



**HAL**  
open science

## Comparative Analysis of Transcriptomes from Secondary Reproductives of Three Reticulitermes Termite Species

Franck Dedeine, Lucy A Weinert, Diane Bigot, Thibaut Josse, Marion Ballenghien, Vincent Cahais, Nicolas Galtier, Philippe Gayral

### ► To cite this version:

Franck Dedeine, Lucy A Weinert, Diane Bigot, Thibaut Josse, Marion Ballenghien, et al.. Comparative Analysis of Transcriptomes from Secondary Reproductives of Three Reticulitermes Termite Species. PLoS ONE, 2015, 10.1371/journal.pone.0145596 . hal-01286070

**HAL Id: hal-01286070**

**<https://hal.science/hal-01286070>**

Submitted on 10 Mar 2016

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RESEARCH ARTICLE

# Comparative Analysis of Transcriptomes from Secondary Reproductives of Three *Reticulitermes* Termite Species

Franck Dedeine<sup>1</sup>, Lucy A. Weinert<sup>2\*</sup>, Diane Bigot<sup>1</sup>, Thibaut Josse<sup>1</sup>, Marion Ballenghien<sup>2</sup>, Vincent Cahais<sup>2</sup>, Nicolas Galtier<sup>2</sup>, Philippe Gayral<sup>1\*</sup>

**1** Institut de Recherche sur la Biologie de l'Insecte, UMR 7261, CNRS—Université François Rabelais, 37200, Tours, France, **2** Institut des Sciences de l'Evolution, UMR 5554, Université de Montpellier—CNRS—IRD—EPHE, Montpellier, France

✉ Current address: University of Cambridge, Department of Veterinary Medicine, Madingley Road, Cambridge, CB3 0ES, United Kingdom

\* [philippe.gayral@univ-tours.fr](mailto:philippe.gayral@univ-tours.fr)



CrossMark  
click for updates

OPEN ACCESS

**Citation:** Dedeine F, Weinert LA, Bigot D, Josse T, Ballenghien M, Cahais V, et al. (2015) Comparative Analysis of Transcriptomes from Secondary Reproductives of Three *Reticulitermes* Termite Species. PLoS ONE 10(12): e0145596. doi:10.1371/journal.pone.0145596

**Editor:** Erjun Ling, Institute of Plant Physiology and Ecology, CHINA

**Received:** February 27, 2015

**Accepted:** December 7, 2015

**Published:** December 23, 2015

**Copyright:** © 2015 Dedeine et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Reads used for this study originated from a previous work [31] on *R. flavipes* (Sequence Read Archive accession no. SRX565295 and SRX565296) and *R. grassei* (SRA accession no. SRX565297 to SRX565305), and from the present study for *R. lucifugus* (SRA accession no. SRX565306 and SRX565307).

**Funding:** Termite collection was supported by a French ANR grant to FD (EvoSymTer 08-JCJC-007601). Transcriptome sequencing was supported by a European Research Council (ERC) grant to NG (ERC PopPhyl 232971).

## Abstract

Termites are eusocial insects related to cockroaches that feed on lignocellulose. These insects are key species in ecosystems since they recycle a large amount of nutrients but also are pests, exerting major economic impacts. Knowledge on the molecular pathways underlying reproduction, caste differentiation or lignocellulose digestion would largely benefit from additional transcriptomic data. This study focused on transcriptomes of secondary reproductive females (nymphoid neotenic). Thirteen transcriptomes were used: 10 of *Reticulitermes flavipes* and *R. grassei* sequenced from a previous study, and two transcriptomes of *R. lucifugus* sequenced for the present study. After transcriptome assembly and read mapping, we examined interspecific variations of genes expressed by termites or gut microorganisms. A total of 18,323 orthologous gene clusters were detected. Functional annotation and taxonomic assignment were performed on a total of 41,287 predicted contigs in the three termite species. Between the termite species studied, functional categories of genes were comparable. Gene ontology (GO) terms analysis allowed the discovery of 9 cellulases and a total of 79 contigs potentially involved in 11 enzymatic activities used in wood metabolism. Altogether, results of this study illustrate the strong potential for the use of comparative interspecific transcriptomes, representing a complete resource for future studies including differentially expressed genes between castes or SNP analysis for population genetics.

## Introduction

Termites (Blattodea, Termitoidea) constitute an ecologically and evolutionary diversified group of social insects (>2600 species) that share a common ancestry with cockroaches [1]. The ecological success of termites is often attributed to the combination of their sophisticated social organization with their unique ability to feed on recalcitrant plant matters such as wood

**Competing Interests:** The authors are also grateful to the Genotoul bioinformatics 405 platform Toulouse Midi-Pyrenees for providing computing and storage resources. This does not alter the authors' adherence to PLOS ONE policies on sharing data and Materials.

(lignocellulose) [2]. Lignocellulose digestion relies on a complex enzymatic system which is synthesized by termites and a diverse intestinal microbial community composed of numerous prokaryotes and, in some termites, unicellular eukaryotes (flagellated protists) [3]. Termites are major decomposers in many tropical and subtropical ecosystems and therefore, are crucial for recycling organic matter [2]. Conversely, some termites are pests, causing serious damage to human-built structures and woody plant crops.

Aside from a specialized nutritional regime, another characteristic of termites allowing them to successfully diversify worldwide is their sophisticated social organization. As in other social insects, such as social Hymenoptera (ants, some bees and wasps), termites live in complex societies where individuals are morphologically, physiologically and behaviorally specialized into distinct castes. The castes work together to accomplish specific and complementary tasks within a colony. Division of labor among castes is the key to efficient colony development, survival and reproduction. The social organization of termites also represents a primary reason why termite infestations can be difficult to control and eradicate. Therefore, a detailed understanding of the expressed genes of termites is not only interesting for academic research but is also essential in the development of new termite-specific insecticides [4].

The acquisition of genetic data in termites and their gut microbial community has been of recent interest to the scientific community. This is mostly due to the development and accessibility of new sequencing technologies such as 454 pyrosequencing and Illumina sequencing [4]. To date, (meta-) genomic and (meta-) transcriptomic studies in termites have been principally aimed at identifying host and/or symbiont genes underlying lignocellulose digestion [5–12], caste differentiation [13–17], reproduction [18] or defense [19]. Large-scale EST libraries have also been constructed in a few distantly related termite species belonging to different families for comparative purpose [13]. Despite these efforts and the recent publication of the first two termite genomes [20,21], genetic data are only available for a limited number of termite species and comparative studies remain scarce. The diversity and identity of genes expressed have rarely been compared between closely related species.

*Reticulitermes* (Rhinotermitidae) represent an important genus of termites with multiple pest species, particularly in temperate regions [22]. They have cryptic nesting habits and form complex colonies with diffuse nests and multiple feeding sites connected by underground tunnels [23]. Termite colonies are typically founded by a single pair of winged reproductives (i.e., the “queen” and the “king”) following a nuptial flight. However, reproduction is not always reserved only to the primary couple within colonies. Another type of reproductives can indeed differentiate among the offspring of the primary couple. Such secondary reproductives are called ‘nymphoid neotenic’ when they differentiate from nymphs and ‘ergatoid neotenic’ when they differentiate from workers. Although these two types of neotenic are morphologically different, both are wingless and have no pigmentation; they stay in their native colonies where they may replace or supplement the primary couple’s reproduction. The presence of productive neotenic within colonies has tremendous genetic and dynamic impacts on colonies [23], and some authors have argued that the acquisition of this caste has played a major role in the evolution of social life [24,25]. In *Reticulitermes* termites, the number of neotenic is extremely variable among and within species, and several studies have argued that a high number of neotenic could improve the capacity of colonies to develop and disperse in urban areas [23,26,27]. Despite their importance, the conditions under which neotenic reproductives differentiate within colonies as well as the molecular mechanisms underlying such a differentiation remain unknown. In comparison with other castes (workers, soldiers, primary reproductives), only a few studies focus on determining the genes expressed in neotenic reproductives of termites [15,28,29].

The present study compares the transcriptome content obtained from 13 nymphoid neotenic females of three *Reticulitermes* species: two West European species, *R. grassei* and *R. lucifugus* [30], and one North American species, *R. flavipes*, that has been introduced to France [26]. Eleven out of the thirteen transcriptomes analyzed in this study (i.e., those obtained in *R. flavipes* and *R. grassei*) were obtained in a previous study [31]. This dataset was used for SNP detection and population genomics inference only [32,33] and no gene content and functional analysis was performed. Using the same protocol, two additional transcriptomes were generated from nymphoid female neotenic collected in two distinct colonies of *R. lucifugus*. BLAST analysis allowed functional and taxonomic assignment of the predicted contigs, assembled from the whole species from pools of individuals and compared between termite species. Functional analysis of contigs associated with wood degradation was performed using Gene Ontology (GO) term analysis and contributed to the characterization of *Reticulitermes* transcriptomes.

## Materials and Methods

### Termite samples

We sampled 13 termite colonies representative of 3 *Reticulitermes* species: 9 colonies of *R. grassei*, 2 colonies of *R. flavipes* and 2 colonies of *R. lucifugus*. All samples were collected in 2010 in France from 11 locations (Table 1). None of these termite species are endangered or protected, and no specific permission was required for collecting them since they were taken from unprotected areas. Five hundred to 1000 individuals per colony were brought to the laboratory where they were maintained for 90 days under controlled conditions in their original wood piece at 25°C and 80% humidity. For each colony, the transcriptome of a single fecund nymphoid neotenic female was obtained using the protocol described below. The transcriptomes obtained from *R. grassei* and *R. flavipes* females were obtained and assembled in a previous study [31]. In this study, we used a single nymphoid neotenic female isolated from 2 different colonies of *R. lucifugus* for further RNA isolation, cDNA library construction and transcriptome sequencing (see below).

