



OPEN ACCESS

EDITED BY

Rafael Rivilla,
Autonomous University of Madrid, Spain

REVIEWED BY

Cristian Beza-Beza,
North Carolina State University, United States
Zaki Saati Santamaría,
University of Salamanca, Spain

*CORRESPONDENCE

Juan José González Plaza
✉ gonzalezplazajuanjose@gmail.com

SPECIALTY SECTION

This article was submitted to
Terrestrial Microbiology,
a section of the journal
Frontiers in Microbiology

RECEIVED 04 November 2022

ACCEPTED 13 February 2023

PUBLISHED 14 March 2023

CITATION

González Plaza JJ and Hradecký J (2023) The
tropical cookbook: Termite diet and
phylogenetics—Over geographical
origin—Drive the microbiome and functional
genetic structure of nests.
Front. Microbiol. 14:1089525.
doi: 10.3389/fmicb.2023.1089525

COPYRIGHT

© 2023 González Plaza and Hradecký. This is
an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

The tropical cookbook: Termite diet and phylogenetics—Over geographical origin—Drive the microbiome and functional genetic structure of nests

Juan José González Plaza* and Jaromír Hradecký

Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, Prague, Czechia

Termites are key decomposers of dead plant material involved in the organic matter recycling process in warm terrestrial ecosystems. Due to their prominent role as urban pests of timber, research efforts have been directed toward biocontrol strategies aimed to use pathogens in their nest. However, one of the most fascinating aspects of termites is their defense strategies that prevent the growth of detrimental microbiological strains in their nests. One of the controlling factors is the nest allied microbiome. Understanding how allied microbial strains protect termites from pathogen load could provide us with an enhanced repertoire for fighting antimicrobial-resistant strains or mining for genes for bioremediation purposes. However, a necessary first step is to characterize these microbial communities. To gain a deeper understanding of the termite nest microbiome, we used a multi-omics approach for dissecting the nest microbiome in a wide range of termite species. These cover several feeding habits and three geographical locations on two tropical sides of the Atlantic Ocean known to host hyper-diverse communities. Our experimental approach included untargeted volatile metabolomics, targeted evaluation of volatile naphthalene, a taxonomical profile for bacteria and fungi through amplicon sequencing, and further diving into the genetic repertoire through a metagenomic sequencing approach. Naphthalene was present in species belonging to the genera *Nasutitermes* and *Cubitermes*. We investigated the apparent differences in terms of bacterial community structure and discovered that feeding habits and phylogenetic relatedness had a greater influence than geographical location. The phylogenetic relatedness among nests' hosts influences primarily bacterial communities, while diet influences fungi. Finally, our metagenomic analysis revealed that the gene content provided both soil-feeding genera with similar functional profiles, while the wood-feeding genus showed a different one. Our results indicate that the nest functional profile is largely influenced by diet and phylogenetic relatedness, irrespective of geographical location.

KEYWORDS

metabolomics, metabarcoding, metagenomic sequencing, termite nest microbiome, termite diet, phylogenetic relatedness

1. Introduction

Termites represent one of the most interesting stories of adaptation in the evolution of the animal kingdom since they obtain their energetic resources from lignocellulose, a recalcitrant material. The use of lignocellulosic sources depends largely on the symbiotic relationships with a wide range of microorganisms (Brune, 1998; Ohkuma and Brune, 2010). Termites interact with these microorganisms for energetic purposes but are also deeply related to them because they live on their surfaces, in their nest material, or in their food substrates. Nests are the Achilles' heel of termites since their whole social structure revolves around their central place of existence. That is so because termites live in populated colonies with a series of intricate galleries where the conditions of humidity (Wiltz et al., 1998) and temperature create an ideal setting for the growth of pathogenic organisms such as fungi. Not exclusively because of these two, but also due to the high content of nutrients in termite nests in comparison with the nutrient-poor surrounding soil environment (López-Hernández, 2001; van der Sande et al., 2018), yet rich in nutrient cycling (Jordan and Herrera, 2015). Among the reasons for the high content of nutrients in nests and their immediate surrounding areas is the way in which nests are built. Nests are a product of a considerable amount of feces as a cementing tool (Jose et al., 1989; Chouvenec et al., 2018) and soil particles that are transported from the surrounding environment (Mizumoto et al., 2021). Considering these limited availabilities of food substrates in the surrounding soil, the conditions are set for a competitive race for survival in the microbial moiety, with a continuous input of external strains potentially detrimental and threatening the existence of the whole colony. These external microorganisms thrive in an environment that could limit the presence of grazing organisms (Clarholm, 1981; Rönn et al., 2002), which also has a high availability of nutrients provided by the input of termite feces when they maintain and build their nests (Chouvenec et al., 2013). However, termites are very successful in their strategies for limiting the growth of detrimental microorganisms, which is an important aspect of their ecological success. Among them, it has been established that fumigation with certain types of compounds, such as naphthalene, aids in controlling pathogens (Chen et al., 1998a; Wright et al., 2000). The role of the nest microbiome in maintaining a pathogen-free environment is not clear, though studies have addressed its potential (Chouvenec et al., 2018).

In this context, we carried out a study to better characterize the microbial communities living in termite nests among several termite species in several geographical locations. Before generalizing into how the microbiome takes part in the defense, we need to understand which factors drive communities into one composition or another. The defense strategies may vary according to the input of nutrients that they receive (termite diet), the lifestyle of the termite (genus as a product of the evolutionary history that determined the termite's interaction with the environment), or what type of microbial communities and ecological features exist (represented by the geographical origin of the termite). Our questions aim to determine the metabolite, taxonomic, and metagenomic composition. We hypothesize that the nutrient input plays an important role and that the

microbiome will show a different pattern regardless of the geographical origin.

We sampled nests from termites with two feeding strategies, namely, wood and soil, originating on both sides of the Atlantic, including rainforest and savanna representatives.

To evaluate the presence of volatile compounds, we used a coupled untargeted gas chromatography-mass spectrometry (GC-MS) analyte identification strategy, followed by a targeted naphthalene analysis. These metabolomics approaches were complemented with a taxonomic characterization of the microbial communities existing on the same material to try to understand the potential co-occurrences of certain identified metabolites with bacterial or fungal genera. Finally, we further performed shotgun metagenomic sequencing from selected termite colonies for a deeper understanding of the genomic diversity of nests and what types of genes could be involved in biocontrol strategies or, to a certain extent, in microbiological warfare.

Our results showed a much lower amount of naphthalene in termite nests than those previously reported in the literature. A very interesting observation found throughout the three omics approaches (metabolomic, amplicon sequencing, and metagenomic sequencing) is that termite genera and especially the feeding habits determine the relationships between the different microbial communities, where bacterial communities are mostly affected by host phylogenetic relatedness and fungi by diet, with a lower impact of the geographical origin for both cases.

2. Material and methods

2.1. Sample collection and initial processing

Nests from 28 termite colonies were collected during 2019 and 2020 at different sampling campaigns in South America and Africa (Table 1). The first sampling site is located in French Guiana, Petit-Saut Dam area, which is located in the Sinnamary River (5°03' N, -53°02.76' E), and is characterized by hot and humid weather throughout the year with two dry seasons (February to March and July to November) and two rainy seasons (Colas et al., 2020).

In Africa, we collected samples in Ebogo village (Cameroon, 3°23.86'N, 11°28.19'E), located in the Mbalmayo Forest Reserve. This area is characterized by a bimodal precipitation pattern (Knoben et al., 2019), with a dense evergreen forest subjected to perturbations on approximately a third of the territory due to agricultural practices (Mey and Gore, 2021). The dry periods span from December to February and July to August (Zapfack and Engwald, 2008). The second African sampling site was Nyika National Park (Malawi, -10°32.92'N, 33°53.47'E), characterized mostly by miombo woodland, and to a lesser extent, montane dambos and grasslands, with mean average temperatures of 23°C (Allingham and Harvey, 2013), and a dry season that lasts mostly from June to October, with rainfalls occur chiefly during November to March (van Velden et al., 2020). Two of the samples originate in the breeds kept at the Faculty of Forestry and Wood Sciences (Czech University of Life Sciences Prague, Czech Republic) at humid constant temperatures of 28°C.

Nest samples were processed in clean, hexane-wiped metal trays under a laminar flow hood to avoid contamination. Large

TABLE 1 Species identification after COII method.

Genus	Species	Feeding habits	Origin	Termite feeding group	Code	Sampling campaign	Accession (16S)	Accession (ITS)	Accession (shotgun)
<i>Coptotermes</i>	<i>formosanus</i>	Wood	BR	I	Copfor BR I	July 2019	ERX9702657	ERX9703253	
<i>Nasutitermes</i>	sp.	Wood	BR	II	Nassp BR II	July 2019	ERX9702672	ERX9703268	
<i>Nasutitermes</i>	<i>lujae</i>	Wood	CM	II	Naslu 1 CM II	July 2019	ERX9702667	ERX9703263	ERX9702718
<i>Nasutitermes</i>	<i>lujae</i>	Wood	CM	II	Naslu 2 CM II	July 2019	ERX9702668	ERX9703264	ERX9702719
<i>asutitermes</i>	<i>lujae</i>	Wood	CM	II	Naslu 3 CM II	July 2019	ERX9702669	ERX9703265	ERX9702720
<i>Nasutitermes</i>	sp.	Wood	CM	II	Nassp CM II	July 2019	ERX9702673	ERX9703269	ERX9702721
<i>Microcerotermes</i>	sp.	Wood	CM	II	Micsp CM II	January 2020	ERX9702666	ERX9703262	
<i>Cephalotermes</i>	<i>rectangularis</i>	Wood	CM	II	Cefre CM II	July 2019	ERX9702655	ERX9703251	
<i>Constrictotermes</i>	<i>cavifrons</i>	Lichens	FG	II	Conca FG II	January 2020	ERX9702656	ERX9703252	
<i>Nasutitermes</i>	<i>similis</i>	Wood	FG	II	Nassi 1 FG II	May 2019	ERX9702670	ERX9703266	
<i>Nasutitermes</i>	<i>similis</i>	Wood	FG	II	Nassi 2 FG II	May 2019	ERX9702671	ERX9703267	
<i>Anoplotermes</i>	<i>banksii</i>	Soil	FG	III	Anba 1 FG III	May 2019	ERX9703247	ERX9702651	
<i>Anoplotermes</i>	<i>banksii</i>	Soil	FG	III	Anba 2 FG III	May 2019	ERX9702652	ERX9703248	ERX9702711
<i>Anoplotermes</i>	<i>banksii</i>	Soil	FG	III	Anba 3 FG III	May 2019	ERX9702653	ERX9703249	ERX9702712
<i>Anoplotermes</i>	<i>banksii</i>	Soil	FG	III	Anba 4 FG III	May 2019	ERX9702654	ERX9703250	ERX9702713
<i>Embiratermes</i>	<i>neotenicus</i>	Soil	FG	III	Embne FG III	January 2020	ERX9702664	ERX9703260	
<i>Neocapritermes</i>	<i>taracua</i>	Soil	FG	III	Neota 1 FG III	January 2019	ERX9702675	ERX9703271	
<i>Neocapritermes</i>	<i>taracua</i>	Soil	FG	III	Neota 2 FG III	May 2019	ERX9702676	ERX9703272	
<i>Neocapritermes</i>	<i>taracua</i>	Soil	FG	III	Neota 3 FG III	May 2019	ERX9702677	ERX9703273	
<i>Silvestritermes</i>	<i>heyeri</i>	Soil	FG	III	Silvhe FG III	January 2020	ERX9702678	ERX9703274	
<i>Termes</i>	<i>fatalis</i>	Soil	FG	III	Terfa FG III	January 2020	ERX9702679	ERX9703275	
<i>Cubitermes</i>	<i>planifrons</i>	Soil	CM	IV	Cubpl CM IV	February 2020	ERX9702661	ERX9703257	ERX9702714
<i>Cubitermes</i>	<i>sulcifrons</i>	Soil	CM	IV	Cubsu CM IV	February 2020	ERX9702662	ERX9703258	ERX9702715
<i>Labiotermes</i>	<i>labralis</i>	Soil	FG	IV	Labla FG IV	January 2020	ERX9702665	ERX9703261	
<i>Cubitermes</i>	<i>aff. ugandensis</i>	Soil	MW	IV	Cubug MW IV	July 2019	ERX9702663	ERX9703259	
<i>Cubitermes</i>	<i>inclitus</i>	Soil	MW	IV	Cubin MW IV	July 2019	ERX9702658	ERX9703254	ERX9702716
<i>Cubitermes</i>	<i>muneris</i>	Soil	MW	IV	Cubmu 1 MW IV	July 2019	ERX9702659	ERX9703255	ERX9702717
<i>Cubitermes</i>	<i>muneris</i>	Soil	MW	IV	Cubmu 2 MW IV	July 2019	ERX9702660	ERX9703256	

The feeding habits of termites, their geographical origin, and classification according to the termite-feeding group are shown as well. For metagenomic data analysis after sequencing, we have used the generic Ban1, Ban2, and Ban3 for simplicity purposes and have the following correspondence: Ban1 = Anba 2 FG III, Ban2 = Anba 3 FG III, and Ban3 = Anba 4 FG III. Breed samples were collected in our termite collection at the Czech University of Life Sciences, Prague, Czech Republic.

N/A, not applicable; BR, breeds; CM, Cameroon; FG, French Guiana; MW, Malawi. The sequencing data have been deposited in the ENA archive under project accession PRJEB55779.

pieces of nests devoid of termites were broken into small fragments and stored at -80°C until processing. Samples were further homogenized in a Retsch oscillatory mill. The hexane-wiped steel 10-ml chambers were filled with nest material, sealed, and submerged in liquid nitrogen until stabilized. Samples contained in steel chambers were then homogenized to a fine powder for 5 min at 30 Hz frequency, recovered, and stored at -80°C until further analysis.

2.2. Metabolite profiling

For initial headspace solid-phase microextraction (HS-SPME) volatiles profiling, 200 mg of previously homogenized samples were placed into different 10-ml screw-top vials for headspace analysis and were sealed with a magnetic cap. Sample incubation and the following metabolite extraction were performed at 50°C in the agitator of the Gerstel MPS2 autosampler (Gerstel, SUI) for 10 and 30 min, respectively. For separation and detection, a gas chromatograph coupled with a time-of-flight mass spectrometer (GC-TOF-MS; Leco Pegasus 4D, Leco, USA) was used. The temperature-programmed injector was operated in splitless mode at 275°C . For separation, a 30-m (0.25 mm i.d., 0.25 μm film thickness) Rxi-5Sil MS (Restec, USA) column was used. The temperature program for the 1D oven was as follows: 40°C for 1 min, then ramped at a rate of $10^{\circ}\text{C}/\text{min}$ to 210°C , and finally at $20^{\circ}\text{C}/\text{min}$ to 300°C with a hold time of 3 min. The total GC run time was 26 min. The mass spectrometer was operated in the mass range of 35–500 m/z with an acquisition speed of 10 Hz.

Peak find, mass spectral deconvolution, and peak alignment were performed in ChromaTOF software (LECO, USA). Compounds with an automatically selected quantification mass and a signal-to-noise ratio (S/N) higher than 50 were selected for alignment. The mass spectral similarity of signals in different samples, to be considered as the same compound, must be higher than 60% and the retention time difference must be lower than 5 s. A tentative identification based on the spectral similarity of the deconvoluted spectrum with spectra stored in the NIST mass spectral library (NIST17) was provided for each aligned compound. Before statistical analysis, data were normalized using constant row sum (each signal quantification mass area in one sample is divided by the sum of all signal quantification areas in the respective sample) and then transformed using the centered log-ratio (clr) transformation. All samples were analyzed using principal component analysis (PCA) to find the separation and grouping of samples, while only samples with more than three representatives per genus were processed *via* partial least squares discriminant analysis (PLS-DA). Based on the variable importance plot (VIP) in PLS-DA models, interesting compounds underwent further confirmation of identity *via* comparison of the calculated retention index with retention indices from the NIST lib.

For targeted naphthalene analysis, an SPME method based on Tsimeli et al. (2008) and Cao (2012) was used, employing standard addition quantification. In total, 100 mg of homogenized sample were placed into a 10-ml HS vial, and 0.5-ml of water (UHPLC-MS purity, Supelco, USA) was added before sealing with a magnetic screw cap. Four vials per sample were prepared. A volume of 10 μl

of methanol (HPLC Plus purity, Sigma-Aldrich, USA) was injected through the septum into the first two vials, while 10 μl of methanol containing different amounts of naphthalene was injected into the remaining two vials. The sample was heated under agitation at 50°C for 10 min, and then volatiles from the headspace were collected onto an SPME fiber for 5 min.

A linear response of naphthalene was observed when a blank nest sample was spiked at levels between 1 and 1,100 ng/g. Method accuracy and reproducibility were checked by measuring a certified reference material (CRM) of polyaromatic hydrocarbons in soil (CRM170, Lot: LRAC8900, Sigma-Aldrich, USA). Measuring six replications of CRM, the relative standard deviation (RSD) of the entire determination was 12%, resulting in an uncertainty ($U = 2 \times \text{RSD}$) of 25%. Comparing obtained naphthalene content of $563 \text{ ng/g} \pm 140 \text{ ng/g}$ (average $\pm U$) to the certified value of $573 \pm 45 \text{ ng/g}$ showed good accuracy of the presented method. The only modification for CRM measurement was that due to the high naphthalene concentration, the sample weight was reduced 10-fold to 10 mg per vial.

2.3. Species identification

DNA from five termite workers belonging to each colony was isolated using the Qiagen Blood and Tissue Kit according to the manufacturer's instructions. Isolated DNA was used for amplification of the cytochrome oxidase II gene (COII) with the following program: 1 min at 94°C , 30 cycles of 94°C for 15 s, 61°C for 1 min, 72°C for 35 s, and a final step of elongation at 72°C for 10 min. PCR products were submitted to Microsynth AG (Switzerland) for Sanger sequencing. Results were compared to the NCBI nr database for retrieving BLAST results (Altschul et al., 1990), where the result with the highest identity percentage was selected. The sequencing primers were previously described by Benjamino and Graf (2016).

2.4. Microbial DNA isolation

Powder-homogenized samples were used for the isolation of DNA. A quantity of 100 mg of soil was used as starting material using the NucleoSpin Soil DNA Kit (Macherey-Nagel, Germany). The whole procedure was carried out under laminar flow using sterile material, and kit components were exclusively opened and manipulated under sterile conditions. Samples were quantified by a Qubit 2.0 Fluorometer (Thermo Scientific, USA), and quality was checked through spectrophotometric methods.

2.5. Sequencing

All sequencing was carried out following standard company procedures (Novogene, Hong Kong, China). Amplicon sequencing was carried out on all samples, while shotgun metagenomic sequencing was performed on selected samples (both indicated in Table 1).

2.5.1. 16S rRNA gene sequencing

Primers 515F (GTGCCAGCMGCCGCGGTAA) and 806R (GGACTACHVGGGTWTCTAAT) were used to sequence the 16S V4 region. In brief, a 292-bp amplicon was obtained through PCR with the above-mentioned specific primers. PCR products were quantified, purified (Qiagen Gel Extraction Kit, Qiagen, Germany), and used for library preparation (NEBNext Ultra DNA Library Pre-Kit, Illumina, San Diego, CA, USA). Libraries were sequenced following a paired-end 250-bp sequencing strategy on NovaSeq 6000 Illumina instruments.

2.5.2. Internal transcribed spacer (ITS) sequencing

In brief, 1 ng of DNA was used to amplify fungal ITS genes from the ITS2 region with a specific set of primers, namely, ITS3 (GCATCGATGAAGAACGCAGC) and ITS4 (TCCTCCGCTTATTGATATGC), with an amplicon size of 386 bp. The overall procedure with amplicons follows the one explained for 16S.

