Ecological specificity of the metagenome in a set of lower termite species supports contribution of the microbiome to adaptation of the host

Lena Waidele, Judith Korb, Christian R Voolstra, Franck Dedeine, and Fabian Staubach

Author affiliations
For L. Waidele, J. Korb, and F. Staubach see corresponding author address

Christian R Voolstra
Red Sea Research Center, Division of Biological and Environmental Science and Engineering (BESE), King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia
christian.voolstra@kaust.edu.sa

Franck Dedeine
Institut de Recherche sur la Biologie de l'Insecte, UMR 7261, CNRS - Université François-Rabelais de Tours, Tours, France
franck.dedeine@univ-tours.fr

Corresponding author
Dr. Fabian Staubach
Biologie I
Universitaet Freiburg
Hauptstr. 1
79104 Freiburg
Germany
phone: +49-761-203-2911
fax: +49-761-203-2644
email: fabian.staubach@biologie.uni-freiburg.de

Running title
lower termite metagenomes

Originality-Significance Statement

This is the first study that applies metagenomic shotgun sequencing in a multispecies framework of lower termites to gain insight into the contribution of eukaryote and bacterial microbes to host adaptation. Our study design employs a selection of host species that span different ecologies, reared under identical conditions, to tease apart the relative contribution of acclimation vs. adaptive differences. Functional changes of the microbiome that aligned with ecology were highly specific. We found the most prominent changes in the bacterial metabolic metagenome, suggesting a role of the microbiome in dietary adaptation.

Summary

Elucidating the interplay between hosts and their microbiomes in ecological adaptation has become a central theme in evolutionary biology. The microbiome mediated adaptation of lower termites to a wood-based diet. Lower termites have further adapted to different ecologies. Substantial ecological differences are linked to different termite life types. Termites of the wood-dwelling life type never leave their nests and feed on a uniform diet. Termites of the foraging life type forage for food outside the nest, access a more diverse diet, and grow faster. Here we reasoned that the microbiome that is directly involved in food substrate breakdown and nutrient acquisition might contribute to adaptation to these ecological differences. This should leave ecological imprints on the microbiome. To search for such imprints, we applied metagenomic shotgun sequencing in a multispecies framework covering both life types. The microbiome of foraging species was enriched with genes for starch digestion, while the metagenome of wood-dwelling species was enriched with genes for hemicellulose utilization. Furthermore, increased abundance of genes for nitrogen sequestration, a limiting factor for termite growth, aligned with faster growth.

These findings are consistent with the notion that a specific subset of functions encoded by the microbiome contributes to host adaptation.

Introduction

The importance of microbes for the evolution of higher organisms is starting to be realized (McFall-Ngai *et al*., 2013; Bang *et al*., 2018). Metazoan evolution is not only driven by pathogenic microbes, as reflected by fast evolution of immune genes (Clark *et al*., 2007). Rather, microbes often are facilitators of metabolic and environmental adaptations (Jaenike *et al*., 2010; Douglas, 2015; Bang *et al*., 2018).

The gut microbial communities of wood-feeding roaches and termites facilitated thriving on a wood diet that is difficult to digest and poor in nitrogen. Nitrogen fixation and the digestion of wood depend on the termite gut microbiome (Brune, 2014; Brune and Dietrich, 2015; Bang *et al*., 2018). In lower termites, lignocellulose degradation was initially mainly attributed to unicellular eukaryotes (protists) in the gut (Cleveland, 1923). Recently, it has become evident that lignocellulose degradation is a synergistic effort of the termite, its associated protists, and bacteria (Scharf *et al*., 2011; Peterson *et al*., 2015; Peterson and Scharf, 2016). Bacteria are also considered the major agents in reductive acetogenesis (Leadbetter *et al*., 1999). In this process bacteria use hydrogen that is an abundant by-product of cellulose degradation (Pester and Brune, 2007) as an energy source, resulting in the production of acetic acid. Acetic acid, in return, serves the termite host as an energy source (Brune, 2014). In addition to their role in acetic acid production, bacteria are also essential for the fixation of atmospheric nitrogen (Lilburn *et al*., 2001; Desai and Brune, 2012) and the recycling of nitrogen from urea or uric acid (Potrikus and Breznak, 1981; Hongoh *et al*., 2008b). By using genome sequencing and pathway reconstruction, these processes have been assigned to four major bacterial phyla in the

termite gut: Proteobacteria (*Desulfovibrio* (Kuwahara *et al*., 2017)), Spirochetes (*Treponema* (Warnecke *et al*., 2007; Ohkuma *et al*., 2015)), Bacteroidetes (*Azobacteroides*) (Hongoh *et al*., 2008b) , and Elusimicrobia (*Endomicrobium* (Hongoh *et al*., 2008a; Zheng *et al*., 2016)).

Many bacteria in the termite gut live in tight association with protists, where they sit on the surface (Hongoh *et al*., 2007; Yuki *et al*., 2015), in invaginations of the cell membrane (Kuwahara *et al*., 2017), or even inside the protist cells (Brune, 2012). Such tight associations lead to frequent vertical transmission of bacteria between protist generations. In return, protists and bacteria are vertically transmitted between termite generations via proctodeal trophallaxis during colony foundation (Shimada *et al*., 2013). Vertical transmission has led to co-speciation between bacteria and their protist hosts, and sometimes even the termite hosts (Noda *et al*., 2007; Ohkuma, 2008; Ikeda-Ohtsubo and Brune, 2009; Desai *et al*., 2010). Evidence for horizontal transfer of protists between termite species, so called transfaunations, is limited to a few exceptions (Radek *et al*., 2018). Hence, the termite host species association is rather strict, leading to strong phylogenetic imprints on protist community structure (Yamin, 1979; Kitade, 2004; Waidele *et al*., 2017). In comparison, the bacterial microbiome is more flexible, frequently transferred between termite host species (Bourguignon *et al*., 2018), and affected by diet (Boucias *et al*., 2013; He *et al*., 2013; Dietrich *et al*., 2014; Mikaelyan *et al*., 2015; Rahman *et al*., 2015; Rossmassler *et al*., 2015; Tai *et al*., 2015; Waidele *et al*., 2017).

There is evidence that the gut microbiome of termites has contributed to the adaptation of different termite species to their specific ecologies (Zhang and Leadbetter, 2012; He *et al*., 2013; Waidele *et al*., 2017; Duarte *et al.* 2018, Liu *et al*., 2018). There are pronounced ecological differences between the so-called termite life types (Abe, 1987; Korb, 2007). Termite species of the wood-dwelling life type never leave their nest except

for the mating flight. They feed on a uniform bonanza resource, that is the piece of wood they built their nest in, and grow rather slowly (Lenz, 1994; Korb, Hoffmann, *et al*., 2012). On the other hand, foraging species leave their nest to forage for food, can access a more diverse diet, and grow faster (Lenz, 1994; Lainé *et al*., 2003). Different growth rates and diets likely impose different selection pressures on the termite holobiont, in particular with regard to nutrient uptake. Because the microbiome is directly involved in nutrient uptake, it seems reasonable to hypothesize that it may also play a role in adaptation to life type related ecological differences. In this scenario, one would expect the life types to leave an imprint on microbiome structure and function. As such, searching for microbial imprints of a given life type can possibly provide us with a lead for microbiome-mediated adaptation.

One potential pitfall of such an endeavor is that microbiomes may bear imprints from transient microbes that were ingested from the environment. Transient microbes rarely form evolutionary relevant relationships with the host (Douglas, 2011; Ma and Leulier, 2018). Instead, they reflect the microbiome of the local environment the termites were collected from. When the local environment correlates with other ecological differences these imprints could be falsely interpreted as potentially adaptive changes of the microbiome. Therefore, it is essential to reduce the impact of these microbes on the analysis. In order to explore potential ecological imprints on the microbiome, we focused on an evolutionary switch between wood-dwelling and foraging life types in the Rhinotermitidae (Figure 1). *Reticulitermes* species are of the foraging life type, while *Prorhinotermes simplex* is wood-dwelling. If the microbiome was affected by life type specific ecology, we would expect that the microbiome of *Prorhinotermes simplex* was similar to that of the other wood-dwelling species (*Cryptotermes*) although these are from a different family (Kalotermitidae). At the same time, the microbiome of the foraging *Reticulitermes* species should bear distinct features. Alternatively, if there was no

ecological imprint, we would expect the microbiome to follow a phylogenetic pattern, with the Rhinotermitidae *Prorhinotermes* and *Reticulitermes* forming a cluster and the *Cryptotermes* species (Kalotermitidae) forming a second cluster. Using this experimental setup, we recently showed that protist community composition aligned with phylogeny, but bacterial communities aligned more strongly with wood-dwelling and foraging life types (Waidele *et al*., 2017).

To further explore this, we investigated whether changes in microbiome composition are also reflected by changes in microbiome function, as would be expected if the microbiome played a role in adaptation. For instance, we would expect dietary adaptations to be reflected by changes to pathways involved in substrate breakdown and that effective provisioning of limiting nutrients such as nitrogen is more important for faster growing foraging *Reticulitermes* species. In order to test whether and which changes in the functional repertoire align with life type and could be involved in potential adaptation to different ecologies, we characterized the metagenome of two foraging species; *Reticulitermes flavipes* and *Reticulitermes grassei.* We compared their functional repertoire to that of three wood-dwelling species *Prorhinotermes simplex*, *Cryptotermes secundus*, and *Cryptotermes domesticus.* Because there can be substantial variation in microbial communities between colonies (Hongoh *et al*., 2005, 2006; Benjamino and Graf, 2016), we analyzed five *C. domesticus*, eight *C. secundus*, seven *P. simplex*, five *R. flavipes*, and four *R. grassei* replicate colonies. We focused on the persistent microbiome by controlling the influx of transient microbes through feeding on a common diet of sterile *Pinus* wood for several weeks prior to sample collection.

Results

We analyzed a total of ~440 million metagenomic shotgun sequences. Between 974,176 and 8,949,734 sequences per sample were of microbial origin (Table S1). Sequences were subsampled (rarefied) to 1,386,882 bacterial and 2,781 protist annotated sequences per sample. For annotation, sequences were aligned to a reference database of clusters of orthologous groups of genes (COGs) with known function. These COGs represent the lowest level of the eggNOG hierarchical annotation. At the next higher level, the COGs are grouped into pathways (Figure S1, Figure 2), and at the third and highest level, the pathways are grouped into three categories; “information storage and processing”, “cellular process and signaling”, and “metabolism”. We adhere to this definition of the eggNOG hierarchical terms throughout the study.

Genes with signaling functions are more common in the protist metagenome and metabolic genes are more common in the bacterial metagenome

Functional annotation assigned 64.3% of the protist metagenome to “cellular process and signaling” (Figure 2A). In this functional category, the largest fraction of sequences (22.8% of the total) were assigned to signal transduction pathways. The proportion of genes in the protist metagenome in this category was significantly higher than in bacteria (4.2% of the total in bacteria, Wilcoxon rank sum test: $p = 2.1e-15$). In the signal transduction pathways, two of the three most abundant COGs (COG1131, COG1132) in protists were ABC transporters, suggesting that the ability for transmembrane transport differs between protist and bacterial metagenomes.

In the bacterial metagenomes, sequences encoding proteins with metabolic function were more common than in protists (49.7% of the total in bacteria, Wilcoxon rank sum test: $p = 2.1e-15$, Figure 2B). In this group, amino acid (12.6%) and carbohydrate metabolism

(10.0%) pathways were the most abundant. The latter included many genes involved in lignocellulose degradation, and the former, genes involved in nitrogen fixation and amino acid synthesis. Both are central functions of the termite gut microbiome (Brune and Dietrich, 2015). Furthermore, genes involved in energy production (9.0%) and inorganic ion transport (7.6%) contributed considerably to the bacterial metagenome. The pathway with the largest fraction of sequences was “replication, recombination and repair”. A closer inspection revealed that most sequences in this pathway originated from transposases (30.7%, Table S2).

“information storage and processing” differentiates the protist metagenomes of wood-dwelling and foraging lower termite species

In our previous study (Waidele *et al*., 2017) on the identical samples, the protist communities of the Rhinotermitidae *Prorhinotermes* and *Reticulitermes* clustered together, supporting a phylogenetic imprint on community composition. Here, we tested whether this pattern was also reflected by the functions encoded by the protist metagenome. Therefore, we compared the functional metagenome profiles, using Bray-Curtis-Dissimilarity (Bray and Curtis, 1957). This index considers abundance of functional categories, thus avoiding arbitrary coverage cutoffs.

The functional repertoire of protist origin clustered according to host family (Figure 3A), thus showing a dominant phylogenetic imprint. This was supported by Redundancy analysis (RDA). A model including host family explained more variance in the functional repertoire and produced lower AICs than a model based on life type (Table 1). For a more detailed view, we analyzed the three categories at the highest level in the eggNOG hierarchical annotation (Figure S1) separately. Cluster analysis of the categories “cellular process and signaling“ and “metabolism” supported the notion that phylogenetic

relatedness is an important factor for functional similarity (Supplement Figures S4B and D). In contrast, the portion of the metagenome assigned to “information storage and processing” (Figure 3B) clustered primarily by life type. The stronger effect of life type than phylogeny on this functional category was supported by higher explanatory power and lower AICs in RDA (Table 1).

Identification of the functions that differentiate the protist metagenomes of the wood-dwelling and foraging species can hold clues with regard to the nature of potentially adaptive phenotypes in the protist metagenome. In order to do so, we performed a linear discriminant analysis (LEfSe: Segata *et al*., 2011). This analysis identified 22 over-represented COGs in foraging and 14 in wood-dwelling species (Figure 4, Table S3, $p < 0.05$, $q < 0.05$, LDA score > 2, Figure S6).

The pathway “replication, recombination, and repair” was over-represented in the foraging species (Figure 4, Table S3, $p = 0.0001$, $q = 0.002$). The over-represented COGs in this pathway included a DNA dependent DNA-polymerase (COG0470) and five helicases (COG0514, COG0553, COG1199, COG1204, ENOG410XNUT, see Table S3 for p- and q-values). In wood-dwelling species, the pathway “transcription” was over-represented ($p = 0.0004$, $q = 0.003$). The over-represented COGs in this pathway contained DNA binding domains and were supposedly involved in transcriptional regulation (COG5147, ENOG4111SAB).

<u>The bacterial metabolic metagenome aligns with host ecology</u>

In our previous study (Waidele *et al*., 2017), the bacterial community composition of termite hosts clustered primarily by life type. Following the rationale above, we tested whether this pattern was reflected by the functions encoded by the metagenome.

Against the expectation from our previous study on community composition, the functional bacterial profiles showed no life type, but a phylogenetic imprint which is in line with the protist functional profiles. Most samples clustered according to host family (Figure 3C). Analyzing the three high-level eggNOG functional categories separately provided more detailed insight. The categories “cellular process and signaling“ and “information storage and processing” supported the notion of strong phylogenetic effects on metagenome function (Supplement Figures S5B and C). In contrast, the metabolic metagenomes (Figure 3D) clustered primarily according to host life type. Host life type was also a better predictor for metabolic functions than host family in RDA (Table 1).

Aside from these general patterns, several samples stood out. Samples Rg2 and Rg4 of *R. grassei* were on long branches in the dendrograms (Figure 3 and Supplement Figure S5), suggesting unusual functional profiles. Notably, these samples already stood out in our previous study (Waidele *et al*., 2017) because of their unusual abundance of microbial taxa potentially due to infection with pathogens. This unusual composition was confirmed by taxonomic annotation in this study (see Supplement Figure S3). Sample Cs7 (*C. secundus*) also clustered separately from the other samples. This was mainly driven by abundant transposases in this sample (53.1% of sequences) (for example COG1662, COG3385, or ENOG410XT1T, see Table S2), accompanied by an increase in the frequency of *Bacteroides* (Figure S3) that are rich in conjugative transposons (Whittle *et al*., 2002; Liu *et al*., 2013). We performed all analyses with and without these samples and found no qualitative differences (data not shown).