### RNA isolation

Total RNA was isolated independently from the whole body of a single neotenic using an adapted protocol using Guanidinium Thiocyanate-Phenol solution supplemented with

**Table 1. Information and sequencing results of nymphoid neotenic female termites used in this study.**

Identification name	<i>Reticulitermes</i> sp.	Locality	Latitude	Longitude	No of illumina reads (x10 <sup>6</sup> )
GA15A	<i>R. grassei</i>	Pissos	44.307069	-0.778946	14.95
GA15B	<i>R. grassei</i>	Poitiers	46.580224	0.340375	44.77
GA15C	<i>R. grassei</i>	Gatseau	45.826772	-1.208771	27.21
GA15D	<i>R. grassei</i>	Gatseau	45.826772	-1.208771	37.83
GA15E	<i>R. grassei</i>	Ustaritz	43.395936	-1.45459	22.20
GA15F	<i>R. grassei</i>	Eguilles	43.5686589	5.354234	28.57
GA15G	<i>R. grassei</i>	Eguilles	43.5686589	5.354234	27.60
GA15H	<i>R. grassei</i>	Cugnaux	43.537141	1.344995	24.52
GA15I	<i>R. grassei</i>	Petit Bôo	44.353499	-1.042781	23.30
GA15K	<i>R. flavipes</i>	Olonne	46.536402	-1.772826	58.86
GA15L	<i>R. flavipes</i>	Oléron	45.9159335	-1.2716743	46.98
GA15M	<i>R. lucifugus</i>	Allauch	43.336148	5.4825739	43.65
GA15N	<i>R. lucifugus</i>	Le Rove	43.369912	5.252998	44.80

doi:10.1371/journal.pone.0145596.t001

glycogen [34]. Quality and quantity of total RNA were determined using agarose gel electrophoresis, NanoDrop spectrophotometry and analysis on Agilent bioanalyzer 2100 system using the Eukaryote Total RNA Nano assay. RNA isolation for the two other species *R. grassei* and *R. flavipes* used in this study was obtained with the same protocol [31].

### Transcriptome sequencing

For each sample, 5 µg of total RNA of *R. lucifugus* were used to build 3'-primed, non-normalized cDNA libraries. Although prokaryotic RNA might be present in the libraries since the gut was not removed before extraction, these transcripts were not specifically targeted. Oligo(dT)-primed first-strand synthesis and cap-primed second-strand synthesis were performed using the SMART cDNA library construction kit (Clontech, Mountain View, CA, USA). Libraries were sequenced using Genome Analyzer II (Illumina) with 5 tagged libraries pooled per lane. Fifty bp single-end reads were produced. The cDNA library construction and sequencing were performed by GATC biotech company (Conzanz, Germany). After tag-removing, low quality bases, adaptors and primers were removed with SeqClean software (<http://compbio.dfci.harvard.edu/tgi/>) with default parameters. Reads used for this study originated from a previous work [31] on *R. flavipes* (Sequence Read Archive accession no. SRX565295 and SRX565296) and *R. grassei* (SRA accession no. SRX565297 to SRX565305), and from the present study for *R. lucifugus* (SRA accession no. SRX565306 and SRX565307).

### Transcriptome assembly

Transcriptomes were assembled by pooling the reads obtained from individuals belonging to the same species (N = 9 for *R. grassei* and N = 2 for both *R. flavipes* and *R. lucifugus*). Assemblies were performed using ABYSS [35] with Kmer set at 40, followed by two consecutive runs of CAP3 [36] as described in [37]. Contigs shorter than 100 bases were discarded.

### ORF detection

Complete and 5'- or 3'-truncated open reading frames (ORF) were detected using Prodigal software for metagenomic data [38] using standard genetic code. ORF with a stretch of N (undetermined nucleotides) inside the sequence were not discarded. When several ORF were detected on the same contig, only the longest was kept since it would more likely correspond to a true protein. The software Cd-hit [39] was used to remove ORF redundancy from our dataset by detecting sequences showing 100% identity in homologous regions.

### Orthology prediction

BLAST-based pairwise orthologs relationships between *R. grassei*-*R. flavipes*, *R. grassei*-*R. lucifugus*, and *R. flavipes*-*R. lucifugus* species pairs were first assessed by InParanoid 4.1 program using default parameters [40]. Non-redundant translated contigs of the 3 species displaying length > 100 bases were used for that purpose. MultiParanoid [41] was then used to analyze the orthology relationship of gene clusters between the 3 species.

### Taxonomic assignation

BLAST results from the annotation step were parsed to retrieve sequence identifier (GI) and NCBI taxonomic identifier (TaxID) from the NCBI database (<ftp://ftp.ncbi.nih.gov/pub/taxonomy>). Each contig was assigned to 5 taxonomic ranks (i.e., superkingdom, kingdom, phylum, genus and species). Contigs were assigned to Bacteria when the indicated superkingdom was 'Bacteria'; they were assigned to Termites when the superkingdom was 'Eukaryota' and the

kingdom was 'Metazoa'; they finally were assigned to protists when the superkingdom was 'Eukaryota' and when the kingdom was not 'Fungi, Viridiplantae and/or metazoa'. For protists, downstream taxonomic ranks (phylum, genus and species) were inspected manually to verify that each of them corresponded to a protist taxon.

## Functional annotation

Amino acid homologies of the non-redundant predicted ORFs were analyzed using BLASTp program [42] against the Genbank nr database (March 1, 2011). The first BLAST hits were kept and the minimum E-value was set at 0.001. HSP length cut off was 33 and the lower capacity filter was enabled. GI identifiers from BLAST results were used to retrieve UniProt IDs from the PIR Protein Sequence Database, the latter served to associate GO terms with our predicted ORFs. In addition, known protein signatures were detected using the software InterProScan [43] based on InterPro collection databases. Annotation steps were performed with Blast2go software V.2.2.6 [44] using default parameters. GO terms analysis was performed on the three termite species independently. Based on the KEGG metabolism pathways database V.64.0 [45,46] implemented in Blast2go, GO terms of biological processes were used to retrieve contigs belonging to the starch and sucrose metabolism process (map00500). GO results were displayed at GO level 2. A GO level referred to the hierarchical structure of the Gene Ontology as the number of GO terms between a given term and the Root Term of the ontology. Contigs assigned to a function related to the starch and sucrose metabolism process were further analyzed for the presence of hallmarks of Carbohydrate-Active Enzymes (CAZ) domains using CAZymes Analysis Toolkit (CAT) [47] with the CAZy database updated on 09/20/2013 [48]. In this software, the assignment method is based on both similarity search on proteins sequences and the presence of Pfam conserved domains (option 'Pfam rules based annotation').

## Quantification of transcript abundance

For each species, a pool of individual reads were mapped to the species transcriptome using BWA [49] with default parameters and served for FPKM (i.e., Fragment Per Kilobase per Millions fragments mapped) calculation using Cufflinks [50]. To avoid biases due to differences in read numbers across individual data sets, a random subsampling of reads was performed before pooling. A total of 14.95, 49.98 and 43.65 million reads were subsampled for each individual of *R. grassei*, *R. flavipes* and *R. lucifugus*, respectively. This number corresponded to the number of reads of the individual displaying the lower number of reads for the species. Only the 250 most abundant contigs per species (with higher FPKM values) with a Blast2GO hit were kept. As previously described, the functional annotation of these contigs was performed using Blast2GO. Preliminary analyses indicated that GO term levels 2 and 3 were not appropriated due to a limited number of contigs producing too few functional categories. The final analysis was therefore conducted at the GO term level 4.

## Results and Discussion

### Sequencing, assembly and annotation of transcriptomes

The numbers of illumina reads obtained for all of the 11 termite samples are indicated in Table 1. For each termite species, reads from conspecific individuals were pooled together before assembly ( $N = 9$  for *R. grassei*,  $N = 2$  for *R. flavipes* and *R. lucifugus*). The obtained species-level assemblies were comparable among species and exhibited high N50 values (Table 2). In total, 64,328, 65,814 and 79,404 non-redundant ORF were predicted in *R. flavipes*, *R.*



**Table 2. Assemblies of transcriptomes, ORF predictions and functional annotations for interspecific comparisons and Gene Ontology analyses.** Contigs < 100 bases were removed from analyses.