2.5.3. Metagenomic sequencing

We selected two genera, comprising four samples each, from those where naphthalene was produced (*Cubitermes* and *Nasutitermes*). For comparison purposes, a third group with no naphthalene presence in nests was selected (*Anoplotermes*). After quality control in standard agarose gels and quantity concentration estimation with Qubit 2.0 (ThermoFisher, USA), DNA was randomly sheared using sonication. Fragments (150 bp) were end-polished, A-tailed, and then the Illumina adapters (5' adapter: 5'-AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT-3'; 3' adapter: 5'-GATCGGAAGAGCACACGTCTGAACTCCAGTCAC-3') were ligated. PCR amplification was carried out using P5- and P7-indexed oligonucleotides and products were purified with the AMPure XP system. The size distribution of libraries was evaluated using the Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA) and quantified using real-time PCR for achieving equimolar concentrations. Sequencing (paired-end) was subsequently carried out on NovaSeq 6000 Illumina instruments.

2.6. Microbiome data processing and analysis

2.6.1. Amplicon sequencing data: 16S and ITS

The assignment of paired-end reads to samples was based on their unique barcodes, which were later truncated by cutting off the barcode and primer sequence. Overlaps of paired-end reads served for merging reads using FLASH (Magoč and Salzberg, 2011). Reads were quality filtered to obtain high-quality clean tags (Bokulich et al., 2013) using QIIME (V1.7.0). Reads were then compared to a reference database (Gold DB, http://drive5.com/uchime/uchime_download.html) with the UCHIME algorithm (Edgar et al., 2011) for the detection and removal of chimera sequences. Clean reads were stored in individual files in fastq format. These files were used as input in a common tool used in microbiome studies for the bioinformatic treatment of clean, demultiplexed reads, QIIME2

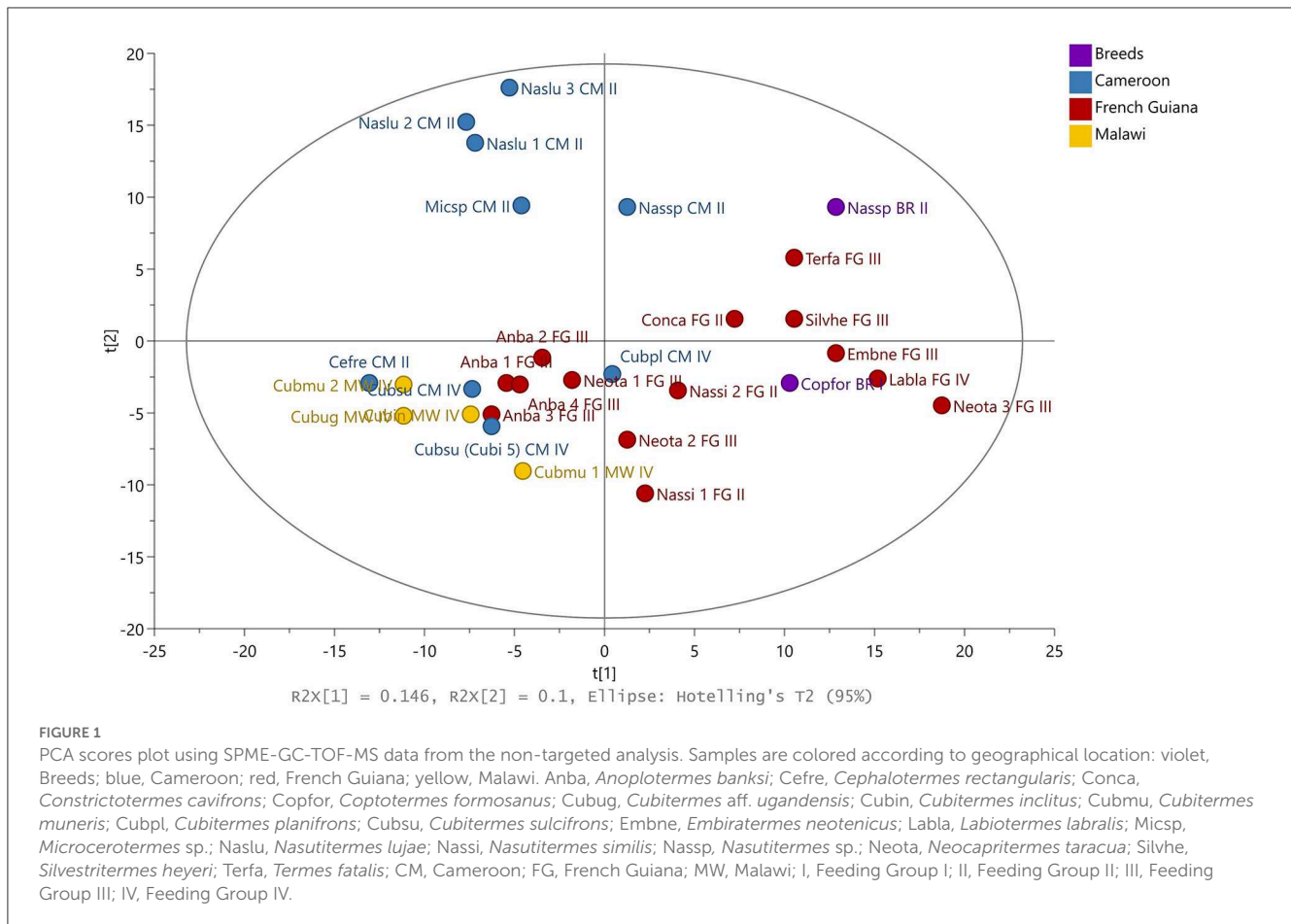
(release 2021.4; Bolyen et al., 2019). This wrap-up tool was used to import (q2-import, SingleEndFastqManifestPhred33V2) the clean demultiplexed single fastq files corresponding to our samples with the import plugin. Denoising was carried out with DADA2 (via q2-dada2-denoise single; Callahan et al., 2016), which allows for the identification of all observed amplicon sequence variants. ASVs were aligned using mafft (Katoh et al., 2002; via q2-alignment) and used to build a phylogeny (fasttree2), via q2-phylogeny (Price et al., 2010).

After DADA2, sequences that were found in the negative control ".fastq" file were removed from the project through quality-control exclude-seqs plugins.

Core metrics were calculated with the core-metrics-phylogenetic plugin, where rarefaction was performed using a sampling depth of 10,000 for 16S data. Among metrics, we calculated Faith's phylogenetic diversity (Faith, 1992), beta diversity such as UniFrac (Lozupone et al., 2007), unweighted UniFrac (Lozupone and Knight, 2005), Jaccard distance, and Bray-Curtis dissimilarity. We analyzed the microbial communities in the nests with a quick overview of the bacterial community compositions through the calculation of the Bray-Curtis dissimilarity metric (Sorensen, 1948; Bray and Curtis, 1957). This index provides an idea of similarity/dissimilarity due to the product of present taxa and their relative abundances (van Rensburg et al., 2015). The ACE index of α -diversity provides an overview of the microbial richness within one sample, without considering the ASVs abundance (Qiao et al., 2017). At β -diversity, the weighted UniFrac distance can represent the differences between samples according to the evolutionary history (Lozupone and Knight, 2005; Lozupone et al., 2007, 2011), and how different are two microbial communities. Plugins for 16S amplicon data were core-metrics-phylogenetic, alpha-group-significance, and beta-group-significance, where the metadata column of choice was "geography," "genus," or "feeding-group." We calculated non-phylogenetic "ace" diversity for ITS data with the "diversity alpha" plugin and a beta diversity through "diversity beta" ("Bray-Curtis"). Significance was tested through the "diversity beta-group-significance" (ANOSIM test).

Taxonomy was assigned with the q2-feature classifier classifysklearn naïve Bayes taxonomy classifier (Bokulich et al., 2018; Kaehler et al., 2019); for the 16S, we used the weighted Greengenes 13_8 99% OTUs full-length sequences (DeSantis et al., 2006), and for the ITS, we used UNITE (<https://doi.plutof.ut.ee/doi/10.15156/BIO/2483915>). Taxonomy plots were drawn in R through the ampvis2 package, using the tutorials available at <https://sites.google.com/a/ciad.mx/bioinformatica/home/metagenomica/visualizacion/ampvis2> and <https://madsalbertsen.github.io/ampvis2/articles/ampvis2.html>, where we used the Heillinger transformation in order to represent the PCA (Legendre and Gallagher, 2001). We used the ASV annotated data in R through the ampvis2 package to create PCA plots.

Once taxonomy was assigned and frequency was calculated per sample, we performed a least discriminant analysis (LDA) effect size (LEfSe; Segata et al., 2011). This algorithm discovers biomarkers that identify features of the tested biological conditions. A text-tab-separated file in metadata-columnar format was loaded at a Galaxy instance (<https://huttenhower.sph.harvard.edu/galaxy/>). In the case of 16S data, the OTU table was filtered to include only those that satisfied frequency > 0.001 in at least one of the samples.



2.6.2. Metagenomics bioinformatics

Sequenced reads were first cleaned using the Trimmomatic (Bolger et al., 2014) and then subjected to a quality control process (FASTX Toolkit 0.0.14. FASTQ/A short-reads pre-processing tools). The Kalamazoo Protocol for metagenome assembly, khmer 0.8.4 (Brown et al., 2013), was then used. We used the digital normalization protocol as stated (Brown et al., 2012, 2015) and used the normalized reads as input for SqueezeMeta v1.4.0, v. May 2021 (Tamames and Puente-Sánchez, 2019), in order to carry out a first assembly using Megahit (Li et al., 2015). We carried out that first assembly using the *seqmerge* mode but did not use the statistics or any other results. After carrying out the assembly, the generated contigs were used as external assembly references (ext-assembly) in SqueezeMeta, and thus, no assembly was performed at this second step. The input for this second round was trimmed and quality control passed, and (in-house script) files had no spaces and contained separate forward and reverse pairs, gzipped. The pipeline followed the usual procedure for SqueezeMeta, with the following tools.

Assembly was carried out with Megahit (Li et al., 2015), contig statistics with Prinseq (Schmieder and Edwards, 2011), and redundant contig removal by CD-HIT (Schmieder and Edwards, 2011). The merging of contigs was carried out with Minimus2 (Treangen et al., 2011), RNA prediction with Barrnap (Seemann, 2014), taxonomy classification of 16S rRNA sequences by the RDP classifier (Wang et al., 2007), tRNA/tmRNA prediction by Aragorn

(Laslett and Canback, 2004), and Prodigal for open reading frames (ORF) prediction (Hyatt et al., 2010). Diamond (Buchfink et al., 2014) served as the tool to search for similar patterns at GenBank (Clark et al., 2016), eggNOG (Huerta-Cepas et al., 2016), and KEGG (Kanehisa and Goto, 2000). Then HMM homology searches were performed by HMMER3 (Eddy, 2011) for the Pfam database (Finn et al., 2016). Reads were mapped against contigs with Bowtie2 (Langmead and Salzberg, 2012), and binning was done using MaxBin2 (Wu et al., 2016) and Metabat2 (Kang et al., 2019), where binning results were combined using the DAS Tool (Sieber et al., 2018). MiniPath (Ye and Doak, 2009) was used against the KEGG (Kanehisa and Goto, 2000) and MetaCyc (Caspi et al., 2018) databases for pathway prediction. If not indicated, databases were accepted by default in the SqueezeMeta pipeline.

At the end of the pipeline, tables for results processing in R were created by means of the *smq2tables.py* script provided in the SqueezeMeta package. Results were loaded in R through the use of SQMtools (Puente-Sánchez et al., 2020), and charts were produced according to the SQMtools manual. Some of the R packages used were “SQMtools,” “ggplot2,” “reshape2,” “pathview,” and “data.table.”

Non-metric multidimensional scaling (NMDS) was carried out in RStudio (RStudio Team, 2020), using Transcripts per Million (TPM) calculated values obtained from the SqueezeMeta pipeline for those genes with KEGG annotation, through the “vegan” package. This ordination collapses the information from all genes

TABLE 2 Ten most decisive analytes from PLS-DA analysis (Supplementary Figures 1, 2) selected according to VIP.

A (class: diet; scores plot presented on Supplementary Figure 1A)										
Primary ID	Variable name	RT (s)	RI exp	RI lit	VIP	VIP SE 2.44693	Average soil feeders (%)	Average wood feeders (%)		
15	2,3-Butanedione	369	619	612	1.86	1.60	0.04	0.74		
18	2-Butanol	375	626	610	2.05	2.05	0.59	0.07		
37	2,3-Pentanedione	435	694	696	3.24	1.42	0.02	0.13		
97	cis 1-Ethyl-2-methyl-cyclohexane,	641	916	919	1.88	1.67	0.10	0.02		
98	Methyl hexanoate	644	919	925	2.23	1.17	0.00	0.03		
127	Methyl 2-methyl-3-furancarboxylate	723	998	999	2.63	0.89	0.00	0.02		
132	Beta-Phellandrene	734	1,010	1,015	2.08	0.69	0.03	0.11		
138	Limonene	756	1,035	1,030	1.83	0.35	1.05	22.41		
154	Terpinolene	807	1,093	1,089	2.60	1.16	0.01	1.61		
229	Unknown (RI 1495)	1,120	1,495	-	1.80	1.63	0.89	1.88		
B (class: genera; scores plot presented on Supplementary Figure 2A)										
ID	Variable name	RT (s)	RI exp	RI lit	VIP	VIP SE 2.44693	Average Ban (%)	Average Cub (%)	Average Nas (%)	Average Neo (%)
15	2,3-Butanedione	369	582	581	1.63	0.69	0.1	0.0	0.7	0.0
37	2,3-Pentanedione	435	694	696	2.04	1.38	0.0	0.0	0.1	0.0
47	Methyl Isobutyl Ketone	473	737	739	1.57	1.80	0.0	0.0	0.0	0.5
81	1-Hexanol	591	865	867	1.59	0.89	2.7	5.6	0.2	0.3
91	1-Ethyl-4-methylcyclohexane	621	895	888	1.91	0.90	0.0	0.2	0.0	0.0
103	Propyl-cyclohexane	663	938	941	1.66	1.21	0.0	0.2	0.0	0.0
114	1-Octen-3-ol	696	971	979	1.70	0.85	6.1	0.1	0.1	0.4
127	Methyl 2-methyl-3-furancarboxylate	723	998	999	1.61	1.49	0.0	0.0	0.0	0.0
154	Terpinolene	807	1,093	1,089	1.70	1.28	0.0	0.0	1.6	0.0
235	δ-Cadinene	1,147	1,537	1,527	1.64	1.18	0.0	0.0	0.1	0.2

(Continued)

TABLE 2 (Continued)

C (class: geographical origin; scores plot presented on Supplementary Figure 2B)										
ID	Variable name	RT (s)	RI exp	RI lit	VIP	VIP SE 2.44693	Average CM (%)	Average FG (%)	Average MW (%)	
15	2,3-Butanedione	369	619	612	1.77	2.03	0.7	0.0	0.0	
90	2-Butylfuran	617	891	897	1.76	1.99	0.2	0.2	0.0	
101	Unknown (RI 931)	656	931	-	2.04	1.91	0.0	1.5	0.4	
103	Propylcyclohexane,	663	938	941	1.87	2.15	0.1	0.0	0.3	
122	Unknown (RI 988)	713	988	-	1.94	1.60	0.1	0.2	0.1	
185	2-Methyl-dodecane,	946	1,255	1,265	1.93	1.20	0.3	0.4	0.0	
213	Ledane	1,039	1,375	1,373	2.07	1.66	0.0	0.1	0.0	
218	β -Copaene	1,070	1,420	1,420	1.92	1.49	0.0	0.0	0.0	
229	Unknown (RI 1495)	1,120	1,495	1,468	1.96	2.23	0.1	2.5	0.2	
232	α -Bulnesene	1,136	1,521	1,512	1.73	1.31	0.1	0.7	0.1	

Classes used in PLS-DA are (A) diet, (B) genera, and (C) geographical origin.

VIP, variable importance in projection; RT, retention time; RI exp, measured retention index; RI lit, literature; VIP SE, VIP standard error; ID, identifier; Ban, Anoplotermes; Cub, Cubitermes; Nas, Nasutitermes; Neo, Neocapritermes; CM, Cameroon; FG, French Guiana; MW, Malawi.

% refers to the average relative abundance in respective classes.

vs. the samples to visualize the relationships between samples. The PFAM annotation was interrogated for retrieving functions (through the grep command) and annotated with the following keywords: “antibiotic,” “cellulose,” “phenol,” or “resistance.” TPM values were recorded, and bar plots were built to obtain a quick comparison for those annotated genes among samples.

We further used diamond (Buchfink et al., 2014) in order to search for genes and patterns of interest in our final contig list with Open Reading Frames, where the databases of reference were AnHyDeg (Callaghan and Wawrik, 2016), AromaDeg (Duarte et al., 2014), BacMet (Pal et al., 2014), bactibase (Hammami et al., 2010), CARD (protein homology model; Alcock et al., 2020), and CAZy (Carbohydrate Active enZYmes database, <http://www.cazy.org/>; Lombard et al., 2014). We used by default an e-value threshold of 0.005, a minimum identity of $\geq 40\%$, and manually curated the results, discarding hits with a bit-score of < 30 . An arbitrary threshold of $\geq 90\%$ identity was selected first for considering genes as known, while genes below this threshold were considered novel.

We evaluated the overall differences among samples in terms of the presence of genes (“differentially present genes”) through DESeq2 (Love et al., 2014), according to the methodology explained in the SURFBIO transcriptomics module (https://surfbio.eu/training_wet_dry/), where a gene was considered differentially present when it fulfilled the following conditions: fold change (FC) ≥ 2 or $|\log_2\text{FoldChange}| \geq 1$; DESeq2 p-adj ≤ 0.05 .

3. Results

3.1. Identification of termite species belonging to each collected colony

Freshly killed termites belonging to each of the collected nests used in this study were taxonomically identified through molecular methods (COII amplification and sequencing). Results are shown in Table 1. Sequences from this project have been uploaded to the ENA public repository with ref. no. PRJEB55779.

We classified the nests according to the termite species inhabiting them, and only one colony belonged to Group I. Other 10 colonies were classified in Group II from three locations, namely, our laboratory breeds (BR), Cameroon (CM), and French Guiana (FG). Another group comprising 10 colonies originating from French Guiana was classified as Group III. Finally, Group IV comprised colonies originating from Cameroon, French Guiana, and Malawi (MW; Table 1).

3.2. Untargeted metabolite profiling

Among the wood-feeding representatives, we observe a clearer separation in the genus *Nasutitermes* between CM and FG (Figure 1). Three samples belonging to the nests of the same species (Nasl 1-3) were grouped together and located in the same quadrant as another CM wood feeder nest (Nassp CM II). The samples originating from MW were grouped together with close to four nest samples belonging to the soil-feeding species (*Anoplotermes banksi*) from FG, along with most of the representatives from feeding groups III and IV.

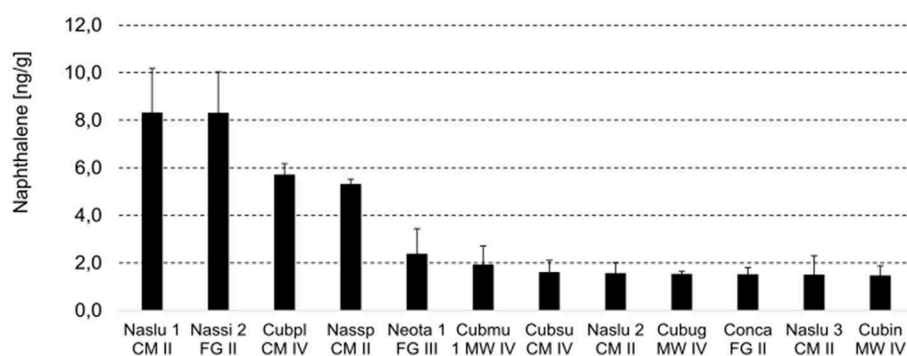


FIGURE 2

Naphthalene analysis: results from targeted metabolite profiling. Naphthalene concentration is expressed in ng per gram of nest material. Naslu, *Nasutitermes lujae*; Nassi, *Nasutitermes similis*; Cubpl, *Cubitermes planifrons*; Nassp, *Nasutitermes* sp.; Cubmu, *Cubitermes muneris*; Cubsu, *Cubitermes sulcifrons*; Cubug, *Cubitermes aff. ugandensis*; Conca, *Constrictotermes cavifrons*; Cubin, *Cubitermes inclitus*; CM, Cameroon; FG, French Guiana; MW, Malawi.