Bacterial metabolic functions that differentiated wood-dwelling from foraging species, were identified using linear discriminant analysis (LEfSe). 151 metabolic COGs were over-represented in the foraging species, while 105 were over-represented in the wood-dwelling species (Table S4, $p < 0.05$, $q < 0.05$, LDA score > 2, Figure S7). All COGs

described as over-represented or enriched in the following were subject to these p-value, q-value and LDA cutoffs.

In the foraging species, COGs involved in nitrite reduction (COG5557, COG1251) and nitrate reduction (COG5013, COG2181, COG5557), two crucial steps in nitrogen fixation, were enriched. Nitrogen is also recycled from uric acid and urea in the termite gut, involving glutamate synthesis (Potrikus and Breznak, 1981; Hongoh *et al*., 2008a; Burnum *et al*., 2011; Brune, 2014). Several COGs that are potentially involved in nitrogen recycling by urea utilization or glutamate synthesis were over-represented in the foraging species (COG0504, COG0174, COG1246, COG01364, COG0067, COG0069, COG0347 and COG0804, Table S4, Figure 5B). A substantial proportion of sequences (76.9% in total) in these COGs could be assigned to *Treponema* and *Desulfovibrio*. Apart from nitrogen fixation and recycling, specific genes involved in carbohydrate degradation were enriched in the foraging species (COG3858, COG0366, Table S4). The more abundant of the two (COG0366) was an amylase. Amylases break down starch. After starch break down, bacteria can uptake the resulting maltose with the phosphotransferase system (PTS) (Kotrba *et al*., 2001; Gao *et al*., 2016). Accordingly, PTSs were also over-represented in the foraging species (COG1080, COG1264, COG1455, COG2190, COG4668). Mostly members of the genus *Treponema* and *Desulfovibrio* were affiliated with these functions (see Table S4). Finally, twelve COGs that putatively play a role in hydrogen utilization were enriched in the foraging species. Among them was a molybdoprotein oxidoreductase (COG0243), five [NiFe] hydrogenases that bind hydrogen (COG0374, COG1740, COG3261, COG1969, COG0680) as well as iron-sulfur based hydrogenase – cofactors (COG0247, COG0437, COG1139, COG1142, COG1145). Genes for producing these cofactors (COG0822) were enriched as well. A diverse set of bacteria encoded the over-represented dehydrogenases and their putative cofactors (Table S4). These included

treponemes, *Desulfovibrio, Gordonibacter*, *Eggerthella*, Beta- (*Azovibrio*) and Gammaproteobacteria (*Enterobacter)* and Actinobacteria. Hydrogen that is formed during lignocellulose degradation in the termite gut (Pester and Brune, 2007) is an important energy source for termite gut microbes.

In the wood-dwelling species COGs with functions in the metabolism and synthesis of specific amino acids were more common. These comprised the essential amino acids tryptophan (COG3033, COG1350) and histidine (COG0241, COG2986), as well as the semi-essential amino acid cysteine (COG0031, COG0626, COG0520) and glycine (COG0404, COG1003). Furthermore, one putative cellulase (COG1472) and ten putative hemicellulases (COG1449, COG1482, COG3250, COG3507, COG3534, COG3537, COG3661, ENOG410XNW5, ENOG410XPHB, ENOG410YBRR) were more abundant, suggesting that hemicellulose digestion might be more important for wood-dwelling than for foraging species. All of these were affiliated with Bacteroidetes (mostly members of the genus *Bacteroides*) or the genus *Treponema*. Additional support for the increased importance of hemicellulose utilization in the wood-dwelling species, was provided by the over-representation of twelve COGs annotated as TonB-dependent receptors in their respective metagenomes (ENOG410XNNV, ENOG410XNPQ or COG4206, see Table S4). Apart from other substrates, these receptors are important for the uptake of plant-derived hemicellulose (Déjean *et al*., 2013; Hottes *et al*., 2004). All functions annotated as TonB-dependent receptors (or TonB-dependent associated receptor plugs) were affiliated with the genus *Bacteroides* (see Table S4).

Discussion

In this study, we assessed functional differences of termite metagenomes across an evolutionary switch from wood-dwelling to foraging. A set of five termite species (two

foraging, three wood-dwelling species) was chosen to test whether the functional profiles of the termite gut microbiome followed a phylogenetic pattern or aligned with ecology. We hypothesized that alignment of the termite life type with microbiome function is consistent with contributions of the microbiome to adaptations of the termite holobiont to different ecologies.

A potential pitfall of this approach is that an alignment of the termite microbiome with life type-related ecology could also be caused by transient microbes that are ingested from the environment. These microbes rather reflect the local environmental microbiome at the collection site and are unlikely to contribute to adaptation. Microbes that mediate adaptation of their host are expected to be fostered by the host and to persist (Foster *et al*., 2017). For persistent microbes, beneficial effects on the host, such as adaptive functions, are favored by evolution through long term and repeated interactions (Sachs, 2006; Ebert, 2013). Therefore, one must focus on the microbiome that persists in the termite gut. In order to focus on the persistent microbes and reduce the transient microbiome, we kept all termites under extensively tested and near natural common conditions, feeding them sterile wood (see material and methods section). The underlying assumption was that ecological imprints on the microbiome that remain under these common controlled conditions, reflect functions of microbes that persist and are more likely to play a role in adaptive processes. We consider these imprints more robust to false positive associations of ecology and the microbiome that might otherwise be caused by transient environmental microbes.

<u>Functional potential of protist and bacterial metagenomes</u>

The largest group of functions encoded by the protist metagenome were related to signaling (Figure 2A). In particular ABC-transporters were abundant. ABC-transporters can

transport a large variety of substrates in and out of the cell (Rees *et al*., 2009). In parasitic protists, ABC transporters are essential for drug resistance and nutrient salvation (Dean *et al*., 2014). They also play an important role in several insect-microbe-symbioses, in particular, when metabolic pathways are split between partners and intermediate metabolites have to be shuttled between organisms (Peterson and Scharf, 2016). Shuttling of molecules is of particular importance to protists in the termite gut because they occupy a mediator position between their intracellular endosymbiotic bacteria and the termite host. Metabolites and other molecules cannot be directly exchanged between termites and intracellular bacteria, instead transport through the protist is required. This in return requires transmembrane transporters and might explain their abundance.

In contrast to the prevalence of transporters in protists, the largest fraction of bacterial sequences had metabolic functions. The most common metabolic category was 'amino acid metabolism' including nitrogen uptake. The microbiome plays a central role in provisioning of amino acids and in nitrogen uptake because the host's primary food source, wood, is poor in nitrogen (Potrikus and Breznak, 1981; Burnum *et al*., 2011; Brune and Dietrich, 2015). The second most common pathway was carbohydrate metabolism. In this pathway, we identified 28 different families of glycoside hydrolases (Table S3 and S4). These are the primary enzymes responsible for lignocellulose degradation. Thus, our results support a growing body of evidence that bacteria play a fundamental role in lignocellulose degradation that was formerly mainly attributed to protists in lower termites (Tokuda and Watanabe, 2007; Do *et al*., 2014; Peterson *et al*., 2015; Yuki *et al*., 2015).

<u>Increased potential for replication in the protists of foraging termite species</u>

Besides differences in the functional repertoire of protists and bacteria, we detected differences between the wood-dwelling and foraging termite species. In the foraging

species, genes involved in replication were more abundant. High replication rates are expected to be more frequently under positive selection during recolonization of the gut with protists, when the gut environment has not yet reached carrying capacity (Baum *et al*., 2013). *Reticulitermes* guts have to be recolonized more frequently because they molt more frequently; the intermolt periods in *Reticulitermes* are about two weeks long (Lainé *et al*., 2003), while they average almost two months in *Cryptotermes* (Korb *et al*., 2012). During molting the protists are lost and the guts have to be recolonized through proctodeal trophallaxis from nest mates (McMahan, 1969).

Increased potential for nitrogen uptake in foraging species

While genes involved in replication differentiated the protist metagenomes of wood-dwelling and foraging species in our study, metabolic genes differentiated the bacterial metagenomes. The foraging *Reticulitermes* species had a higher potential for nitrogen uptake and recycling. This coincides with faster growth rates, as reflected by shorter intermolt periods, and increased fecundity when compared to wood-dwelling species (Lenz, 1994; Korb, 2010a; Korb, 2010b). Because a lack of nitrogen can limit termite growth (Lenz, 1994) and fecundity (Brent and Traniello, 2002), this suggests that increased growth and fecundity of the foraging species might be related to the increased nitrogen uptake ability by the microbiome.

The majority of sequences with functions in nitrogen fixation or recycling could be assigned to *Treponema* supporting the findings of previous studies (Lilburn *et al*., 2001; Ohkuma *et al*., 2015). Interestingly, sequences attributed to *Desulfovibrio* seemingly contributed to nitrogen fixation in this study. In comparision, Kuwahara et al. (Kuwahara *et al*., 2017) found no enzymes for nitrogen fixation, nor transporters for urea or ammonia in *Desulfovibrio* strain Rs-N31 from *Reticulitermes speratus*. We found a nitrate reductase

(COG2181) and a glutamate synthases (COG0067) that could be involved in fixing nitrogen, suggesting that the functional repertoire might differ between *Desulfovibrio* strains from *R. speratus* and *R. flavipes*. In the wood-dwelling species, we found evidence for increased potential to synthesize the essential amino acids tryptophan and histidine, as well as the semi-essential amino acids cysteine and glycine. Amino acid synthesis has long been attributed to bacterial symbionts in the termite gut (Brune, 2014). Wood-dwelling species with their uniform, amino acid poor wood diet might rely more heavily on their resident microbiome for acquisition of essential and rare amino acids. Foraging species, on the other hand, are able to tap into other sources for specific amino acids in the environment, e.g. ingested bacteria and fungi (Zoberi and Grace, 1990), and generally feed on a more diverse range of substrates (Waller, 1991). Although their primary food source in most habitats is also wood, *Reticulitermes,* the genus of foraging termites analyzed here, occurs in the prairie that is devoid of wood where it feeds on grasses and even roots of life plants (Brown *et al*., 2008). *Reticulitermes* was also isolated from horse dung (Rahman *et al*., 2015). Furthermore, *R. flavipes* can acquire nutrients from soil (Janzow and Judd, 2015) and actively balance mineral uptake by food choice (Judd et al., 2017).

<u>Larger functional potential for hemicellulose utilization in wood-dwelling and for starch utilization in foraging termite species</u>

Consistent with differences in their respective diets, the metagenomes of foraging and wood-dwelling species in our study differed by their potential for carbohydrate utilization. Several glycoside hydrolases that have hemicellulolytic function were over-represented in the metagenomes of wood-dwelling species (GH families 2, 3, 18, 43, 57, mannosidases, furanosidases, glucoronidases, see Table S4). This suggests a more important role for

hemicellulose degradation in the metabolism of the wood-dwelling species in our study. Accordingly, TonB dependent transporters were over-represented. These receptors can shuttle hemicellulose and its building-blocks, in particular xylans and xylose through bacterial membranes (Blanvillain *et al*., 2007; Schauer *et al*., 2008). A large fraction of both, hemicellulases and putative TonB transporters were attributed to the genus *Bacteroides.* In *Bacteroides*, TonB dependent transporters are often co-localized and co-regulated with enzymes for polysaccharide degradation like hemicellulases (Koebnik, 2005; Liu *et al*., 2013). This suggests a partnership of enzymes and transporters in polysaccharide degradation. *Bacteroides* species from the human gut are also hemicellulose degraders (Martens *et al*., 2011), suggesting a distinctive role for the genus in hemicellulose degradation in termites as well.

In contrast to the higher potential for hemicellulase utilization in wood-dwelling species, the foraging *Reticulitermes* species in our study showed an elevated abundance of amylases and potential transporters of amylase metabolites (PTSs) in their metagenomes. Amylases degrade starch to maltose. In return, maltose can be transported into the bacterial cell and degraded into glucose by a PTS, as has been shown by a PTS from *Coptotermes formosanus* (Gao *et al*., 2016).

The above identified changes in functional potential are suggestive of adaptations to a dietary shift from substrates richer in hemicellulose to substrates containing more starch. Hemicellulose as well as starch content differs between wood species and tree organs (Fengel and Grosser, 1975; Pettersen, 1984; Bellasio *et al*., 2014). Hardwood contains elevated hemicellulose levels when compared to softwood. This might reflect that the wood-dwelling *Cryptotermes* species in our study are mostly found in hard-wood mangroves (Korb and Lenz, 2004). The other wood-dwelling genus in our study, *Prorhinotermes,* lives in similar coastal habitats with a similar arboreal flora (Scheffrahn,

2007). On the other hand, *Reticulitermes* species originated in inland habitats (Dedeine *et al*., 2016) and prefer soft woods like pine (Smythe and Carter, 1969; Cárdenas *et al*., 2018) that have on average lower hemicellulose levels. The higher potential for starch digestion in *Reticulitermes* might reflect a preference for feeding on roots that have a higher starch content (Bellasio *et al*., 2014). This seems reasonable because *Reticulitermes* are subterranean termites, thus roots of dead plants are a readily accessible food-source.

Dehydrogenases in foraging termite species

Hydrogen is an abundant intermediate metabolite of cellulose degradation in the termite gut (Pester and Brune, 2007). It is primarily consumed by acetogenic bacteria through reductive acetogenesis from carbon dioxide via dehydrogenases. Dehydrogenases in the termite gut are often fused to signaling domains (Warnecke *et al*., 2007). Proteins combining dehydrogenase with signaling domains can be used for chemotaxis along hydrogen gradients by bacteria. Strong hydrogen gradients have been found in the gut of *Reticulitermes*, but do not exist in *Cryptotermes* (Pester and Brune, 2007). Hence, a higher need for sensing differences in hydrogen concentration in the gut of *Reticulitermes* might explain the increase in dehydrogenases.

Phylogeny and ecology align with metagenome-encoded functions

The higher abundance of dehydrogenases as well as the increased propensity for replication of protists in the gut of *Reticulitermes* are consistent with the hypothesis that the resident microbiome adapts to host species specific conditions in the gut (hydrogen gradients, shorter intermolt periods). On the other hand, differences in the propensity for nitrogen uptake are more likely to reflect differences in growth rates and diet of the termite

host. Given differences in diet between the species that represent the different life types, it seems also reasonable to suggest that the changes in the repertoire of hemicellulases and amylases reflect adaptations of the microbiome to diets with different hemicellulose and starch contents. The finding that this manifested specifically in the metabolic functional repertoire, may suggest that potential selection acts in particular on metabolic functions.

Microbiome mediated adaptation to different diets can happen in two ways. First, acquisition of new microbes with adaptive functions could lead to adaptive changes of the microbiome. Second, genome evolution of microbes that are already associated with the host could lead to adaptation. Microbes that were already present before the onset of lineage specific adaptation are likely to be shared among host species. By contrast, newly acquired microbes are expected to be host lineage specific. We found that the bacterial groups that contributed most to the differentiation of metabolic functions are shared among all five host species (*Treponema*, *Desulfovibrio*, *Dysgomonas*, Table S4, Figure S3). This supports that genome evolution of microbes that were already associated with the host contributed to potential adaptation in our model system.

In conclusion, we applied metagenomics techniques and a controlled experimental setup to study a potential role of the microbiome in adaptation of the host to different ecologies that parallel the evolutionary switch from wood dwelling to foraging life types. While the overall patterns of microbiome variation were governed by a phylogenetic signal, specific parts of the microbiome aligned with host-ecology. The specificity of the changes allowed us to hypothesize that the microbiome contributed to dietary adaptations. This hypothesis can now be tested, assessing host fitness under different dietary conditions. Such experiments will be crucial to disentangle adaptive functional changes from selectively neutral functional turnover or side effects of other adaptive events.

