in wood- and humus-feeding higher termites. *FEMS Microbiology Ecology* 93:fiw210. DOI: 10.1093/femsec/fiw210.

Mikaelyan A, Strassert JFH, Tokuda G, Brune A. 2014. The fibre-associated cellulolytic bacterial community in the hindgut of wood-feeding higher termites (*Nasutitermes* spp.). *Environmental Microbiology* 16:2711–2722. DOI: 10.1111/1462-2920.12425.

Moreira EA, Alvarez TM, Persinoti GF, Paixão DAA, Menezes LR, Cairo JPF, Squina FM, Costa-Leonardo AM, Carrijo T, Arab A. 2018. Microbial communities of the gut and nest of the humus- and litter-feeding termite *Procornitermes araujoi* (Syntermitinae). *Current Microbiology*:1–10. DOI: 10.1007/s00284-018-1567-0.

Neumann AP, McCormick CA, Suen G. 2017. Fibrobacter communities in the gastrointestinal tracts of diverse hindgut-fermenting herbivores are distinct from those of the rumen. *Environmental Microbiology* 19:3768–3783. DOI: 10.1111/1462-2920.13878.

Neumann AP, Suen G. 2018. The phylogenomic diversity of herbivore-associated *Fibrobacter* spp. is correlated to lignocellulose-degrading potential. *mSphere* 3:e00593-18. DOI: 10.1128/mSphere.00593-18.

Odelson DA, Breznak JA. 1983. Volatile fatty acid production by the hindgut microbiota of xylophagous termites. *Applied and Environmental Microbiology* 45:1602–13.

Ohkuma M, Iida T, Kudo T. 1999. Phylogenetic relationships of symbiotic spirochetes in the gut of diverse termites. *FEMS Microbiology Letters* 181:123–129.

Ohkuma M, Noda S, Hattori S, Iida T, Yuki M, Starns D, Inoue J, Darby AC, Hongoh Y. 2015. Acetogenesis from H2 plus CO2 and nitrogen fixation by an endosymbiotic spirochete of a termite-gut cellulolytic protist. *Proceedings of the National Academy of Sciences* 112:10224–10230. DOI: 10.1073/pnas.1423979112.

Ohkuma M, Noda S, Kudo T. 1999. Phylogenetic diversity of nitrogen fixation genes in the symbiotic microbial community in the gut of diverse termites. *Applied and environmental microbiology* 65:4926–34.

Ottesen EA, Leadbetter JR. 2011. Formyltetrahydrofolate synthetase gene diversity in the guts of higher termites with different diets and lifestyles. *Applied and Environmental Microbiology*

77:3461–3467. DOI: 10.1128/AEM.02657-10.

Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nature Biotechnology*. DOI: 10.1038/nbt.4229.

Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome research* 25:1043–55. DOI: 10.1101/gr.186072.114.

Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW. 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature Microbiology* 2:1533–1542. DOI: 10.1038/s41564-017-0012-7.

Paster BJ, Dewhirst FE, Cooke SM, Fussing V, Poulsen LK, Breznak JA. 1996. Phylogeny of not-yet-cultured spirochetes from termite guts. *Applied and environmental microbiology* 62:347–52.

Prosser JI. 2015. Dispersing misconceptions and identifying opportunities for the use of "omics" in soil microbial ecology. *Nature Reviews Microbiology* 13:439–446. DOI: 10.1038/nrmicro3468.

Qiu YL, Kuang XZ, Shi XS, Yuan XZ, Guo RB. 2014. *Paludibacter jiangxiensis* sp. nov., a strictly anaerobic, propionate-producing bacterium isolated from rice paddy field. *Archives of Microbiology* 196:149–155. DOI: 10.1007/s00203-013-0951-1.

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research* 41:D590-6. DOI: 10.1093/nar/gks1219.

R Development Core Team. 2015. R: A Language and Environment for Statistical Computing.

Le Roes-Hill M, Rohland J, Burton S. 2011. Actinobacteria isolated from termite guts as a source of novel oxidative enzymes. *Antonie van Leeuwenhoek* 100:589–605. DOI: 10.1007/s10482-011-9614-x.

Rossmassler K, Dietrich C, Thompson C, Mikaelyan A, Nonoh JO, Scheffrahn RH, Sillam-Dussès D, Brune A. 2015. Metagenomic analysis of the microbiota in the highly compartmented hindguts of six wood- or soil-feeding higher termites. *Microbiome* 3:56. DOI: 10.1186/s40168-015-0118-1.

Sabree ZL, Moran NA. 2014. Host-specific assemblages typify gut microbial communities of related insect species. *SpringerPlus* 3:138. DOI: 10.1186/2193-1801-3-138.

Schloss PD, Girard RA, Martin T, Edwards J, Thrash JC. 2016. Status of the Archaeal and

PeerJ Preprints NOT PEER-REVIEWED

Bacterial census: An update. *mBio* 7:e00201-16. DOI: 10.1128/mBio.00201-16.