According to the PLS-DA (Supplementary Figure 1A), a clear separation exists between nests from soil or wood feeders, with influence from genera (Supplementary Figure 1B) and geographical location (Supplementary Figure 1C).

The most responsible volatiles for the separation were used as a class in the PLS-DA models (Supplementary Figure 2), which also used genera [R2X(cum) 0.53, R2Y(cum) 0.99, and Q2(cum) 0.65] and geographical location [R2X(cum) 0.25, R2Y(cum) 0.78, and Q2(cum) 0.5]. Using the variable importance plot (VIP), the 10 most decisive compounds for separations presented in Supplementary Figure 2 are identified in Table 2.

3.3. Targeted analysis: Naphthalene content

Naphthalene is a polyaromatic hydrocarbon observed in the volatile profiling during the non-targeted metabolite analysis. The highest relative abundances were recorded in samples from the genus *Nasutitermes*, *Coptotermes* from BR, and some of the *Cubitermes*.

To determine the exact concentration of naphthalene in nest material, we applied a quantitative method, where the concentration ranged from 1.5 to 8 $\mu\text{g}/\text{kg}$ (Figure 2). These results indicate that naphthalene is not generally present in the studied species belonging to Group III, except for the nest material from one of the colonies of *Neocapritermes taracua* (colony No. 1). We observed the presence in species belonging to feeding Groups II and IV; in detail, in three samples from the genus *Nasutitermes* (Group II, both Cameroon and French Guiana) and the five species belonging to the genus *Cubitermes* (Group IV). Naphthalene appeared in the colony nest material from *Constrictotermes cavifrons*, belonging to Group II (French Guiana). The samples belonging to Group III do not show detectable amounts of naphthalene (except for one of the *Neocapritermes* samples).

3.4. Bacterial community structure

We calculated more than 130,000 effective tags after removing chimera sequences from our samples. These values range from

more than 120,000 tags (*Coptotermes*, BR) to more than 140,000 non-chimeric tags (NasLu3, CM). In our DNA isolation process, which was carried out with sterility-opened reagents inside a laminar flow hood throughout the entire isolation process, we included a negative DNA extraction control and yielded more than 36,000 tags after chimera removal.

When comparing the assigned ASVs, 531 belong to the negative extraction control, from a total of 19,753 ASVs in our project. The overlap with other samples was calculated, and from these 532 ASVs, we found to overlap with other samples in a total of 106 ASVs. These relationships are represented as a Venn diagram (Supplementary Figure 3) with selected samples.

Filtering sequences after “dada2” removes the majority of control sequences, leaving a total of 112 tags (from a total of 36,406 before filtering). We observed a decrease in sequences in the rest of the samples, where the minimum number of tags belongs to Neota2FG.III with 38,777 tags.

3.4.1. Subset of samples with at least three representatives per genus

We carried out a sub-selection of samples, as we did not find clear information when using all samples. From a statistical point of view (ACE Index, where the direction is always Group 1 for the first described element and Group 2 for the second), we found that at a Kruskal-Wallis pairwise comparison, there are geographical differences between FG and MW (p -value = 0.031), although not significant when taking all groups together (p -value = 0.077). Comparing feeding groups, we found significant differences when comparing II vs. III (p -value = 0.045) and III vs. IV (p -value = 0.015), and these differences are statistically significant among all groups (p -value = 0.028). For general, we found differences in *Anoplotermes* vs. *Neocapritermes* (p -value = 0.034), *Cubitermes* vs. *Neocapritermes* (p -value = 0.02), and *Nasutitermes* vs. *Neocapritermes* (p -value = 0.034), existing significant differences among groups (p -value = 0.03).

The β -diversity (unweighted UniFrac distance) showed significance between feeding groups (pairwise PerMANOVA; differences between groups II and III, p -value = 0.029; groups III

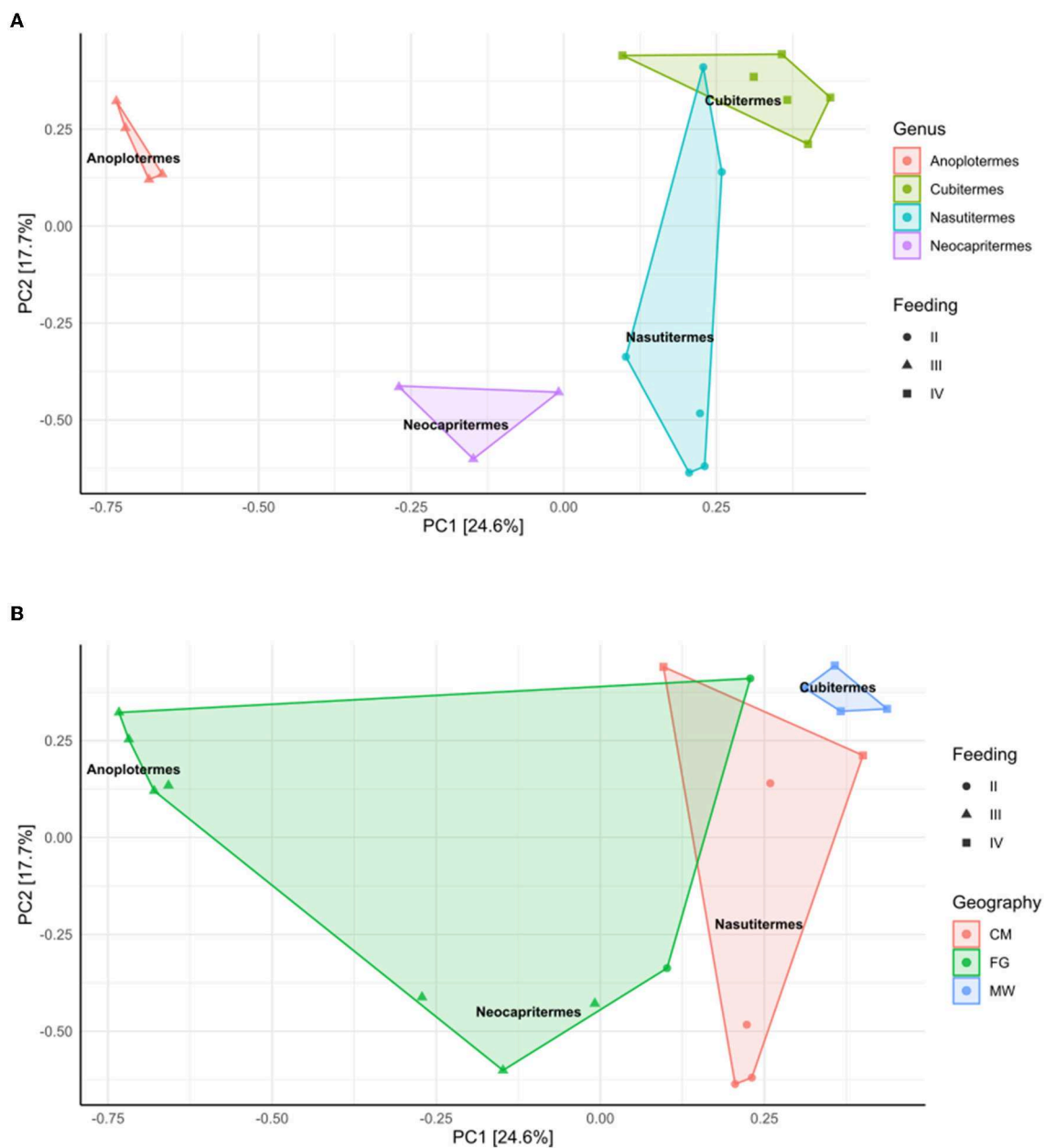


FIGURE 3

PCA analysis with 16S data after elimination of samples with <3 representatives per genus, and the tags from the negative sample control. (A) Samples were subjected to Hellinger transformation, sample constraint, and color were "Genus," and shape by feeding strategy. (B) Samples were subjected to Hellinger transformation and the sample constraint was "Genus," while sample color was geography and shaped by feeding strategy. Genera indications are kept as shown in Figure 4.

and IV, p -value = 0.005; differences among all groups, p -value = 0.001). In the case of geography, differences were observed between FG and MW (p -value = 0.045), while differences among all groups were of p -value = 0.02. Data with at least three representatives per genus could be compared by genera as well. In this case, we found statistical differences between *Anoplotermes* and *Cubitermes* (p -value = 0.006) and between *Anoplotermes* and *Nasutitermes* (p -value = 0.006). Finally, the differences between groups had a p -value = 0.001. Further analysis through ANOSIM revealed that all three considered variables (diet, genus, and geographical

location) influenced the community following the order (R -value): 1 > Genus (0.67) > Diet (0.65) > Geography (0.42) > 0.

The PCA plot shows a clear stratification of samples both by genera and diet (Figure 3A), while the stratification is less clear when grouping samples by geographical location (Figure 3B). The two principal components of the analysis explain 24.6 and 17.7% of the variance.

Table 3 shows the phylogenetic assignment of taxa with the most abundant orders, grouping samples by termite genera with a distinction of geographies. We observe an overall representation of

TABLE 3 Taxonomic overview of the samples retained for further analyses (with more than two representatives per genus).

	Anba (FG)	Cub (CM)	Cub (MW)	NasLu (CM)	Nas (FG)	NasSp (CM)	Neo (FG)
p_Actinobacteria;c_Actinobacteria; o_Actinomycetales	32.79	39.88	31.61	37.50	24.00	44.50	34.61
Unidentified bacteria	19.22	6.42	4.69	6.13	11.55	3.82	6.64
p_Actinobacteria;c_Thermoleophila; o_Solirubrobacterales	2.35	5.09	5.07	7.10	3.50	4.95	6.78
p_Actinobacteria;c_Acidimicrobiia; o_Acidimicrobiales	3.11	5.87	3.13	7.32	3.21	9.86	8.58
p_Firmicutes;c_Clostridia; o_Clostridiales	10.50	5.61	4.98	4.73	5.60	10.46	2.49
p_Planctomycetes;c_Planctomycetia; o_Gemmatales	2.74	2.66	6.16	1.16	5.26	0.31	5.35
p_Spirochaetes;c_Spirochaetes; o_Spirochaetales	3.07	2.71	1.81	5.97	8.58	1.79	1.45
p_Proteobacteria;c_Alphaproteobacteria; o_Rhodospirillales	2.80	4.50	3.67	4.73	4.98	3.28	5.72
p_Chloroflexi;c_Ktedonobacteria; o_Ktedonobacterales	0.82	0.78	3.48	0.65	0.71	0.00	0.08
p_Firmicutes;c_Bacilli; o_Bacillales	1.19	0.93	1.58	0.81	0.74	1.00	2.76
p_Proteobacteria;c_Alphaproteobacteria; o_Sphingomonadales	0.03	1.33	1.62	1.64	0.66	0.31	0.44
p_Actinobacteria	1.02	0.74	0.11	0.07	0.08	0.13	0.10
p_Proteobacteria;c_Alphaproteobacteria; o_Rickettsiales	3.57	1.08	2.13	0.82	4.16	0.11	1.23
p_Firmicutes;c_Bacilli; o_Turicibacterales	1.71	0.87	0.78	0.40	1.14	0.66	1.18
p_Proteobacteria;c_Gammaproteobacteria; o_Aeromonadales	0.00	0.00	0.00	0.00	0.00	2.27	0.00
p_Chloroflexi;c_Ktedonobacteria; o_Thermogemmatissporales	0.08	0.26	1.86	0.06	0.11	0.00	0.21
p_Proteobacteria;c_Betaproteobacteria; o_Burkholderiales	1.39	0.54	1.28	1.38	1.98	0.57	2.81
p_Acidobacteria;c_Acidobacteriia; o_Acidobacteriales	1.50	1.67	2.07	1.16	2.99	0.76	2.36
p_Firmicutes;c_Bacilli; o_Lactobacillales	0.31	0.56	0.42	0.78	0.33	0.61	0.48
p_Verrucomicrobia;c_[Spartobacteria]; o_[Chthoniobacterales]	0.13	0.25	0.63	0.10	0.25	0.01	0.16
p_Proteobacteria;c_Gammaproteobacteria; o_Xanthomonadales	0.27	1.49	1.99	1.80	2.50	0.78	1.56
p_Proteobacteria;c_Alphaproteobacteria; o_Ellin329	0.23	0.28	0.27	0.22	0.50	0.13	0.49

(Continued)

TABLE 3 (Continued)

	Anba (FG)	Cub (CM)	Cub (MW)	NasLu (CM)	Nas (FG)	NasSp (CM)	Neo (FG)
<i>P_Synergistetes;c__Synergistia;o__Synergistales</i>	0.53	0.13	0.18	0.18	0.24	0.09	0.09
<i>p__Bacteroidetes;c__Bacteroidii;o__Bacteroidales</i>	2.13	1.40	1.01	1.13	1.34	1.02	0.33
<i>p__Acidobacteria;c__Solibacteres;o__Solibacterales</i>	0.38	0.73	0.81	0.55	1.40	0.62	0.95
<i>p__Proteobacteria;c__Alphaproteobacteria;o__Rhizobiales</i>	0.19	0.94	1.97	1.63	1.31	0.69	1.91

Taxa with more than 5% representation have been shaded and bolded.

FG, French Guiana; CM, Cameroon; MW, Malawi; Anba, *Anoplotermes banksi*; Cub, *Cubitermes*; NasLu, *Nasutitermes lujai*; Nas, *Nasutitermes*; NasSp, *Nasutitermes* sp.; Neo, *Neocapritermes taracua*.

Actinomycetales, *Acidimicrobiales*, *Clostridiales*, *Gemmatales*, and *Solirubrobacterales*, being *Spirochaetales* well-represented category.

The class Betaproteobacteria, order IS_44, characterized *Neocapritermes* nests in a LEfSe analysis (Figure 4). Representatives from the genus *Streptomyces*, down to species *lanatus*, dominate in *Nasutitermes* nests. *Cubitermes* nests are dominated by members of the genus *Rhodococcus* and the family Intrasporangiaceae, whereas *Anoplotermes* nests are dominated by *Chitinophaga*.

3.5. Fungal community structure

The whole group of samples showed inconclusive results; therefore, we analyzed a smaller dataset with more than two representatives per genus. At the alpha diversity by the ACE Index (non-phylogenetic, the direction always from first element to second introduced element), we found only statistical differences among feeding groups at a Kruskal-Wallis pairwise comparison ($\alpha \leq 0.05$). Particularly when comparing II vs. III (p -value = 0.003), II vs. IV (p -value = 0.01), and all groups (p -value = 0.006). Further analysis through ANOSIM revealed that all three considered variables (diet, genus, and geographical location) influenced the community following the order (R -value): 1 > Diet (0.627) > Genus (0.592) > Geography (0.422) > 0. These results show that fungi are less influenced by the phylogenetic relatedness of the nest host than their bacterial counterparts.

A PCA plot was built in R through the ampvis2 package (Figure 5). Results show a clear stratification of samples, especially by genera, and then feeding strategy (Figure 5A), while the stratification is less clear when grouping samples by geographical location (Figure 5B). The two principal components of the analysis explain 10.6 and 9.5% of the variance.

Taxonomical assignment produced hits mostly to unidentified species, and in this case, we did not proceed with the LEfSe analysis.

3.6. Metagenomic sequencing

3.6.1. Cleaning, assembly, and SqueezeMeta pipeline

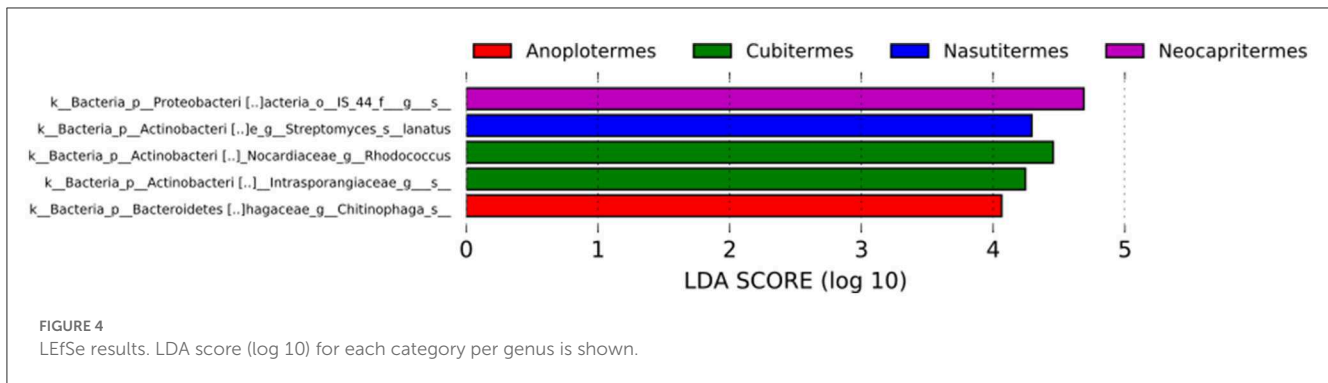
The first assembly yielded more than 5.5 million contigs, with an average size of more than 720 bp, and N50 of 776 bp.

The reads were assigned taxonomically, and the most abundant hits per sample are shown in Supplementary Table 2, while the statistics on open reading frames (ORFs) are presented in Supplementary Table 3.

3.6.2. Overall characterization of samples

We used the TPM values for all KEGG-annotated genes to evaluate relationships between samples through non-metric multidimensional scaling (NMDS; Figure 6). All *Nasutitermes* samples cluster together, while soil-feeder nests are in another area.

We can observe that the four samples belonging to the genus *Nasutitermes* (all originating in CAM) have a similar profile at the genus level, while the two soil-feeding genera (*Cubitermes*, from MW and CAM, and *Anoplotermes*, from FG), despite differences



at the species level, present a closer profile (Figure 7A). An outstanding exception is a biological replicate Ban1, with a lower representation of Proteobacteria than the rest of samples from this genus (all samples also belong to the species *Anoplotermes banksii*), potentially owed to issues during transportation. For *Nasutitermes*, we find differences between the three *Nasutitermes lujae* replicates and *Nasutitermes* sp. The samples belonging to the genus *Cubitermes* have a more variable profile. Among those, CubPl and CubSu belong to the CM sampling site, while CubIn and CubMu originate in MW.

We further carried out a subset of the present genera belonging to the “aromatic AA” representation (“phenylalanine, tyrosine, and tryptophan biosynthesis”; Figure 7B). The most important genera for soil feeders seem to lie within an unclassified *Streptosporangiales* genus with higher abundance for *Cubitermes* and *Anoplotermes*, while *Nasutitermes* had a higher representation of *Microclunatus*, an unclassified genus within *Microbacteriaceae*, and *Gryllotalpicola*.

We followed a “quick interrogation” step to retrieve from the PFAM-annotated set contigs with the keywords described in the methodology section. We constructed four-bar plots (Supplementary Figures 4, 5) where we observe a different profile between soil and wood feeders. One of the most important categories in *Nasutitermes* seems to be cellulose synthases (Supplementary Figure 4A), while they largely lack some genes annotated as “cadmium resistance transporters” or the “tellurite resistance protein.”