Experimental Procedures

Termite samples

All termites were collected from typical natural habitats (see Waidele *et al.* 2017). They were kept under constant conditions (27°C, 70% humidity) on autoclaved *Pinus radiata* wood from the same source for at least six weeks prior to the experiment. The feeding of *Pinus* represents a natural or near natural treatment; *Pinus* is a natural food source of *P. simplex* and *Reticulitermes. Cryptotermes* growth and behavior on *Pinus* recapitulates that on natural substrate (Korb and Lenz, 2004). The time of the acclimation period was chosen to lie well beyond the gut passage time of 24 h in lower termites (Breznak, 1984; Li *et al*., 2017) and following Huang et al., (Huang *et al*., 2013), who showed that six weeks are sufficient for the microbiota to adjust to a new diet. That way, all excretable material like remaining food, transient microbes taken up from the environment that have no mechanisms to persist in the gut, and microbial DNA taken up before the experiment was made sure to be excreted. The samples were identical to those analyzed in our previous study, (Waidele *et al*., 2017) where detailed information about animal collection, keeping, and cytochrome oxidase II based species identification and a phylogeny can be found.

DNA extraction and Shotgun sequencing

DNA was extracted from a pool of three worker guts per colony using bead beating, chloroform extraction and isopropanol precipitation (see Supplementary material and methods file S7). Each of the 29 colony samples went through independent metagenomic shotgun library preparation and sequencing on an Illumina HiSeq platform (150 bp paired end reads).

Analysis

We employed a double filtering strategy to remove host DNA from our analysis. First, sequences were removed that mapped to available host genomes from *C. secundus* (Harrison *et al*., 2018) and transcriptomes from *P. simplex* (Misof *et al*., 2014) and *R. flavipes*, provided by the 1KITE consortium (www.1kite.org, BioSample SAMN04005235) using BBMap (Bushnell, BBMap - sourceforge.net/projects/bbmap/) (for detailed workflow and more detailed information about used genomes and transcriptomes see Supplementary Figure S2 and file S8). Of note, the sequences were not assembled, but individual reads were directly annotated. In a second step we used taxonomic and functional annotations with Megan6 (Huson *et al.*, 2007), to retrieve only sequences that could be unambiguously assigned to either bacteria or protists. In order to compare the bacterial and protist data sets of all samples, they were rarefied to the number of sequences in the sample with lowest coverage, resulting in 1,386,882 and 2,781 sequences per sample, respectively. Sample Cs4 was excluded from the analysis for insufficient sequence coverage (974,176 sequences), so was Cs5 from the protist data. Sample Ps5 did not pass the analysis pipeline and was also excluded.

Functional annotation with the eggNOG database resulted in the highest number of annotated sequences (21,215,480 annotated sequences in total) and was chosen for further functional analysis. Bray-Curtis distances of functional abundances were clustered with the pvClust package in R (Suzuki and Shimodaira, 2015). Multivariate modeling was performed via RDA (Redundancy Analysis) and AICs as well as values for the proportion of variance explained were derived with the model selection tool ordistep and ordiR2step, as implemented in the R vegan package (Oksanen *et al*., 2017). Models were compared to the null-model via ANOVA To identify over-represented functions associated with the two termite life types, a Linear Discriminant Analysis (LDR) was performed using LEfSe

(Segata *et al*., 2011) and visualized using graphlan (Asnicar *et al*., 2015). A detailed workflow for full reproducibility can be found in Supplementary file S8.

Acknowledgments

We thank Jan Sobotnik for kindly providing *P. simplex* colonies, as well as Charles Darwin University (Australia), especially S. Garnett and the Horticulture and Aquaculture team for providing logistic support to collect *C. secundus*. The Parks and Wildlife Comission, Northern Territory, Department of the Environment, Water, Heritage and the Arts gave permission to collect (Permit number 59044) and export (Permit PWS2016-AU-001559) the termites. This work was funded by DFG grants STA1154/2-1 and KO1895/16-1. We thank Craig Michell and the KAUST Bioscience Core Lab for sequence library generation and sequencing. This study was supported by King Abdullah University of Science and Technology (KAUST) and the High Performance and Cloud Computing Group at the Zentrum fuer Datenverarbeitung of the University of Tuebingen, the state of Baden-Wuerttemberg through bwHPC and the German Research Foundation (DFG) through grant no INST 37/935-1 FUGG. We thank Karen Meusemann and the 1KITE consortium, in particular the 1KITE Blattodea group, Alexander Donath, Lars Podsiadlowski, Bernhard Misof, Xin Zhou for granting access of the transcriptome data of *R. santonsensis* syn. *R. flavipes*.

Data availability

The raw data has been uploaded to the NCBI short-read archive (BioProject ID PRJNA509211, Accession: SAMN10573992 – SAMN10574019).

576 **References**

Abe, T. (1987) Evolution of life types in termites. In, Kawano, S., Connell, J.H., and Hidaka, T. (eds), *Evolution and Coadaptation in Biotic Communities*. University of Tokyo Press, pp. 125–146.

Asnicar, F., Weingart, G., Tickle, T.L., Huttenhower, C., and Segata, N. (2015) Compact graphical representation of phylogenetic data and metadata with GraPhlAn. *PeerJ* **3**:.

Bang, C., Dagan, T., Deines, P., Dubilier, N., Duschl, W.J., Fraune, S., et al. (2018) Metaorganisms in extreme environments: do microbes play a role in organismal adaptation? *Zoology* **127**: 1–19.

Baum, D.A., Futuyma, D.J., Hoekstra, H.E., Lenski, R.E., Moore, A.J., Peichel, C.L., et al. (2013) The Princeton Guide to Evolution. Princeton University Press.

Bushnell, B. BBMap - sourceforge.net/projects/bbmap/.

Bellasio, C., Fini, A., and Ferrini, F. (2014) Evaluation of a High Throughput Starch Analysis Optimised for Wood. *PLOS ONE* **9**: e86645.

Benjamino, J. and Graf, J. (2016) Characterization of the Core and Caste-Specific Microbiota in the Termite, *Reticulitermes flavipes*. *Front Microbiol* **7**:.

Blanvillain, S., Meyer, D., Boulanger, A., Lautier, M., Guynet, C., Denancé, N., et al. (2007) Plant Carbohydrate Scavenging through TonB-Dependent Receptors: A Feature Shared by Phytopathogenic and Aquatic Bacteria. *PLOS ONE* **2**: e224.

Boucias, D.G., Cai, Y., Sun, Y., Lietze, V.-U., Sen, R., Raychoudhury, R., and Scharf, M.E. (2013) The hindgut lumen prokaryotic microbiota of the termite *Reticulitermes flavipes* and its responses to dietary lignocellulose composition. *Molecular Ecology* **22**: 1836–1853.

Bourguignon, T., Lo, N., Dietrich, C., Šobotník, J., Sidek, S., Roisin, Y., et al. (2018) Rampant Host Switching Shaped the Termite Gut Microbiome. *Current Biology* **28**: 649-654.e2.

Bray, J.R. and Curtis, J.T. (1957) An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecological Monographs* **27**: 325–349.

Brent, C.S. and Traniello, J.F.A. (2002) Effect of Enhanced Dietary Nitrogen on Reproductive Maturation of the Termite *Zootermopsis angusticollis* (Isoptera: Termopsidae). *Environmental Entomology* **31**: 313–318.

Breznak, J.A. (1984) Biochemical aspects of symbiosis between termites and their intestinal microbiota. In, Anderson,J.M. (ed), *Invertebrate - Microbial Interaction*. Cambridge University Press, Cambridge, pp. 173–203.

Brown, K.S., Broussard, G.H., Kard, B.M., Smith, A.L., and Smith, M.P. (2008) Colony Characterization of *Reticulitermes flavipes* (Isoptera: Rhinotermitidae) on a Native Tallgrass Prairie. *The American Midland Naturalist* **159**: 21.

Brune, A. (2012) Endomicrobia: intracellular symbionts of termite gut flagellates. *Endocytobiosis & Cell Research* **23**: 11–15.

Brune, A. (2014) Symbiotic digestion of lignocellulose in termite guts. *Nat Rev Micro* **12**: 168–180.

Brune, A. and 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.

Burnum, K.E., Callister, S.J., Nicora, C.D., Purvine, S.O., Hugenholtz, P., Warnecke, F., et al. (2011) Proteome insights into the symbiotic relationship between a captive colony of *Nasutitermes corniger* and its hindgut microbiome. *The ISME Journal* **5**: 161–164.

Cárdenas, A.M., Gallardo, P., and Toledo, D. (2018) Suitability of multiple Mediterranean oak species as a food resource for *Reticulitermes grassei* Clément (Isoptera: Rhinotermitidae). *Bulletin of Entomological Research* **108**: 532–539.

Clark, A.G., Eisen, M.B., Smith, D.R., Bergman, C.M., Oliver, B., Markow, T.A., et al. (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**: 203–218.

Cleveland, L.R. (1923) Symbiosis between termites and their intestinal protozoa. *Proceedings of the National Academy of Sciences of the United States of America* **9**: 424.

Dean, P., Major, P., Nakjang, S., Hirt, R.P., and Embley, T.M. (2014) Transport proteins of parasitic protists and their role in nutrient salvage. *Front Plant Sci* **5**:.

Dedeine, F., Dupont, S., Guyot, S., Matsuura, K., Wang, C., Habibpour, B., et al. (2016) Historical biogeography of *Reticulitermes* termites (Isoptera: Rhinotermitidae) inferred from analyses of mitochondrial and nuclear loci. *Molecular Phylogenetics and Evolution* **94**, **Part B**: 778–790.

Déjean, G., Blanvillain-Baufumé, S., Boulanger, A., Darrasse, A., Bernonville, T.D. de, Girard, A.-L., et al. (2013) The xylan utilization system of the plant pathogen *Xanthomonas campestris* pv *campestris* controls epiphytic life and reveals common features with oligotrophic bacteria and animal gut symbionts. *New Phytologist* **198**: 899–915.

Desai, M.S. and Brune, A. (2012) *Bacteroidales* ectosymbionts of gut flagellates shape the nitrogen-fixing community in dry-wood termites. *ISME J* **6**: 1302–1313.

Desai, M.S., Strassert, J.F.H., 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). *Environmental Microbiology* **12**: 2120–2132.

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.

Do, T.H., Nguyen, T.T., Nguyen, T.N., Le, Q.G., Nguyen, C., Kimura, K., and Truong, N.H. (2014) Mining biomass-degrading genes through Illumina-based de novo sequencing and metagenomic analysis of free-living bacteria in the gut of the lower termite *Coptotermes gestroi* harvested in Vietnam. *Journal of Bioscience and Bioengineering* **118**: 665–671.

Douglas, A.E. (2011) Lessons from Studying Insect Symbioses. *Cell Host & Microbe* **10**: 359–367.

Douglas, A.E. (2015) Multiorganismal insects: diversity and function of resident microorganisms. *Annu. Rev. Entomol.* **60**: 17–34.

Duarte, S., Nobre, T., Borges, P.A.V., and Nunes, L. (2018) Symbiotic flagellate protists as cryptic drivers of adaptation and invasiveness of the subterranean termite *Reticulitermes grassei* Clément. *Ecol Evol* **8**: 5242–5253.

Ebert, D. (2013) The Epidemiology and Evolution of Symbionts with Mixed-Mode Transmission. *Annu. Rev. Ecol. Evol. Syst.* **44**: 623–643.

Fengel, D. and Grosser, D. (1975) Chemische Zusammensetzung von Nadel- und Laubholzern. Eine Literaturuebersicht. *Holz als Roh- und Werkstoff*.

Foster, K.R., Schluter, J., Coyte, K.Z., and Rakoff-Nahoum, S. (2017) The evolution of the host microbiome as an ecosystem on a leash. *Nature* **548**: 43–51.

Gao, G., Wang, A., Gong, B.-L., Li, Q.-Q., Liu, Y.-H., He, Z.-M., and Li, G. (2016) A novel metagenome-derived gene cluster from termite hindgut: Encoding

phosphotransferase system components and high glucose tolerant glucosidase. *Enzyme Microb. Technol.* **84**: 24–31.

Harrison, M.C., Jongepier, E., Robertson, H.M., Arning, N., Bitard-Feildel, T., Chao, H., et al. (2018) Hemimetabolous genomes reveal molecular basis of termite eusociality. *Nature Ecology & Evolution* **2**: 557–566.

He, S., Ivanova, N., Kirton, E., Allgaier, M., Bergin, C., Scheffrahn, R.H., et al. (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., Inoue, T., Moriya, S., Trakulnaleamsai, S., Ohkuma, M., et al. (2005) Intra- and Interspecific Comparisons of Bacterial Diversity and Community Structure Support Coevolution of Gut Microbiota and Termite Host. *Appl. Environ. Microbiol.* **71**: 6590–6599.

Hongoh, Y., Ekpornprasit, L., Inoue, T., Moriya, S., Trakulnaleamsai, S., Ohkuma, M., et al. (2006) Intracolony variation of bacterial gut microbiota among castes and ages in the fungus-growing termite *Macrotermes gilvus*. *Molecular Ecology* **15**: 505–516.

Hongoh, Y., Sato, T., Noda, S., Ui, S., Kudo, T., and Ohkuma, M. (2007) Candidatus *Symbiothrix dinenymphae*: bristle-like *Bacteroidales* ectosymbionts of termite gut protists. *Environmental Microbiology* **9**: 2631–2635.

Hongoh, Y., Sharma, V.K., Prakash, T., Noda, S., Taylor, T.D., Kudo, T., et al. (2008a) Complete genome of the uncultured Termite Group 1 bacteria in a single host protist cell. *PNAS* **105**: 5555–5560.

Hongoh, Y., Sharma, V.K., Prakash, T., Noda, S., Toh, H., Taylor, T.D., et al. (2008b) Genome of an Endosymbiont Coupling N2 Fixation to Cellulolysis Within Protist Cells in Termite Gut. *Science* **322**: 1108–1109.

Hottes, A.K., Meewan, M., Yang, D., Arana, N., Romero, P., McAdams, H.H., and Stephens, C. (2004) Transcriptional Profiling of *Caulobacter crescentus* during Growth on Complex and Minimal Media. *J. Bacteriol.* **186**: 1448–1461.

Huang, X.-F., Bakker, M.G., Judd, T.M., Reardon, K.F., and Vivanco, J.M. (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.

Huson, D.H., Auch, A.F., Qi, J., and Schuster, S.C. (2007) MEGAN analysis of metagenomic data. *Genome Res* **17**: 377–386.

Ikeda-Ohtsubo, W. and Brune, A. (2009) Cospeciation of termite gut flagellates and their bacterial endosymbionts: *Trichonympha* species and 'Candidatus *Endomicrobium trichonymphae*.' *Molecular Ecology* **18**: 332–342.

Jaenike, J., Unckless, R., Cockburn, S.N., Boelio, L.M., and Perlman, S.J. (2010) Adaptation via Symbiosis: Recent Spread of a *Drosophila* Defensive Symbiont. *Science* **329**: 212–215.

Janzow, M.P. and Judd, T.M. (2015) The Termite *Reticulitermes flavipes* (Rhinotermitidae: Isoptera) Can Acquire Micronutrients from Soil. *Environmental Entomology* **44**: 814–820.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2017) vegan: Community Ecology Package.

Judd, T., Landes, J., Ohara, H., and Riley, A. (2017) A Geometric Analysis of the Regulation of Inorganic Nutrient Intake by the Subterranean Termite *Reticulitermes flavipes* Kollar. *Insects* **8**: 97.