Species	<i>R. flavipes</i>	<i>R. grassei</i>	<i>R. lucifugus</i>
No of contigs	93,420	111,549	91,280
N50	998	1,041	1,067
No. of predicted ORF	64,342	79,640	65,855
Non-redundant ORF	64,328	79,404	65,814
No of ORF with BLASTp hit	19,375 (30.1%)	21,671 (27.3%)	20,377 (31%)
No of annotated ORF (GO term)	5,389 (8.4%)	4,922 (6.2%)	5,214 (7.9%)

doi:10.1371/journal.pone.0145596.t002

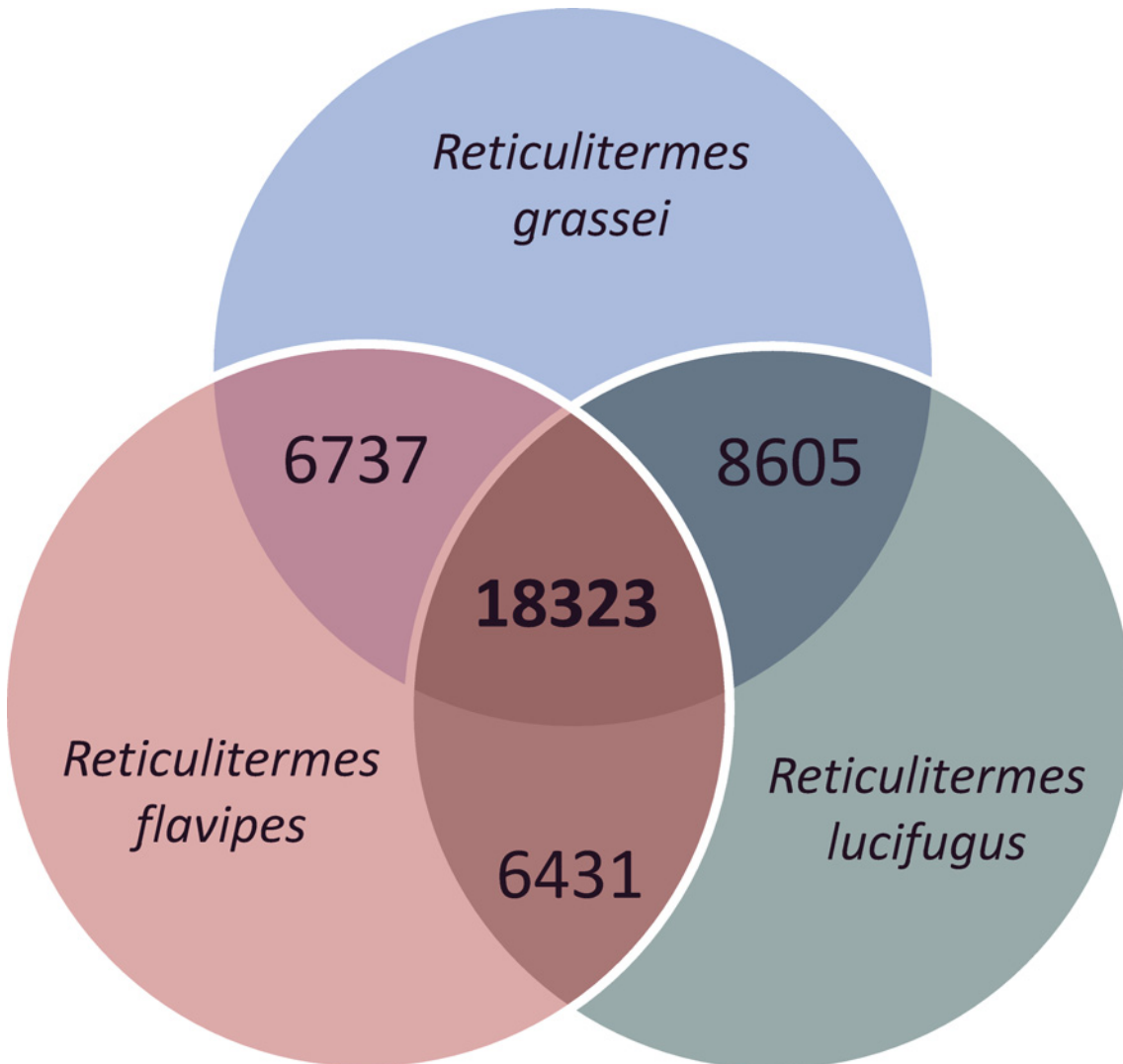
*lucifugus* and *R. grassei*, respectively. As expected with non-model organisms lacking complete and well-annotated genome sequences, only a third of ORF showed a significant BLAST hit in the nr protein database. After the functional annotation performed with Blast2Go program, only a small fraction (6–8%) of the initial ORF set could be assigned with one or more GO terms (Table 2). Whatever the number of individual reads sets pooled together (2 or 9), the number and quality of the obtained contigs were similar among species, suggesting that pooling reads from 2 individuals produced satisfactory results.

### Orthology relationships

Gene clusters were identified within each species from transcript data. The presence of homologous clusters between species (i.e. orthologues) was assessed to better understand the genetic relationships between the 3 termite species. For each species, transcript clusters were composed of representative unigenes (i.e. alternative transcripts derived from a single locus) or transcripts derived from young duplicated genes (i.e. paralogues). Interspecific comparison showed that a large number (18,323) of orthologous genes clusters (i.e. homologous genes clusters identified in other species) were detected in the 3 species (Fig 1). Orthologous genes were more abundant between *R. grassei* and *R. lucifugus* (8,605) than between the two other pairs of species (6,737 between *R. grassei* and *R. flavipes*, and 6,431 between *R. lucifugus* and *R. grassei*). This result is in accordance with the phylogenetic relationships between these termite species [30]. The North American species, *R. flavipes*, is indeed distantly related to the two European species, *R. grassei* and *R. lucifugus*, which are closely related and probably sister species. Transcripts of gut microbiota may also reinforce this phylogenetic relationship between the taxa, albeit to a lesser extent since they do not contribute much in terms of contig numbers (see next paragraph). This hypothesis is supported by recent studies showing how microbial communities are usually more similar between closely related species than distant species [51–54].

### Taxonomic assignment

Table 3 shows the taxonomic distribution of the contigs displaying a significant BLAST hit of their coding sequence against the nr protein database. In total, 41,287 contigs were assigned. Most of them (94.9%) were assigned to the termite genome, whereas the remaining contigs were assigned to diverse lineages of microorganisms (i.e. protists, bacteria, archaea, virus) (3.9%), fungi (0.9%) or plants (0.4%). Contigs assigned to fungi and plants most likely represent environmental contaminations since no endosymbiotic association has been described so far between these organisms and *Reticulitermes* termites. Most contigs assigned to microorganisms were probably expressed by diverse microorganisms living in the hindgut of *Reticulitermes* termites. These well-known microbial communities are composed of protists (two main lineages: Parabasalia and Oxymonadida), Bacteria (the most abundant: Spirochaetes,



**Fig 1. Orthology relationships between *R. grassei*, *R. flavipes* and *R. lucifugus* contigs.** The number of orthologous gene clusters is indicated inside Venn diagram.

doi:10.1371/journal.pone.0145596.g001

Bacteroidetes, Firmicutes and Elusimicrobia), methanogenic Archaea (Methanobacteriaceae family) and bacteriophage virus infecting Spirochaetes [3,55,56].

Several reasons could explain the low proportion of microbial contigs in our dataset. First, contigs assigned to prokaryotes were particularly scarce (< 1% for the 3 termite species), probably because the cDNA library protocol underwent an enrichment of mRNA based on the existence of poly-A tails, which are mostly absent in prokaryotic transcripts. Second, as expected with non-model organisms, which lack complete and well-annotated genome sequences, expressed genes in gut microbial communities of termites have not been fully characterized yet. Therefore, the genomic database is likely incomplete and could thus result in an underestimation of microbial genes. Third, all transcriptomes were generated from nymphoid neotenic females. Like other castes or developmental stages in subterranean termites, reproductive castes do not necessarily feed on the wood themselves and instead are fed by workers who provide them with nutrient-rich salivary trophallactic transfers [57]. Since the gut microbiota might be not essential for extracting nutrients from wood in these secondary reproductives,



microorganisms could be less abundant in the hindgut of reproductives compared to that of wood-feeding castes. This hypothesis is supported by previous work in primary reproductives (alates) in 3 species of *Reticulitermes* [58], as well as even earlier work in *R. flavipes* [15] which suggests a reduced microbiota in neotenic reproductives.

### Description of gene ontologies of expressed genes

The function of assembled contigs was evaluated by retrieving GO terms according to their termite, protist or bacterial origin. The 3 descriptive ontologies ‘cellular components’, ‘molecular function’ and ‘biological process’ were analyzed. For the cellular components ontology, 10 localizations were found. Transcripts products mainly localized in cell (37–38% depending on termite species), membrane (24–27%), organelle (19–20%) and macromolecular complex (11–13%) (Fig 2). Twenty-one classes of biological processes were found with 2 dominant classes corresponding to metabolic (30–32%) and cellular processes (31–32%) (Fig 3). Finally, 13 GO terms were found in ‘molecular function’, among which putative catalytic activity (40–42%) and binding (40%) were the most abundant (Fig 4). A very similar distribution of GO terms for the 3 types of ontologies was observed between the 3 termite species, suggesting unbiased transcriptome assembly. The relative taxonomic distribution (termite, protist, bacteria) was consistent among the 3 ontologies and with the total number of contigs (Table 3). GO analysis showed that protists accounted for a significant part of cellular components (mainly cell, organelle and macromolecular complex) of biological processes (metabolic and cellular process) and molecular functions (catalytic activity, binding and structural molecule activity).

The putative biological functions of the 250 most expressed contigs were studied for each of the 3 termite species (Fig 5). These 750 contigs corresponded to highly expressed transcripts, having a FPKM value ranging from 5,940 in *R. grassei* to 3,448,530 in *R. lucifugus*. Eighty-three functional categories were assigned in total among the 3 termite species. The 31 functional categories displaying the most numerous contigs were related to metabolic process whereas the 52 remaining categories corresponded to diverse other biological functions. This result appeared consistent with the global analysis of transcriptomes (Fig 4), and suggests that the most expressed genes have metabolism-related functions.