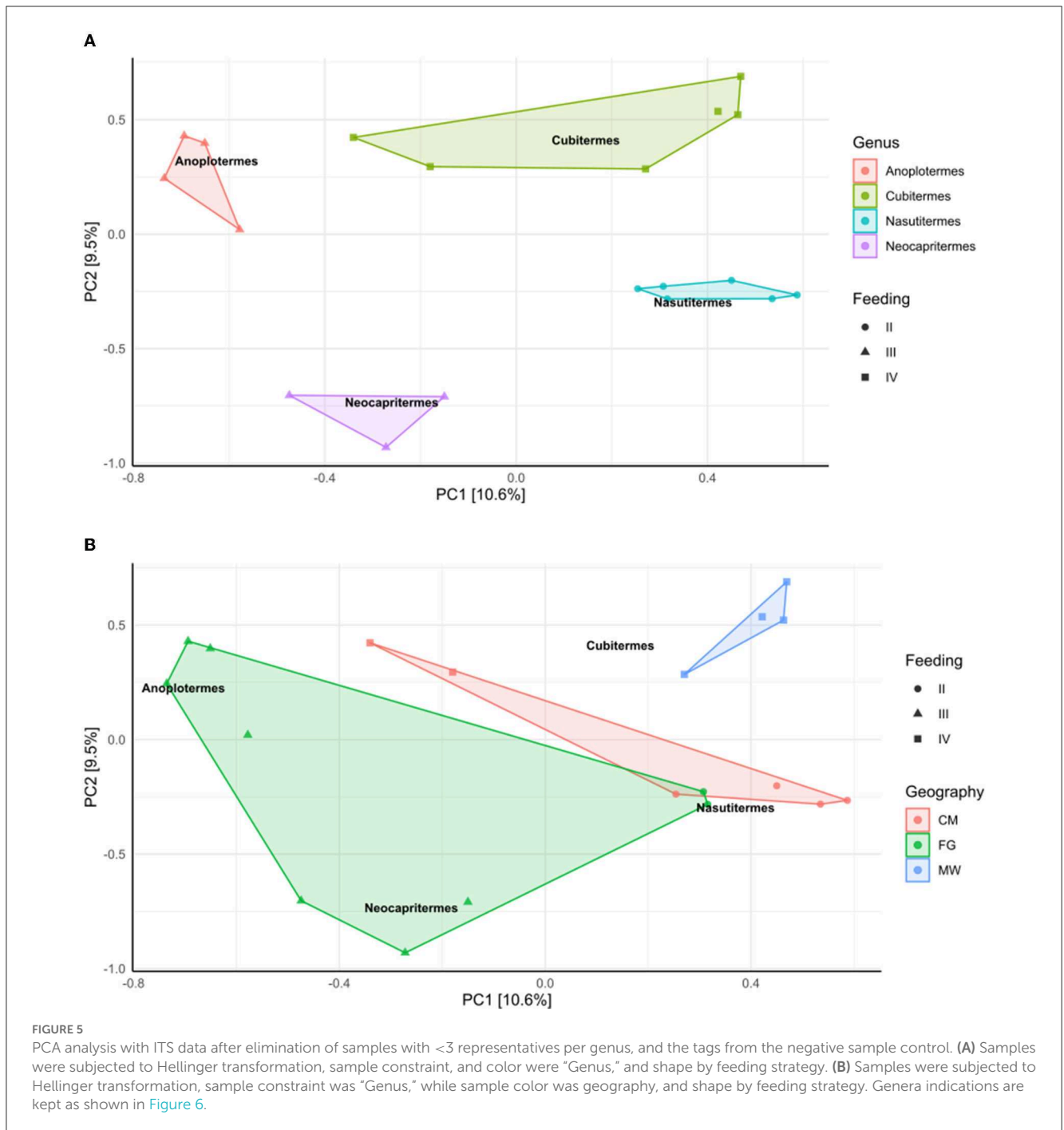
An LDA effective size (LEfSe) was calculated using the PFAM annotation with differences among the three represented genera. We obtained 61 classes of annotated genes that are specific biomarkers for each (Figure 8, Supplementary Table 5). *Anoplotermes* was characterized by several biomarkers, among them the first PF00196: bacterial regulatory proteins luxR family, PF13424: tetratricopeptide repeat, PF03704: bacterial transcriptional activator domain, and PF13191: AAA ATPase domain. *Cubitermes* had PF00872: transposases, mutator family, and PF0084: sulfatase as the only biomarker categories found. One of the most represented biomarker categories in the genus *Nasutitermes* was PF00296: luciferase-like monooxygenase, followed by PF02515: CoA transferase family III, PF00873: AcrB/AcrD/AcrF family, and PF00171: aldehyde dehydrogenase family.

3.6.3. Differential gene content analysis

Hierarchical clustering renders two clearly separated nodes (Figure 9A), one with all *Nasutitermes* samples (*Nasutitermes*, CAM) and the other with the soil-feeding genera. In the second group, there are two nodes; one comprises Ban1 and Cubin, and the other is subdivided into *Anoplotermes* (Ban2 and Ban3) and *Cubitermes* samples. Among them, in an inner position, we find the *Cubitermes* samples originating in CAM. These differences can be clearly observed through a correlation heatmap (Figure 9B), while the differentially present genes (DPGs) appear in a PFAM heatmap (Supplementary Figure 6).

The main differences existing between pairs of samples in terms of DPG have been displayed through volcano plots (Supplementary Figure 7). Using COG annotation, at a fold change of 2, we found a total of 2,099 DPGs: 1,147 upregulated genes (more present in *Cubitermes*), when comparing the two soil feeders (*Cubitermes* vs. *Anoplotermes*) and 952 COG annotated genes, with higher presence in *A. banksii* samples (Supplementary Figure 7A). When comparing *Nasutitermes* and *Cubitermes*, we found 6,771 upregulated and 7,871 downregulated DPGs (a statistically significant higher presence in the soil feeder; Supplementary Figure 7B). The last comparison was carried out between *Nasutitermes* and *Anoplotermes* (Supplementary Figure 7C), with 7,152 upregulated and 7,478 downregulated genes.

The functional profiles of soil genera nests had 184 PFAM-annotated contigs with a statistically significant over-presence in *Cubitermes*, while *Anoplotermes* had 265 PFAM-annotated contigs with over-presence. We could equally state that these 265 genes have a decreased presence in the microbial communities from the *Cubitermes* genus, in comparison with *Anoplotermes*. We found some genes with extreme differences in the presence (higher for *Cubitermes*), such as PF03945: delta endotoxin, N-terminal domain ($\log_2FC = 23.96$; $p\text{-adj} = 1.55 \cdot 10^{-07}$) and PF14470: bacterial PH domain ($\log_2FC = 22.41$; $p\text{-adj} = 1.34 \cdot 10^{-15}$). On the contrary, PF02183: homeobox-associated leucine zipper ($\log_2FC = -11.78$; $p\text{-adj} = 9.10 \cdot 10^{-14}$) has a higher presence in *Anoplotermes*. The comparisons between the microbial communities from the soil-feeding termites with *Nasutitermes*' nest provided again a higher number of DPGs. For *Nasutitermes* vs. *Cubitermes*, we found a total of 1,908 PFAM-annotated contigs with statistically significant over-presence in *Nasutitermes*, while *Cubitermes* had

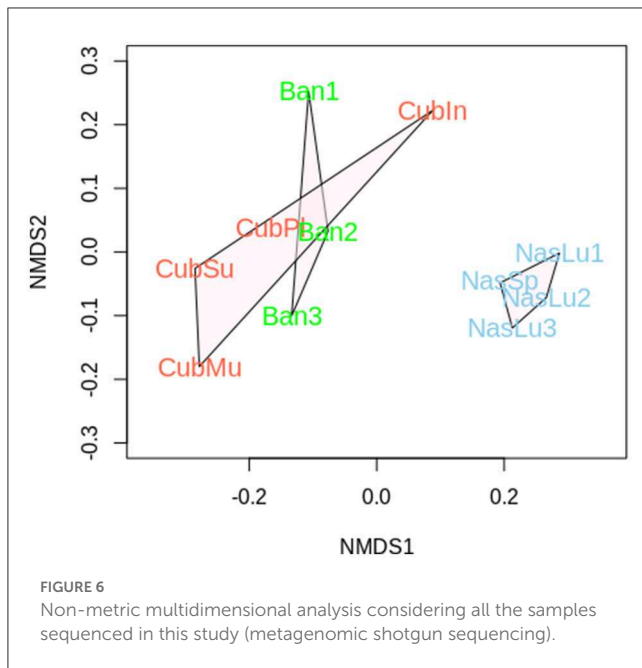


918 PFAM-annotated contigs with over-presence. We could equally state that these genes have a decreased presence in the microbial communities from *Nasutitermes* than in the microbial communities on those nests from *Cubitermes*. Among the functions most present in *Nasutitermes*, we found PF16937: type III secretion system translocator protein, HrpF ($\log_2FC = 12.16$; $p\text{-adj} = 7.20 \cdot 10^{-13}$), PF17667: fungal protein kinase, HrpF ($\log_2FC = 10.99$; $p\text{-adj} = 2.26 \cdot 10^{-07}$), and PF05420: cellulose synthase operon protein C C-terminus (BCSC_C; $\log_2FC = 10.92$; $p\text{-adj} = 1.01 \cdot 10^{-15}$). For instance, PF03945: delta endotoxin, N-terminal domain ($\log_2FC = -28.16$, $p\text{-adj} = 8.66 \cdot 10^{-14}$) is an example of a highly present PFAM function in *Cubitermes* (Supplementary Table 6).

The comparison between *Nasutitermes* and *Anoplotermes* yielded 1,962 upregulated and 1,016 downregulated genes (Supplementary Table 6).

3.6.4. Annotation of ORFs in specialized databases

We carried out further annotation of the ORFs retrieved by SqueezeMeta with several specialized databases of interest (see the "Material and Methods" section). The first characterization of these results was to count known genes over a 90% identity threshold. We used a blasting approach with a lower identity threshold of 40% but focused on the bit-score for annotation (especially \geq



50), based on Pearson (2013), to discover homologous genes. We have used this approach because we found cases where a 100% BLAST identity had a length of alignment under 50 amino acids. Using the maximum bit-score values obtained through DIAMOND Blast (Buchfink et al., 2014), the e-value was zero in 28 hits at the BacMet list, with alignment lengths varying from 605 to 1,118 amino-acids for those 28 genes, and identities ranging from 55 to 99%. As a result, and in accordance with the study by Enault et al. (2016), we suggest that the use of bit-score is a much better indication than identity for discovering homology to described genes. We used the annotation obtained through diamond Blastx with three databases, namely, AromaDeg, CARD, and CAZy. These lists were used as input for differential gene content analyses.

The AromaDeg-annotated genes served as an input for a heatmap displaying the gene content profile in all samples (Figure 10). We found clear differences and tested them for statistical significance. There were 8,132 AromaDeg-annotated genes, from which the comparison of *Cubitermes* vs. *Anoplotermes* yielded 14 upregulated and 21 downregulated DPGs. The comparison of *Nasutitermes* vs. *Cubitermes* was carried out, and 468 genes were upregulated or significantly more present in the *Nasutitermes* genus, while 82 were downregulated or had a higher presence in the *Cubitermes* genus. For the comparison of *Nasutitermes* vs. *Anoplotermes*, we found 224 upregulated and 12 downregulated genes.

The comparison between the soil feeders yielded 54 upregulated genes (*Cubitermes*), and 241 genes had a higher presence in *Anoplotermes*. The correlation among samples and hierarchical clustering clearly divided all samples by genera. The comparison of *Nasutitermes* vs. *Cubitermes* resulted in 1,587 CARD-annotated genes with a higher presence in the wood feeder nest community, while the soil feeder nest community had 988 upregulated genes. When *Nasutitermes* was compared with *Anoplotermes*, we found

1,403 upregulated and 1,745 downregulated genes (higher presence in the soil feeder).

We carried out the analysis with the CAZy DB-annotated genes. In the comparison between both soil feeders, we found no statistical differences. *Nasutitermes* vs. *Cubitermes* yielded 287 upregulated and 2 downregulated genes. We found a larger number of genes for the *Nasutitermes* vs. *Anoplotermes* comparison. This one provided 519 upregulated and 218 downregulated genes.

Some of the most deregulated genes for each of the three analyses are displayed in Table 4.

4. Discussion

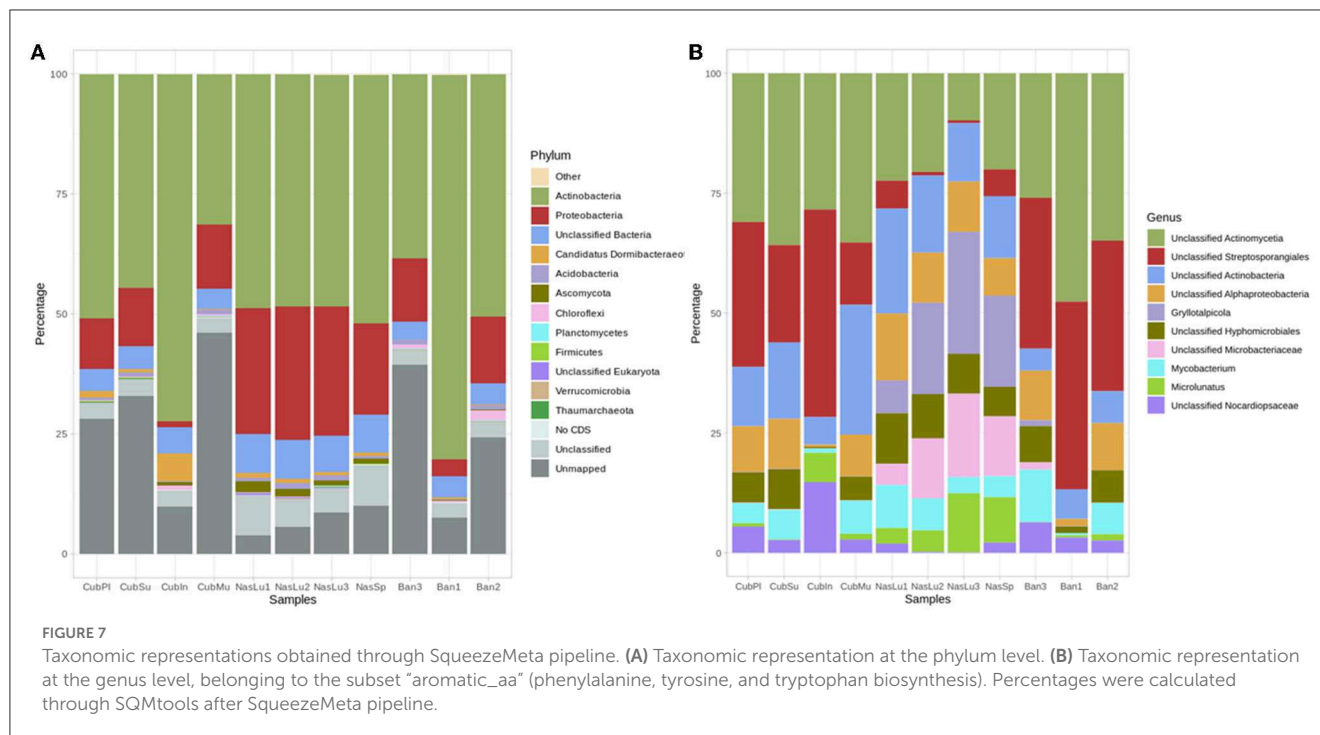
We evaluated a wide range of termite nests originating on both sides of the Atlantic Ocean, including soil and wood feeders. Sampling locations encompassed tropical rainforest and savannah habitats. Our first aim was to better characterize the drivers of microbiome differences and to provide insights into the population structure and functions as they could be reflected in the metabolite and gene content.

4.1. Nest metabolomic composition

The metabolite untargeted approach stratified the samples according to feeding strategies, with a certain influence of the geographical location and genera (Figure 1). Separated PLS-DA models could differentiate between genera, geographical location, and feeding group (Supplementary Figures 1, 2). The 10 most decisive analytes were mainly alcohols, their derivatives, and terpenic compounds.

One of the most important compounds for the separation of samples according to feeding habits was 2-butanol (Table 2A), which can originate from the decomposition of organic matter in the soil. Its methyl derivative, 2-methyl-butanol, was previously reported to have antifungal activity (Matsuura and Matsunaga, 2015). Another two important compounds for this separation are diacetyl (2,3-butanedione) and 2,3-pentanedione, potentially originating in higher amounts in wood feeders' nests because of fermentation (Yamaguchi et al., 1994). Other compounds important for separation and more abundant in wood feeders' nests were mostly terpenes.

Terpenes are a group of aromatic compounds that have been related to the mediation of information transmission (Gershenzon and Dudareva, 2007) or have been tested specifically as a toxic compound against termites (Seo et al., 2014). Among them, we established terpinolene as an analyte with decision power. Terpenes have been reported repeatedly in termite literature as a chemical component of the defensive secretions in *Nasutitermes* soldiers (Prestwich et al., 1981), not related to feeding habits as explained by its absence in workers. Besides, terpenes have been appointed as an antimicrobial substance, as confirmed by Mitaka et al. (2017). Terpinolene was first identified in the cephalic secretion from *Amitermes herbertensis* (Moore, 1968); and later in small quantities of a *Nasutitermitinae*, *Grallatitermes africanus* (Prestwich, 1979); in the defensive secretion from *Nasutitermes ephratae* soldiers (Valterová et al., 1986); soldier frontal glands



in *Termitogeton planus* (Dolejšová et al., 2014); and significant amounts in *Constrictotermes cyphergaster* soldiers (De Mello et al., 2021). We found this compound mostly in the nests from *Nasutitermes*, with neglectable concentrations in others.

Our data indicated a lower level of naphthalene in comparison with published literature (Chen et al., 1998b; Wilcke et al., 2000). Overall, two genera concentrate the readings of naphthalene (*Nasutitermes* and *Cubitermes*), while the rest of the samples had a scattered representation. We did not record naphthalene in the nests from *Anoplotermes*. The type of sample preparation could have influenced, although a soil standard indicated correctness in our method. In some cases, the results could have been affected by the time elapsed from sampling to processing, a matter of a few days due to traveling and transportation. Nevertheless, we could still observe clear differences between samples at the volatile level that were further confirmed at the other omic approaches.