Kitade, O. (2004) Comparison of Symbiotic Flagellate Faunae between Termites and a Wood-Feeding Cockroach of the Genus *Cryptocercus*. *Microbes and Environments* **19**: 215–220.

Koebnik, R. (2005) TonB-dependent trans-envelope signalling: the exception or the rule? *Trends in Microbiology* **13**: 343–347.

Korb, J. (2007) Termites. *Current Biology* **17**: R995-R999.

Korb, J. (2010a) Social insects, major evolutionary transitions and multilevel selection. In, *Animal Behaviour: Evolution and Mechanisms*. Springer Press, Heidelberg, pp. 179–211.

Korb, J. (2010b) Termites: Social Evolution. In, Breed, M.D. and Moore, J. (eds), *Encyclopedia of Animal Behavior*. Academic Press, Oxford, pp. 394–400.

Korb, J., Hoffmann, K., and Hartfelder, K. (2012) Molting dynamics and juvenile hormone titer profiles in the nymphal stages of a lower termite, *Cryptotermes secundus* (Kalotermitidae) – signatures of developmental plasticity. *Journal of Insect Physiology* **58**: 376–383.

Korb, J. and Lenz, M. (2004) Reproductive decision-making in the termite, *Cryptotermes secundus* (Kalotermitidae), under variable food conditions. *Behav Ecol* **15**: 390–395.

Kotrba, P., Inui, M., and Yukawa, H. (2001) Bacterial phosphotransferase system (PTS) in carbohydrate uptake and control of carbon metabolism. *Journal of Bioscience and Bioengineering* **92**: 502–517.

Kuwahara, H., Yuki, M., Izawa, K., Ohkuma, M., and Hongoh, Y. (2017) Genome of '*Ca.* Desulfovibrio trichonymphae', an H_2-oxidizing bacterium in a tripartite symbiotic system within a protist cell in the termite gut. *The ISME Journal* **11**: 766–776.

Lainé, L.V., Lainé, L.V., and Wright, D.J. (2003) The life cycle of *Reticulitermes* spp (Isoptera: Rhinotermitidae): what do we know? *Bull. Entomol. Res.* **93**: 267–378.

Leadbetter, J.R., Schmidt, T.M., Graber, J.R., and Breznak, J.A. (1999) Acetogenesis from H2 Plus CO2 by Spirochetes from Termite Guts. *Science* **283**: 686–689.

Lenz, M. (1994) Food resources, colony growth and caste development in wood-feeding termites. In: Nourishment and evolution in insect societies. *Westview Press* 159–209.

Li, H., Yelle, D.J., Li, C., Yang, M., Ke, J., Zhang, R., et al. (2017) Lignocellulose pretreatment in a fungus-cultivating termite. *PNAS* **114**: 4709–4714.

Lilburn, T.G., Kim, K.S., Ostrom, N.E., Byzek, K.R., Leadbetter, J.R., and Breznak, J.A. (2001) Nitrogen Fixation by Symbiotic and Free-Living Spirochetes. *Science* **292**: 2495–2498.

Liu, N., Li, H., Chevrette, M.G., Zhang, L., Cao, L., Zhou, H., et al. (2018) Functional metagenomics reveals abundant polysaccharide-degrading gene clusters and cellobiose utilization pathways within gut microbiota of a wood-feeding higher termite. *The ISME Journal* 1.

Liu, N., Zhang, L., Zhou, H., Zhang, M., Yan, X., Wang, Q., et al. (2013) Metagenomic Insights into Metabolic Capacities of the Gut Microbiota in a Fungus-Cultivating Termite (*Odontotermes yunnanensis*). *PLoS One* **8**:.

Ma, D. and Leulier, F. (2018) The importance of being persistent: The first true resident gut symbiont in *Drosophila*. *PLOS Biology* **16**: e2006945.

Martens, E.C., Lowe, E.C., Chiang, H., Pudlo, N.A., Wu, M., McNulty, N.P., et al. (2011) Recognition and Degradation of Plant Cell Wall Polysaccharides by Two Human Gut Symbionts. *PLOS Biology* **9**: e1001221.

McFall-Ngai, M., Hadfield, M.G., Bosch, T.C.G., Carey, H.V., Domazet-Lošo, T., Douglas, A.E., et al. (2013) Animals in a bacterial world, a new imperative for the life sciences. *PNAS* **110**: 3229–3236.

McMahan, E.A. (1969) Feeding relationships and radio-isotope techniques. In, Krishna K. and Weesner, F.M. (eds), *Biology of Termites* vol.1. Academic Press, New York, London, pp. 387-406.

Mikaelyan, A., Dietrich, C., Köhler, T., Poulsen, M., Sillam-Dussès, D., and Brune, A. (2015) Diet is the primary determinant of bacterial community structure in the guts of higher termites. *Mol Ecol* **24**: 5284–5295.

Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., et al. (2014) Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**: 763–767.

Noda, S., Kitade, O., Inoue, T., Kawai, M., Kanuka, M., Hiroshima, K., et al. (2007) Cospeciation in the triplex symbiosis of termite gut protists (*Pseudotrichonympha* spp.), their hosts, and their bacterial endosymbionts. *Molecular Ecology* **16**: 1257–1266.

Ohkuma, M. (2008) Symbioses of flagellates and prokaryotes in the gut of lower termites. *Trends in Microbiology* **16**: 345–352.

Ohkuma, M., Noda, S., Hattori, S., Iida, T., Yuki, M., Starns, D., et al. (2015) Acetogenesis from H2 plus CO2 and nitrogen fixation by an endosymbiotic spirochete of a termite-gut cellulolytic protist. *PNAS* **112**: 10224–10230.

Pester, M. and Brune, A. (2007) Hydrogen is the central free intermediate during lignocellulose degradation by termite gut symbionts. *The ISME Journal* **1**: 551–565.

Peterson, B.F. and Scharf, M.E. (2016) Metatranscriptome analysis reveals bacterial symbiont contributions to lower termite physiology and potential immune functions. *BMC Genomics* **17**: 772.

Peterson, B.F., Stewart, H.L., and Scharf, M.E. (2015) Quantification of symbiotic contributions to lower termite lignocellulose digestion using antimicrobial treatments. *Insect Biochemistry and Molecular Biology* **59**: 80–88.

Pettersen, R.C. (1984) The Chemical Composition of Wood. In, Rowell,R. (ed), *The Chemistry of Solid Wood*. American Chemical Society, Washington, DC, pp. 57–126.

Potrikus, C.J. and Breznak, J.A. (1981) Gut bacteria recycle uric acid nitrogen in termites: A strategy for nutrient conservation. *Proc Natl Acad Sci U S A* **78**: 4601–4605.

Radek, R., Meuser, K., Strassert, J.F.H., Arslan, O., Teßmer, A., Šobotník, J., et al. (2018) Exclusive Gut Flagellates of Serritermitidae Suggest a Major Transfaunation Event in Lower Termites: Description of *Heliconympha glossotermitis* gen. nov. spec. nov. *Journal of Eukaryotic Microbiology* **65**: 77–92.

Rahman, N.A., Parks, D.H., Willner, D.L., Engelbrektson, A.L., Goffredi, S.K., Warnecke, F., et al. (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.

Rees, D.C., Johnson, E., and Lewinson, O. (2009) ABC transporters: The power to change. *Nat Rev Mol Cell Biol* **10**: 218–227.

Rossmassler, K., Dietrich, C., Thompson, C., Mikaelyan, A., Nonoh, J.O., Scheffrahn, R.H., et al. (2015) Metagenomic analysis of the microbiota in the highly compartmented hindguts of six wood- or soil-feeding higher termites. *Microbiome* **3**: 56.

Sachs, J.L. (2006) Cooperation within and among species. *Journal of Evolutionary Biology* **19**: 1415–1418.

Scharf, M.E., Karl, Z.J., Sethi, A., and Boucias, D.G. (2011) Multiple Levels of Synergistic Collaboration in Termite Lignocellulose Digestion. *PLOS ONE* **6**: e21709.

Schauer, K., Rodionov, D.A., and de Reuse, H. (2008) New substrates for TonB-dependent transport: do we only see the 'tip of the iceberg'? *Trends in Biochemical Sciences* **33**: 330–338.

Scheffrahn, R. (2007) Cuban subterranean termite - *Prorhinotermes simplex* (Hagen). http://entnemdept.uft.edu/creatures/urban/termites/p_simplex.htm Accessed 12 Jul 2018.

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S., and Huttenhower, C. (2011) Metagenomic biomarker discovery and explanation. *Genome Biol* **12**: R60.

Shimada, K., Lo, N., Kitade, O., Wakui, A., and Maekawa, K. (2013) Cellulolytic Protist Numbers Rise and Fall Dramatically in Termite Queens and Kings during Colony Foundation. *Eukaryotic Cell* **12**: 545–550.

Smythe, R.V. and Carter, F.L. (1969) Feeding Responses to Sound Wood by the Eastern Subterranean Termite, *Reticulitermes flavipes*. *Annals of the Entomological Society of America* **62**: 335–337.

Suzuki, R. and Shimodaira, H. (2015) pvclust: Hierarchical Clustering with P-Values via Multiscale Bootstrap Resampling. R package version 2.0-0. http://CRAN.R-project.org.

Tai, V., James, E.R., Nalepa, C.A., Scheffrahn, R.H., Perlman, S.J., and Keeling, P.J. (2015) The Role of Host Phylogeny Varies in Shaping Microbial Diversity in the Hindguts of Lower Termites. *Appl. Environ. Microbiol.* **81**: 1059–1070.

Tokuda, G. and Watanabe, H. (2007) Hidden cellulases in termites: revision of an old hypothesis. *Biol Lett* **3**: 336–339.

Waidele, L., Korb, J., Voolstra, C.R., Künzel, S., Dedeine, F., and Staubach, F. (2017) Differential Ecological Specificity of Protist and Bacterial Microbiomes across a Set of Termite Species. *Front. Microbiol.* **8**:.

Waller, D.A. (1991) Feeding by *Reticulitermes* spp. *Sociobiology* **19**: 91–99.

Warnecke, F., Luginbühl, P., Ivanova, N., Ghassemian, M., Richardson, T.H., Stege, J.T., et al. (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**: 560–565.

Whittle, G., Shoemaker, N.B., and Salyers, A.A. (2002) The role of *Bacteroides* conjugative transposons in the dissemination of antibiotic resistance genes. *Cell. Mol. Life Sci.* **59**: 2044–2054.

Yamin, M., A. (1979) Flagellates of the orders Trichomonadida Kirby, Oxymonadida Grasse´, and Hypermastigida Grassi & Foa reported from lower termites (Isoptera families Mastotermitidae, Kalotermitidae, Hodotermitidae, Termopsidae, Rhinotermitidae, and Serritermitidae) and from the wood-feeding roach *Cryptocercus* (Dictyoptera: Cryptocercidae). *Sociobiology* **4**: 5–119.

Yuki, M., Kuwahara, H., Shintani, M., Izawa, K., Sato, T., Starns, D., et al. (2015) Dominant ectosymbiotic bacteria of cellulolytic protists in the termite gut also have the potential to digest lignocellulose. *Environmental Microbiology* **17**: 4942–4953.

Zhang, X. and Leadbetter, J.R. (2012) Evidence for Cascades of Perturbation and Adaptation in the Metabolic Genes of Higher Termite Gut Symbionts. *mBio* **3**: e00223-12.

Zheng, H., Dietrich, C., Radek, R., and Brune, A. (2016) *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. *Environmental Microbiology* **18**: 191–204.

Zoberi, M.H. and Grace, J.K. (1990) Fungi Associated with the Subterranean Termite *Reticulitermes flavipes* in Ontario. *Mycologia* **82**: 289–294.

Figures and Tables

Table 1: Models of the effects of life type and host family (phylogeny) on functional community profiles. Effects were analyzed with Redundancy analysis using the functional abundance table as response variable. Host family was the best explanatory variable for the combination of all protist functions, however life type explained more variance (R^2) and produced a lower AIC for the category "information storage and processing" than host family. For all bacterial functions taken together, host family again explained a larger proportion of the variance, while life type was the best explanatory variable in the category "metabolism".

		model	AIC	R2	*p*
protists	all functions	null	-41.8		
		host family	-49.0	0.27	0.001
		host lifetype	-45.4	0.16	0.001
	information storage and processing	null	-36.8		
		host lifetype	-39.0	0.11	0.001
		host family	-38.4	0.09	0.001
bacteria	all functions	null	-85.0		
		host family	-92.7	0.27	0.001
		host lifetype	-90.1	0.20	0.001
	metabolism	null	-92.4		
		host lifetype	-98.1	0.22	0.001
		host family	-96.8	0.18	0.001

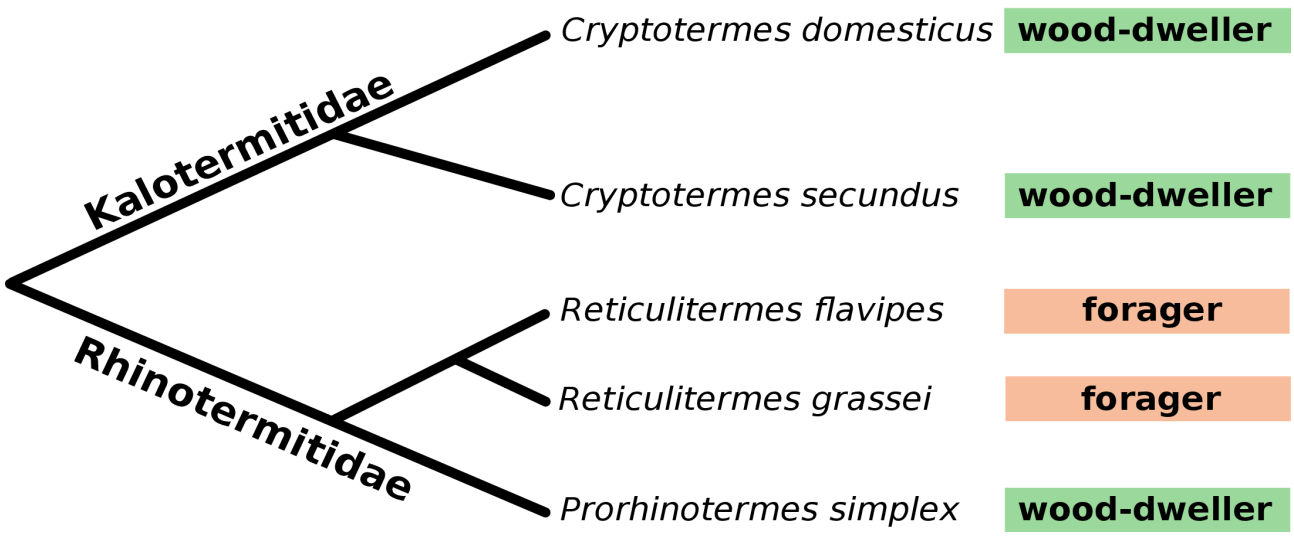


Figure 1: Schematic phylogeny of the five lower termite species used in this study from (Waidele *et al*., 2017)**.** Branch length not drawn to scale. Colored boxes indicate the life type.