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, Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology* 75:7537–7541. DOI: 10.1128/AEM.01541-09.

Schmidt O, Horn MA, Kolb S, Drake HL. 2015. Temperature impacts differentially on the methanogenic food web of cellulose-supplemented peatland soil. *Environmental Microbiology* 17:720–734. DOI: 10.1111/1462-2920.12507.

Schulz F, Eloe-Fadrosh EA, Bowers RM, Jarett J, Nielsen T, Ivanova NN, Kyrpides NC, Woyke T. 2017. Towards a balanced view of the bacterial tree of life. *Microbiome* 5:140. DOI: 10.1186/s40168-017-0360-9.

Sczyrba A, Hofmann P, Belmann P, Koslicki D, Janssen S, Dröge J, Gregor I, Majda S, Fiedler J, Dahms E, Bremges A, Fritz A, Garrido-Oter R, Jørgensen TS, Shapiro N, Blood PD, Gurevich A, Bai Y, Turaev D, DeMaere MZ, Chikhi R, Nagarajan N, Quince C, Meyer F, Balvočiūtė M, Hansen LH, Sørensen SJ, Chia BKH, Denis B, Froula JL, Wang Z, Egan R, Don Kang D, Cook JJ, Deltel C, Beckstette M, Lemaitre C, Peterlongo P, Rizk G, Lavenier D, Wu Y-W, Singer SW, Jain C, Strous M, Klingenberg H, Meinicke P, Barton MD, Lingner T, Lin H-H, Liao Y-C, Silva GGZ, Cuevas DA, Edwards RA, Saha S, Piro VC, Renard BY, Pop M, Klenk H-P, Göker M, Kyrpides NC, Woyke T, Vorholt JA, Schulze-Lefert P, Rubin EM, Darling AE, Rattei T, McHardy AC. 2017. Critical Assessment of Metagenome Interpretation—a benchmark of metagenomics software. *Nature Methods* 14:1063–1071. DOI: 10.1038/nmeth.4458.

Shi Y, Huang Z, Han S, Fan S, Yang H. 2015. Phylogenetic diversity of Archaea in the intestinal tract of termites from different lineages. *Journal of Basic Microbiology* 55:1021–1028. DOI: 10.1002/jobm.201400678.

Söllinger A, Tveit AT, Poulsen M, Noel SJ, Bengtsson M, Bernhardt J, Frydendahl Hellwing AL, Lund P, Riedel K, Schleper C, Højberg O, Urich T. 2018. Holistic assessment of rumen microbiome dynamics through quantitative metatranscriptomics reveals multifunctional redundancy during key steps of anaerobic feed degradation. *mSystems* 3:e00038-18. DOI: 10.1128/mSystems.00038-18.

Spring S, Bunk B, Spröer C, Schumann P, Rohde M, Tindall BJ, Klenk H-P. 2016. Characterization of the first cultured representative of Verrucomicrobia subdivision 5 indicates the proposal of a novel phylum. *The ISME Journal* 10:2801–2816. DOI: 10.1038/ismej.2016.84.

Stewart RD, Auffret MD, Warr A, Wiser AH, Press MO, Langford KW, Liachko I, Snelling TJ, Dewhurst RJ, Walker AW, Roehe R, Watson M. 2018. Assembly of 913 microbial genomes from metagenomic sequencing of the cow rumen. *Nature Communications* 9:870. DOI: 10.1038/s41467-018-03317-6.

Sujada N, Sungthong R, Lumyong S. 2014. Termite nests as an abundant source of cultivable Actinobacteria for biotechnological purposes. *Microbes and Environments* 29:211–219. DOI: 10.1264/jsme2.ME13183.

Svartström O, Alneberg J, Terrapon N, Lombard V, de Bruijn I, Malmsten J, Dalin A-M, EL Muller E, Shah P, Wilmes P, Henrissat B, Aspeborg H, Andersson AF. 2017. Ninety-nine de novo assembled genomes from the moose (Alces alces) rumen microbiome provide new insights into microbial plant biomass degradation. *The ISME Journal* 11:2538–2551. DOI: 10.1038/ismej.2017.108.

Tokuda G, Lo N, Watanabe H, Arakawa G, Matsumoto T, Noda H. 2004. Major alteration of the expression site of endogenous cellulases in members of an apical termite lineage. *Molecular Ecology* 13:3219–3228. DOI: 10.1111/j.1365-294X.2004.02276.x.

Tokuda G, Mikaelyan A, Fukui C, Matsuura Y, Watanabe H, Fujishima M, Brune A. 2018. Fiber-associated spirochetes are major agents of hemicellulose degradation in the hindgut of wood-feeding higher termites. *Proceedings of the National Academy of Sciences* 115:E11996–E12004. DOI: 10.1073/pnas.1810550115.

Tokuda G, Tsuboi Y, Kihara K, Saitou S, Moriya S, Lo N, Kikuchi J. 2014. Metabolomic profiling of 13C-labelled cellulose digestion in a lower termite: insights into gut symbiont function. *Proceedings. Biological sciences / The Royal Society* 281:20140990. DOI: 10.1098/rspb.2014.0990.