4.2. Nest microbial composition and structure

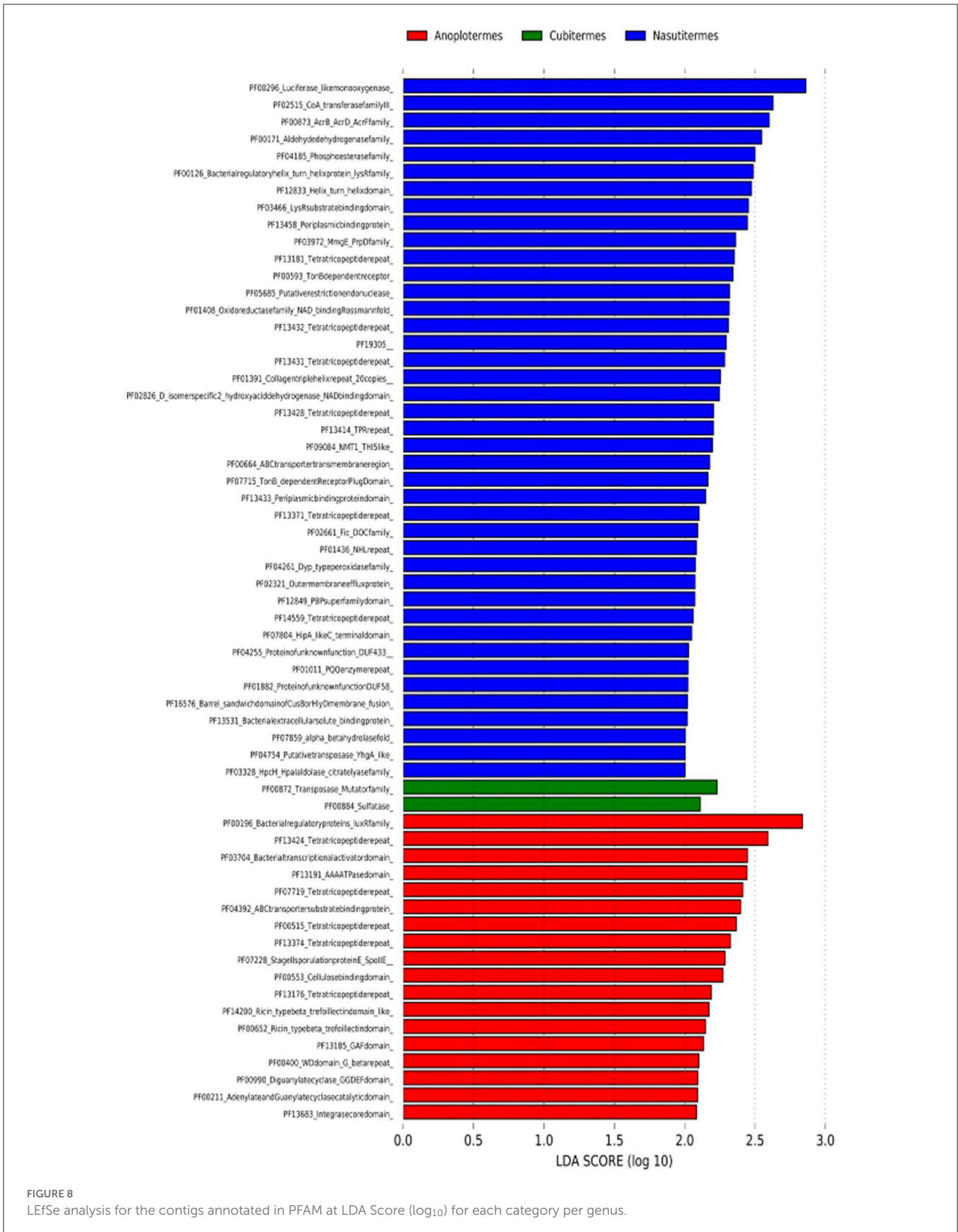
4.2.1. Amplicon sequencing: Bacteria

The whole set of samples showed similar results to the overall untargeted metabolite profile without a clear stratification (Supplementary Figure 1). We observe an effect of phylogenetic relatedness and an influence of the geographical origin in the richness of the microbial communities associated with different nests. The highest values were those from breeds, potentially owed to an artificial input of strains from samples and colonies kept in similar laboratory conditions and proximity, and the potential input of human microbiome. The second highest level

was that of MW samples. We hypothesize that the savannah conditions could pose a higher degree of stress over soil microbial communities, which could find a better shelter in these nests. Our β -diversity results show significant statistical differences between feeding Groups III and IV and between FG and MW.

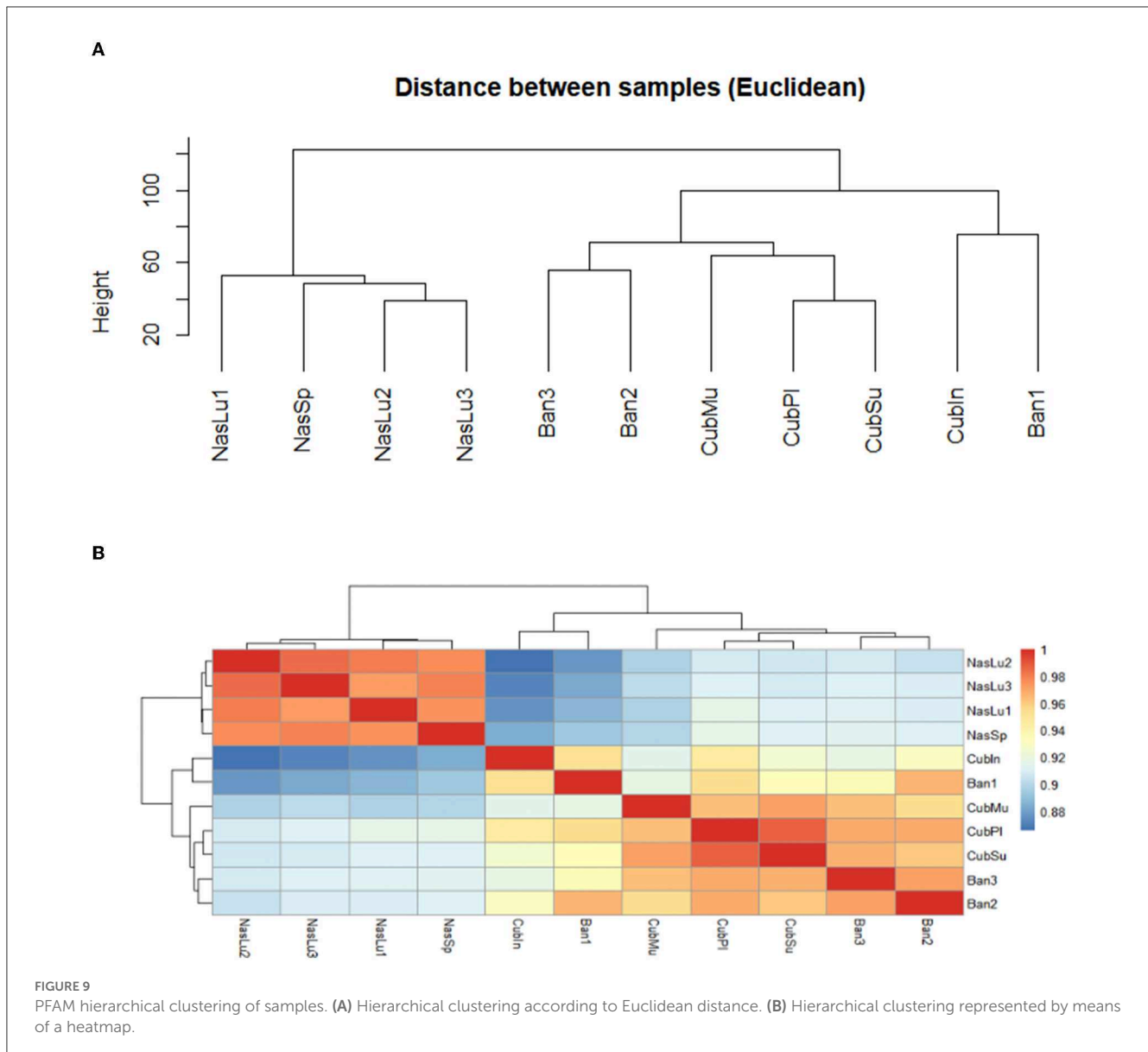
At the ASVs-annotated PCA plot, samples with at least three representatives, we confirmed that the differences between microbial communities are determined strongly by the genera and diet (Figure 3A). While *Anoplotermes* and *Neocapritermes* belong to the same feeding group, they occupy distant locations in the plot, which provides the idea that genus is a stronger component than feeding strategy. The second part of the plot (Figure 3B) indicates that the geographical location is weaker when explaining the variability among nest bacterial communities based on 16S amplicon data. The statistical analysis through ANOSIM agrees with the previous analysis.

Our results agree with the published termite literature regarding the bacterial taxonomic composition of nests. *Actinomycetales* were reported as dominant in termite nests (Krishanti et al., 2018). These are Gram-positive bacteria capable of forming branching hyphae, related to nutrient recycling roles (Goodfellow and Williams, 2003), and have been used for the production of secondary metabolites, accounting for a very large fraction of all bioactive secondary metabolites in use in industry (Olano et al., 2008). This group has been also reported in the study by Chouvenec et al. (2018), where authors studied the origins of the acquisition of *Streptomyces*, the largest genus within *Actinobacteria* (El-Naggar, 2021), in *Coptotermes* nests. Authors have reported recruitment from surrounding soil. Similar results regarding *Streptomyces* were reported in the nests of *Reticulitermes flavipes* (Aguero et al., 2021). *Clostridiales* are Gram-positive



anaerobic bacteria (Bowman, 2011) that can form resistant spores (Paredes-Sabja et al., 2011), described as thermophilic ethanologens capable of fermenting lignocellulosic biomass (Arora et al., 2019),

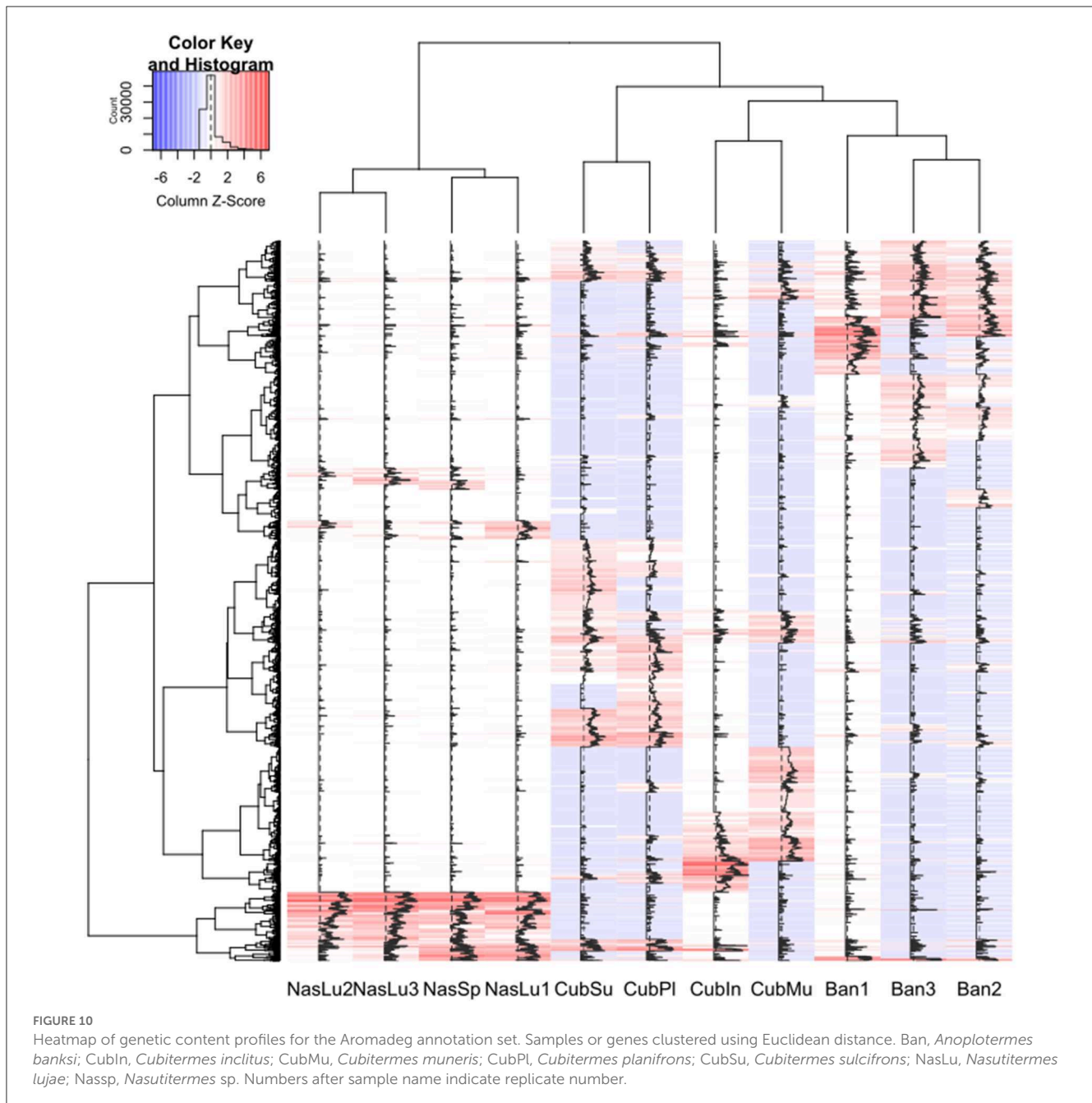
including species that are lignocellulolytic (Auer et al., 2017). This order has been associated with the termite gut community of soil feeders (Ohkuma and Brune, 2010), and also *Reticulitermes*



santonensis, through the presence of *Clostridia*-related clones within *Firmicutes* (Yang et al., 2005) and was appointed as one of the groups carrying out recycling of uric acid (Thong-On et al., 2012). In *Odontotermes yunnanensis*, Long et al. (2010) reported the presence of a wide representation of bacteria from the *Clostridium* genus in fungus comb structures. We found that this group was of the highest proportion in two of the soil feeders (*Anoplotermes* and *Cubitermes*) but also in nests from *Nasutitermes*. Regarding *Acidimicrobiales*, we have found no specific information in termite literature. *Solirubrobacterales* is an order with a few species described as Gram-positive mesophilic bacteria growing in a moderate range of temperatures (Whitman and Suzuki, 2015). Our results are in agreement with those described by Enagbonma et al. (2020), who report this group at termite mounds from *Coptotermes* species. Uncultured representatives of *Solirubrobacterales* have been associated with the degradation of lignin in soils (Wilhelm et al., 2018; Silva et al., 2021),

also previously reported in sugarcane farm soils (Ogola et al., 2021).

Spirochaetales are Gram-negative that can comprise both anaerobic and strict aerobic bacteria (Smibert and Johnson, 1973). They have been reported in increased numbers when manure is applied to soil (Rieke et al., 2018). This group was also reported in *R. santonensis* gut (Yang et al., 2005) and gust from *Coptotermes gestroi* (Do et al., 2014) and have been appointed as active cellulose degraders (Xia et al., 2014), with abundance in wood-feeding termites (Tokuda et al., 2018; Hu et al., 2019). *Gemmatales* are Gram-negative chemoheterotrophic bacteria (Dedysh et al., 2020), which could act as an indicator of elevated phosphorus content in soil (Mason et al., 2021), and in a complex microbial community dedicated to the utilization of cellulose could be providing endocellulases and β -glucosidases activities (McDonald et al., 2019). Another group of bacteria notably described in the literature is that of *Rhizobiales*, Gram-negative bacteria belonging



to the *Alphaproteobacterial* class, and with many members capable of fixing nitrogen (Beekmans and Xie, 2015) and associated with many plants, lichens, and mosses (Erlacher et al., 2015). This order was also reported as enriched in termite nests of *Coptotermes testaceus*, *Heterotermes tenuis*, and *Nasutitermes octopilis* (Soukup et al., 2021). In our case, the proportion is limited.

Biomarkers analysis (LEfSe, Figure 4) indicated that certain genera were representative of specific samples. *Streptomyces* was associated with *Nasutitermes*, in agreement with Chouvenec et al. (2018) and Agüero et al. (2021), and could be indicative of a defensive role of this bacterial order in the termite nest. *Neocapritermes* with β -proteobacteria are described as characteristic cellulolytic bacteria in acidic forest soils in temperate areas (Štursová et al., 2012). Belonging to this class, the genera

Burkholderia and *Collimonas* were characterized as capable of carrying out mineral weathering (Lepleux et al., 2012). Nest microbial communities from *Cubitermes* were associated with the genus *Rhodococcus*, comprising aerobic Gram-positive bacteria. Members from this genus have been associated with the degradation of nitroaromatic compounds (Subashchandrabose et al., 2018), but more importantly with lignin degradation activities (Chong et al., 2018). This observation is aligned with the fact that lignin degradation is not important in wood feeders (Ohkuma, 2003; Griffiths et al., 2013) and could be an indication that members of this genus carry this important task in the *Cubitermes*-associated communities. *Chitinophaga* has been described as a chitinolytic and ligninolytic genus (Funicelli et al., 2021), besides *Spirochaeta*, where one of the species was reported to degrade complex plant polysaccharides and depolymerize lignin

TABLE 4 Genes with differential presence obtained from the specific annotations with AromaDeg, CARD, and CAZy databases.

DB	ID	log2FC			Annotation	Bit-score
		Nas_Ban	Nas_Cub	Cub_Ban		
AromaDeg	Merged2_1077050_827-1954	n/s	n/s	22.66	WP_018158791 Smaragdicoccus_niigatensis Non	523.5
AromaDeg	Merged2_2904114_1-831	-22.63	-20.1	n/s	ZP_05485171 Streptomyces_sp_SPB78 Non	497.3
AromaDeg	Merged2_367852_494-1618	13.4	11.89	n/s	YP_004333941 Pseudonocardia_dioxanivorans_CB1190 Non	458.4
AromaDeg	Merged2_4295503_689-1222	n/s	22.27	-23.56	WP_017934574 Nocardioides_sp_Iso805N Non	201.4
AromaDeg	Merged2_4370622_615-1943	-10.24	-5.48	n/s	YP_004331506 Pseudonocardia_dioxanivorans_CB1190 Non	589.3
AromaDeg	Merged2_464734_376-1311	13.44	12.84	n/s	YP_004333941 Pseudonocardia_dioxanivorans_CB1190 Non	302
AromaDeg	Merged2_5228169_723-1589	-9.36	-6.2	n/s	WP_020545991 Nonomuraea_coxensis Non	512.3
AromaDeg	Merged2_5256423_2-274	n/s	20.74	n/s	YP_004333721 Pseudonocardia_dioxanivorans Non	152.9
AromaDeg	Merged2_534212_2705-3736	24.23	8.14	16.09	WP_009739265 Frankia_sp_QA3 Non	491.5
AromaDeg	Merged2_5430661_32-883	-7.99	-5.77	n/s	WP_009739265 Frankia_sp_QA3 Non	427.6
AromaDeg	Merged2_55299_4711-5784	13.77	10.06	n/s	YP_005022825 Vibrio_sp_EJY3 Non	42.7
AromaDeg	Merged2_696712_630-1328	13.58	8.5	n/s	YP_004978866 Burkholderia_sp_YI23 Non	253.4
AromaDeg	Merged2_79161_68-1423	14.14	11.21	n/s	ZP_01011465 Maritimibacter_alkaliphilus_HTCC2654 Non	268.5
AromaDeg	Merged2_80050_2-889	13.38	10.94	n/s	YP_004331506 Pseudonocardia_dioxanivorans_CB1190 Non	276.9
AromaDeg	Merged2_817923_3-584	24.33	36.37	n/s	YP_005043863 Burkholderia_sp_YI23 Non	91.3
AromaDeg	Merged2_840796_5933-7240	24.69	39.6	n/s	EPZ52966 Alicyclobacillus_acidoterrestris_ATCC_49025 Non	555.8
AromaDeg	Merged2_840796_7419-8720	25.81	41.95	n/s	YP_004331506 Pseudonocardia_dioxanivorans_CB1190 Non	566.2
CARD	Merged2_1347390_441-2417	-17.53	-25.21	n/s	gb AAF67494.2 ARO:3002522 novA	469.5
CARD	Merged2_2236544_1-2061	n/s	-28.73	22.13	gb AAC73564.1 ARO:3000216 acrB	499.2
CARD	Merged2_2796197_372-1982	n/s	-33.33	22.03	gb AAG05913.1 ARO:3004072 OpmB	261.2
CARD	Merged2_4197078_2-1540	-12.36	-6.48	n/s	gb AEY83581 ARO:3000510 mupB	451.8
CARD	Merged2_4263479_2-3136	-12.61	n/s	-7.72	gb ACL82957.1 ARO:3002928 vanRM	89
CARD	Merged2_4296752_27868-29202	-26.08	-14.57	-11.51	gb CAA45050.1 ARO:3002828 smB	49.7
CARD	Merged2_4339136_17551-18582	-25.76	-19.42	n/s	gb APB03214.1 ARO:3003980 tetA(48)	225.7
CARD	Merged2_4803523_2-655	-25.43	-18.27	n/s	gb ABX00624.1 ARO:3002881 lmcC	227.3
CARD	Merged2_5024718_1074-1937	-26.34	n/s	-26.42	gb NP_862226.1 ARO:3004639 Corynebacterium	47.4
CARD	Merged2_5052312_58-822	-25.91	n/s	-25.99	gb NP_388442.1 ARO:3004476 vmlR	60.5
CARD	Merged2_515582_339-1277	12.29	10.52	n/s	gb ADP36409.1 ARO:3003730 Bifidobacteria	257.3
CARD	Merged2_520858_3-854	22.82	30.91	n/s	gb AAV85982.1 ARO:3000535 macB	171.4
CARD	Merged2_5287493_2434-3165	-25.79	n/s	-25.86	gb AAC73788.1 ARO:3003841 kdpE	131
CARD	Merged2_5440049_805-1578	-25.83	-17.58	-8.25	gb BAA03674.1 ARO:3001300 myrA	38.1
CAZy	Merged2_205627_1-1206	13.79	10.61	n/s	AYG04081.1 GH3	648.7
CAZy	Merged2_281608_187-1065	12.61	12.71	n/s	AYG02734.1 GT51	407.1
CAZy	Merged2_290204_1-612	12.63	12	n/s	AYG02714.1 GH0	417.9
CAZy	Merged2_360077_1354-2079	11.54	12.6	n/s	AYG04270.1 GT2	432.6
CAZy	Merged2_4502565_874-2007	-10.25	n/s	n/s	QUQ64728.1 CBM35 GH31	566.2
CAZy	Merged2_463292_2-682	12.74	10.4	n/s	AYG02303.1 GH77	382.9
CAZy	Merged2_4991804_2-2488	-8.96	-7.66	n/s	QZA17584.1 GT53	1,398.60
CAZy	Merged2_5007895_2370-3887	-10.49	n/s	n/s	ACU76286.1 CBM5	387.9
CAZy	Merged2_5052611_3-3326	-10.3	n/s	n/s	QYN20448.1 GH106	1,817.70
CAZy	Merged2_529559_541-1530	13.82	10.7	n/s	AYG05512.1 GT4	568.2
CAZy	Merged2_56028_2-997	13.33	10.29	n/s	AYG02501.1 GH5_22	637.9
CAZy	Merged2_56983_1-1659	12.64	10.23	n/s	QGN32350.1 GT35	1033.1
CAZy	Merged2_687490_1-1236	13.74	10.38	n/s	AYG03523.1 GH65	740.7
CAZy	Merged2_818860_1-942	12.81	12.18	n/s	AYG02303.1 GH77	535
CAZy	Merged2_839819_337-1104	12.57	12.67	n/s	AYG02714.1 GH0	46.2
CAZy	Merged2_852959_1-867	12.58	9.17	n/s	AYG03103.1 GH2	539.3

Bit-score values below 50 have been indicated in red-color font. FC, fold change; DB, database; ID, identifier; Nas, *Nasutitermes*; Ban, *Anoplotermes*; Cub, *Cubitermes*; n/s, not significant. Red shading, under-represented or under-expressed genes. Green-shading, over-represented or over-expressed genes.