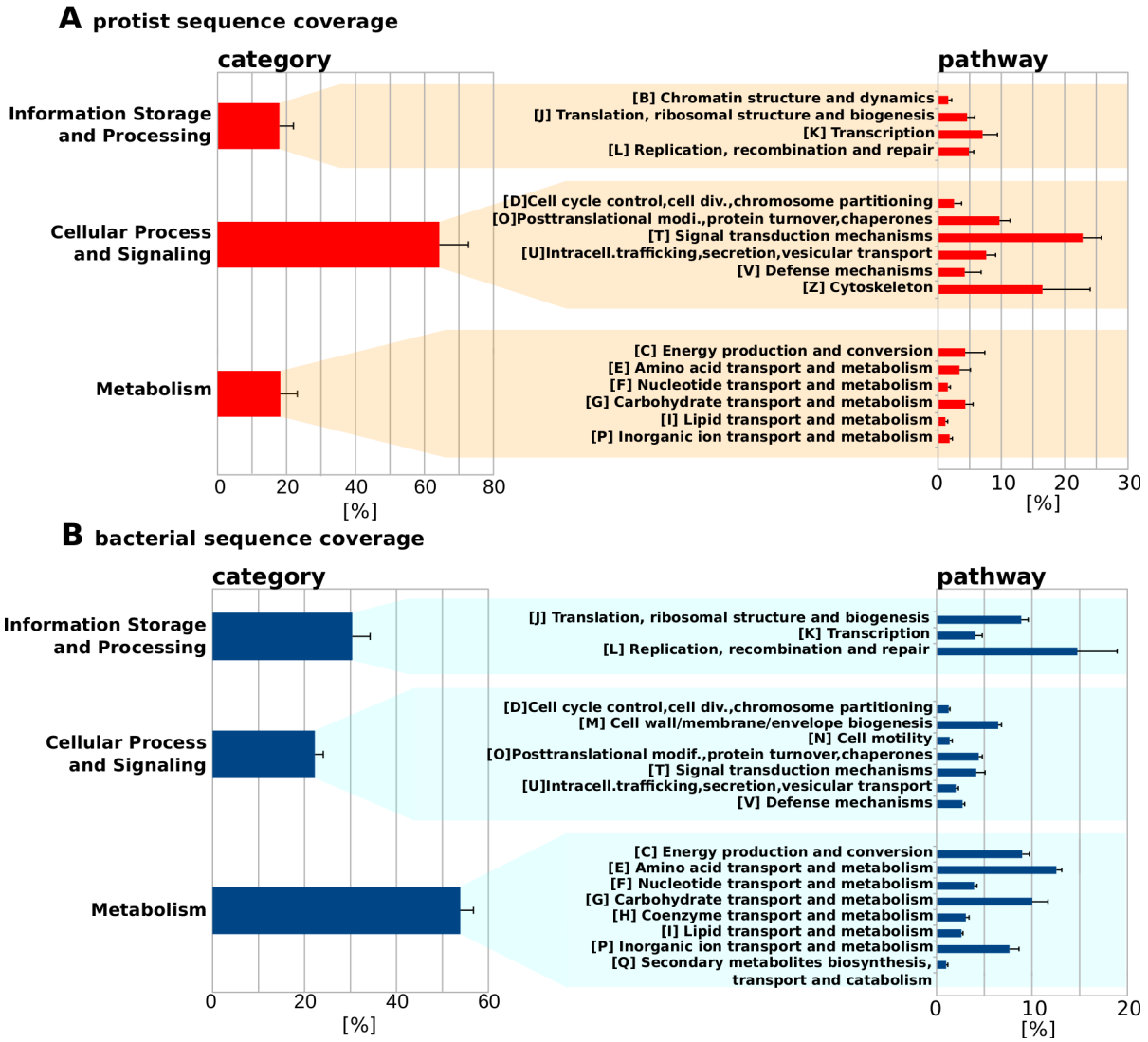


Figure 2: Metagenome sequence coverage of functional categories. A) protist sequences. B) bacterial sequences. Pathways with less than 1% coverage were removed for clarity. Background highlighting represents grouping of pathways into categories according to the eggNOG hierarchy (Figure S1). More than half (64.3%) of the protist metagenome were assigned to “cellular process and signaling”. In the bacterial metagenome sequences with annotations related to metabolic functions comprised the major group (49.7%). Error bars represent standard deviation across all samples.

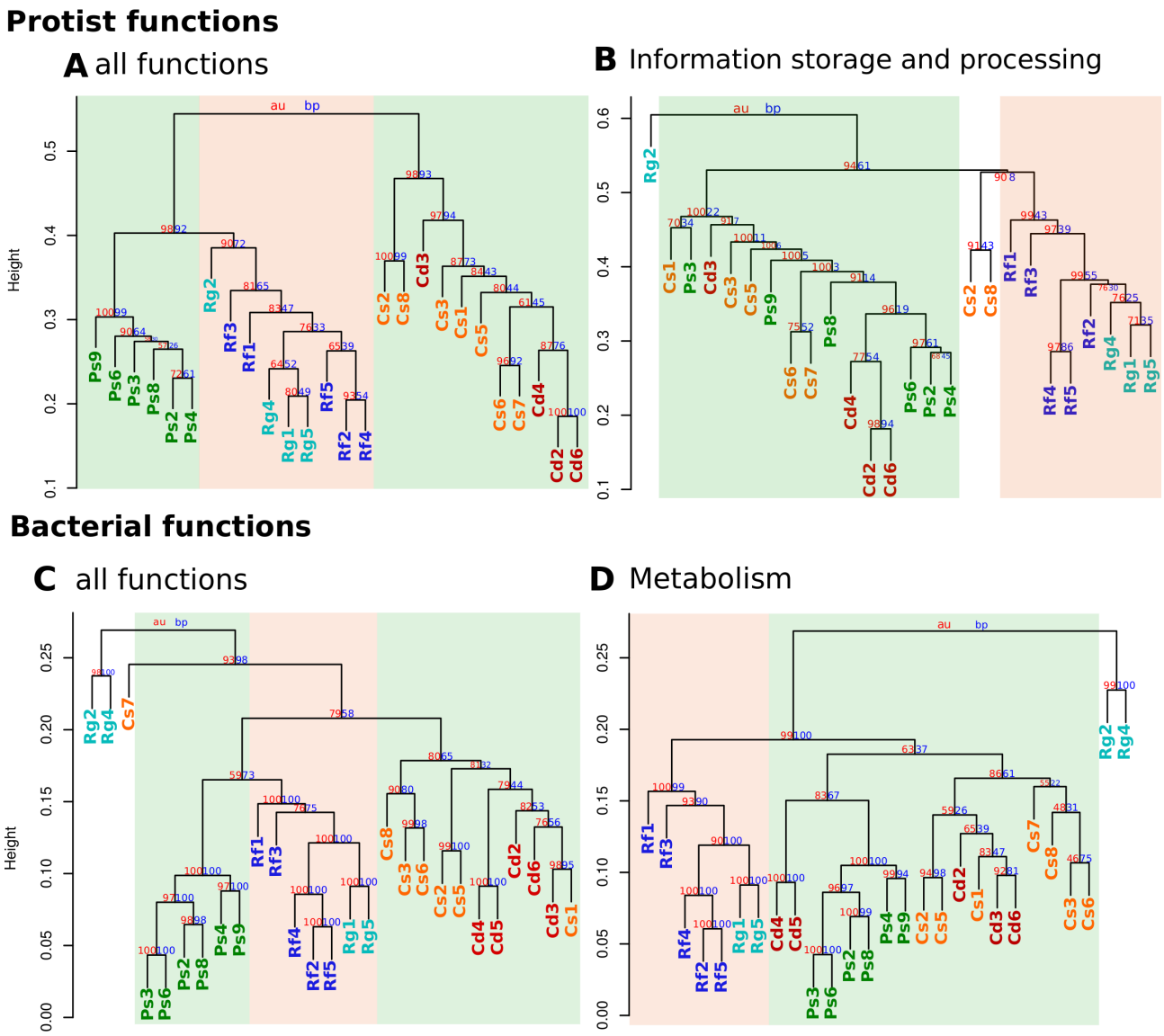
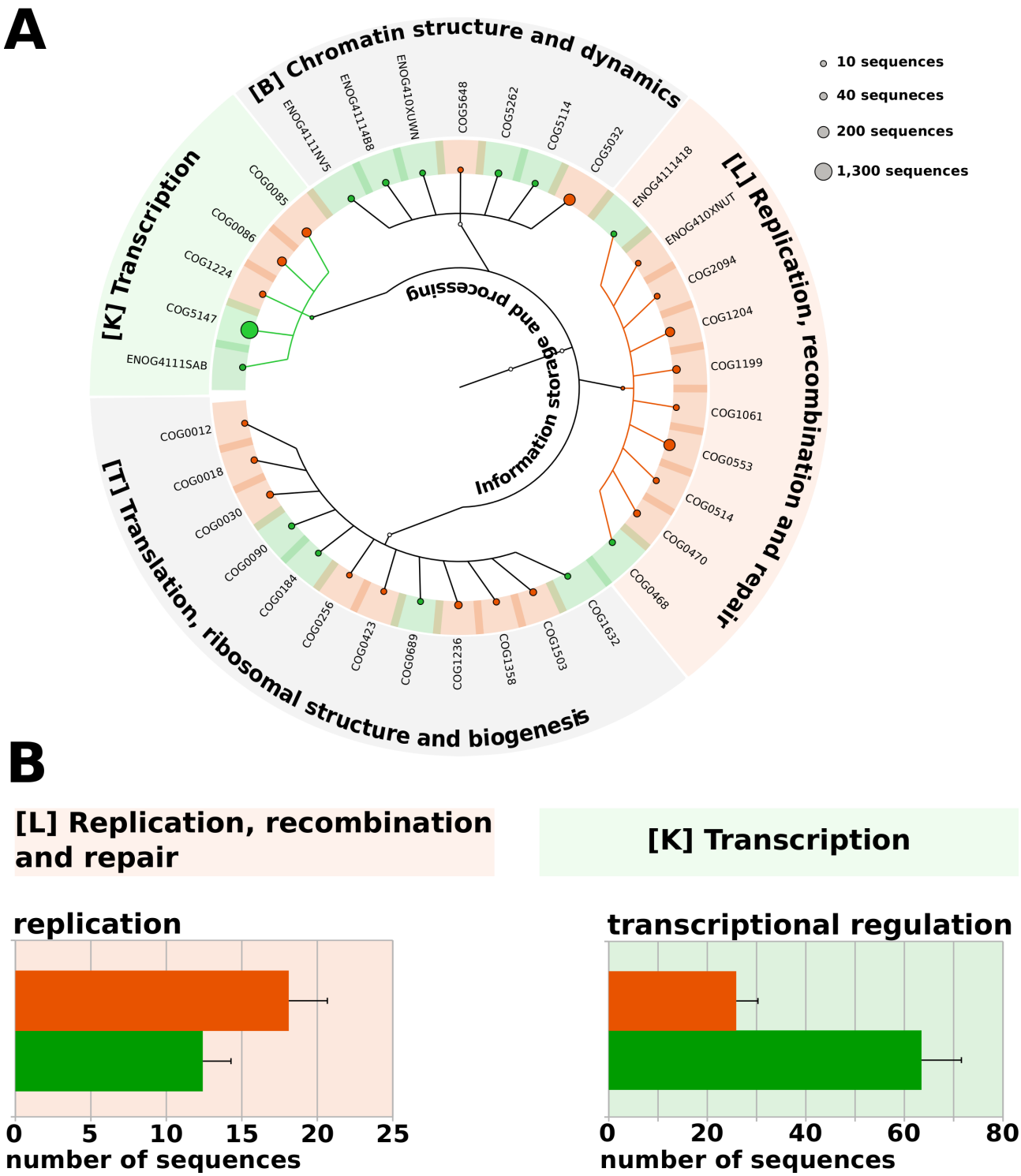


Figure 3: Cluster dendrograms of the functional profiles of the protist and bacterial community. Community distances are based on Bray-Curtis Dissimilarities of A) all functions of the protist community (25,795 sequences), B) The category "information storage and processing" of the protist community (4,527 sequences), C) all functions of the bacterial community (21,215,480 sequences), and D) the category "metabolism" of the bacterial community (10,586,058 sequences). Cd (red) = *C. domesticus* colonies; Cs (orange) = *C. secundus* colonies; Ps (green) = *P. simplex* colonies; Rf (blue) = *R. flavipes* colonies; Rg (lightblue) = *R. grassei* colonies. Green background = wood-dwelling

life types; orange background = foraging life type. For protist functions involved in "information storage and processing" in the protist community, samples clustered according to life type. Similarly, the bacterial metabolic metagenomes clustered according to life type. Cluster dendrograms of all functional categories of protist and bacterial communities can be found in Supplementary Figure S4 and S5.



672 **Figure 4: Differences in the functional content of the protist metagenomes of wood-**
673 **dwelling and foraging species.** A) Circular dendrogram/hierarchy of all over-represented
674 COGs in the category "information storage and processing" in wood-dwelling species
675 (green) or foraging species (orange). Circle size at the edges scales with abundance of the

COG. Colored branches indicate over-represented pathways. Over-representation was detected with LEfSe (Segata *et al*., 2011) ($p < 0.05$, $q < 0.05$, LDA > 2). A Venn diagram visualizing the total number, and the differentially abundant number of functions in each of the five pathways that constitute the category “information storage and processing” can be found in Supplementary Figure S6. B) Sequence coverage of wood-dwelling (green) and foraging (orange) species of examples of overrepresented COGs in the pathways within the category “information storage and processing”. Error bars represent 95% confidence intervals across replicate colonies.

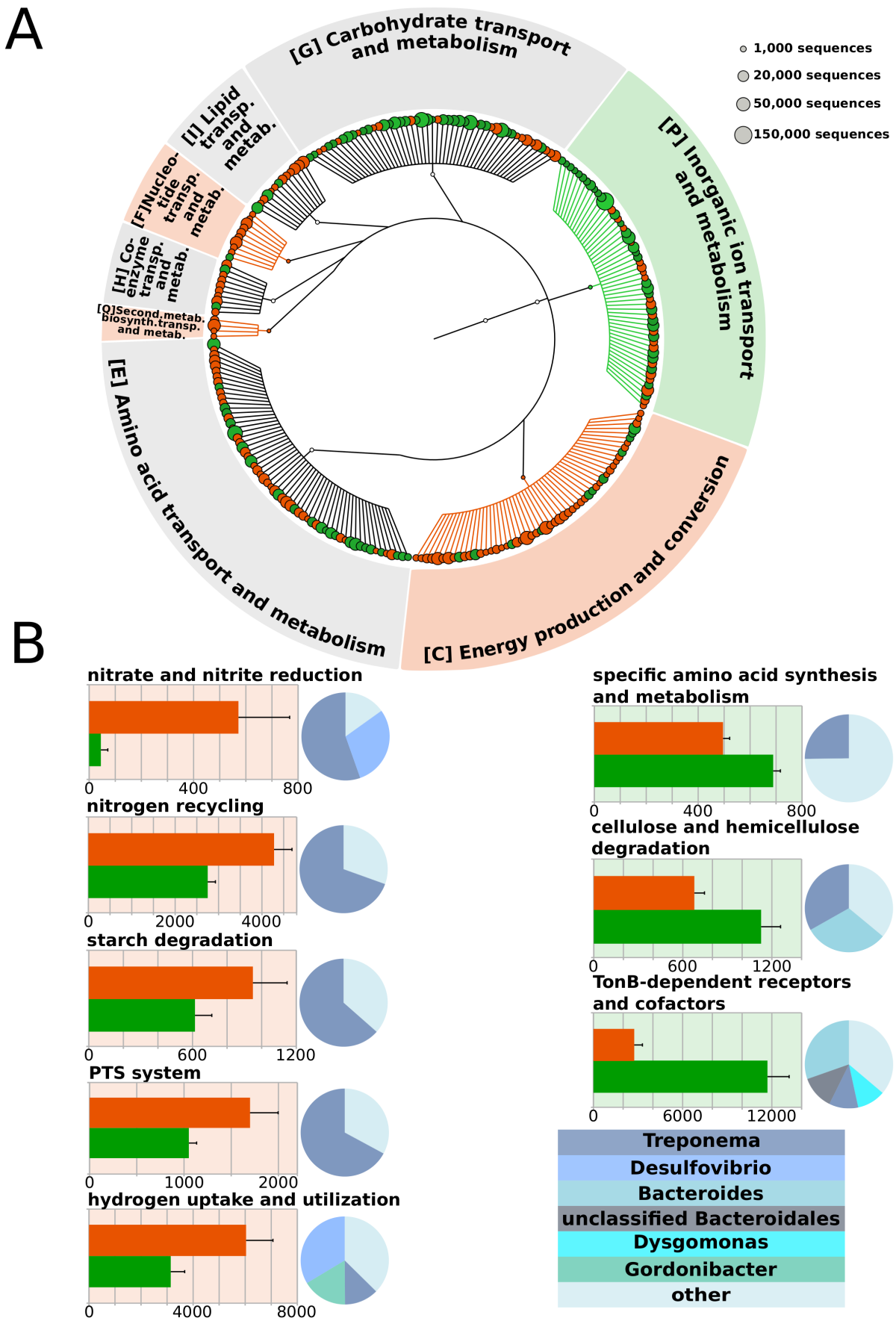

A
[G] Carbohydrate transport and metabolism
[P] Inorganic ion transport and metabolism
[C] Energy production and conversion
[E] Amino acid transport and metabolism
[O]Second.metab. biosynth.transp. and metab.
[H] Co-enzyme transp. and metab.
[F]Nucleo-tide transp. and metab.
[I] Lipid transp. and metab.
1,000 sequences
20,000 sequences
50,000 sequences
150,000 sequences
B
nitrate and nitrite reduction
0
400
800
nitrogen recycling
0
2000
4000
starch degradation
0
600
1200
PTS system
0
1000
2000
hydrogen uptake and utilization
0
4000
8000
specific amino acid synthesis and metabolism
0
400
800
cellulose and hemicellulose degradation
0
600
1200
TonB-dependent receptors and cofactors
0
6000
12000
Treponema
Desulfovibrio
Bacteroides
unclassified Bacteroidales
Dysgomonas
Gordonibacter
other