Ueki A, Akasaka H, Suzuki D, Ueki K. 2006. *Paludibacter propionicigenes* gen. nov., sp. nov., a novel strictly anaerobic, Gram-negative, propionate-producing bacterium isolated from plant residue in irrigated rice-field soil in Japan. *International Journal of Systematic and Evolutionary Microbiology* 56:39–44. DOI: 10.1099/ijs.0.63896-0.

Utami YD, Kuwahara H, Igai K, Murakami T, Sugaya K, Morikawa T, Nagura Y, Yuki M, Deevong P, Inoue T, Kihara K, Lo N, Yamada A, Ohkuma M, Hongoh Y. 2019. Genome analyses of uncultured TG2/ZB3 bacteria in '*Margulisbacteria*' specifically attached to ectosymbiotic spirochetes of protists in the termite gut. *The ISME Journal* 13:455–467. DOI: 10.1038/s41396-018-0297-4.

Utami YD, Kuwahara H, Murakami T, Morikawa T, Sugaya K, Kihara K, Yuki M, Lo N, Deevong P, Hasin S, Boonriam W, Inoue T, Yamada A, Ohkuma M, Hongoh Y. 2018. Phylogenetic diversity and single-cell genome analysis of "*Melainabacteria*", a non-

PeerJ Preprints NOT PEER-REVIEWED

photosynthetic cyanobacterial group, in the termite gut. *Microbes and Environments* 33:50–57. DOI: 10.1264/jsme2.ME17137.

Wang Y, Su L, Huang S, Bo C, Yang S, Li Y, Wang F, Xie H, Xu J, Song A. 2016. Diversity and resilience of the wood-feeding higher termite *Mironasutitermes shangchengensis* gut microbiota in response to temporal and diet variations. *Ecology and Evolution* 6:8235–8242. DOI: 10.1002/ece3.2497.

Wertz JT, Kim E, Breznak JA, Schmidt TM, Rodrigues JLM. 2012. Genomic and physiological characterization of the Verrucomicrobia isolate *Diplosphaera colitermitum* gen. nov., sp. nov., reveals microaerophily and nitrogen fixation genes. *Applied and Environmental Microbiology* 78:1544–1555. DOI: 10.1128/AEM.06466-11.

Wickham H. 2016. *ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. DOI: 10.1007/978-3-319-24277-4.

Woyke T, Doud DFR, Schulz F. 2017. The trajectory of microbial single-cell sequencing. *Nature Methods* 14:1045–1054. DOI: 10.1038/nmeth.4469.

Yamada A, Inoue T, Wiwatwitaya D, Ohkuma M, Kudo T, Abe T, Sugimoto A. 2005. Carbon mineralization by termites in tropical forests, with emphasis on fungus combs. *Ecological Research* 20:453–460.

Yuki M, Kuwahara H, Shintani M, Izawa K, Sato T, Starns D, Hongoh Y, Ohkuma M. 2015. Dominant ectosymbiotic bacteria of cellulolytic protists in the termite gut also have the potential to digest lignocellulose. *Environmental Microbiology* 17:4942–4953. DOI: 10.1111/1462-2920.12945.

Yuki M, Sakamoto M, Nishimura Y, Ohkuma M. 2018. *Lactococcus reticulitermitis* sp. nov., isolated from the gut of the subterranean termite *Reticulitermes speratus*. *International Journal of Systematic and Evolutionary Microbiology* 68:596–601. DOI: 10.1099/ijsem.0.002549.