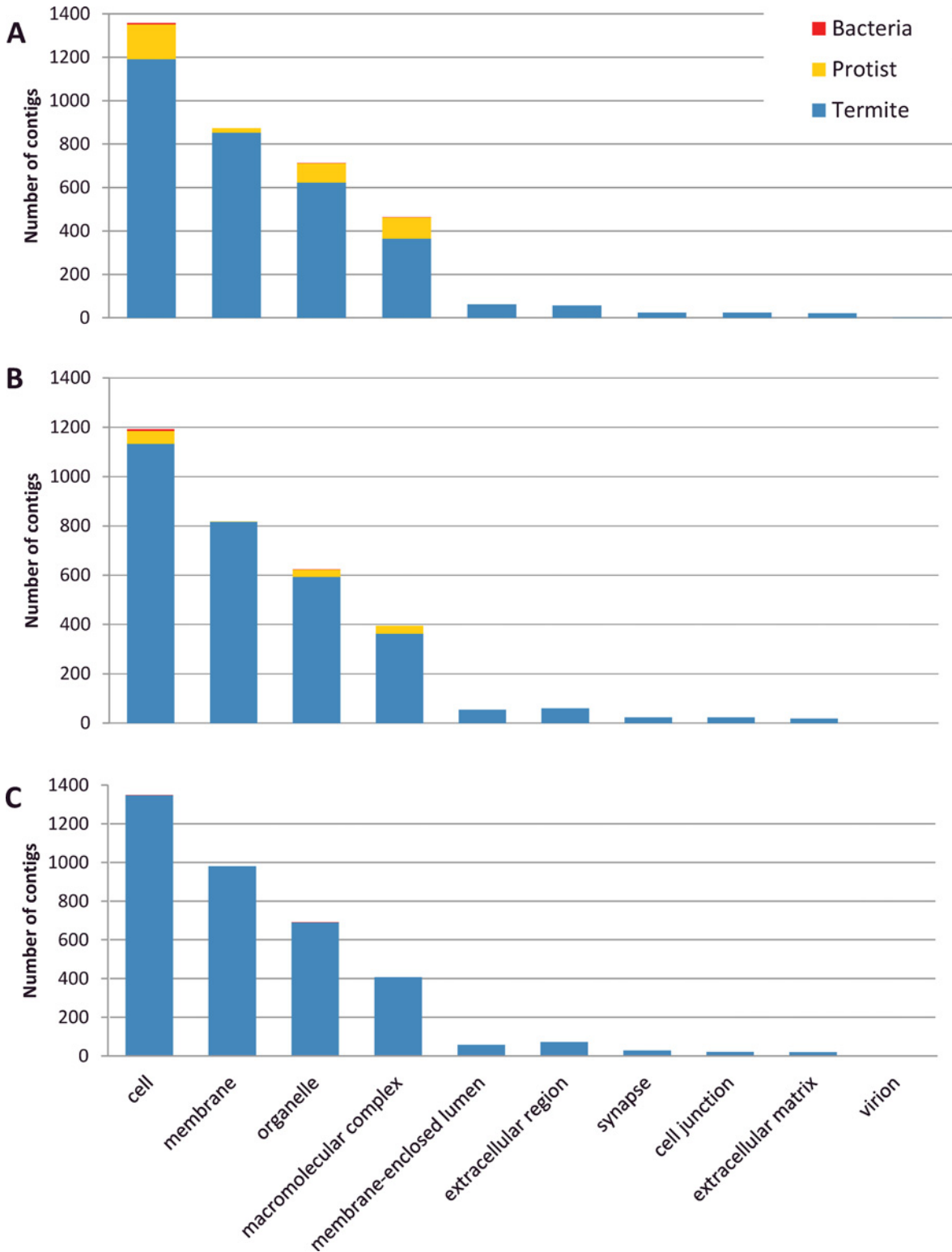
### Contigs putatively associated with wood-degradation enzymes

Transcriptome annotation was used to retrieve contigs associated with functions belonging to the starch and sucrose metabolism. Some of these may be involved in cellulose degradation

**Table 3. Taxonomic assignation of contigs obtained in the three termite species.**

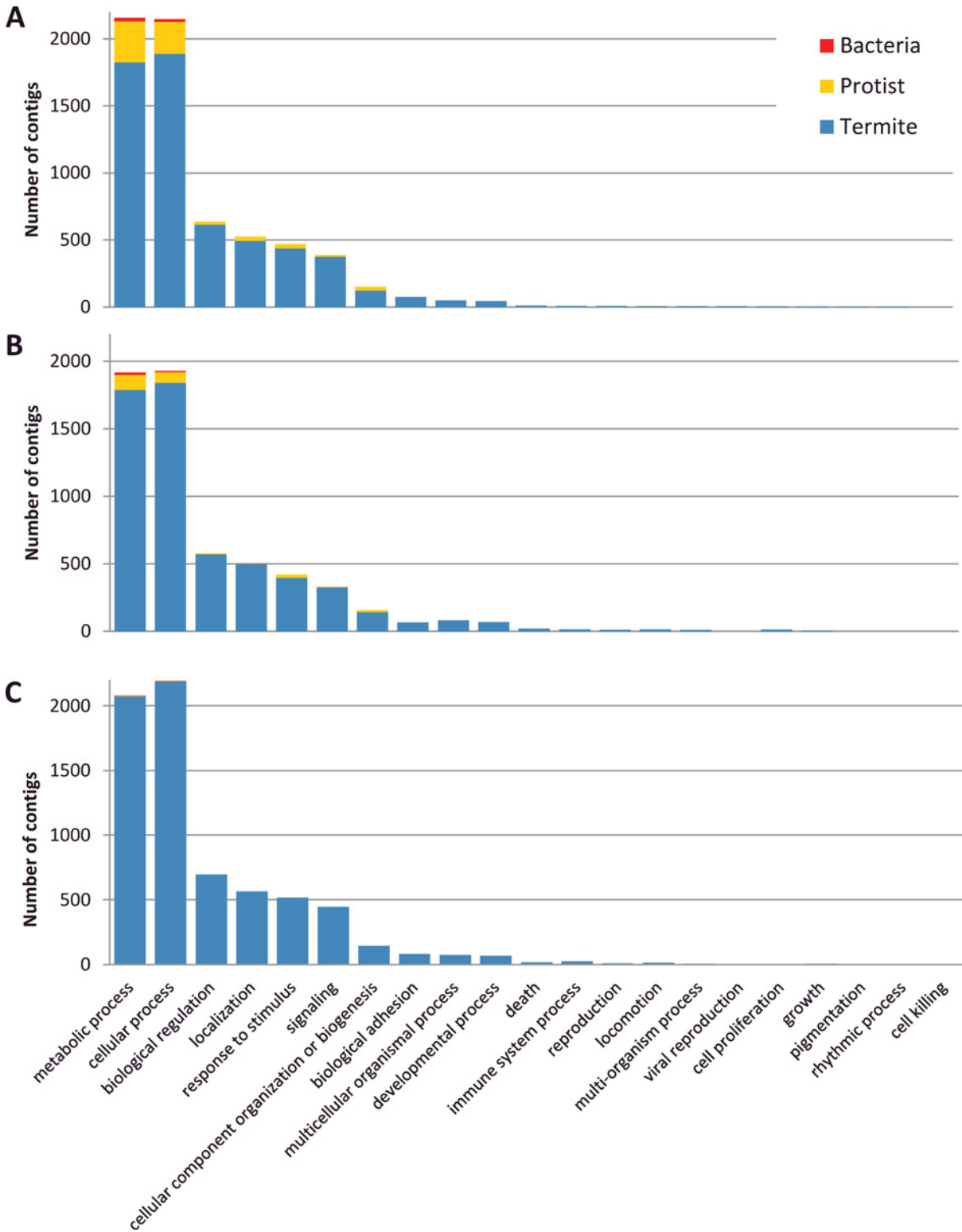
Species		<i>R. flavipes</i>	<i>R. grassei</i>	<i>R. lucifugus</i>	Total
No. of transcriptomes		2	9	2	13
No. of contigs (%)	All	13,889 (100.00)	14,453 (100.00)	12,945 (100.00)	41,287 (100)
	Termites	12,806 (92.11)	13,620 (94.05)	12,798 (94.05)	39,224 (94.87)
	Microorganisms <sup>a</sup>	1,006 (7.23)	438 (3.02)	95 (0.73)	1,539 (3.92)
	Protists	893 (6.42)	300 (2.07)	39 (0.30)	1,232 (3.14)
	Bacteria	101 (0.72)	120 (0.82)	46 (0.35)	267 (0.68)
	Archeae	6 (0.04)	9 (0.62)	3 (0.02)	18 (0.04)
	Viruses	6 (0.04)	9 (0.62)	7 (0.05)	22 (0.05)
	Fungi	39 (0.28)	302 (2.08)	26 (0.20)	367 (0.93)
	Viridiplantae	38 (0.27)	93 (0.64)	26 (0.20)	157 (0.40)

<sup>a</sup>Contigs of microorganisms include those of Protists, Bacteria, Archeae and Viruses.