Figure 5: Differences in the functional content of the bacterial metagenomes of wood-dwelling and foraging species. A) Circular dendrogram/hierarchy of all COGs in the category “metabolism” over-represented in wood-dwelling species (green) or foraging species (orange). Circle size at the leafs scales with the abundance of the COG. Colored branches indicate over-represented pathways. Over-representation was detected with LefSe (Segata *et al*., 2011) ($p < 0.05$, $q < 0.05$, LDA > 2). Pathways “[C] energy production and conversion”, “[Q] secondary metabolites biosynthesis, transport and catabolism” and “[F] nucleotide transport and metabolism” were over-represented in foraging species. The pathway “[P] inorganic ion transport and metabolism” was over-represented in wood-dwelling species. A Venn diagram visualizing the total number, and the differentially abundant number of functions in each of the five pathways that constitute the category “metbolism” can be found in Supplementary Figure S7. B) Sequence coverage of wood-dwelling (green) and foraging (orange) species of examples of over-represented COGs in the pathways within the category “metabolism”. Error bars represent 95% confidence intervals across replicate colonies. Pie charts show taxonomic composition of bacteria found to be associated with the over-represented COGs.

Supplementary figures, tables and information

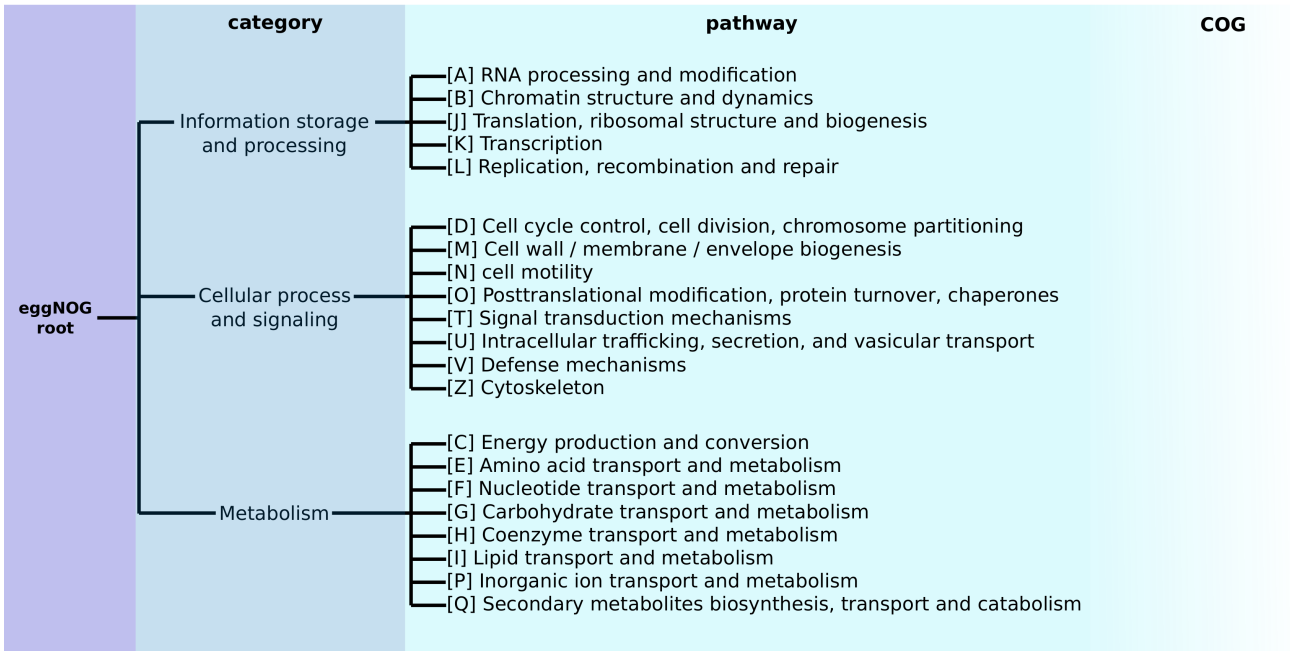


Figure S1: eggNOG annotation structure. The eggNOG annotation (Huerta-Cepas *et al.*, 2016) is divided into three categories and 21 pathways into which the COGs are grouped.

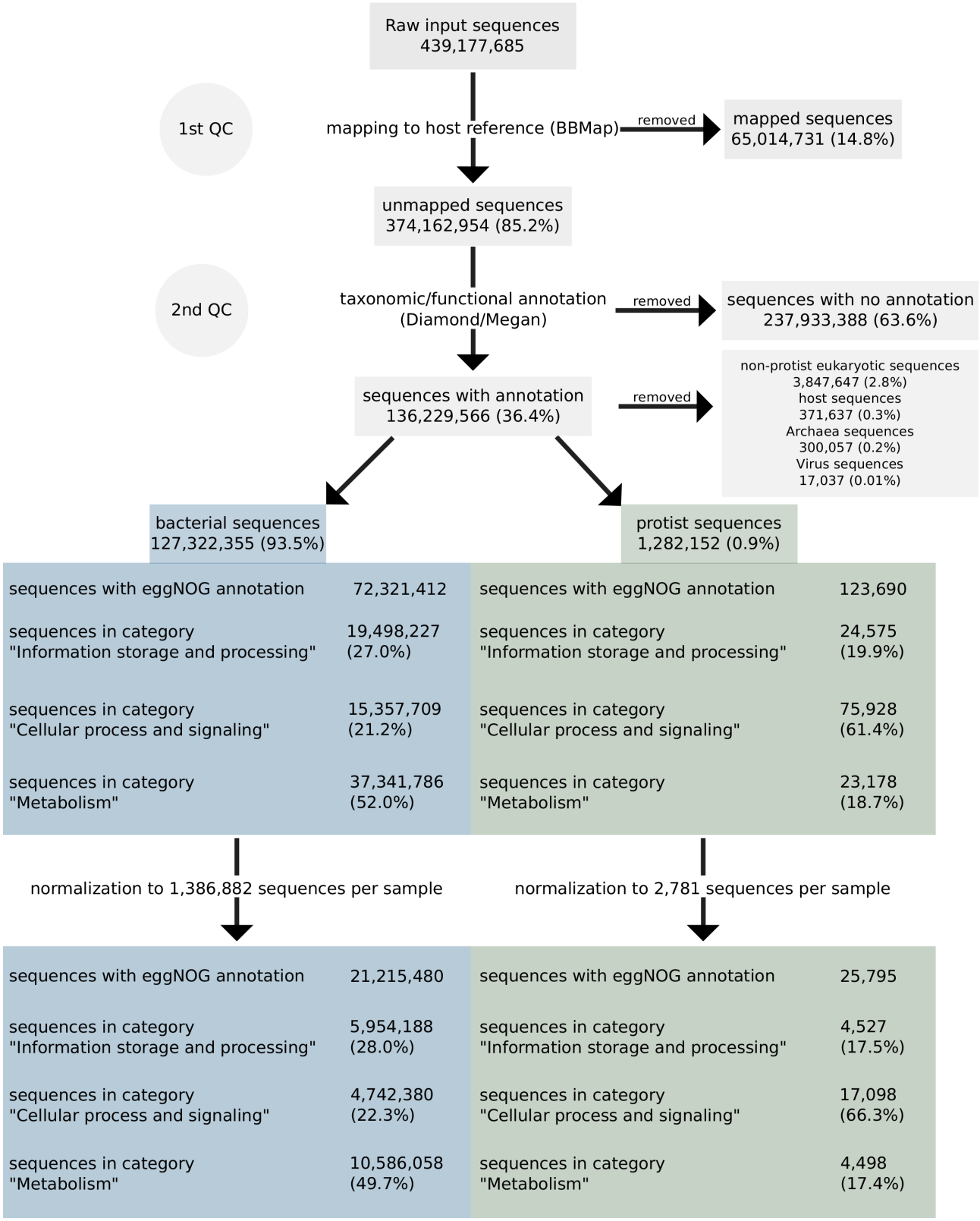

Raw input sequences
439,177,685
1st QC
mapping to host reference (BBMap)
removed
mapped sequences
65,014,731 (14.8%)
unmapped sequences
374,162,954 (85.2%)
2nd QC
taxonomic/functional annotation
(Diamond/Megan)
removed
sequences with no annotation
237,933,388 (63.6%)
sequences with annotation
136,229,566 (36.4%)
removed
non-protist eukaryotic sequences
3,847,647 (2.8%)
host sequences
371,637 (0.3%)
Archaea sequences
300,057 (0.2%)
Virus sequences
17,037 (0.01%)
bacterial sequences
127,322,355 (93.5%)
protist sequences
1,282,152 (0.9%)
sequences with eggNOG annotation 72,321,412
sequences in category "Information storage and processing" 19,498,227 (27.0%)
sequences in category "Cellular process and signaling" 15,357,709 (21.2%)
sequences in category "Metabolism" 37,341,786 (52.0%)
sequences with eggNOG annotation 123,690
sequences in category "Information storage and processing" 24,575 (19.9%)
sequences in category "Cellular process and signaling" 75,928 (61.4%)
sequences in category "Metabolism" 23,178 (18.7%)
normalization to 1,386,882 sequences per sample
normalization to 2,781 sequences per sample
sequences with eggNOG annotation 21,215,480
sequences in category "Information storage and processing" 5,954,188 (28.0%)
sequences in category "Cellular process and signaling" 4,742,380 (22.3%)
sequences in category "Metabolism" 10,586,058 (49.7%)
sequences with eggNOG annotation 25,795
sequences in category "Information storage and processing" 4,527 (17.5%)
sequences in category "Cellular process and signaling" 17,098 (66.3%)
sequences in category "Metabolism" 4,498 (17.4%)

Figure S2: Analysis workflow. Of note, sequences were not assembled, since resulting contigs were not significantly longer than individual reads, likely due to the complexity of the microbiomes (data not shown). As a first filtering step, raw sequences were mapped against a host reference, using BBMap (version 37.02, (Bushnell, BBMap - sourceforge.net/projects/bbmap/)), with standard settings. *C. secundus* and *C. domesticus* samples were mapped against the *C. secundus* (Harrison *et al*., 2018) genome. *P. simplex* samples were mapped against the *P. simplex* transcriptome (Misof *et al.*, 2014) , *R. flavipes* and *R. grassei* samples against the *R. santonensis* transcriptome (provided by the 1KITE consortium (www.1kite.org, BioSample SAMN04005235 (for details on sequencing and removal of non-termite contaminants and cross-contaminants from the 1kite data, see (Peters *et al.*, 2017)). Only sequences that did not match the host reference ("unmapped sequences") were used for further analysis. Taxonomic and functional annotation of unmapped sequences was performed using Diamond (version 0.8.37.99, (Buchfink *et al.*, 2015)) and Megan (MEGAN Community Edition (version 6.7.18), (Huson *et al.*, 2016)). As a second filtering step, only sequences with taxonomic classification "bacteria" or "parabasalia" were exported into two separate files for each sample. Therefore, contaminants such as left over host sequences, archaea, or virus sequences were removed from the datasets. In order to compare the bacterial and protist metagenomes of all samples, they were normalized to 1,386,882 and 2,781 sequences per sample, respectively. See also Supplementary file S8 below for detailed analysis steps.