(Pandit et al., 2016). That observation agrees with the type of diet in *Anoplotermes*, which will find lignin-enriched sources in the soil, and by the construction of the nest, this type of bacteria can be enriched. Additionally, we could speculate that this bacteria is in charge of eliminating the growth of hyphae from fungi due to its ability to grow on fungal material (McKee et al., 2019). Perhaps, the role of this genus in the

Anoplotermes nest could be related to the recycling of dead material from fungi.

4.2.2. Amplicon sequencing: Fungi

The ACE index calculations informed of a lack of differences in the geographical origin and species. Differences exist according

to diet and phylogenetic relatedness, which can drive a differential richness in the nest fungal communities.

The PCA plot built in R (Figure 5) indicated a stronger effect of genera and feeding strategy. We believe that genera were even stronger in sample stratification than the type of diet because even two genera in the same feeding group are clearly separated (*Anoplotermes* and *Neocapritermes*). Finally, while it existed, the geographical influence was weaker than the other two. These results are in overall agreement with our observations of the bacterial communities. In keeping with these results, ANOSIM showed a stronger influence of diet and genus over geographical origin.

4.2.3. Shotgun metagenomic sequencing

We have used different ways of presenting these data due to the massive amount of information that could be extracted. Feeding group and genera stratify samples with limited influence from the geographical location (Figure 6). Despite small variability between nests from species belonging to the same genus, even communities from soil-feeding termite nests show closer genomic and taxonomic composition at both sides of the Atlantic, as revealed by the NMDS analysis that takes into account the full metagenome.

At the phylum level (Figure 7A), we observe a comparable situation, where all *Nasutitermes* have a closer profile among themselves than those of *Cubitermes* or *Anoplotermes*. There are still differences between samples within the same genera (except for *Nasutitermes*). One of the samples from *Anoplotermes banksi* shows large differences with the other two samples from this species, probably owing to the worst state of that sample when arriving in our laboratory. The three samples from *Nasutitermes lujae* are very stable in terms of taxonomic representation. We can conclude that the taxonomic composition of the nests reflects the genera (where it seems to be stable) and especially eating habits. Taxonomic metabarcoding through 16S amplicon sequencing shows similar trends with the metagenomic bacterial annotations, with a notable presence of *Actinobacteria*. However, that is not the case for *Proteobacteria*. We found that the most abundant order in many of the sequenced soil feeders was *Streptosporangiales*, whereas *Nasutitermes* belong to the class *Actinomycetia*, either to genera *Grylotalpicola* in three cases or *Mycobacterium* in the last one. Differences in the results of both sequencing approaches have been previously addressed in the literature (Rausch et al., 2019).

Metagenomics can provide very interesting insights into the microbial community by dissecting the functional potential. Given a certain function, we could observe if at that level exists a differential composition in terms of taxa. We carried out a subset of functions (Puente-Sánchez et al., 2020) related to “phenylalanine, tyrosine, and tryptophan biosynthesis.” These three aromatic α -amino acids are involved in the metabolism of secondary metabolites (Parthasarathy et al., 2018). We found apparent differences according to the feeding strategies (Figure 7B). Once more, both soil-feeding genera resemble each other, while the *Nasutitermes* keep a closer profile among them. This indicates that the differences are not only at the metabolomic level but also at the overall taxonomic structure of the community (both amplicon and metagenomic sequencing data). The differences go beyond this, and certain functions are carried out by different

microorganisms. While nests from soil-feeding termites display similar profiles on both sides of the Atlantic, those from wood and soil feeding are more dissimilar. Considering that, the geographical component seems to have a weaker contribution than the feeding habits in driving the functional differences among communities. There seems to be a small reflection of a component that relates communities from the African environment, as we observe a group of unclassified *Actinobacteria* with a similar proportion between *Nasutitermes* and *Cubitermes*.

Our explanation here is that the feeding habits of termites and possibly some termite-active antimicrobial strategies are creating a set of conditions that allow the growth of certain types of genera, which will carry the required set of community functions, in this case, the metabolism of aromatic amino acids. We mentioned the termite active strategies because our study included terpene as a selective metabolite, which has antimicrobial properties and is secreted by *Nasutitermes*, as both are documented in the literature. The defensive or protective component of feces was greatly reviewed by Cole et al. (2021). In our study, we inferred that the combination of active strategies and passive constraints imposed by the feeding strategy created a specific growth medium for selected microorganisms. We could establish parallelism with microbiology work in the laboratory and state that termites with different feeding habits have formulated a growth medium for their allied microbiome to thrive. Several iterations of this process through trial and error in evolutionary time have probably fine-tuned these strategies, where the most successful in ensuring the nest and colony survival have been fixed. To such an extent that microbial communities associated with two different genera of soil feeders resemble in structure on both sides of the Atlantic Ocean, where the different flora composition imposes a different soil community (Lima-Perim et al., 2016). However, despite these stated differences even at short distances belonging to two clearly distinct biotopes, the differences between the *Cubitermes* from CM and MW are minor. Thus, this is one more observation that helps us to associate a lower discrimination power to the geographical location. The recruitment of external or surrounding strains reported in the literature can be different in terms of specific strains, but they will be fundamentally equivalent in the functions that they carry out. Soil feeders “will recruit” microorganisms that are performing or carrying the same genes. We want to highlight that “recruit” may not be an active termite program, but a consequence of the complex crosstalk between active antimicrobial strategies, and the effect of diet and phylogenetic relatedness.

Among the communities with a higher presence in *Nasutitermes*, the genus *Grylotalpicola* has been found in larval galleries from the pine beetle *Monochamus alternatus* collected in the Chinese Jiangxi province (Chen et al., 2020) and was previously isolated from the guts of the wood-feeding *Reticulitermes chinensis* Snyder, sampled in Wuhan, China (Fang et al., 2015). The interesting connection is that members from this genus have been found in insects with wood-feeding habits at distant locations. *Streptosporangiales* is an order of bacteria that can obtain carbon from green waste (Cai et al., 2018), especially at the middle composting stage (Li et al., 2021). They can secrete cellulases and hemicellulases for lignocellulosic degradation, besides being producers of antibiotics that suppress pathogens (Gomes et al., 2017; Waglechner et al., 2019). These organisms are

highly abundant in soil (McCarthy and Williams, 1992). Thus, since soil-feeding genera use soil as their substrate, *Streptosporangiales* may originate in their surrounding environment and be enriched due to the substrate type and nutrients present. They could provide a cleaner environment at the termite nest due to the production of antimicrobial compounds. This strategy, which we cannot ascertain if it is an active strategy or one derived from the termite activity, repeats itself on both sides of the Atlantic Ocean. The genus *Microclunatus* (*Actinobacteria*) has been previously reported in *Mycocepurus smithii* ants (Kellner et al., 2015), a fungus-farming ant reported to acquire its microbiome from the surrounding environment. Bacteria from this genus have been found in deep-sea sediments (Jroundi et al., 2020) to rhizospheric soil (Wang et al., 2008), and it has been indicated that they have the potential for pollution management (Zhang et al., 2017), especially for capabilities of phosphorus accumulation (Zhong et al., 2018). *Microbacteriaceae* is a family of Gram-positive bacteria with members that have been reported among the dominant cellulase-producing strains isolated from yellow stem borers (*Scirpophaga incertulas*; Bashir et al., 2013), but in much less proportion in the guts of termites at the same study identified as *Odontotermes hainanensis*, which were feeding on decomposed trees and leaf litter.

The LEFSE analysis on the ORFs annotated at PFAM has found that each of the genera has a specific signature (Figure 8), which was also a conclusion that we could get after the quick interrogation of PFAM annotation (Supplementary Figures 4, 5). There are differences when we seek terms that are related to key aspects of the microbial lifestyle (e.g., cellulose-annotated genes). But are these apparent differences in specific functions translated into differential content of certain genes beyond a LEFSE analysis or our quick approach? Has the microbial community gained or is being enriched in certain functions? To answer these questions, we applied a transcriptomics statistical test for finding differentially expressed genes (DGEs); in our case, we adopted the term DPGs. Testing for significant statistical differences will illustrate if a certain function is highly abundant in a nest microbial community when compared with another. We must observe that in a metagenomics context, we will not get an idea of “active use” but more of “potential use.” While not all present genes will be expressed, we can conclude that a content increase in certain functions is relevant and not random. That is the rationale behind using qPCRs to quantify the differences in gene content due to pharmaceutical pollution among different sites (González-Plaza et al., 2019; Milaković et al., 2019).

While TPM is a normalization method generally used in RNAseq experiments and its data treatment, according to Puente-Sánchez et al. (2020), it can be very useful in metagenomic experiments. In that regard, we found it to be a very important value that allows for sample comparison and general overviews (Figure 6). However, because we understand the limitations of TPM use (Zhao et al., 2020), we have carried out additional normalizations for comparison purposes using procedures that use raw count values and are normalized differently (González Plaza, in preparation).

Comparison based on the normalized count values for the PFAM-annotated ORFs clearly separates *Nasutitermes* from the soil feeders (Figure 9), again a division by the

termite diet and phylogenetic relatedness. We found that the differences between the microbial communities in wood-feeding termites' nests and those in soil-feeding termites' nests are much stronger. We carried out a simple analysis, extracting terms related to the different metabolisms of the communities (e.g., cellulose; Supplementary Figures 4, 5), where we observed genes such as cellulose synthases that play a strong role in *Nasutitermes*.

This led us to carry out a differential gene content analysis with the results obtained in specialized databases and observed a similarity among soil feeders and great differences to the wood feeder genus. The differences in the use of CAZy-annotated genes were small and absent between both soil feeders. We must note that this could be a resemblance to the smaller number of genes that we used for this analysis due to computing limitations. If proportions are kept regarding what we observe in the other comparisons, a higher number of DPGs between *Nasutitermes* and each of the soil feeders, it is logical to think that we did not have enough number of “events” for statistical differences to appear. We found interesting differences, as three of the most differentially present genes in the wood-feeder nest belong to the glycosyl transferase family (GT 2, 4, and 51), previously reported in woodchip communities but not related to the cellulose hydrolysis (Nnadozie et al., 2017).

The rationale for this additional genetic prospecting is that these databases can reflect different aspects of the microbial communities, which can be manifold and related to metabolism as a cornerstone (Schmidt et al., 2019). Genes annotated in AromaDeg could serve for protection purposes when organic antimicrobial compounds such as terpenes are used, but they could also be part of a metabolism that could drive social microbial interactions. The degradation of aromatic hydrocarbon compounds was shown in the guts of wood-feeding termites (Ke et al., 2011), while in our study, the microbial communities could be utilizing lignin-related aromatic compounds. The differences observed between wood and soil feeders (in their nests) are owed to the different inputs of compounds due to the diet, as soil feeders have a great input of polyaromatic compounds (Ohkuma, 2003). Antibiotic resistance can be used among members of the community to defend themselves against antibiotics, either from those produced by them or from competing organisms (Scherlach and Hertweck, 2020). Additionally, at the usual concentrations of antibiotics found in natural communities, these genes can serve for communication purposes rather than survival (Romero et al., 2011). Besides, we found it interesting to query the CARD database because the antibiotic resistance genes harbored in wildlife-associated microbial communities have not been properly addressed (Dolejska and Literak, 2019). The low number of described genes at BactiBase or AnhyDeg indicates the potential of these termite nest microbial communities for being a source of novel genes that can be used in fighting antimicrobial resistance (bacteriocins). MacB was one of the most dominant genes in *Nasutitermes*, a gene reported in several cultivation ecosystems (Fan et al., 2021).

Hu et al. (2019) suggested the linkage between diet-microbiome in termite guts, a widespread observation in other animal models (Colman et al., 2012; David et al., 2013; Baker et al., 2014; Flint et al., 2015; Otani et al., 2019; Leite-Mondin et al., 2021). However, through three omics approaches, we state that these differences

go beyond the gut. Diet shapes the nest community structure because of the building strategy, without forgetting that there is a contribution of phylogenetic relatedness fine-tuned by the geographical location.

5. Conclusion

We found differences in the overall metabolite profile indicating a stronger effect of the feeding habits and phylogenetic relatedness than the geographical location, where some selective metabolites may have an antimicrobial role. Naphthalene appeared in nests from *Cubitermes* and *Nasutitermes*.

The bacterial and fungal taxonomical profiles point toward a stratification mainly owed to diet and phylogenetic relatedness, which is potentially explained since nest material is a product of termite feces and surrounding soil. Even the differences between soil-feeding termites' nests from both African sites were minor.

Metagenomic sequencing supports the above results, where the major drivers for genetic functional differences are diet (greater effect over fungi) and phylogenetic relatedness (greater effect over bacteria). The genetic differences between both soil-feeding termites were of a more limited entity than those between the wood-feeding ones. Our mining approaches with specialized databases show a vast genetic richness of functions and genes for biotechnological applications. This is only one example of the virtually unlimited genetic diversity that tropical rainforests harbor and should draw attention to the necessity of preserving this resource that could be applied in fields such as the bioremediation of organic compounds or the synthesis of biofuels.

Many studies have indicated that termites recruit microbial members from the surrounding environment. Our results also indicate that the microbiome genetic functions in soil-feeding termites are similar across the Atlantic, even in savannah and rainforest regions. What ultimately fixes the functional genomic content is diet, the nutrients that are available, where diet is a product of evolutionary history.

Translated into our world, controlling the nutrients that microbial communities receive or are exposed to leads to dramatic changes in the composition and, more importantly, the functional genetic content. These results can inspire methodologies to drive the composition and functions of microbial communities and limit the spread of pathogens.

Further studies considering the different microbial niches (gut, environment, food substrate, and nest) are necessary to understand the relationship between the microbial communities under the influence of termites and those in the surrounding environment.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: <https://www.ebi.ac.uk/ena>, PRJEB55779.

Ethics statement

Sampling was carried out in accordance with the Nagoya guidelines for biodiversity sampling ([Convention on Biological Diversity, 2010](#)), with the required research permits to Jan Šobotník and other members of FFWS, and in accordance with the ethical behavior toward local communities.

Author contributions

JG and JH designed the study. JG participated in sample collection, designed the sampling strategy, isolated DNA and COII-identified many specimens used in this study, processed nest material to powder, carried out all experiments regarding nest DNA isolation, sequencing data processing, analysis, and wrote the main body of the manuscript. JH performed the metabolite profiling setup, experiments, data processing, writing of the corresponding metabolite sections in the manuscript, and participated in article writing and data interpretation. All authors contributed to the article and approved the submitted version.

Funding

This project has been funded by grants IGA No. B_19_04 from the Faculty of Forestry and Wood Sciences (FFWS, Czech University of Life Sciences, Prague); and EVA4.0: Advanced research supporting the forestry and wood-processing sector's adaptation to global change and the 4th industrial revolution, OP RDE, Ministry of Education, Youth, and Sports of Czechia, Grant No. CZ.02.1.01/0.0/0.0/16_019/0000803.

Acknowledgments

We acknowledge María Suárez at Wageningen University (the Netherlands) for support with scripts. We acknowledge the role of Tereza Beranková for the identification of termites through COII sequencing and Jan Šobotník for constructive comments on our study. We thank the EVA4.0 management team at FFWS.

Conflict of interest

JG discloses an International PCT Application No. PCT/CZ2022/050072, entitled, A bacteriocin composition for coding Linocin-M18-like protein with antimicrobial activity and its usage, filed on 5 August 2022 and is related to the results of this study.