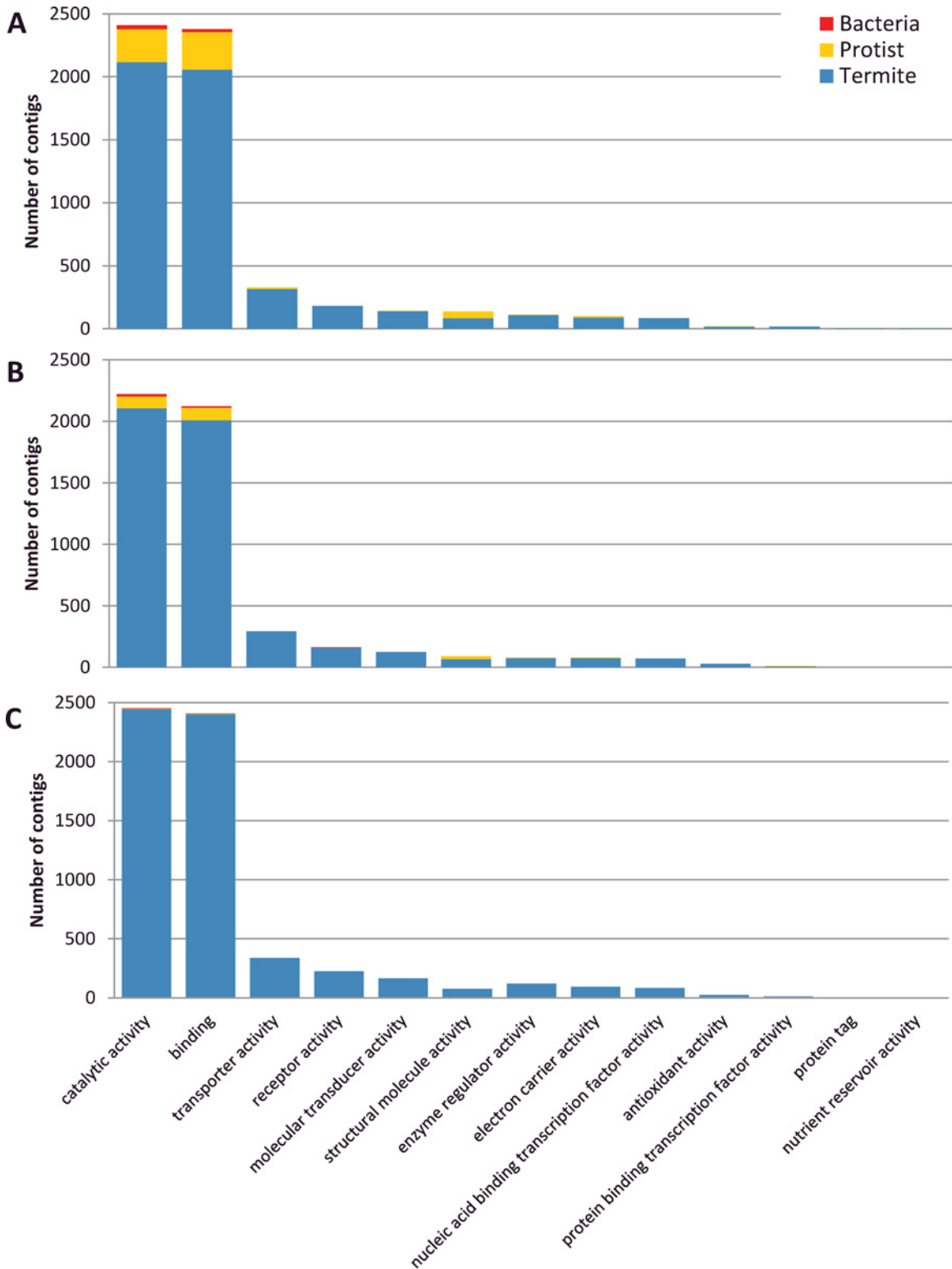
**Fig 2. Gene Ontology analysis of Cellular Components (GO level 2).** A: *R. flavipes*. B: *R. grassei*, C: *R. lucifugus*.

doi:10.1371/journal.pone.0145596.g002



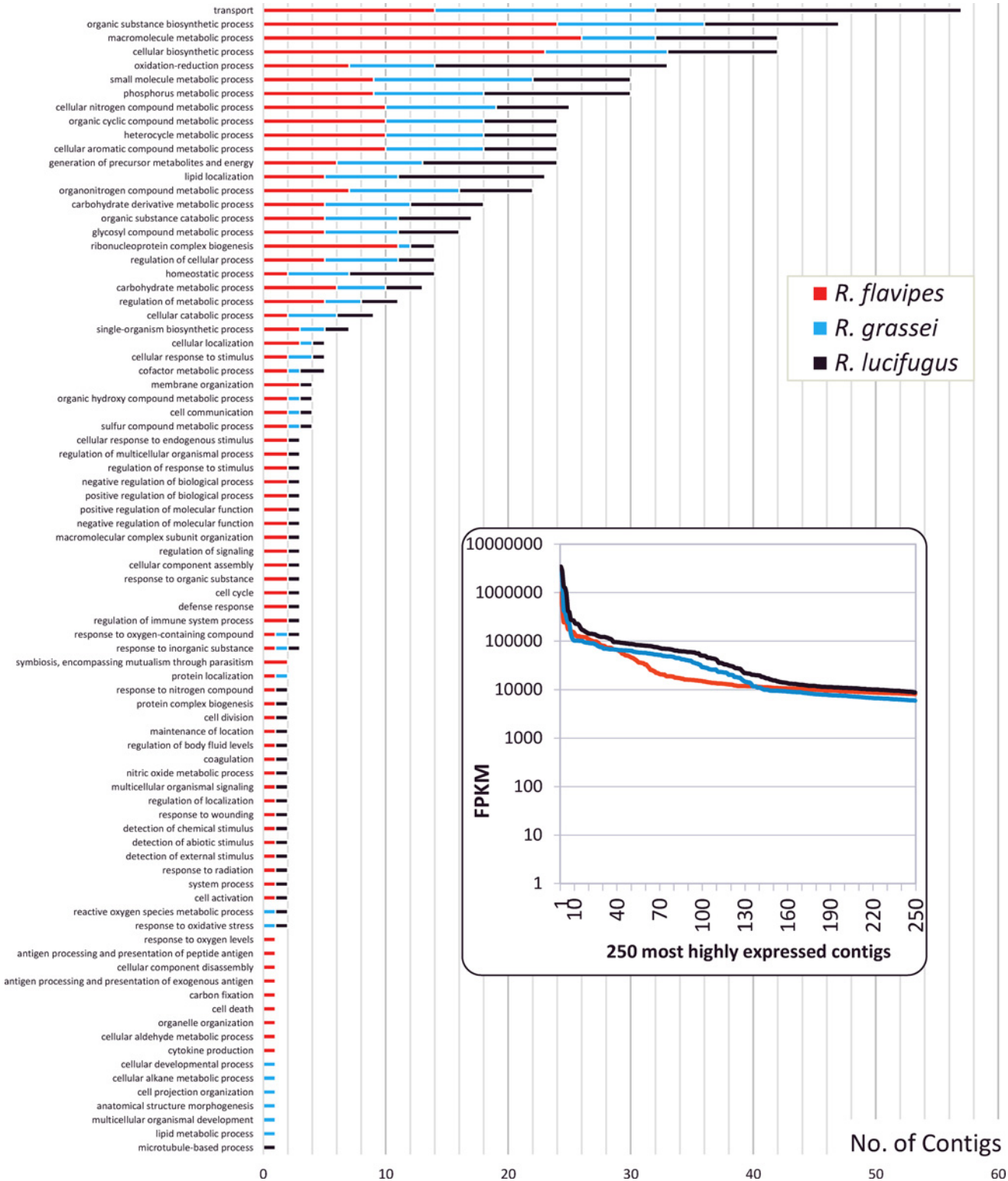
**Fig 3. Gene Ontology analysis of Biological Processes (GO level 2).** A: *R. flavipes*. B: *R. grassei*, C: *R. lucifugus*.

doi:10.1371/journal.pone.0145596.g003



**Fig 4. Gene Ontology analysis of Molecular Functions (GO level 2).** A: *R. flavipes*. B: *R. grassei*, C: *R. lucifugus*.

doi:10.1371/journal.pone.0145596.g004



**Fig 5. Distribution of the 250 most expressed contigs across their biological functions detected by GO term analysis at the GO level 4.** The relative abundance of contigs estimated by FPKM is shown in the box.

doi:10.1371/journal.pone.0145596.g005

[59]. The complete list of contigs displaying functions related to starch and sucrose metabolism and their amino acid sequences are shown in supplementary information (S1 Table and S1 Dataset). Fig 6 shows the enzymatic processes detected in our dataset and plots them on starch and sucrose metabolism map. A total of 11 functions putatively associated with known enzymatic activities were found in the assemblies. Among these functions, we found contigs putatively associated with cellulytic activities. Cellulase is a general term for cellulytic enzymes, of which three main classes are recognized on the basis of the mode of enzymatic actions and substrate specificities: endoglucanases (EGs; EC 3.2.1.4), cellobiohydrolases (CBHs; EC 3.2.1.91) and  $\beta$ -glucosidases (BGs; EC 3.2.1.21). These three categories of enzymes work synergistically to efficiently degrade chains of cellulose. Whereas EGs and BGs are quite common in microorganisms, animals and plants, CBHs are apparently more rare and appear to be restricted to bacteria, fungi and protists [59,60]. In our dataset, 9 contigs were assigned to putative EGs and 11 contigs were assigned to putative BGs (Fig 6). However, we found no evidence for the presence of genes encoding CBHs.