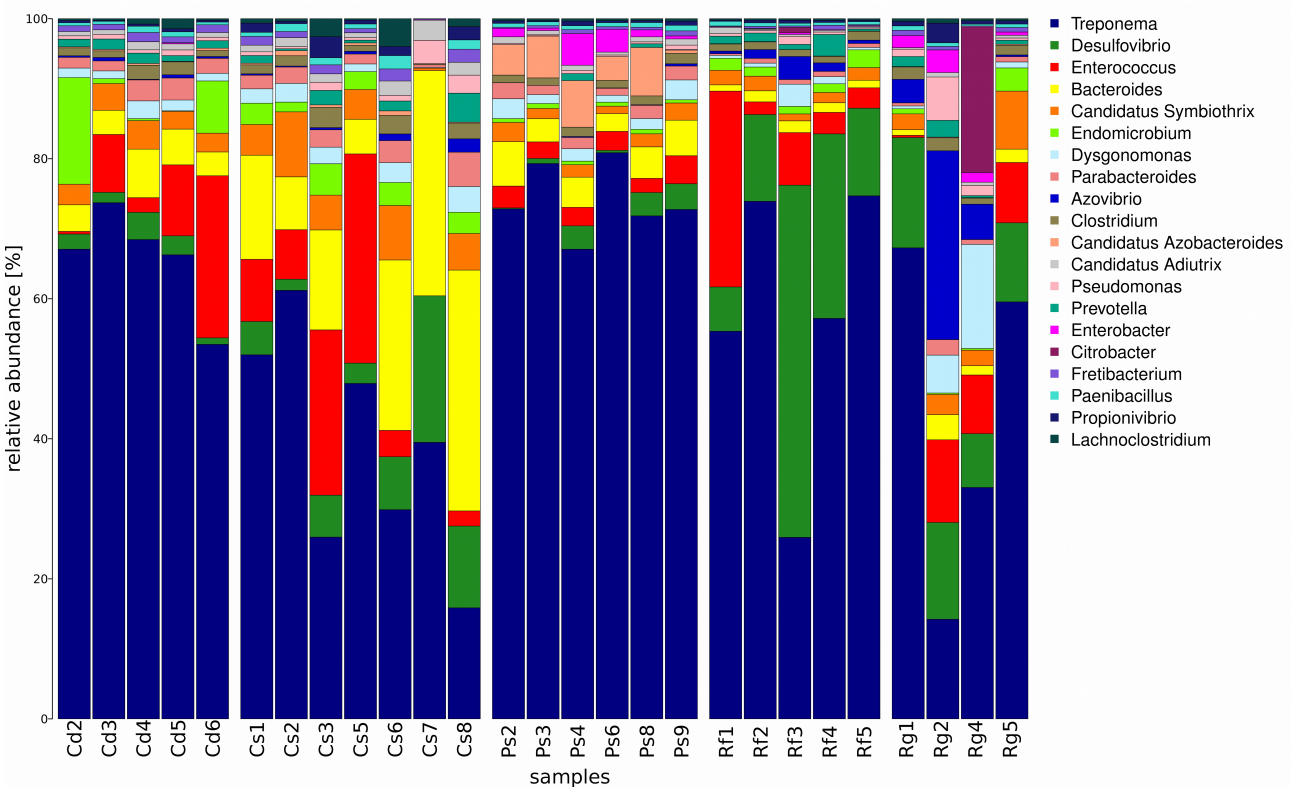
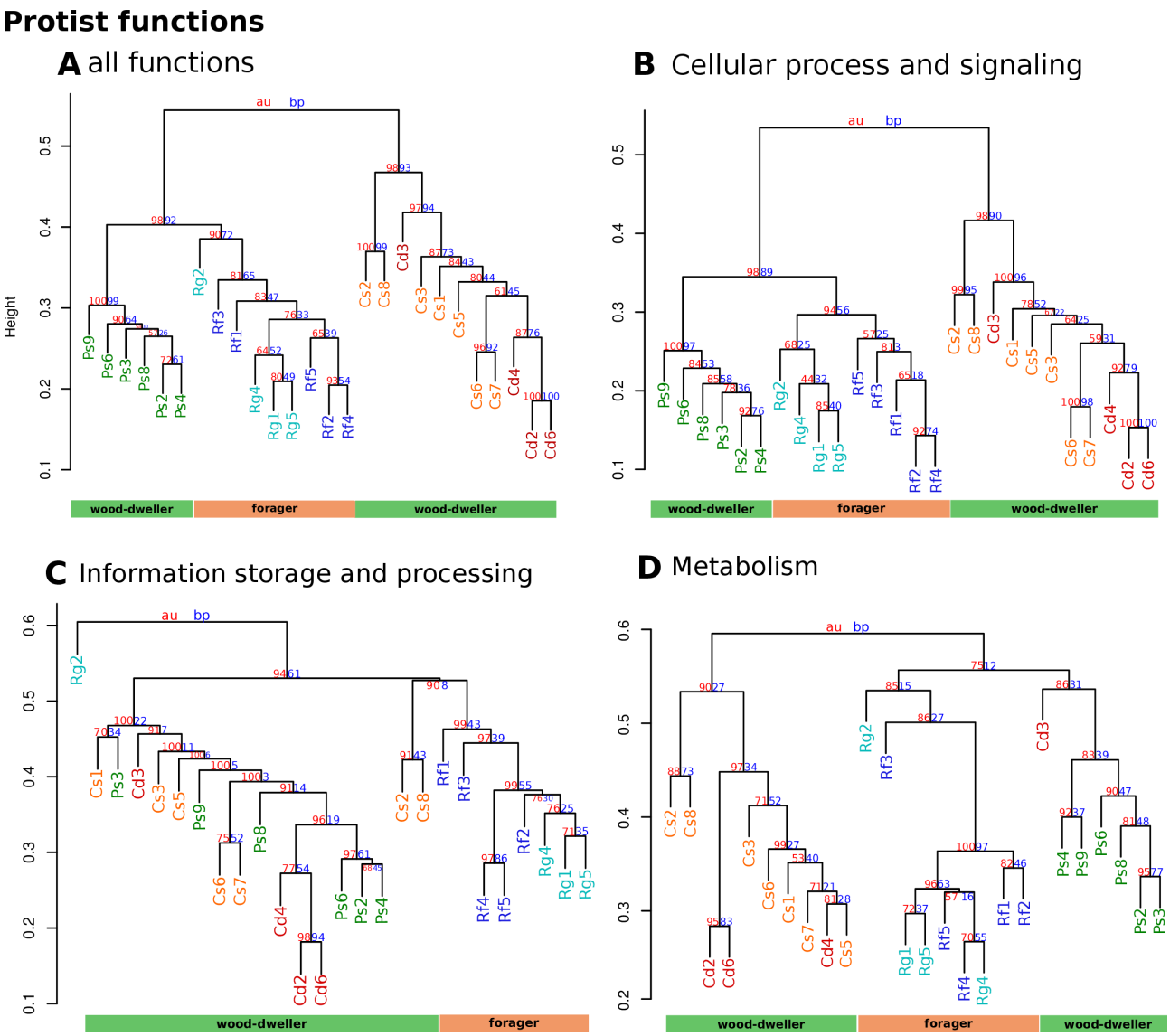


Figure S3: Frequency of bacterial genera. Cd = colony replicates of *C. domesticus*; Cs = colony replicates of *C. secundus*; Ps = colony replicates of *P. simplex*; Rf = colony replicates of *R. flavipes*; Rg and = colony replicates of *R. grassei*. Shown are only the 20 bacterial genera with the highest sequence coverage. *Treponema* was the most abundant genus (14.2% (Rg2) – 80.9% (Ps6)), followed by *Desulfovibrio* (0.1% (Ps2) - 50.3% (Rf3)), *Enterococcus* (0% (Cs7) – 29.9% (Cs5)) and *Bacteroides* (0.8% (Rf1) – 34.4% (Cs8)). Like already observed in a previous study with the same samples (Waidele *et al*., 2017), Rg2 and Rg4 showed an unusual community structure.



784 **Figure S4: Cluster dendrograms of the functional profiles of the protist community.**
785 Community distances are based on Bray-Curtis Dissimilarities. A) all functions (25,795
786 sequences). B) category “cellular process and signaling” (17,098 sequences). C) category
787 “information storage and processing” (4,527 sequences). D) category “metabolism” (4,498
788 sequences). Cd (red) *C. domesticus* colonies; Cs (orange) *C. secundus* colonies; Ps
789 (green) *P. simplex* colonies; Rf (blue) *R. flavipes* colonies; Rg (lightblue) *R. grassei*
790 colonies. All functional profiles show a strong phylogenetic imprint. However, functions
791 involved in “information storage and processing” were more similar between hosts of the
792 same ecological life type.
793

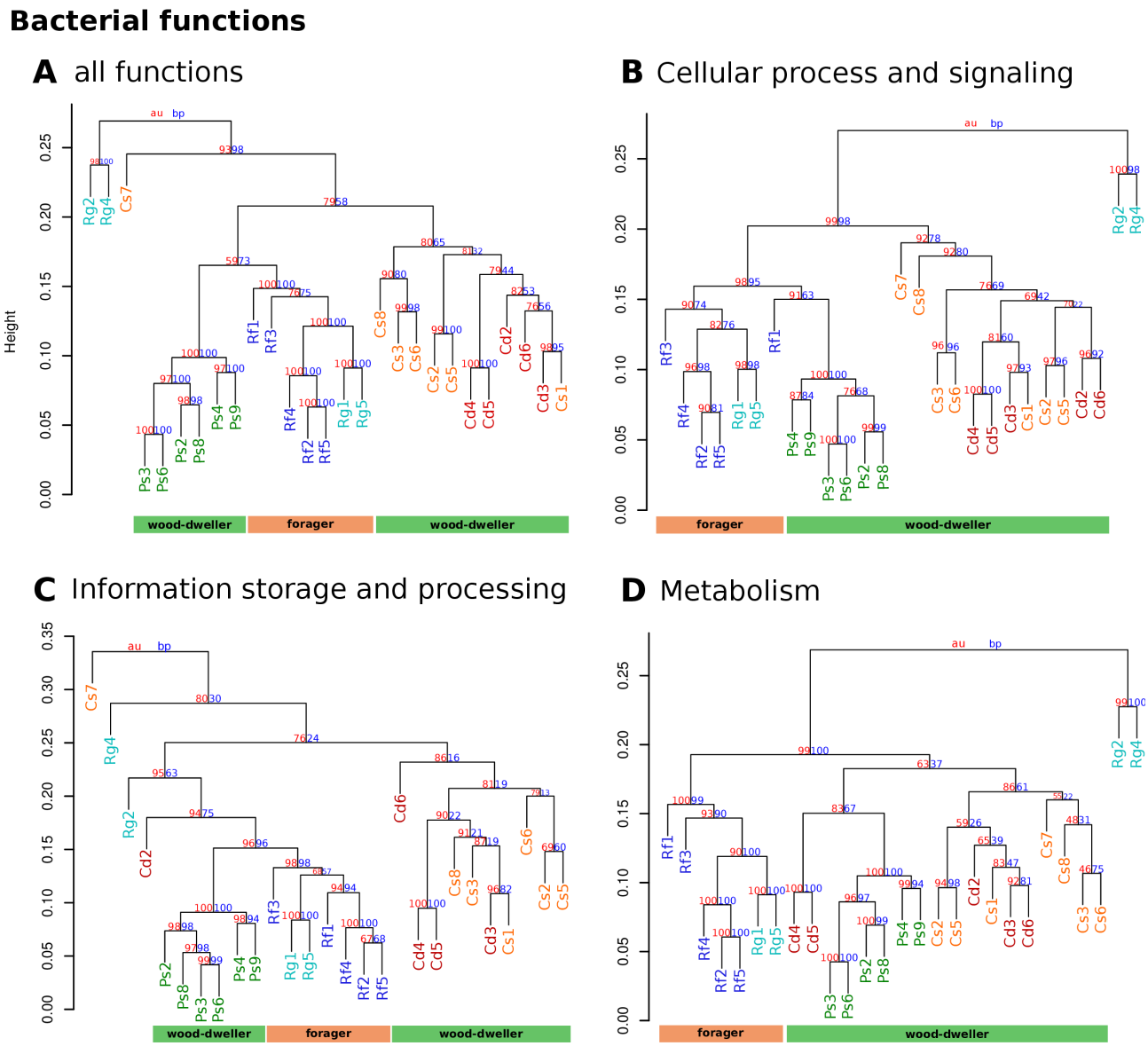


Figure S5: Cluster dendrograms of the functional profiles of the bacterial community. Community distances are based on Bray-Curtis Dissimilarities. A) all functions (21,215,480 sequences). B) category "cellular process and signaling" (4,742,380 sequences). C) category "information storage and processing" (5,954,188 sequences). D) category "metabolism" (10,586,058 sequences). Cd (red) *C. domesticus* colonies; Cs (orange) *C. secundus* colonies; Ps (green) *P. simplex* colonies; Rf (blue) *R. flavipes* colonies; Rg (lightblue) *R. grassei* colonies. Similar to the protist set, bacterial functional metagenome showed a strong phylogenetic imprint. However, the metabolic metagenomes clustered according to host ecology (life type).

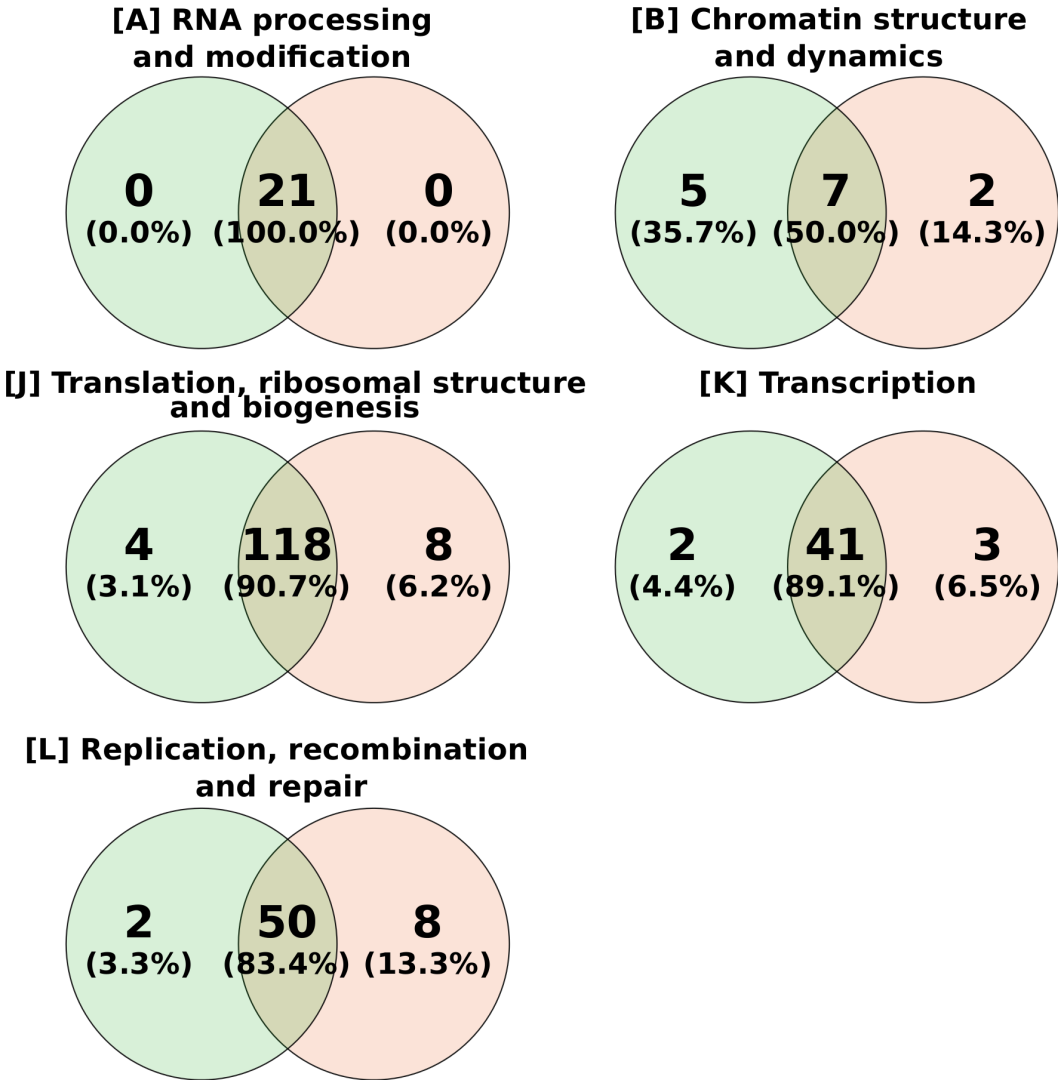


Figure S6: Venn diagram of COGs in the pathways of the eggNOG category "information storage and processing" in the protist metagenoms. Shown are numbers and percentages of over-represented and shared COGs found in the protist metagenomes of wood-dwelling (green) and foraging (orange) termite species.