The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of

their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1089525/full#supplementary-material>

References

- Aguero, C. M., Eyer, P. A., Crippen, T. L., and Vargo, E. L. (2021). Reduced environmental microbial diversity on the cuticle and in the galleries of a subterranean termite compared to surrounding soil. *Microb. Ecol.* 81, 1054–1063. doi: 10.1007/s00248-020-01664-w
- Alcock, B. P., Raphenya, A. R., Lau, T. T. Y., Tsang, K. K., Bouchard, M., Edalatmand, A., et al. (2020). CARD 2020: Antibiotic resistance surveillance with the comprehensive antibiotic resistance database. *Nucl. Acids Res.* 48, D517–D525. doi: 10.1093/nar/gkz935
- Allingham, S. M., and Harvey, M. (2013). Effects of different fire regimes on amphibian communities in the Nyika National Park. *Malawi* 32, 1–8. doi: 10.5358/hsj.32.1
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Arora, R., Sharma, N. K., Kumar, S., and Sani, R. K. (2019). Lignocellulosic ethanol: Feedstocks and bioprocessing. *Bioethanol Prod.* 6, 165–185. doi: 10.1016/B978-0-12-813766-6.00009-6
- Auer, L., Lazuka, A., Sillam-Dussès, D., Miambi, E., O'Donohue, M., and Hernandez-Raquet, G. (2017). Uncovering the potential of termite gut microbiome for lignocellulose bioconversion in anaerobic batch bioreactors. *Front. Microbiol.* 8, 2623. doi: 10.3389/fmicb.2017.02623
- Baker, R., Buckland, A., and Sheaves, M. (2014). Fish gut content analysis: Robust measures of diet composition. *Fish Fish* 15, 170–177. doi: 10.1111/faf.12026
- Bashir, Z., Kondapalli, V. K., Adlakha, N., Sharma, A., Bhatnagar, R. K., Chandel, G., et al. (2013). Diversity and functional significance of cellulolytic microbes living in termite, pill-bug and stem-borer guts. *Sci. Rep.* 3, 1–11. doi: 10.1038/srep02558
- Beeckmans, S., and Xie, J. P. (2015). Glyoxylate cycle. *Ref. Modul. Biomed. Sci.* 5, 2440. doi: 10.1016/B978-0-12-801238-3.02440-5
- Benjamino, J., and Graf, J. (2016). Characterization of the core and caste-specific microbiota in the Termite, *Reticulitermes flavipes*. *Front. Microbiol.* 7, 171. doi: 10.3389/fmicb.2016.00171
- Bokulich, N. A., Kaehler, B. D., Rideout, J. R., Dillon, M., Bolyen, E., Knight, R., et al. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome* 6, 1–17. doi: 10.1186/s40168-018-0470-z
- Bokulich, N. A., Subramanian, S., Faith, J. J., Gevers, D., Gordon, J. I., Knight, R., et al. (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nat. Methods* 10, 57–59. doi: 10.1038/nmeth.2276
- Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120. doi: 10.1093/bioinformatics/btu170
- Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., et al. (2019). Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat. Biotechnol.* 37, 852–857. doi: 10.1038/s41587-019-0209-9
- Bowman, J. P. (2011). Protein-based analysis and other new and emerging non-nucleic acid based methods for tracing and investigating foodborne pathogens. *Tracing Pathog. Food Chain* 2011, 292–341. doi: 10.1533/9780857090508.3.292
- Bray, J. R., and Curtis, J. T. (1957). An ordination of the upland forest communities of Southern Wisconsin. *Ecol. Monogr.* 27, 325–349. doi: 10.2307/1942268
- Brown, C. T., Crusoe, M. R., Alameldin, H. F., Awad, S., Boucher, E., Caldwell, A., et al. (2015). The khmer software package: Enabling efficient nucleotide sequence analysis. *F1000Research* 4, 900. doi: 10.12688/f1000research.6924.1
- Brown, C. T., Howe, A., Zhang, Q., Pyrkosz, A. B., and Brom, T. H. (2012). A Reference-Free Algorithm for Computational Normalization of Shotgun Sequencing Data. Ithaca, NY: Cornell University.
- Brown, C. T., Scott, C., Crusoe, M., Sheneman, L., Rosenthal, J., and Howe, A. (2013). KHMER-Protocols Documentation. *Figshare*. Available online at: https://figshare.com/articles/dataset/khmer_protocols_0_8_3_documentation/878460
- Brune, A. (1998). Termite guts: The world's smallest bioreactors. *Trends Biotechnol.* 16, 16–21. doi: 10.1016/S0167-7799(97)01151-7
- Buchfink, B., Xie, C., and Huson, D. H. (2014). Fast and sensitive protein alignment using DIAMOND. *Nat. Methods* 12, 59–60. doi: 10.1038/nmeth.3176
- Cai, L., Gong, X., Sun, X., Li, S., and Yu, X. (2018). Comparison of chemical and microbiological changes during the aerobic composting and vermicomposting of green waste. *PLoS ONE* 13, e0207494. doi: 10.1371/journal.pone.0207494
- Callaghan, A. V., and Wawrik, B. (2016). *AnHyDeg: A Curated Database of Anaerobic Hydrocarbon Degradation Genes*. Available online at: <https://github.com/AnaerobesRock/AnHyDeg/tree/v1.0>
- Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., and Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nat. Methods* 13, 581–583. doi: 10.1038/nmeth.3869
- Cao, X. L. (2012). A review recent development on analytical methods for determination of bisphenol a in food and biological samples. *J. Liq. Chromatogr. Relat. Technol.* 35, 2795–2829. doi: 10.1080/10826076.2012.720325
- Caspi, R., Billington, R., Fulcher, C. A., Keseler, I. M., Kothari, A., Krummenacker, M., et al. (2018). The MetaCyc database of metabolic pathways and enzymes. *Nucl. Acids Res.* 46, D633–D639. doi: 10.1093/nar/gkx935
- Chen, H., Hao, D., Wei, Z., Wang, L., and Lin, T. (2020). Bacterial communities associated with the pine wilt disease insect vector *Monochamus alternatus* (Coleoptera: Cerambycidae) during the Larvae and Pupae stages. *Insects* 11, 376. doi: 10.3390/insects11060376
- Chen, J., Henderson, G., Grimm, C. C., Lloyd, S. W., and Laine, R. A. (1998a). Naphthalene in formosan subterranean termite carton nests. *J. Agric. Food Chem.* 46, 2337–2339. doi: 10.1021/jf9709717
- Chen, J., Henderson, G., Grimm, C. C., Lloyd, S. W., and Laine, R. A. (1998b). Termites fumigate their nests with naphthalene. *Nature* 392, 558–559. doi: 10.1038/33305
- Chong, G. G., Huang, X. J., Di, J. H., Xu, D. Z., He, Y. C., Pei, Y. N., et al. (2018). Biodegradation of alkali lignin by a newly isolated *Rhodococcus pyridinivorans* CCZU-B16. *Bioprocess Biosyst. Eng.* 41, 501–510. doi: 10.1007/s00449-017-1884-x
- Chouvenc, T., Efstathion, C. A., Elliott, M. L., and Su, N.-Y. (2013). Extended disease resistance emerging from the faecal nest of a subterranean termite. *Proc. R. Soc. B Biol. Sci.* 280, 20131885. doi: 10.1098/rspb.2013.1885
- Chouvenc, T., Elliott, M. L., Šobotník, J., Efstathion, C. A., and Su, N.-Y. (2018). The termite fecal nest: A framework for the opportunistic acquisition of beneficial soil streptomycetes (Actinomycetales: Streptomycetaceae). *Environ. Entomol.* 47, 1431–1439. doi: 10.1093/ee/nvy152
- Clarholm, M. (1981). Protozoan grazing of bacteria in soil—impact and importance. *Microb. Ecol.* 7, 343–350. doi: 10.1007/BF02341429
- Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., and Sayers, E. W. (2016). GenBank. *Nucl. Acids Res.* 44, D67–D72. doi: 10.1093/nar/gk/v1276
- Colas, F., Chanudet, V., Daufresne, M., Buchet, L., Vigouroux, R., Bonnet, A., et al. (2020). Spatial and temporal variability of diffusive CO₂ and CH₄ fluxes from the Amazonian Reservoir Petit-Saut (French Guiana) reveals the importance of allochthonous inputs for long-term C emissions. *Glob. Biogeochem. Cycles* 34, e2020GB006602. doi: 10.1029/2020GB006602
- Cole, M. E., Ceja-Navarro, J. A., and Mikaelyan, A. (2021). The power of poop: Defecation behaviors and social hygiene in insects. *PLoS Pathog.* 17, e1009964. doi: 10.1371/journal.ppat.1009964
- Colman, D. R., Toolson, E. C., and Takacs-Vesbach, C. D. (2012). Do diet and taxonomy influence insect gut bacterial communities? *Mol. Ecol.* 21, 5124–5137. doi: 10.1111/j.1365-294X.2012.05752.x
- Convention on Biological Diversity (2010). “The strategic plan for biodiversity 2011–2020,” in Tenth Meeting of the Conference of the Parties to the Convention on Biological Diversity, Nagoya.

- David, L. A., Maurice, C. F., Carmody, R. N., Gootenberg, D. B., Button, J. E., Wolfe, B. E., et al. (2013). Diet rapidly and reproducibly alters the human gut microbiome. *Nature* 505, 559–563. doi: 10.1038/nature12820
- De Mello, A. P., De Moraes, M. M., Da Câmara, C. A. G., and Vasconcellos, A. (2021). Effect of spatial variation on defensive substances of constrictotermes cyphogaster soldiers (Blattaria, Isoptera). *J. Chem. Ecol.* 47, 544–551. doi: 10.1007/s10886-021-01271-0
- Dedysh, S. N., Kulichevskaya, I. S., Beletsky, A. V., Ivanova, A. A., Rijpstra, W. I. C., Damsté, J. S. S., et al. (2020). *Lacipirellula parvula* gen. nov., sp. nov., representing a lineage of planctomycetes widespread in low-oxygen habitats, description of the family *Lacipirellulaceae* fam. nov. and proposal of the orders *Pirellulales* ord. nov., *Gemmatales* ord. nov. and *Isophaerales* ord. nov. *Syst. Appl. Microbiol.* 43, 126050. doi: 10.1016/j.syapm.2019.126050
- DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., et al. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl. Environ. Microbiol.* 72, 5069–5072. doi: 10.1128/AEM.03006-05
- Do, T. H., Nguyen, T. T., Nguyen, T. N., Le, Q. G., Nguyen, C., Kimura, K., et al. (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. *J. Biosci. Biotech.* 118, 665–671. doi: 10.1016/j.jbiosc.2014.05.010
- Dolejska, M., and Literak, I. (2019). Wildlife is overlooked in the epidemiology of medically important antibiotic-resistant bacteria. *Antimicrob. Agents Chemother.* 63:e01167-19. doi: 10.1128/AAC.01167-19
- Dolejšová, K., Krasulová, J., Kotalová, K., and Hanus, R. (2014). Chemical alarm in the termite *Termitogeton planus* (Rhinotermitidae). *J. Chem. Ecol.* 40, 1269–1276. doi: 10.1007/s10886-014-0515-0
- Duarte, M., Jauregui, R., Vilchez-Vargas, R., Junca, H., and Pieper, D. H. (2014). AromaDeg, a novel database for phylogenomics of aerobic bacterial degradation of aromatics. *Database* 2014, bau118. doi: 10.1093/database/bau118
- Eddy, S. R. (2011). Accelerated profile HMM searches. *PLoS Comput. Biol.* 7, 1002195. doi: 10.1371/journal.pcbi.1002195
- Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* 27, 2194–2200. doi: 10.1093/bioinformatics/btr381
- El-Naggar, N. E.-A. (2021). Streptomyces-based cell factories for production of biomolecules and bioactive metabolites. *Microb. Cell Factories Eng. Prod. Biomol.* 8, 183–234. doi: 10.1016/B978-0-12-821477-0.00011-8
- Enagbonma, B. J., Ajillogba, C. F., and Babalola, O. O. (2020). Metagenomic profiling of bacterial diversity and community structure in termite mounds and surrounding soils. *Arch. Microbiol.* 202, 2697–2709. doi: 10.1007/s00203-020-01994-w
- Enault, F., Briet, A., Bouteille, L., Roux, S., Sullivan, M. B., and Petit, M. A. (2016). Phages rarely encode antibiotic resistance genes: A cautionary tale for virome analyses. *ISME J.* 11, 237–247. doi: 10.1101/053025
- Erlacher, A., Cernava, T., Cardinale, M., Soh, J., Sensen, C. W., Grube, M., et al. (2015). Rhizobiales as functional and endosymbiotic members in the lichen symbiosis of *Lobaria pulmonaria* L. *Front. Microbiol.* 6, 53. doi: 10.3389/fmicb.2015.00053
- Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biol. Conserv.* 61, 1–10. doi: 10.1016/0006-3207(92)91201-3
- Fan, L., Li, F., Chen, X., Liping, Q., Dong, X., Hu, G., et al. (2021). Metagenomics analysis reveals the distribution and communication of antibiotic resistance genes within two different red swamp crayfish *Procambarus clarkii* cultivation ecosystems. *Environ. Pollut.* 285, 117144. doi: 10.1016/j.envpol.2021.117144
- Fang, H., Lv, W., Huang, Z., Liu, S. J., and Yang, H. (2015). *Gryllotalpica reticulitermitis* sp. nov., isolated from a termite gut. *Int. J. Syst. Evol. Microbiol.* 65, 85–89. doi: 10.1099/ijs.0.062984-0
- Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2016). The Pfam protein families database: Towards a more sustainable future. *Nucl. Acids Res.* 44, D279–D285. doi: 10.1093/nar/gkv1344
- Flint, H. J., Duncan, S. H., Scott, K. P., and Louis, P. (2015). Links between diet, gut microbiota composition and gut metabolism. *Proc. Nutr. Soc.* 74, 13–22. doi: 10.1017/S0029665114001463
- Funnell, M. I. G., Pinheiro, D. G., Gomes-Pepe, E. S., de Carvalho, L. A. L., Campanharo, J. C., Fernandes, C. C., et al. (2021). Metagenome-assembled genome of a *Chitinophaga* sp. and its potential in plant biomass degradation, as well of affiliated *Pandoraea* and *Labrys* species. *World J. Microbiol. Biotechnol.* 37, 1–17. doi: 10.1007/s11274-021-03128-w
- Gershenzon, J., and Dudareva, N. (2007). The function of terpene natural products in the natural world. *Nat. Chem. Biol.* 3, 408–414. doi: 10.1038/nchembio.2007.5
- Gomes, K. M., Duarte, R. S., and Bastos, M. C. F. (2017). Lantibiotics produced by Actinobacteria and their potential applications (A review). *Microbiology* 163, 109–121. doi: 10.1099/mic.0.000397
- González-Plaza, J. J., Blau, K., Milaković, M., Jurina, T., Smalla, K., and Udiković-Kolić, N. (2019). Antibiotic-manufacturing sites are hot-spots for the release and spread of antibiotic resistance genes and mobile genetic elements in receiving aquatic environments. *Environ. Int.* 130, 104735. doi: 10.1016/j.envint.2019.04.007
- Goodfellow, M., and Williams, S. T. (2003). Ecology of actinomycetes. *Annu. Rev. Microbiol.* 37, 189–216. doi: 10.1146/annurev.mi.37.100183.001201
- Griffiths, B. S., Bracewell, J. M., Robertson, G. W., and Bignell, D. E. (2013). Pyrolysis–mass spectrometry confirms enrichment of lignin in the faeces of a wood-feeding termite, *Zootermopsis nevadensis* and depletion of peptides in a soil-feeder, *Cubitermes ugandensis*. *Soil Biol. Biochem.* 57, 957–959. doi: 10.1016/j.soilbio.2012.08.012
- Hammami, R., Zouhir, A., Le Lay, C., Ben Hamida, J., and Fliss, I. (2010). BACTIBASE second release: a database and tool platform for bacteriocin characterization. *BMC Microbiol.* 10, 22. doi: 10.1186/1471-2180-10-22
- Hu, H., da Costa, R. R., Pilgaard, B., Schiøtt, M., Lange, L., and Poulsen, M. (2019). Fungiculture in termites is associated with a mycolytic gut bacterial community. *mSphere* 4, 19. doi: 10.1128/mSphere.00165-19
- 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. *Nucl. Acids Res.* 44, D286–D293. doi: 10.1093/nar/gkv1248
- Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinform.* 11, 1–11. doi: 10.1186/1471-2105-11-119
- Jordan, C. F., and Herrera, R. (2015). Tropical rain forests: Are nutrients really critical? *Am. Nat.* 117, 167–180.
- Jose, J. J. S., Montes, R., Stansly, P. A., and Bentley, B. L. (1989). Environmental factors related to the occurrence of mound-building nasute termites in trachypogon savannas of the orinoco llanos. *Biotropica* 21, 353. doi: 10.2307/2388286
- Jroundi, F., Martinez-Ruiz, F., Merroun, M. L., and Gonzalez-Muñoz, M. T. (2020). Exploring bacterial community composition in Mediterranean deep-sea sediments and their role in heavy metal accumulation. *Sci. Total Environ.* 712, 135660. doi: 10.1016/j.scitotenv.2019.135660
- Kaehler, B. D., Bokulich, N. A., McDonald, D., Knight, R., Caporaso, J. G., and Huttenhower, G. A. (2019). Species abundance information improves sequence taxonomy classification accuracy. *Nat. Commun.* 10, 4643. doi: 10.1038/s41467-019-12669-6
- Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucl. Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27
- Kang, D. D., Li, F., Kirton, E., Thomas, A., Egan, R., An, H., et al. (2019). MetaBAT 2: An adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 2019, e7359. doi: 10.7717/peerj.7359
- Katoh, K., Misawa, K., Kuma, K. I., and Miyata, T. (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucl. Acids Res.* 30, 3059–3066. doi: 10.1093/nar/gkf436
- Ke, J., Singh, D., and Chen, S. (2011). Aromatic compound degradation by the wood-feeding termite *Coptotermes formosanus* (Shiraki). *Int. Biodeterior. Biodegrad.* 65, 16. doi: 10.1016/j.ibiod.2010.12.016
- Kellner, K., Ishak, H. D., Linksvayer, T. A., and Mueller, U. G. (2015). Bacterial community composition and diversity in an ancestral ant fungus symbiosis. *FEMS Microbiol. Ecol.* 91, 73. doi: 10.1093/femsec/fiv073
- Knoben, W. J. M., Woods, R. A., and Freer, J. E. (2019). Global bimodal precipitation seasonality: A systematic overview. *Int. J. Climatol.* 39, 558–567. doi: 10.1002/joc.5786
- Krishanti, N. P. R. A., Zulfinah, D., Wikantyo, B., Zulfitri, A., and Yusuf, S. (2018). Antimicrobial production by an actinomycetes isolated from the termite nest. *J. Trop. Life Sci.* 8, 279–288. doi: 10.11594/jtls.08.03.10
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Laslett, D., and Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucl. Acids Res.* 32, 11–16. doi: 10.1093/nar/gkh152
- Legendre, P., and Gallagher, E. D. (2001). Ecologically meaningful transformations for ordination of species data. *Oecologia* 129, 271–280. doi: 10.1007/s004420100716
- Leite-Mondin, M., DiLegge, M. J., Manter, D. K., Weir, T. L., Silva-Filho, M. C., and Vivanco, J. M. (2021). The gut microbiota composition of *Trichoplusia ni* is altered by diet and may influence its polyphagous behavior. *Sci. Rep.* 11, 1–16. doi: 10.1038/s41598-021-85057-0
- Lepleux, C., Turpault, M. P., Oger, P., Frey-Klett, P., and Uroz, S. (2012). Correlation of the abundance of betaproteobacteria on mineral surfaces with mineral weathering in forest soils. *Appl. Environ. Microbiol.* 78, 7114–7119. doi: 10.1128/AEM.00996-12
- Li, D., Liu, C.-M., Luo, R., Sadakane, K., and Lam, T.-W. (2015). MEGAHIT: An ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31, 1674–1676. doi: 10.1093/bioinformatics/btv033
- Li, L., Guo, X. P., Zhao, T. N., Liu, L., and Li, T. Y. (2021). Identifying the key environmental factors and bacterial communities in humification and their