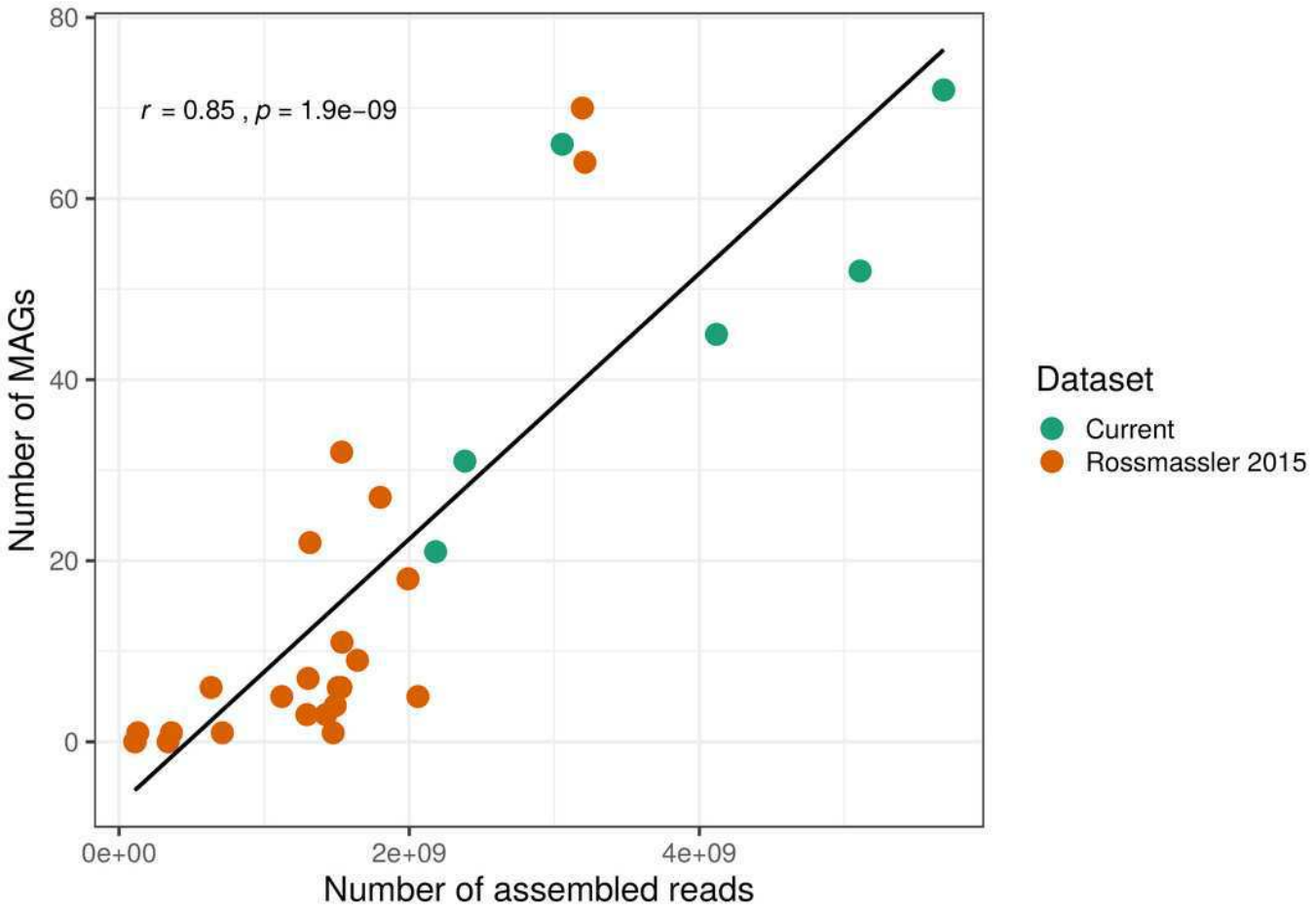
Zheng H, Brune A. 2015. Complete genome sequence of *Endomicrobium proavitum*, a free-living relative of the intracellular symbionts of termite gut flagellates (phylum Elusimicrobia). *Genome Announcements* 3:e00679-15. DOI: 10.1128/genomeA.00679-15.

Zhou Z, Pan J, Wang F, Gu J-D, Li M. 2018. Bathyarchaeota: globally distributed metabolic generalists in anoxic environments. *FEMS Microbiology Reviews*. DOI: 10.1093/femsre/fuy023.

# Figure 1

Relationship between the number of MAGs recovered and the number of assembled reads in the respective metagenomes.