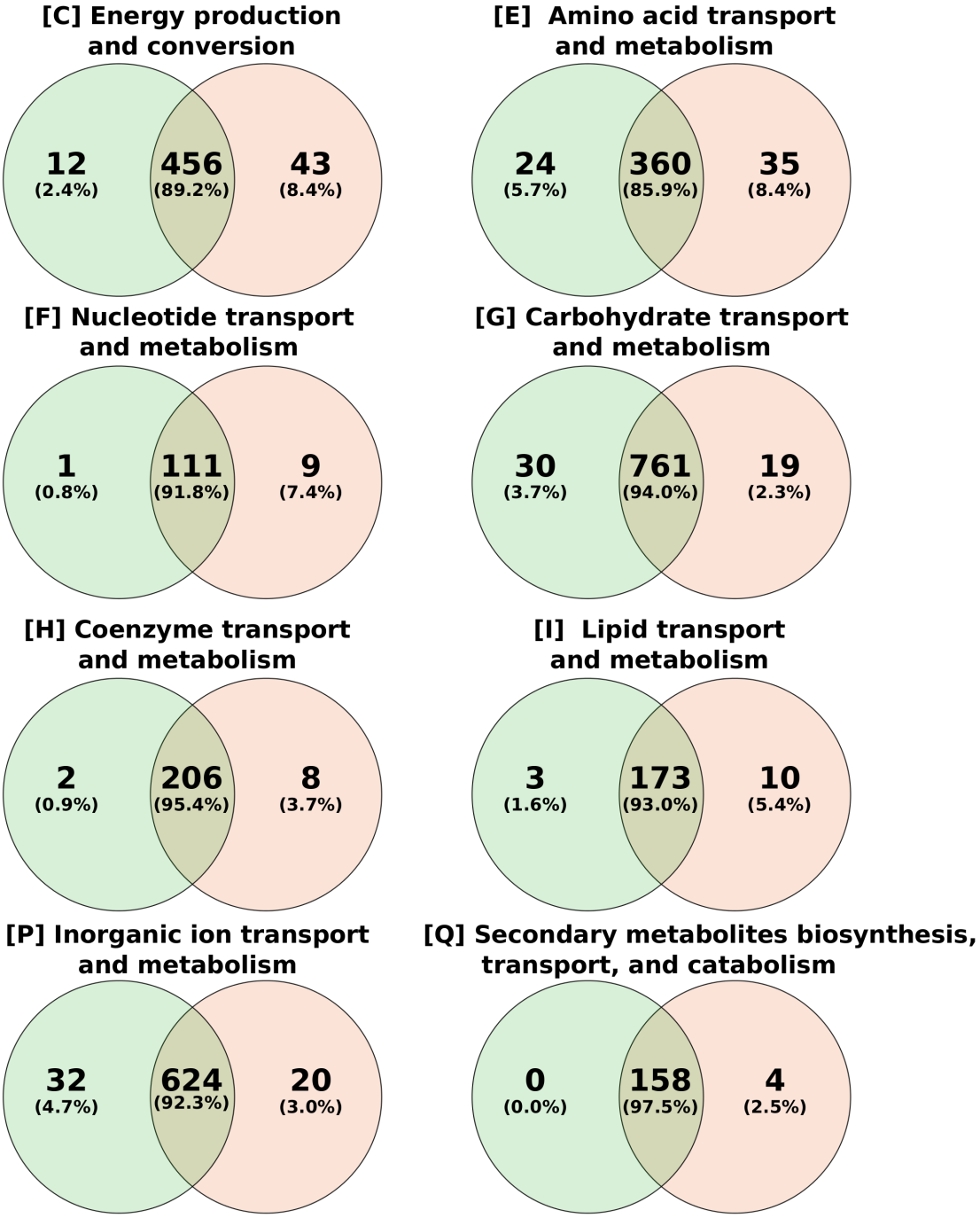


Figure S7: Venn diagram of COGs in the pathways of the eggNOG category "metabolism" in the bacterial metagenoms. Shown are numbers and percentages of over-represented and shared COGs found in the bacterial metagenomes of wood-dwelling (green) and foraging (orange) termite species.

Table S1: Sample Overview. Number of sequences before and after quality filtering, as well as taxonomic and functional annotation statistics.

Table S2: Transposase abundance in the bacterial dataset in the pathway "replication, recombination and repair". COGs annotated as transposase are marked in color.

Table S3: Complete list of results of Linear Discriminant Analysis (LDA) of protist dataset. COGs were grouped into categories and color coded based on their function.

Table S4: Complete list of results of Linear Discriminant Analysis (LDA) of bacterial dataset. COGs were grouped into categories and color coded based on their function.

Supplement file S8. Metagenomic Shotgun Data Analysis Workflow

All computing steps in this script were performed on a High Performance Computing Cluster (bwForCluster BinAC, Eberhard Karls University of Tuebingen, High Performance and Cloud Computing Group at the Zentrum fuer Datenverarbeitung of the University of Tuebingen, the state of Baden-Wuerttemberg through bwHPC and the German Research Foundation (DFG) through grant no INST 37/935-1 FUGG).

1.) Please note that the sequences were not assembled (see Material and Methods main text), but directly annotated using the following workflow.

1.) Mapping raw input sequences against reference genomes with BBMap (version 37.02, (Bushnell, BBMap - sourceforge.net/projects/bbmap/)).

Sequences of *Cryptotermes secundus* and *Cryptotermes domesticus* were mapped against the *C. secundus* genome (Harrison *et al.*, 2018) , sequences of *Prorhinotermes simplex* against the *P. simplex* transcriptome ((Misof *et al.*, 2014) NCBI Bioproject ID: 219597, Assembly version GASE02000000) and sequences of *Reticulitermes flavipes* and *Reticulitermes grassei* against the *R. santonensis* (syn. *R. flavipes* (Austin *et al.*, 2005)) transcriptome provided by the 1KITE consortium (www.1kite.org, BioSample SAMN04005235 (for details on sequencing and removal of non-termite contaminants and cross-contaminants from the 1kite data, see (Peters *et al.*, 2017).

```
>module load devel/java_jdk/1.8.0u112 #load required Java
environment
>cd path/to/directory #change to working directory
>/path/to/software/bbmap.sh
ref=reference_genome_or_transcriptome.fa
in1=Sample1_Read1.fastq.gz in2=Sample1_Read2.fastq.gz
outm1=Sample1_Read1_mapped.fastq.gz
outm2=Sample1_Read2_mapped.fastq.gz
outu1=Sample_Read1_unmapped.fastq.gz
outu2=Sample1_Read2_unmapped.fastq.gz #run BBMap with standard
settings, do for all samples
```

This will produce four output files: Sample_Read1_mapped.fastq.gz, Sample_Read1_unmapped.fastq.gz,Sample_Read2_mapped.fastq.gz, Sample_Read2_unmapped.fastq.gz. Only “unmapped” sequences, that did not match the host reference, were used for further analysis.

2.) SYNTAX for Diamond and Megan6

This script is following the Syntax tutorial as provided by the BinAC HPC. For Diamond (version 0.8.37.99, (Buchfink *et al.*, 2015)), the non-redundant NCBI database was used as reference, for functional classification with Megan6 (MEGAN Community Edition (version 6.7.18, (Huson *et al.*, 2016)), the mapping files prot_acc2tax-May2017.abin.gz,

acc2eggnog-Oct2016X.abin.gz, acc2interpro-Nov2016XX.abin and acc2seed-May2015XX.abin (download details below) were used.

```
>wget ftp://ftp.ncbi.nih.gov/blast/db/FASTA/nr.gz #download NCBI
nr.gz database
> diamond makedb --in nr.gz --db nr #build Diamond index for the
nr database, will generate a new file called nr.dmnd
```

The mapping files for Megan6 were downloaded from the Megan6 download page (http://ab.inf.uni-tuebingen.de/data/software/megan6/download/welcome.html)

```
>mkdir path/to/wanted/directory/00fastq #put unmapped BBMap output
files in here
>mkdir path/to/wanted/directory/10daa #will contain daa files
generated by Diamond
>mkdir path/to/wanted/directory/20rma #will contain rma files
generated by Megan6
```

For each file fastq.gz in the 00fastq directory, run Diamond as follows:

```
>module load bio/diamond/0.8.37 #load diamond environment
>module load bio/megan/6.7.18 #load Megan6 environment
>cd path/to/working/directory #change to working directory
>diamond blastx --query 00fastq/Sample1_Read1_unmapped.fastq.gz
--db nr.dmnd --daa 10daa/Sample1_Read1_unmapped.daa #run Diamond,
will generate a daa Diamond output file in the /10daa directory
```

For each daa file in your 10daa directory, run Megan's daa2rma as follows:

```
>module load devel/java_jdk/1.8.0u112 #load required Java
environment
>module load bio/diamond/0.8.37 #load Diamond environment
>module load bio/megan/6.7.18 #load Megan6 environment
>cd path/to/working/directory # change to working directory
>daa2rma -p -ps 1 -i 10daa/Sample1_Read1_unmapped.daa
10daa/Sample1_Read2_unmapped.daa -o 20rma/Sample1_p_unmapped.rma
-a2t prot_acc2tax-May2017.abin -a2eggnog acc2eggnog-Oct2016X.abin
-a2seed acc2seed-May2015XX.abin -a2interpro2go acc2interpro-
Nov2016XX.abin -fun EGGNOG SEED INTERPRO2GO #the -p function is
used for a paired read input and will generate a merged rma6
output file in the /20rma directory. This file can be used as
input for the Megan6 graphical user interface.
```

Further analysis using the Megan6 graphical user interface was performed locally. Each rma6 file of each sample was imported in Megan6 individually. Then bacterial and protist reads were extracted into separate files using the "Extract reads.." function in Megan6 (see also Material and Method section). To compare the bacterial and protist datasets of all

samples, they were imported into Megan6 using the “Compare...” function. During this step, the bacterial and protist datasets were normalized to 1,386,882 and 2,781 reads respectively.

3.) PvClust Cluster Dendrograms

Cluster dendrograms were generated in Rstudio (version 3.3.1, (R Core Team (2014)) using the PvClust package and an external script (https://github.com/hallamlab/mp_tutorial/blob/master/taxonomic_analysis/code/mp_tutorial_taxonomic_analysis.R) to add the Bray-Curtis dissimilarity index. Functional abundance tables were extracted from Megan6 using the “Export...” function.

In Rstudio, execute:

```
>library(pvclust)
>source('path/to/working/directory/pvclust_bcdist.R')  #source  to
downloaded R script
>setwd("/path/to/working/directory") #set working directory
>read.csv("FunctionalAbundance.csv",header=TRUE,sep=",",row.names
= 1)→dat #read in input
>clust_dat      <-      pvclust(dat,      method.hclust="average",
method.dist="bray-curtis", n=1000) #use pvclust
>plot(clust_dat,cex=0.8) #plot
>pdf(“clust_dat.pdf") #export as pdf
>plot(clust_dat,cex=0.8)
>dev.off()
```

4.) RDA and model selection

Functional abundance tables were exported from Megan6 using “Export...” > “Text (CSV) format...” > “eggnogPath_to_count”. Bacterial functional abundances were rarefied to at least 1,000 sequences per COG, protist functional abundances to at least 10 sequences per COG. Functional abundances were subsequently transformed using Hellinger transformation. RDA and ordistep/ordiR2step for model selection were performed with the

vegan package in R (Oksanen *et al.*, 2017) and compared with ANOVA using the following commands:

```
#set Nullmodel
>rda0 <- rda(functional_abundance_table ~ 1, data = metadata)
#set model with 1 explanatory variable
>rda2 <- rda(functional_abundance_table ~ host_family,data = metadata)
>rda3 <- rda(functional_abundance_table ~ host_lifetype ,data = metadata)
#set model with both explanatory variables for modelselection
>rda1 <- rda(functional_abundance_table ~ host_family + host_lifetype ,data = metadata)
#model selection via ordistep
>model_selection <- ordistep(rda0, scope = formula(rda1))
#model selection via ordiR2step
model_selection_inf <- ordiR2step(rda0_inf, scope = formula(rda1_inf), trace = TRUE, permutations = how(nperm = 499), Pin = 0.05, R2scope = FALSE)
#comparison against Nullmodel
>anova.cca(rda0,rda2)
>anova.cca(rda0,rda3)
```

Removing outlier samples from the datasets (outlier samples were Rg2, Cs2 and Cs8 in the category “information storage and processing” of the protist functional set and Rg2, Rg4 in the category “metabolism” in the bacterial set, and Rg2, Rg4 and Cs7 in all bacterial functions) did not change significant results.

5.) LEfSe

Functional abundance tables were exported from Megan6 using the “Export...” > “Text (CSV) format...” > “eggnogPath_to_count”. To use this file as input for LEfse (Segata *et al.*, 2011) , a row with the “class” “wooddweller” or “forager” was added above the sample names.

```
>python format_input.py FunctionalAbundance.csv FunctionalAbundance.lefse.in -c 1 -s -1 -u 2 -o 1000000 #will create the .in file
>python run_lefse.py FunctionalAbundance.lefse.in FunctionalAbundance.lefse.res #will create the .res file
```

6.) Circular dendrogram of LEfse results using GraPhlAn (Asnicar *et al*., 2015)

Only significant overrespresented fuctions (LDA > 2.0, q-value < 0.05) were used as input.

```
>python export2graphlan.py -i FunctionalAbundance.csv -o FunctionalAbundance.lefse.res -t tree.txt -a annot-txt --title "Functional Abundance" --external_annotations 4 --fname_row 0 --skip_rows 1
>python graphlan_annotate.py --annot test_annot.txt test_tree.txt test_outtree.txt
>python graphlan.py --dpi 150 test_outtree.txt test_outimg.png --external_legends
```

Supplement S9. Supplementary Material and Methods

<u>DNA extraction</u>

Entire guts of three workers per colony were extracted and immediately transferred to 100 µl of CTAB solution (0.75 M NaCl, 0.05 M Tris, 0.01 M EDTA, 2% CTAB). Guts were homogenized (zirconia and glass beats, 3 min at 25*1/s), using a Tissue Lyser II (Qiagen). Additional 400 µl of CTAB solution were added and samples were incubated for 1h at 65°C and 800 rpm on a Thermomixer Comfort (Eppendorf). After addition of 2 µl Proteinase K (Thermoscientific, concentration: ~20 mg/mL) samples were incubated for 2h at 55°C and 800 rpm. The proteinase K reaction was terminated by heating to 98°C for 15 min. DNA was extracted with 500 µl of chloroform:isoamylic alcohol (24:1) followed by centrifugation for 15 min at 10000 rpm. DNA, was precipitated with 325 µl of ice cold isopropanol and overnight incubation at -20°C. Samples were washed with 300 µl of 100% ethanol and twice with 300 µl 70% ethanol, centrifuging for 15 min at 4°C at 14.000 rpm between each washing step. The DNA pellet was air dried and resuspended in 50µl of water.

Supplementary Literature

Asnicar, F., Weingart, G., Tickle, T.L., Huttenhower, C., and Segata, N. (2015) Compact graphical representation of phylogenetic data and metadata with GraPhlAn. *PeerJ* **3**:.

Austin, J.W., Szalanski, A.L., Scheffrahn, R.H., Messenger, M.T., Dronnet, S., and Bagnères, A.-G. (2005) Genetic Evidence for the Synonymy of Two *Reticulitermes* Species: *Reticulitermes flavipes* and *Reticulitermes santonensis*. *Annals of the Entomological Society of America* **98**: 395–401.

Buchfink, B., Xie, C., and Huson, D.H. (2015) Fast and sensitive protein alignment using DIAMOND. *Nat Meth* **12**: 59–60.

Bushnell, B. BBMap - sourceforge.net/projects/bbmap/.

Harrison, M.C., Jongepier, E., Robertson, H.M., Arning, N., Bitard-Feildel, T., Chao, H., et al. (2018) Hemimetabolous genomes reveal molecular basis of termite eusociality. *Nature Ecology & Evolution* **2**: 557–566.

Huerta-Cepas, J., Szklarczyk, D., Forslund, K., Cook, H., Heller, D., Walter, M.C., et al. (2016) eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. *Nucleic Acids Res* **44**: D286–D293.

Huson, D.H., Beier, S., Flade, I., Górska, A., El-Hadidi, M., Mitra, S., et al. (2016) MEGAN Community Edition - Interactive Exploration and Analysis of Large-Scale Microbiome Sequencing Data. *PLOS Computational Biology* **12**: e1004957.

Oksanen, J., Guillaume Blanchet, F., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2017) vegan: Community Ecology Package.

Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., et al. (2014) Phylogenomics resolves the timing and pattern of insect evolution. *Science* **346**: 763–767.

Peters, R.S., Krogmann, L., Mayer, C., Donath, A., Gunkel, S., Meusemann, K., et al. (2017) Evolutionary History of the Hymenoptera. *Current Biology* **27**: 1013–1018.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/. (2014).

Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W.S., and Huttenhower, C. (2011) Metagenomic biomarker discovery and explanation. *Genome Biol* **12**: R60.

Waidele, L., Korb, J., Voolstra, C.R., Künzel, S., Dedeine, F., and Staubach, F. (2017) Differential Ecological Specificity of Protist and Bacterial Microbiomes across a Set of Termite Species. *Front. Microbiol.* **8**:.