Using CAZymes Analysis Toolkit [47], 9 families of putative genes of Carbohydrate-Active Enzymes (CAZy) were detected in the 3 transcriptomes (Table 4): 5 glycoside hydrolases (GH1, GH9, GH13, GH37, GH45), 2 glycoside transferases (i.e., GT3, GT35) and 2 carbohydrate-binding modules families (i.e., CBM6, CBM48). Genes of a same GH family are usually considered to share not only structural motifs and the catalytic machinery, but also an evolutionary origin [61]. Among the putative GH genes detected in our analyses, GH1 represents a single family of BGs, whereas GH9 and GH45 are two families of EGs. Previous studies show GH1 and GH9 to be mostly expressed by the genome of termites either in their salivary glands and/or in the hindgut [62]. However, genes encoding GH45 can be expressed by both the hosts and their symbiotic protists or prokaryotes living in the hindgut. In addition, CBMs are usually considered to be expressed by microorganisms only (Watanabe & Tokuda 2010). Therefore, the detection of putative genes encoding GH45 and CBM in our dataset suggests that the gut microbial community of neotenics may play a role in the synthesis of the enzymatic system involved in the degradation of cellulose.

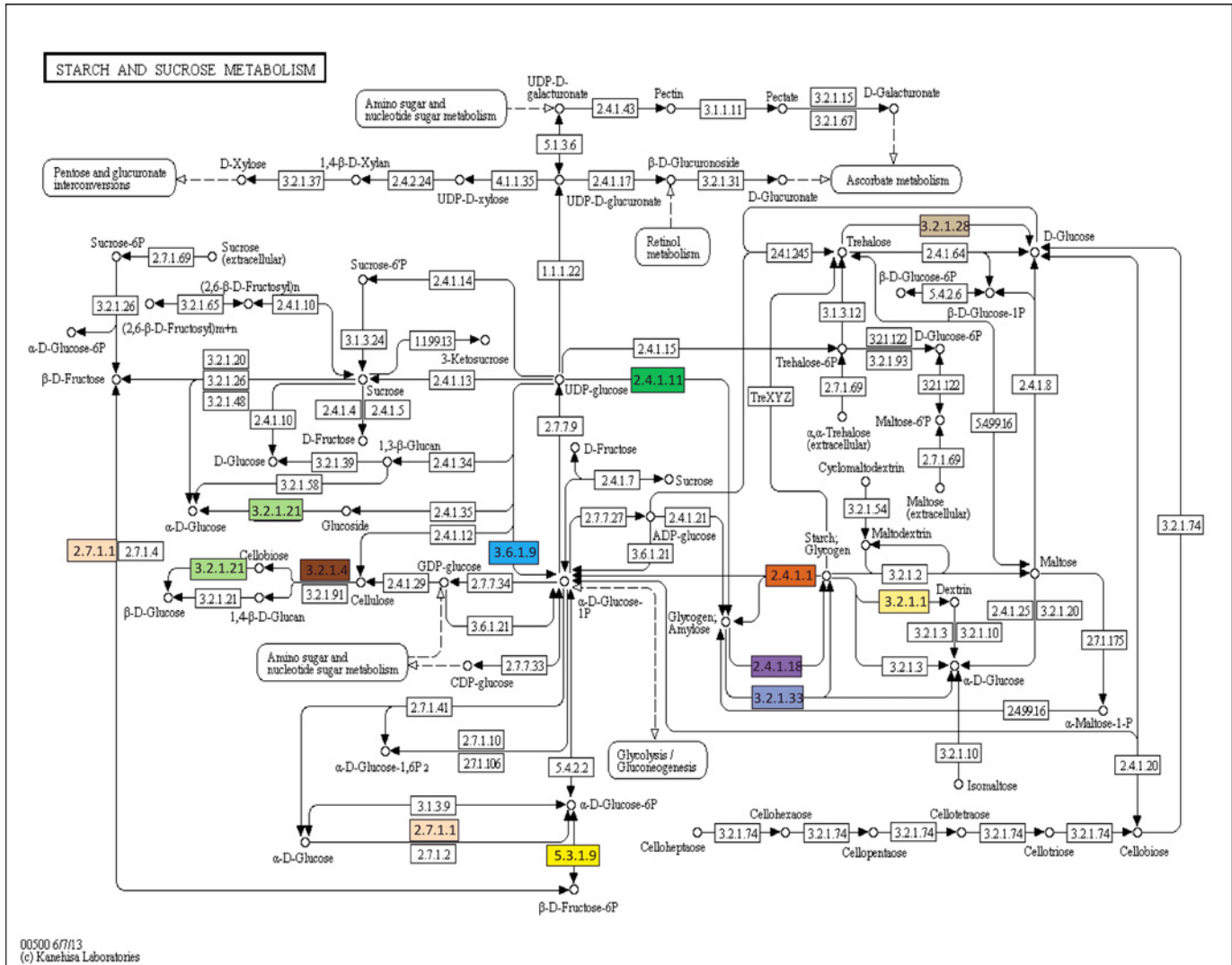
## Interspecific variation of microbial gene expression patterns

The proportion of microbial contigs varies among the three *Reticulitermes* species (Table 3). This variation is particularly evident in the contigs assigned to protists. Representing only 0.3% of the sequences in *R. lucifugus*, protists contigs were more abundant in *R. grassei* (2.1%) and in *R. flavipes* (6.4%). This general pattern has been found in the functional analysis also since the most important proportion of microbial contigs was assigned in *R. flavipes* transcriptome followed by those of *R. grassei* and *R. lucifugus* (Figs 2–4). We cannot exclude that a part of this observed variation results from methodological fluctuations in mRNA isolation, cDNA library construction and sequencing. However, this pattern might also result from variation of gene expression patterns among individual neotenics of different species, either due to transcriptomic noise or associated to a biological function. In any case, our results suggest that gut microbial communities are not totally absent from *Reticulitermes* neotenics, in spite of their feeding lifestyle, which probably does not directly involve lignocellulose digestion. Abundance, role and regulation mechanisms of gut microbial communities in reproductive termites will require further investigations.

## Conclusion

Comparison of a set of assembled transcriptomes of nymphoid neotenic reproductives was performed from 13 colonies belonging to 3 related termite species based on high throughput





00500 6/7/13  
(c) Kanehisa Laboratories

		<i>R. flavipes</i>	<i>R. grassei</i>	<i>R. lucifugus</i>	
	ec:2.4.1.1	phosphorylase	(14)	(4)	(4)
	ec:2.4.1.11	glycogen(starch) synthase	(2)	(1)	(1)
	ec:2.4.1.18	1,4-alpha-glucan branching enzyme	(2)	(0)	(1)
	ec:2.7.1.1	hexokinase	(4)	(4)	(2)
	ec:3.2.1.1	alpha-amylase	(1)	(0)	(0)
	ec:3.2.1.21	beta-glucosidase	(7)	(1)	(3)
	ec:3.2.1.28	alpha,alpha-trehalase	(2)	(2)	(4)
	ec:3.2.1.33	amyl-alpha-1,6-glucosidase	(1)	(0)	(1)
	ec:3.2.1.4	cellulase	(4)	(3)	(2)
	ec:3.6.1.9	nucleotide diphosphatase	(1)	(1)	(0)
	ec:5.3.1.9	glucose-6-phosphate isomerase	(3)	(2)	(2)

**Fig 6. Contigs of *R. flavipes*, *R. grassei* and *R. lucifugus* displaying putative enzymatic activities involved in starch and sucrose metabolism.** Top panel: termite contigs (colored boxes) mapped on the starch and sucrose metabolism KEGG map (black and white boxes). Bottom panel: number of contigs (brackets) associated to the 11 putative functions of starch and sucrose metabolism.

doi:10.1371/journal.pone.0145596.g006

**Table 4. Number of Carbohydrate-Active Enzyme (CAZy) families detected in *Reticulitermes* transcriptomes.**

CAZy Families	Pfam domains	Domain descriptions	Putative enzymatic activities	<i>R. flavipes</i>	<i>R. grassei</i>	<i>R. lucifugus</i>
CBM48 /GH13	CBM_48/Alpha-amylase/Alpha-amylase_C	Carbohydrate-binding module 48 (Isoamylase N-terminal domain)	-	1	0	1
CBM6	Phosphodiesterase	Type I phosphodiesterase / nucleotide pyrophosphatase	-	1	1	0
GH1	Glyco_hydro_1	Glycosyl hydrolase family 1	3.2.1.21	4	0	1
GH13	Alpha-amylase	Alpha amylase, catalytic domain	3.4.1.183.2.1.1	1	0	0
GH37	Trehalase	Trehalase	3.2.1.28	2	2	3
GH45	Glyco_hydro_45	Glycosyl hydrolase family 45	3.2.1.4	1	1	0
GH9	Glyco_hydro_9	Glycosyl hydrolase family 9	3.2.1.4.	1	1	2
GT3	Glycogen_syn	Glycogen synthase	2.4.1.11	2	1	1
GT35	Phosphorylase	Carbohydrate phosphorylase	2.4.1.1	11	3	3

doi:10.1371/journal.pone.0145596.t004

Illumina sequencing. Intraspecific variation was addressed by pooling two to nine individuals per species. As expected with non-model organisms, a large fraction of contigs had no detectable homologs in the public database. The majority of recovered transcripts had a termite origin, although transcripts from microorganisms provided evidence for the presence of an active gut microbiome in this non-wood feeding life stage. These transcripts were indeed over-represented in starch and sucrose metabolism pathways, and some of them are likely to encode enzymes involved in cellulose degradation.