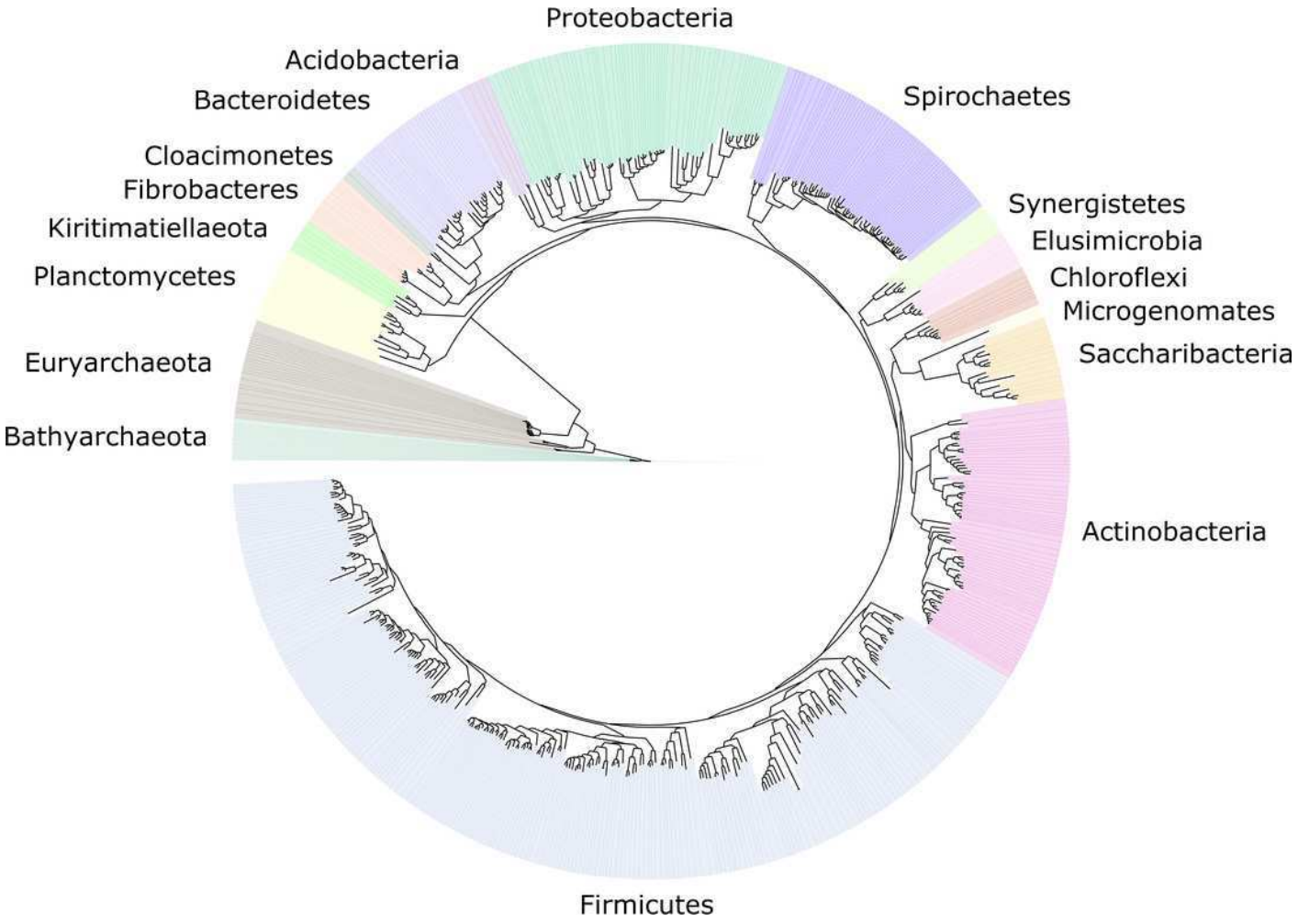
The linear regression line and the Pearson correlation coefficient (r) are shown for the entire dataset.



# Figure 2

Distribution of the 589 MAGs among bacterial and archaeal phyla.

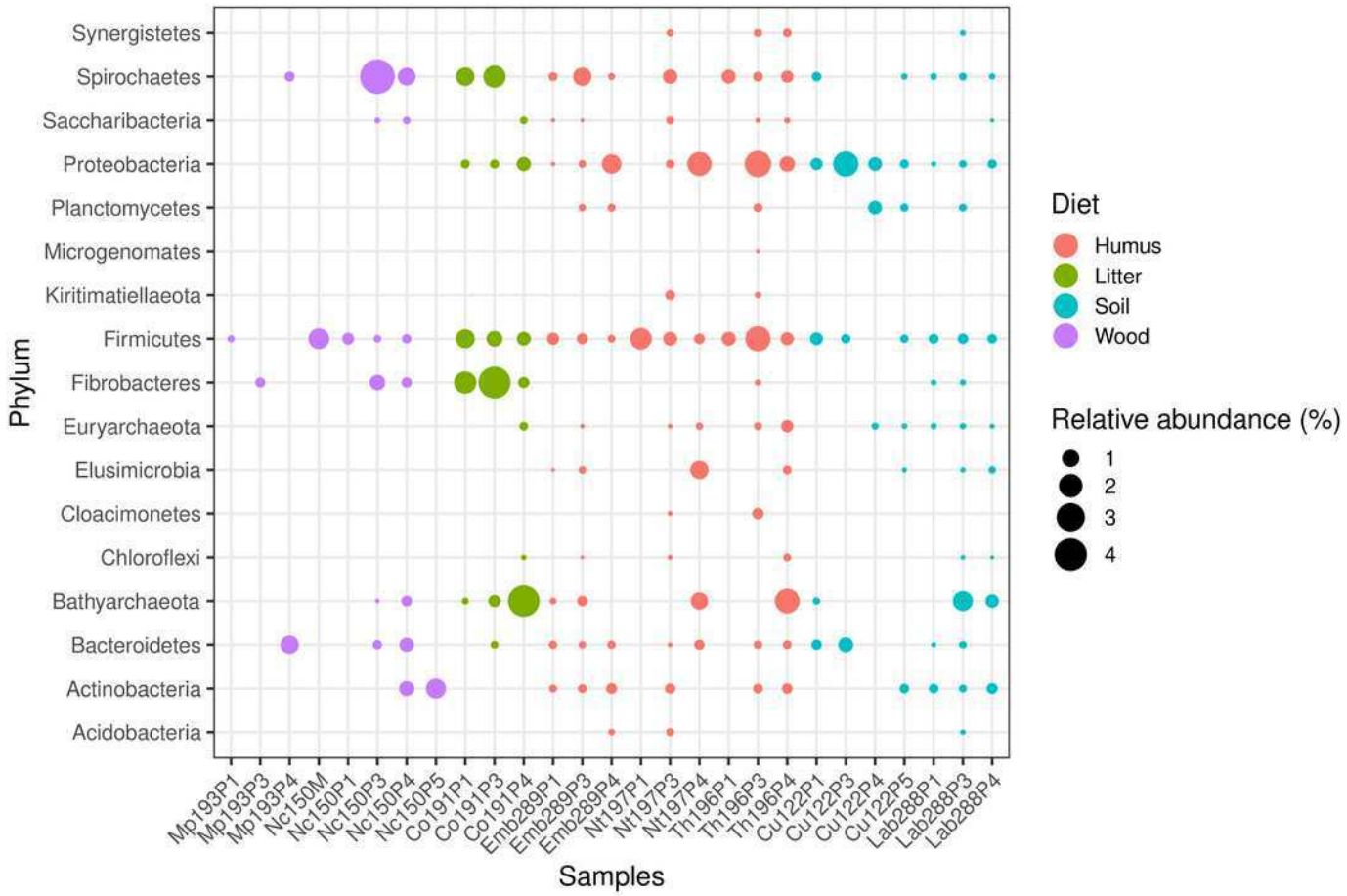
This maximum-likelihood tree was inferred from a concatenated alignment (amino acids) of 43 protein-coding genes using the LG+G+I model of evolution.