- relationships during green waste composting. *Appl. Ecol. Environ. Res.* 19, 45–62. doi: 10.15666/aecer/1901_045062
- Lima-Perim, J. E., Romagnoli, E. M., Dini-Andreote, F., Durrer, A., Dias, A. C. F., and Andreote, F. D. (2016). Linking the composition of bacterial and archaeal communities to characteristics of soil and flora composition in the atlantic rainforest. *PLoS ONE* 11, e0146566. doi: 10.1371/journal.pone.0146566
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The carbohydrate-active enzymes database (CAZy) in 2013. *Nucl. Acids Res.* 42, D490–D495. doi: 10.1093/nar/gkt1178
- Long, Y. H., Xie, L., Liu, N., Yan, X., Li, M. H., Fan, M. Z., et al. (2010). Comparison of gut-associated and nest-associated microbial communities of a fungus-growing termite (*Odontotermes yunnanensis*). *Insect Sci.* 17, 265–276. doi: 10.1111/j.1744-7917.2010.01327.x
- López-Hernández, D. (2001). Nutrient dynamics (C, N and P) in termite mounds of *Nasutitermes ephratae* from savannas of the Orinoco Llanos (Venezuela). *Soil Biol. Biochem.* 33, 747–753. doi: 10.1016/S0038-0717(00)00220-0
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 1–21. doi: 10.1186/s13059-014-0550-8
- Lozupone, C., and Knight, R. (2005). UniFrac: A new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* 71, 8228–8235. doi: 10.1128/AEM.71.12.8228-8235.2005
- Lozupone, C., Lladser, M. E., Knights, D., Stombaugh, J., and Knight, R. (2011). UniFrac: An effective distance metric for microbial community comparison. *ISME J.* 5, 169–172. doi: 10.1038/ismej.2010.133
- Lozupone, C. A., Hamady, M., Kelley, S. T., and Knight, R. (2007). Quantitative and qualitative beta diversity measures lead to different insights into factors that structure microbial communities. *Appl. Environ. Microbiol.* 73, 1576–1585. doi: 10.1128/AEM.01996-06
- Magoč, T., and Salzberg, S. L. (2011). FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* 27, 2957–2963. doi: 10.1093/bioinformatics/btr507
- Mason, L. M., Eagar, A., Patel, P., Blackwood, C. B., and DeForest, J. L. (2021). Potential microbial bioindicators of phosphorus mining in a temperate deciduous forest. *J. Appl. Microbiol.* 130, 109–122. doi: 10.1111/jam.14761
- Matsuura, K., and Matsunaga, T. (2015). Antifungal activity of a termite queen pheromone against egg-mimicking termite ball fungi. *Ecol. Res.* 30, 93–100. doi: 10.1007/s11284-014-1213-7
- McCarthy, A. J., and Williams, S. T. (1992). Actinomycetes as agents of biodegradation in the environment—A review. *Gene* 115, 189–192. doi: 10.1016/0378-1119(92)90558-7
- McDonald, R. C., Watts, J. E. M., and Schreier, H. J. (2019). Effect of diet on the enteric microbiome of the wood-eating catfish *Panaque nigrolineatus*. *Front. Microbiol.* 10, 2687. doi: 10.3389/fmicb.2019.02687
- McKee, L. S., Martínez-Abad, A., Ruthes, A. C., Vilaplana, F., and Brumer, H. (2019). Focused metabolism of β -glucans by the soil bacteroidetes species chitinophaga pinensis. *Appl. Environ. Microbiol.* 85, 18. doi: 10.1128/AEM.02231-18
- Mey, C. B. J., and Gore, M. L. (2021). Biodiversity conservation and carbon sequestration in agroforestry systems of the mbalmayo forest reserve. *J. For. Environ. Sci.* 37, 91–103. doi: 10.7747/JFES.2021.37.2.91
- Milaković, M., Vestergaard, G., González-Plaza, J. J., Petrić, I., Šimatović, A., Senta, I., et al. (2019). Pollution from azithromycin-manufacturing promotes macrolide-resistance gene propagation and induces spatial and seasonal bacterial community shifts in receiving river sediments. *Environ. Int.* 12, 50. doi: 10.1016/j.envint.2018.12.050
- Mitaka, Y., Mori, N., and Matsuura, K. (2017). Multi-functional roles of a soldier-specific volatile as a worker arrestant, primer pheromone and an antimicrobial agent in a termite. *Proc. R. Soc. B Biol. Sci.* 284, 1134. doi: 10.1098/rspb.2017.1134
- Mizumoto, N., Gile, G. H., and Pratt, S. C. (2021). Behavioral rules for soil excavation by colony founders and workers in termites. *Ann. Entomol. Soc. Am.* 114, 654–661. doi: 10.1093/aesa/saaa017
- Moore, B. P. (1968). Studies on the chemical composition and function of the cephalic gland secretion in Australian termites. *J. Insect Physiol.* 14, 33–39. doi: 10.1016/0022-1910(68)90131-5
- Nnadozie, C. F., Lin, J., and Govinden, R. (2017). Selective isolation of a *Eucalyptus* spp. woodchip bacterial community and its taxonomic and metabolic profiling. *Bioenergy Res.* 10, 9. doi: 10.1007/s12155-017-9816-9
- Ogola, H. J. O., Selvarajan, R., and Tekere, M. (2021). Local geomorphological gradients and land use patterns play key role on the soil bacterial community diversity and dynamics in the highly endemic indigenous afrotemperate coastal scarp forest biome. *Front. Microbiol.* 12, 281. doi: 10.3389/fmicb.2021.592725
- Ohkuma, M. (2003). Termite symbiotic systems: Efficient bio-recycling of lignocellulose. *Appl. Microbiol. Biotechnol.* 61, 1–9. doi: 10.1007/s00253-002-1189-z
- Ohkuma, M., and Brune, A. (2010). “Diversity, structure, and evolution of the termite gut microbial community,” in *Biology of Termites: A Modern Synthesis*, eds D. Edward Bignell, Y. Roisin and N. Lo (Springer: Dordrecht), 413–438. doi: 10.1007/978-90-481-3977-4_15
- Olano, C., Lombó, F., Méndez, C., and Salas, J. A. (2008). Improving production of bioactive secondary metabolites in actinomycetes by metabolic engineering. *Metab. Eng.* 10, 281–292. doi: 10.1016/j.ymben.2008.07.001
- Otani, S., Zhukova, M., Koné, N. A., da Costa, R. R., Mikaelyan, A., Sapountzis, P., et al. (2019). Gut microbial compositions mirror caste-specific diets in a major lineage of social insects. *Environ. Microbiol. Rep.* 11, 196–205. doi: 10.1111/1758-2229.12728
- Pal, C., Bengtsson-Palme, J., Rensing, C., Kristiansson, E., and Larsson, D. G. J. (2014). BacMet: Antibacterial biocide and metal resistance genes database. *Nucl. Acids Res.* 42, D737–D743. doi: 10.1093/nar/gkt1252
- Pandit, P. D., Gulhane, M. K., Khardenavis, A. A., and Purohit, H. J. (2016). Mining of hemicellulose and lignin degrading genes from differentially enriched methane producing microbial community. *Bioresour. Technol.* 216, 923–930. doi: 10.1016/j.biortech.2016.06.021
- Paredes-Sabja, D., Setlow, P., and Sarker, M. R. (2011). Germination of spores of *Bacillales* and *Clostridiales* species: Mechanisms and proteins involved. *Trends Microbiol.* 19, 85–94. doi: 10.1016/j.tim.2010.10.004
- Parthasarathy, A., Cross, P. J., Dobson, R. C. J., Adams, L. E., Savka, M. A., and Hudson, A. O. (2018). A three-ring circus: Metabolism of the three proteogenic aromatic amino acids and their role in the health of plants and animals. *Front. Mol. Biosci.* 5, 29. doi: 10.3389/fmolb.2018.00029
- Pearson, W. R. (2013). An introduction to sequence similarity (“homology”) searching. *Curr. Protoc. Bioinform.* 3, bio301s42. doi: 10.1002/0471250953.bio301s42
- Prestwich, G. D. (1979). Defence secretion of the black termite, *Grallatotermes africanus* (Termitidae, Nasutitermitinae). *Insect Biochem.* 9, 563–567. doi: 10.1016/0020-1790(79)90093-3
- Prestwich, G. D., Jones, R. W., and Collins, M. S. (1981). Terpene biosynthesis by nasute termite soldiers (Isoptera: Nasutitermitinae). *Insect Biochem.* 11, 331–336. doi: 10.1016/0020-1790(81)90011-1
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2—approximately maximum-likelihood trees for large alignments. *PLoS ONE* 5, 9490. doi: 10.1371/journal.pone.0009490
- Puente-Sánchez, F., García-García, N., and Tamames, J. (2020). SQMtools: Automated processing and visual analysis of ‘omics data with R and anvi’o. *BMC Bioinform.* 21, 358. doi: 10.1186/s12859-020-03703-2
- Qiao, Q., Wang, F., Zhang, J., Chen, Y., Zhang, C., Liu, G., et al. (2017). The variation in the rhizosphere microbiome of cotton with soil type, genotype and developmental stage. *Sci. Rep.* 7, 1–10. doi: 10.1038/s41598-017-04213-7
- Rausch, P., Rühlemann, M., Hermes, B. M., Doms, S., Dagan, T., Dierking, K., et al. (2019). Comparative analysis of amplicon and metagenomic sequencing methods reveals key features in the evolution of animal metaorganisms. *Microbiome* 7, 1–19. doi: 10.1186/s40168-019-0743-1
- Rieke, E. L., Soupir, M. L., Moorman, T. B., Yang, F., and Howe, A. C. (2018). Temporal dynamics of bacterial communities in soil and leachate water after swine manure application. *Front. Microbiol.* 9, 3197. doi: 10.3389/fmicb.2018.03197
- Romero, D., Traxler, M. F., López, D., and Kolter, R. (2011). Antibiotics as signal molecules. *Chem. Rev.* 111, 5492. doi: 10.1021/cr2000509
- Rønn, R., McCaig, A. E., Griffiths, B. S., and Prosser, J. I. (2002). Impact of protozoan grazing on bacterial community structure in soil microcosms. *Appl. Environ. Microbiol.* 68, 6094–6105. doi: 10.1128/AEM.68.12.6094-6105.2002
- RStudio Team (2020). *RStudio: Integrated Development for R*. Available online at: <http://www.rstudio.com/>
- Scherlach, K., and Hertweck, C. (2020). Chemical mediators at the bacterial-fungal interface. *Annu. Rev. Microbiol.* 74, 267–290. doi: 10.1146/annurev-micro-012420-081224
- Schmidt, R., Ulanova, D., Wick, L. Y., Bode, H. B., and Garbeva, P. (2019). Microbe-driven chemical ecology: past, present and future. *ISME J.* 13, 2656–2663. doi: 10.1038/s41396-019-0469-x
- Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27, 863–864. doi: 10.1093/bioinformatics/btr026
- Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153
- Segata, N., Izard, J., Waldron, L., Gevers, D., Miropolsky, L., Garrett, W. S., et al. (2011). Metagenomic biomarker discovery and explanation. *Genome Biol.* 12, R60. doi: 10.1186/gb-2011-12-6-r60
- Seo, S. M., Kim, J., Kang, J., Koh, S. H., Ahn, Y. J., Kang, K. S., et al. (2014). Fumigant toxicity and acetylcholinesterase inhibitory activity of 4 Asteraceae plant essential oils and their constituents against Japanese termite (*Reticulitermes speratus* Kolbe). *Pestic. Biochem. Physiol.* 113, 55–61. doi: 10.1016/j.pestbp.2014.06.001

- Sieber, C. M. K., Probst, A. J., Sharrar, A., Thomas, B. C., Hess, M., Tringe, S. G., et al. (2018). Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat. Microbiol.* 3, 836–843. doi: 10.1038/s41564-018-0171-1
- Silva, J. P., Ticona, A. R. P., Hamann, P. R. V., Quirino, B. F., and Noronha, E. F. (2021). Deconstruction of lignin: From enzymes to microorganisms. *Molecules* 26, 2299. doi: 10.3390/molecules26082299
- Smibert, R. M., and Johnson, R. C. (1973). Spirochaetales, a review. *Crit. Rev. Microbiol.* 2, 491–552. doi: 10.3109/10408417309108393
- Sorensen, T. A. (1948). A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. *Biol. Skar.* 5, 1–34.
- Soukup, P., Veřtrovský, T., Stiblík, P., Votýpková, K., Chakraborty, A., Sillam-Dussès, D., et al. (2021). Termites ARE ASSOCIATED WITH EXTERNAL SPECIES-SPECIFIC BACTERIAL COMMUNITIES. *Appl. Environ. Microbiol.* 87, 1–13. doi: 10.1128/AEM.02042-20
- Štursová, M., Žifčáková, L., Leigh, M. B., Burgess, R., and Baldrian, P. (2012). Cellulose utilization in forest litter and soil: Identification of bacterial and fungal decomposers. *FEMS Microbiol. Ecol.* 80, 735–746. doi: 10.1111/j.1574-6941.2012.01343.x
- Subashchandrabose, S. R., Venkateswarlu, K., Krishnan, K., Naidu, R., Lockington, R., and Megharaj, M. (2018). *Rhodococcus wratislaviensis* strain 9: An efficient p-nitrophenol degrader with a great potential for bioremediation. *J. Hazard. Mater.* 347, 176–183. doi: 10.1016/j.jhazmat.2017.12.063
- Tamames, J., and Puente-Sánchez, F. (2019). SqueezeMeta, a highly portable, fully automatic metagenomic analysis pipeline. *Front. Microbiol.* 9, 3349. doi: 10.3389/fmicb.2018.03349
- Thong-On, A., Suzuki, K., Noda, S., Inoue, J. I., Kajiwara, S., and Ohkuma, M. (2012). Isolation and characterization of anaerobic bacteria for symbiotic recycling of uric acid nitrogen in the gut of various termites. *Microbes Environ.* 27, 1202010349–1202010349. doi: 10.1264/jisme2.ME11325
- Tokuda, G., Mikaelyan, A., Fukui, C., Matsuura, Y., Watanabe, H., Fujishima, M., et al. (2018). Fiber-associated spirochetes are major agents of hemicellulose degradation in the hindgut of wood-feeding higher termites. *Proc. Natl. Acad. Sci. U. S. A.* 115, E11996–E12004. doi: 10.1073/pnas.1810550115
- Treangen, T. J., Sommer, D. D., Angly, F. E., Koren, S., and Pop, M. (2011). Next generation sequence assembly with AMOS. *Curr. Protoc. Bioinform.* 33, 8–18. doi: 10.1002/0471250953.bi1108s33
- Tsimeli, K., Triantis, T. M., Dimotikali, D., and Hiskia, A. (2008). Development of a rapid and sensitive method for the simultaneous determination of 1,2-dibromoethane, 1,4-dichlorobenzene and naphthalene residues in honey using HS-SPME coupled with GC–MS. *Anal. Chim. Acta* 617, 64–71. doi: 10.1016/j.aca.2008.03.049
- Valterová, I., Vašíčková, S., Buděšínský, M., and Vrkoč, J. (1986). Constituents of frontal gland secretion of peruvian termites *Nasutitermes ephratae*. *Collect. Czechoslov. Chem. Commun.* 51, 2884–2895. doi: 10.1135/cccc19862884
- van der Sande, M. T., Arets, E. J. M. M., Peña-Claros, M., Hoosbeek, M. R., Cáceres-Siani, Y., van der Hout, P., et al. (2018). Soil fertility and species traits, but not diversity, drive productivity and biomass stocks in a Guyanese tropical rainforest. *Funct. Ecol.* 32, 461–474. doi: 10.1111/1365-2435.12968
- van Rensburg, J. J., Lin, H., Gao, X., Toh, E., Fortney, K. R., Ellinger, S., et al. (2015). The human skin microbiome associates with the outcome of and is influenced by bacterial infection. *MBio* 6, e01315–e01315. doi: 10.1128/mBio.01315-15
- van Velden, J. L., Wilson, K., Lindsey, P. A., McCallum, H., Moyo, B. H. Z., and Biggs, D. (2020). Bushmeat hunting and consumption is a pervasive issue in African savannahs: Insights from four protected areas in Malawi. *Biodivers. Conserv.* 29, 1443–1464. doi: 10.1007/s10531-020-01944-4
- Waglechner, N., McArthur, A. G., and Wright, G. D. (2019). Phylogenetic reconciliation reveals the natural history of glycopeptide antibiotic biosynthesis and resistance. *Nat. Microbiol.* 4, 1862–1871. doi: 10.1038/s41564-019-0531-5
- Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microbiol.* 73, 5261–5267. doi: 10.1128/AEM.00062-07
- Wang, Y. X., Cai, M., Zhi, X. Y., Zhang, Y. Q., Tang, S. K., Xu, L. H., et al. (2008). *Microlunatus aurantiacus* sp. nov., a novel actinobacterium isolated from a rhizosphere soil sample. *Int. J. Syst. Evol. Microbiol.* 58, 1873–1877. doi: 10.1099/ijs.0.65518-0
- Whitman, W. B., and Suzuki, K. (2015). Solirubrobacterales. *Bergey's Man. Syst. Archaea Bact.* 25, 1–3. doi: 10.1002/9781118960608.obm00025
- Wilcke, W., Amelung, W., Martius, C., Garcia, M. V. B., and Zech, W. (2000). Biological sources of polycyclic aromatic hydrocarbons (PAHs) in the Amazonian rain forest. *J. Plant Nutr. Soil Sci.* 163, 27–30. doi: 10.1002/(SICI)1522-2624(200002)163:1<27::AID-JPLN27>3.0.CO;2-E
- Wilhelm, R. C., Singh, R., Eltis, L. D., and Mohn, W. W. (2018). Bacterial contributions to delignification and lignocellulose degradation in forest soils with metagenomic and quantitative stable isotope probing. *ISME J.* 13, 413–429. doi: 10.1038/s41396-018-0279-6
- Wiltz, B. A., Henderson, G., and Chen, J. (1998). Effect of naphthalene, butylated hydroxytoluene, dioctyl phthalate, and adipic dioctyl ester, chemicals found in the nests of the formosan subterranean termite (Isoptera: Rhinotermitidae) on a *Saprophytic* *Mucor* sp. (Zygomycetes: Mucorales). *Environ. Entomol.* 27, 936–940. doi: 10.1093/ee/27.4.936
- Wright, M. S., Lax, A. R., Henderson, G., and Chen, J. (2000). Growth response of *Metarhizium anisopliae* to two Formosan subterranean termite nest volatiles, naphthalene and fenchone. *Mycologia* 92, 42–45. doi: 10.1080/00275514.2000.12061128
- Wu, Y. W., Simmons, B. A., and Singer, S. W. (2016). MaxBin 2.0: An automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32, 605–607. doi: 10.1093/bioinformatics/btv638
- Xia, Y., Wang, Y., Fang, H. H. P., Jin, T., Zhong, H., and Zhang, T. (2014). Thermophilic microbial cellulose decomposition and methanogenesis pathways recharacterized by metatranscriptomic and metagenomic analysis. *Sci. Rep.* 4, 1–9. doi: 10.1038/srep06708
- Yamaguchi, M., Ichida, J., Xuan-Xuan, Z., Nakamura, M., and Yoshitake, T. (1994). Determination of glyoxal, methylglyoxal, diacetyl, and 2,3-pentanedione in fermented foods by high-performance liquid chromatography with fluorescence detection. *J. Liq. Chromatogr.* 17, 203–211. doi: 10.1080/10826079408013445
- Yang, H., Schmitt-Wagner, D., Stingl, U., and Brune, A. (2005). Niche heterogeneity determines bacterial community structure in the termite gut (*Reticulitermes santoniensis*). *Environ. Microbiol.* 7, 916–932. doi: 10.1111/j.1462-2920.2005.00760.x
- Ye, Y., and Doak, T. G. (2009). A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. *PLoS Comput. Biol.* 5, e1000465. doi: 10.1371/journal.pcbi.1000465
- Zapfack, L., and Engwald, S. (2008). Biodiversity and spatial distribution of vascular epiphytes in two biotopes of the Cameroonian semi-deciduous rain forest. *Plant Ecol.* 195, 117–130. doi: 10.1007/s11258-007-9308-7
- Zhang, C., Yu, L., and Zhang, Y. (2017). Research progress on the genus *Microlunatus*. *Wei Sheng Wu Xue Bao* 57, 179–187.
- Zhao, S., Ye, Z., and Stanton, R. (2020). Misuse of RPKM or TPM normalization when comparing across samples and sequencing protocols. *RNA* 26, 903–909. doi: 10.1261/rna.074922.120
- Zhong, C., Fu, J., Jiang, T., Zhang, C., and Cao, G. (2018). Polyphosphate metabolic gene expression analyses reveal mechanisms of phosphorus accumulation and release in *Microlunatus phosphovorius* strain JN459. *FEMS Microbiol. Lett.* 365, 34. doi: 10.1093/femsle/fny034