### Supporting Information

**S1 Table. Contigs putatively involved in enzymatic activities linked to the starch and sucrose metabolism pathway and detected by GO terms analysis.**

(DOCX)

**S1 Dataset. Amino acid sequence of contigs putatively involved in enzymatic activities linked to the starch and sucrose metabolism pathway and detected by GO terms analysis in FASTA format.**

(DOCX)

### Acknowledgments

We would like to thank Simon Dupont and Sylvain Guyot for termite sampling, Vincent Ranwez for help with the NCBI taxonomy database and M. Rivera for English revisions. Analyses largely benefited from the ISEM computing cluster platform with the help of Khalid Belkhir. We are also grateful to the Genotoul bioinformatics platform Toulouse Midi-Pyrenees for providing computing and storage resources. This does not alter our adherence to PLOS ONE policies on sharing data and Materials.

### Author Contributions

Conceived and designed the experiments: FD NG PG. Performed the experiments: FD LW MB. Analyzed the data: FD DB VC TJ PG. Wrote the paper: FD PG.

## References

1. Inward DJ, Vogler AP, Eggleton P (2007) A comprehensive phylogenetic analysis of termites (Isoptera) illuminates key aspects of their evolutionary biology. *Mol Phylogenet Evol* 44: 953–967. PMID: [17625919](#)
2. Bignell DE, Roisin Y, Lo N (2011) *Biology of termites: a modern synthesis*: Springer.
3. Brune A (2014) Symbiotic digestion of lignocellulose in termite guts. *Nat Rev Microbiol*.
4. Scharf ME (2015) Omic research in termites: an overview and a roadmap. *Front Genet* 6: 76. doi: [10.3389/fgene.2015.00076](#) PMID: [25821456](#)
5. He S, Ivanova N, Kirton E, Allgaier M, Bergin C, Scheffrahn RH, et al. (2013) Comparative metagenomic and metatranscriptomic analysis of hindgut paunch microbiota in wood- and dung-feeding higher termites. *PLoS ONE* 8: e61126. doi: [10.1371/journal.pone.0061126](#) PMID: [23593407](#)
6. Raychoudhury R, Sen R, Cai Y, Sun Y, Lietze VU, Boucias DG, et al. (2013) Comparative metatranscriptomic signatures of wood and paper feeding in the gut of the termite *Reticulitermes flavipes* (Isoptera: Rhinotermitidae). *Insect Mol Biol* 22: 155–171. doi: [10.1111/imb.12011](#) PMID: [23294456](#)
7. Tartar A, Wheeler MM, Zhou X, Coy MR, Boucias DG, Scharf ME. (2009) Parallel metatranscriptome analyses of host and symbiont gene expression in the gut of the termite *Reticulitermes flavipes*. *Biotechnol Biofuels* 2: 25. doi: [10.1186/1754-6834-2-25](#) PMID: [19832970](#)
8. Todaka N, Inoue T, Saita K, Ohkuma M, Nalepa CA, Lenz M, et al. (2010) Phylogenetic analysis of cellulolytic enzyme genes from representative lineages of termites and a related cockroach. *PLoS ONE* 5: e8636. doi: [10.1371/journal.pone.0008636](#) PMID: [20072608](#)
9. Warnecke F, Luginbuhl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, et al. (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* 450: 560–565. PMID: [18033299](#)
10. Xie L, Zhang L, Zhong Y, Liu N, Long YH, Wang SY, et al. (2012) Profiling the metatranscriptome of the protistan community in *Coptotermes formosanus* with emphasis on the lignocellulolytic system. *Genomics* 99: 246–255. doi: [10.1016/j.ygeno.2012.01.009](#) PMID: [22326742](#)
11. Zhang D, Lax AR, Henrissat B, Coutinho P, Katiya N, Nierman WC, et al. (2012) Carbohydrate-active enzymes revealed in *Coptotermes formosanus* (Isoptera: Rhinotermitidae) transcriptome. *Insect Mol Biol* 21: 235–245. doi: [10.1111/j.1365-2583.2011.01130.x](#) PMID: [22243654](#)
12. Bastien G, Arnal G, Bozonnet S, Laguerre S, Ferreira F, Faure R, et al. (2013) Mining for hemicellulases in the fungus-growing termite *Pseudacanthotermes militaris* using functional metagenomics. *Biotechnol Biofuels* 6.
13. Hayashi Y, Shigenobu S, Watanabe D, Toga K, Saiki R, Shimada K, et al. (2013) Construction and characterization of normalized cDNA libraries by 454 pyrosequencing and estimation of DNA methylation levels in three distantly related termite species. *PLoS ONE* 8.
14. Huang Q, Sun P, Zhou X, Lei C (2012) Characterization of head transcriptome and analysis of gene expression involved in caste differentiation and aggression in *Odontotermes formosanus* (Shiraki). *PLoS ONE* 7: e50383. doi: [10.1371/journal.pone.0050383](#) PMID: [23209730](#)
15. Scharf ME, Wu-Scharf D, Zhou X, Pittendrigh BR, Bennett GW (2005) Gene expression profiles among immature and adult reproductive castes of the termite *Reticulitermes flavipes*. *Insect Mol Biol* 14: 31–44. PMID: [15663773](#)
16. Sen R, Raychoudhury R, Cai YP, Sun YJ, Lietze VU, Boucias DG, et al. (2013) Differential impacts of juvenile hormone, soldier head extract and alternate caste phenotypes on host and symbiont transcriptome composition in the gut of the termite *Reticulitermes flavipes*. *BMC Genomics* 14.
17. Steller MM, Kambhampati S, Caragea D (2010) Comparative analysis of expressed sequence tags from three castes and two life stages of the termite *Reticulitermes flavipes*. *BMC Genomics* 11: 463. doi: [10.1186/1471-2164-11-463](#) PMID: [20691076](#)
18. Husseneder C, McGregor C, Lang RP, Collier R, Delatte J (2012) Transcriptome profiling of female alates and egg-laying queens of the Formosan subterranean termite. *Comp Biochem Physiol Part D Genomics Proteomics* 7: 14–27. doi: [10.1016/j.cbd.2011.10.002](#) PMID: [22079412](#)
19. Hojo M, Maekawa K, Saitoh S, Shigenobu S, Miura T, Hayashi Y, et al. (2012) Exploration and characterization of genes involved in the synthesis of diterpene defence secretion in *Nasute* termite soldiers. *Insect Mol Biol* 21: 545–557. doi: [10.1111/j.1365-2583.2012.01162.x](#) PMID: [22984844](#)
20. Terrapon N, Li C, Robertson HM, Ji L, Meng X, Booth W, et al. (2014) Molecular traces of alternative social organization in a termite genome. *Nat Commun* 5: 3636. doi: [10.1038/ncomms4636](#) PMID: [24845553](#)