# Figure 3

Relative abundance of the MAGs from different phyla among the respective metagenomes.

Circle size indicates the relative abundance of the MAGs among the respective metagenome sample; color indicates host diet.

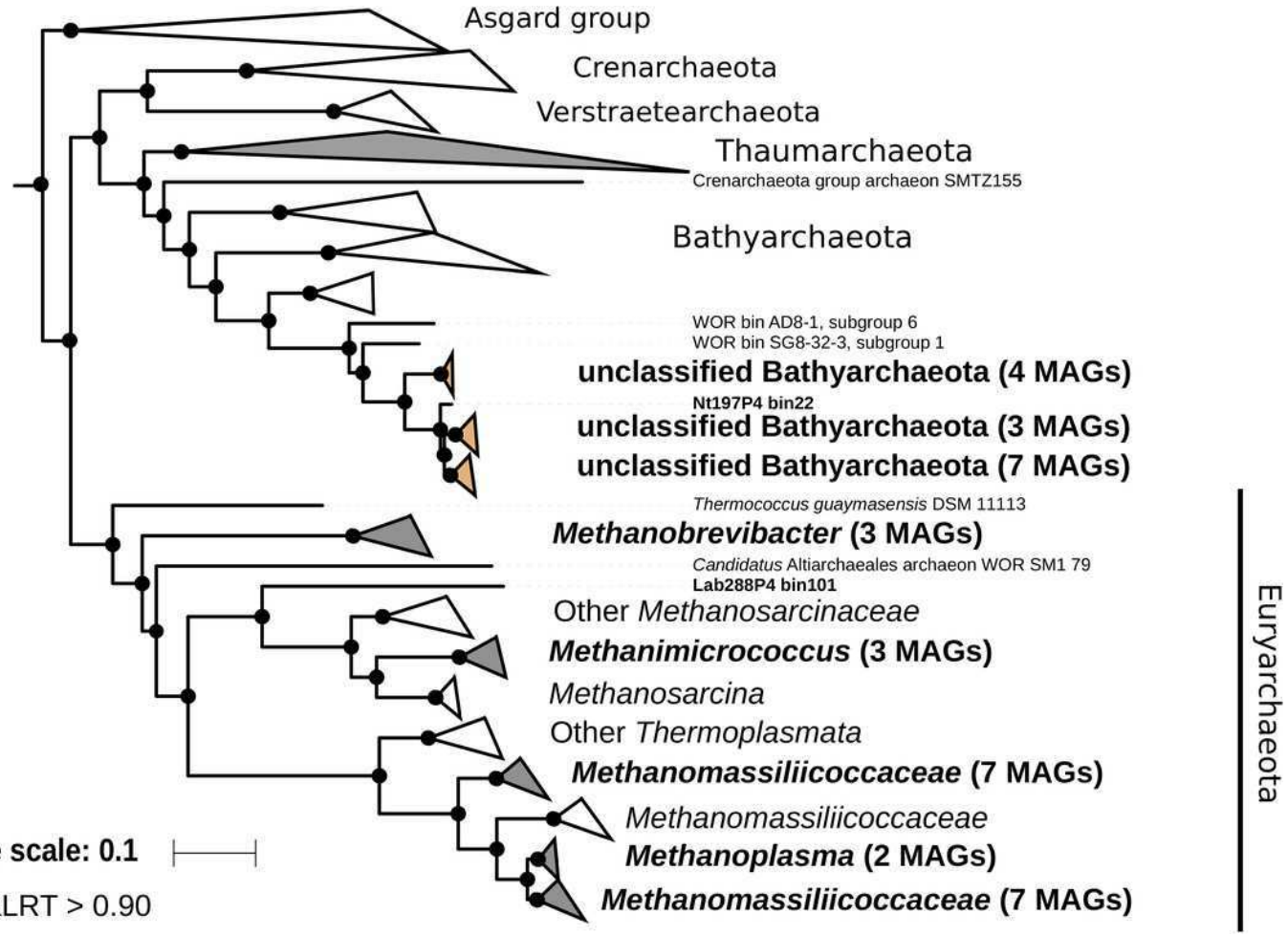


PeerJ Preprints NOT PEER-REVIEWED

# Figure 4

Phylogenomic tree of the archaeal domain.

This maximum-likelihood tree was inferred from a concatenated alignment of 43 proteins using the LG+G+I+F model of amino-acid evolution. Branch supports were calculated using a Chi2-based parametric approximate likelihood-ratio test. Names in bold included MAGs recovered in the present study. Clusters shaded in brown consist exclusively of MAGs from termite guts and clusters shaded in gray contain genomes from termite guts. The Asgard group was used as outgroup.



PeerJ Preprints NOT PEER-REVIEWED

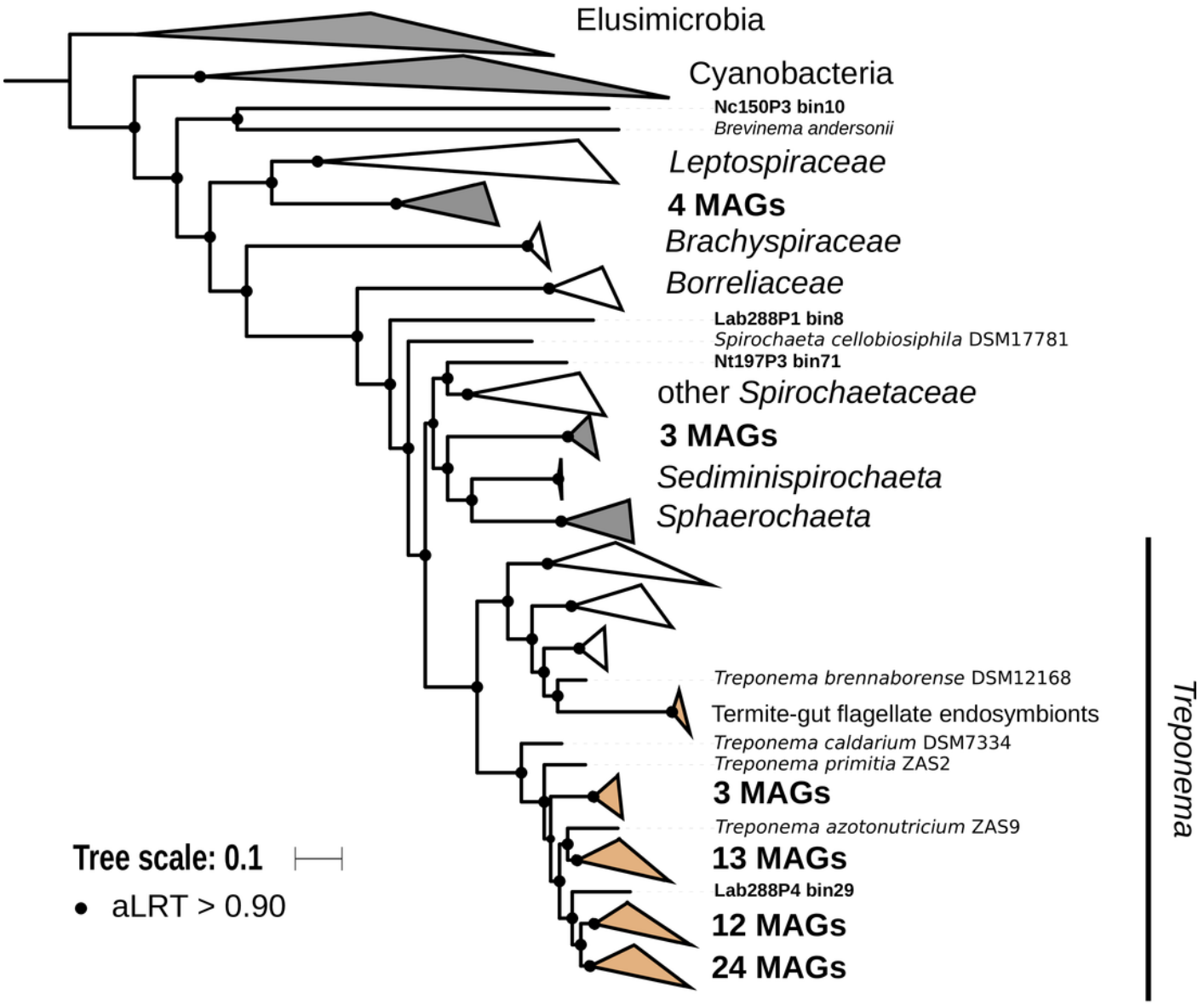
# Figure 5

Phylogenomic tree of the Spirochaetes phylum.