21. Korb J, Poulsen M, Hu H, Li C, Boomsma JJ, Zhang G, et al. (2015) A genomic comparison of two termites with different social complexity. *Front Genet* 6: 9. doi: [10.3389/fgene.2015.00009](https://doi.org/10.3389/fgene.2015.00009) PMID: [25788900](https://pubmed.ncbi.nlm.nih.gov/25788900/)
22. Su NY (2002) Novel technologies for subterranean termite control. *Sociobiology* 40: 95–101.
23. Vargo EL, Husseneder C (2009) Biology of subterranean termites: insights from molecular studies of *Reticulitermes* and *Coptotermes*. *Annu Rev Entomol* 54: 379–403. doi: [10.1146/annurev.ento.54.110807.090443](https://doi.org/10.1146/annurev.ento.54.110807.090443) PMID: [18793101](https://pubmed.ncbi.nlm.nih.gov/18793101/)
24. Howard K, Thorne B (2011) Eusocial evolution in Termites and Hymenoptera. In: Bignell DE, Roisin Y, Lo N, editor. *Biology of termites: a modern synthesis*: Springer. pp. 97–132.
25. Korb J, Hartfelder K (2008) Life history and development—a framework for understanding developmental plasticity in lower termites. *Biol Rev* 83: 295–313. PMID: [18979593](https://pubmed.ncbi.nlm.nih.gov/18979593/)
26. Perdereau E, Bagneres AG, Bankhead-Dronnet S, Dupont S, Zimmermann M, et al. (2013) Global genetic analysis reveals the putative native source of the invasive termite, *Reticulitermes flavipes*, in France. *Mol Ecol* 22: 1105–1119. doi: [10.1111/mec.12140](https://doi.org/10.1111/mec.12140) PMID: [23205642](https://pubmed.ncbi.nlm.nih.gov/23205642/)
27. Perdereau E, Bagneres AG, Vargo EL, Baudouin G, Xu Y, Labadie P, et al. (2015) Relationship between invasion success and colony breeding structure in a subterranean termite. *Mol Ecol* 24: 2125–2142. doi: [10.1111/mec.13094](https://doi.org/10.1111/mec.13094) PMID: [25641360](https://pubmed.ncbi.nlm.nih.gov/25641360/)
28. Weil T, Korb J, Rehli M (2009) Comparison of queen-specific gene expression in related lower termite species. *Mol Biol Evol* 26: 1841–1850. doi: [10.1093/molbev/msp095](https://doi.org/10.1093/molbev/msp095) PMID: [19541881](https://pubmed.ncbi.nlm.nih.gov/19541881/)
29. Weil T, Rehli M, Korb J (2007) Molecular basis for the reproductive division of labour in a lower termite. *BMC Genomics* 8: 198. PMID: [17598892](https://pubmed.ncbi.nlm.nih.gov/17598892/)
30. Clement JL, Bagneres AG, Uva P, Wilfert L, Quintana A, Reinhard J, et al. (2001) Biosystematics of *Reticulitermes* termites in Europe: morphological, chemical and molecular data. *Insect Soc* 48: 202–215.
31. Gayral P, Melo-Ferreira J, Glemin S, Bierne N, Carneiro M, Nabholz B, et al. (2013) Reference-free population genomics from next-generation transcriptome data and the vertebrate-invertebrate gap. *PLoS Genet* 9: e1003457. doi: [10.1371/journal.pgen.1003457](https://doi.org/10.1371/journal.pgen.1003457) PMID: [23593039](https://pubmed.ncbi.nlm.nih.gov/23593039/)
32. Romiguier J, Gayral P, Ballenghien M, Bernard A, Cahais V, Chenail A, et al. (2014) Comparative population genomics in animals uncovers the determinants of genetic diversity. *Nature* 515: 261–263. doi: [10.1038/nature13685](https://doi.org/10.1038/nature13685) PMID: [25141177](https://pubmed.ncbi.nlm.nih.gov/25141177/)
33. Romiguier J, Lourenco J, Gayral P, Faivre N, Weinert LA, Ravel S, et al. (2014) Population genomics of eusocial insects: the costs of a vertebrate-like effective population size. *J Evol Biol* 27: 593–603. doi: [10.1111/jeb.12331](https://doi.org/10.1111/jeb.12331) PMID: [26227898](https://pubmed.ncbi.nlm.nih.gov/26227898/)
34. Gayral P, Weinert L, Chiari Y, Tsagkogeorga G, Ballenghien M, Galtier N (2011) Next-generation sequencing of transcriptomes: a guide to RNA isolation in nonmodel animals. *Mol Ecol Resour*: 650–661.
35. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I (2009) ABySS: a parallel assembler for short read sequence data. *Genome Res* 19: 1117–1123. doi: [10.1101/gr.089532.108](https://doi.org/10.1101/gr.089532.108) PMID: [19251739](https://pubmed.ncbi.nlm.nih.gov/19251739/)
36. Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. *Genome Res* 9: 868–877. PMID: [10508846](https://pubmed.ncbi.nlm.nih.gov/10508846/)
37. Cahais V, Gayral P, Tsagkogeorga G, Melo-Ferreira J, Ballenghien M, Weinert L, et al. (2012) Reference-free transcriptome assembly in non-model animals from next-generation sequencing data. *Mol Ecol Resour* 12: 834–845.
38. Hyatt D, LoCascio PF, Hauser LJ, Uberbacher EC (2012) Gene and translation initiation site prediction in metagenomic sequences. *Bioinformatics* 28: 2223–2230. doi: [10.1093/bioinformatics/bts429](https://doi.org/10.1093/bioinformatics/bts429) PMID: [22796954](https://pubmed.ncbi.nlm.nih.gov/22796954/)
39. Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* 22: 1658–1659. PMID: [16731699](https://pubmed.ncbi.nlm.nih.gov/16731699/)
40. Östlund G, Schmitt T, Forslund K, Köstler T, Messina DN, Roopra S, et al. (2010) InParanoid 7: new algorithms and tools for eukaryotic orthology analysis. *Nucleic Acids Res* 38: D196–D203. doi: [10.1093/nar/gkp931](https://doi.org/10.1093/nar/gkp931) PMID: [19892828](https://pubmed.ncbi.nlm.nih.gov/19892828/)
41. Alexeyenko A, Tamas I, Liu G, Sonnhammer ELL (2006) Automatic clustering of orthologs and inparalogs shared by multiple proteomes. *Bioinformatics* 22: e9–e15. PMID: [16873526](https://pubmed.ncbi.nlm.nih.gov/16873526/)
42. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403–410. PMID: [2231712](https://pubmed.ncbi.nlm.nih.gov/2231712/)
43. Zdobnov EM, Apweiler R (2001) InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17: 847–848. PMID: [11590104](https://pubmed.ncbi.nlm.nih.gov/11590104/)

44. Gotz S, Garcia-Gomez JM, Terol J, Williams TD, Nagaraj SH, Nueda MJ, et al. (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* 36: 3420–3435. doi: [10.1093/nar/gkn176](https://doi.org/10.1093/nar/gkn176) PMID: [18445632](https://pubmed.ncbi.nlm.nih.gov/18445632/)
45. Kanehisa M, Goto S (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28: 27–30. PMID: [10592173](https://pubmed.ncbi.nlm.nih.gov/10592173/)
46. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* 40: D109–114. doi: [10.1093/nar/gkr988](https://doi.org/10.1093/nar/gkr988) PMID: [22080510](https://pubmed.ncbi.nlm.nih.gov/22080510/)
47. Park BH, Karpinets TV, Syed MH, Leuze MR, Uberbacher EC (2010) CAZymes Analysis Toolkit (CAT): web service for searching and analyzing carbohydrate-active enzymes in a newly sequenced organism using CAZy database. *Glycobiology* 20: 1574–1584. doi: [10.1093/glycob/cwq106](https://doi.org/10.1093/glycob/cwq106) PMID: [20696711](https://pubmed.ncbi.nlm.nih.gov/20696711/)
48. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 42: D490–495. doi: [10.1093/nar/gkt1178](https://doi.org/10.1093/nar/gkt1178) PMID: [24270786](https://pubmed.ncbi.nlm.nih.gov/24270786/)
49. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25: 1754–1760. doi: [10.1093/bioinformatics/btp324](https://doi.org/10.1093/bioinformatics/btp324) PMID: [19451168](https://pubmed.ncbi.nlm.nih.gov/19451168/)
50. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28: 511–515. doi: [10.1038/nbt.1621](https://doi.org/10.1038/nbt.1621) PMID: [20436464](https://pubmed.ncbi.nlm.nih.gov/20436464/)
51. Colman DR, Toolson EC, Takacs-Vesbach CD (2012) Do diet and taxonomy influence insect gut bacterial communities? *Mol Ecol* 21: 5124–5137. doi: [10.1111/j.1365-294X.2012.05752.x](https://doi.org/10.1111/j.1365-294X.2012.05752.x) PMID: [22978555](https://pubmed.ncbi.nlm.nih.gov/22978555/)
52. Dietrich C, Köhler T, Brune A (2014) The cockroach origin of the termite gut microbiota: patterns in bacterial community structure reflect major evolutionary events. *Appl Environ Microbiol* 80: 2261–2269. doi: [10.1128/AEM.04206-13](https://doi.org/10.1128/AEM.04206-13) PMID: [24487532](https://pubmed.ncbi.nlm.nih.gov/24487532/)
53. Sabree Z, Moran N (2014) Host-specific assemblages typify gut microbial communities of related insect species. *SpringerPlus* 3: 138. doi: [10.1186/2193-1801-3-138](https://doi.org/10.1186/2193-1801-3-138) PMID: [24741474](https://pubmed.ncbi.nlm.nih.gov/24741474/)
54. Tai V, James ER, Nalepa CA, Scheffrahn RH, Perlman SJ, Keeling PJ (2015) The role of host phylogeny varies in shaping microbial diversity in the hindguts of lower termites. *Appl Environ Microbiol* 81: 1059–1070. doi: [10.1128/AEM.02945-14](https://doi.org/10.1128/AEM.02945-14) PMID: [25452280](https://pubmed.ncbi.nlm.nih.gov/25452280/)
55. Ohkuma M, Brune A (2011) Diversity, structure, and evolution of the termite gut microbial community. In: Bignell DE, Roisin Y, Lo N, editor. *Biology of termites: a modern synthesis*: Springer. pp. 413–438.
56. Tadmor AD, Ottesen EA, Leadbetter JR, Phillips R (2011) Probing individual environmental bacteria for viruses by using microfluidic digital PCR. *Science* 333: 58–62. doi: [10.1126/science.1200758](https://doi.org/10.1126/science.1200758) PMID: [21719670](https://pubmed.ncbi.nlm.nih.gov/21719670/)
57. Su NY, La Fage JP (1987) Initiation of worker-soldier trophallaxis by the Formosan subterranean termite (Isoptera: Rhinotermitidae). *Insect Soc* 34: 229–229.
58. Lewis JL, Forschler BT (2004) Protist communities from four castes and three species of *Reticulitermes* (Isoptera: Rhinotermitidae). *Ann Entomol Soc Am* 97: 1242–1251.
59. Lo N, Tokuda G, Watanabe D (2011) Evolution and function of endogenous termite cellulases. In: Bignell DE, Roisin Y, Lo N, editor. *Biology of termites: a modern synthesis*: Springer. pp. 51–67.
60. Watanabe H, Tokuda G (2010) Cellulolytic systems in insects. *Annu Rev Entomol*. Palo Alto: Annual Reviews. pp. 609–632. doi: [10.1146/annurev-ento-112408-085319](https://doi.org/10.1146/annurev-ento-112408-085319) PMID: [19754245](https://pubmed.ncbi.nlm.nih.gov/19754245/)
61. Henrissat B, Bairoch A (1993) New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochemical Journal* 293: 781–788. PMID: [8352747](https://pubmed.ncbi.nlm.nih.gov/8352747/)
62. Ni JF, Tokuda G (2013) Lignocellulose-degrading enzymes from termites and their symbiotic microbiota. *Biotechnol Adv* 31: 838–850. doi: [10.1016/j.biotechadv.2013.04.005](https://doi.org/10.1016/j.biotechadv.2013.04.005) PMID: [23623853](https://pubmed.ncbi.nlm.nih.gov/23623853/)