This maximum-likelihood tree was inferred from a concatenated alignment of 43 proteins using the LG+G+I+F model of amino-acid evolution. Branch supports were calculated using a Chi2-based parametric approximate likelihood-ratio test. Names in bold included MAGs recovered in the present study. Clusters shaded in brown consist exclusively of genomes from termite guts and clusters shaded in gray contain genomes from termite guts. Elusimicrobia and Cyanobacteria were used as outgroup.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.27929v1 | CC BY 4.0 Open Access | rec: 30 Aug 2019, publ: 30 Aug 2019

PeerJ Preprints NOT PEER-REVIEWED

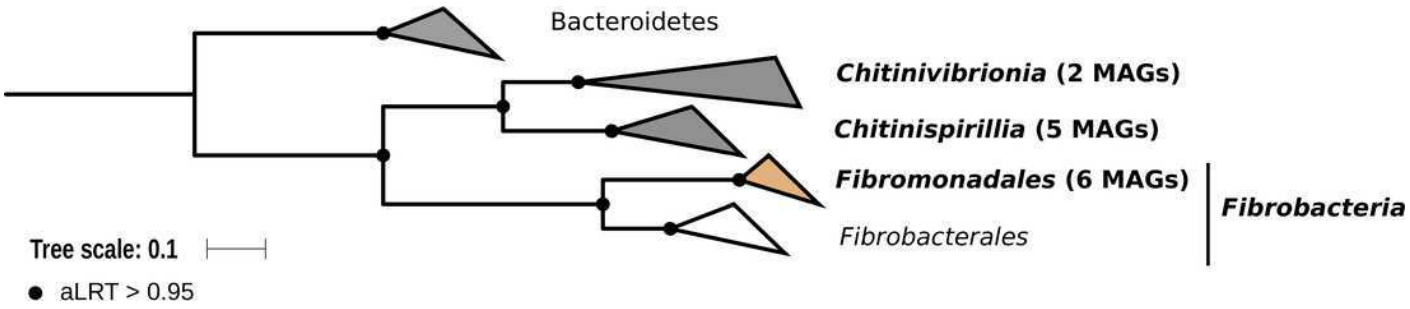


PeerJ Preprints NOT PEER-REVIEWED

# Figure 6

Phylogenomic tree of the Fibrobacteres phylum.

This maximum-likelihood tree was inferred from a concatenated alignment of 43 proteins using the LG+G+I+F model of amino-acid evolution. Branch supports were calculated using a Chi2-based parametric approximate likelihood-ratio test. Names in bold included MAGs recovered in the present study. Clusters shaded in brown consist exclusively of genomes from termite guts and clusters shaded in gray contain genomes from termite guts. Bacteroidetes were used as outgroup.



PeerJ Preprints NOT PEER-REVIEWED

## Table 1(on next page)

Recovery of metagenome-assembled genomes (MAGs) from the 30 termite gut metagenomes analyzed in this study.

The host termite, its mitochondrial genome accession number, dietary preference, and the originating gut compartments are indicated. *C* crop (foregut), *M* midgut, *P1–P5* proctodeal compartments (hindgut). The sample codes used for the metagenomes are the combination of host ID and gut compartment.

PeerJ Preprints NOT PEER-REVIEWED

**Table 1. Recovery of metagenome-assembled genomes (MAGs) from the 30 termite gut metagenomes analyzed in this study.** The host termite, its mitochondrial genome accession number, dietary preference, and the originating gut compartments are indicated. *C* crop (foregut), *M* midgut, *P1–P5* proctodeal compartments (hindgut). The sample codes used for the metagenomes are the combination of host ID and gut compartment.

| Termite species | ID | Mitogenome | Diet | Number of MAGs | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | C | M | P1 | P3 | P4 | P5 | Total |
| *Microcerotermes parvus* | Mp193 | KP091690 | Wood | – [a] | – | 1 | 1 | 4 | – | 6 |
| *Nasutitermes corniger* | Nc150 | KP091691 | Wood | 0 | 1 | 3 | 6 | 9 | 1 | 20 |
| *Cornitermes* sp. | Co191 | KP091688 | Litter | – | – | 32 | 22 | 7 | – | 61 |
| *Neocapritermes taracua* | Nt197 | KP091692 | Humus | – | – | 6 | 70 | 11 | – | 87 |
| *Termes hospes* | Th196 | KP091693 | Humus | – | – | 6 | 64 | 27 | – | 97 |
| *Embiratermes neotenicus* | Emb289 | KY436202 | Humus | – | – | 45 | 52 | 21 | – | 118 |
| *Labiotermes labralis* | Lab288 | KY436201 | Soil | – | – | 66 | 72 | 31 | – | 169 |
| *Cubitermes ugandensis* | Cu122 | KP091689 | Soil | 0 | 0 | 5 | 5 | 3 | 18 | 31 |

[a] Not sequenced.