Heliyon 5 (2019) e02969

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

The *de novo* transcriptome of workers head of the higher group termite *Globitermes sulphureus* Haviland (Blattodea: Termitidae)

Nurul Akmar Hussin [a], Nazalan Najimudin [b], Abdul Hafiz Ab Majid [a,*]

[a] *Household and Structural Urban Entomology Laboratory, Vector Control Research Unit, School of Biological Sciences, Universiti Sains Malaysia, 11800, Minden, Pulau Pinang, Malaysia*

[b] *School of Biological Sciences, Universiti Sains Malaysia, 11800, Minden, Pulau Pinang, Malaysia*

ARTICLE INFO

Keywords:
Agricultural science
Animal science
Genetics
Transcriptomics
Globitermes sulphureus
de novo transcriptome
Trinity
RNA-Seq
Gene expression
Cellulase

ABSTRACT

The subterranean termite *Globitermus sulphureus* is an important Southeast Asian pest with limited genomic resources that causes damages to agriculture crops and building structures. Therefore, the main goal of this study was to survey the *G. sulphureus* transcriptome composition. Here, we performed *de novo* transcriptome for *G. sulphureus* workers' heads using Illumina HiSeq paired-end sequencing technology. A total of 88, 639, 408 clean reads were collected and assembled into 243, 057 transcripts and 193, 344 putative genes. The transcripts were annotated with the Trinotate pipeline. In total, 27, 061 transcripts were successfully annotated using BLASTX against the SwissProt database and 17, 816 genes were assigned to 47, 598 GO terms. We classified 14, 223 transcripts into COG classification, resulting in 25 groups of functional annotations. Next, a total of 12, 194 genes were matched in the KEGG pathway and 392 metabolic pathways were predicted based on the annotation. Moreover, we detected two endogenous cellulases in the sequences. The RT-qPCR analysis showed that there were significant differences in the expression levels of two genes β-glucosidase and endo-β-1,4-glucanase between worker and soldier heads of *G. sulphureus*. This is the first study to characterize the complete head transcriptome of a higher termite *G. sulphureus* using a high-throughput sequencing. Our study may provide an overview and comprehensive molecular resource for comparative studies of the transcriptomics and genomics of termites.

1. Introduction

Termites are a group of eusocial insects found on all continents except Antarctica (Govorushko, 2018). There are about 435 species of termites in Asia (Krishna et al., 2013) and specifically over 175 species in Peninsular Malaysia (Ab Majid and Ahmad, 2011). Termites are important in the ecological ecosystem as they consume and recycle wood, litter, leaf, and dung and release the nutrients back to the ecosystem (Freymann et al., 2008; Lenz et al., 2011). Termites tunneling help in soil fertility and improvements in soil composition (Govorushko, 2018).

Despite the status as ecological engineers, a few species of termites are regarded as pests, infesting and damaging structural woods of buildings, houses as well as agricultural plantations (Kuswanto et al., 2015). Damages caused by termites has led to economic loss; aproximatelybillions of US dollars per year. Rust and Su (2012) revealed that USD32 billion was required for control and damage repairs globally in 2010. In Malaysia, subterranean termite control accounted for about USD8 - 10 million in the year 2000 (Lee, 2002).

The subterranean termite, *Globitermes sulphureus* (Haviland) (Blattodea: Termitidae), is a higher wood-feeding termite that is distributed throughout Southeast Asia, including Singapore, Thailand, and Malaysia (Hussin et al., 2018). This species can be easily recognized by the bright yellow colored abdomen of its soldier (Lee et al., 2002). *G. sulphureus* is an important agricultural pest, especially in coconut and oil palm plantations and is also regarded as a secondary pest where the members of this species infest premises that have been treated with termite bait (Bakaruddin et al., 2018; Lee et al., 2007; Neoh et al., 2011). *G. sulphureus* is commonly found infesting doors, window frames, wooden walls, wooden pillars, and carpet grippers (Ab Majid and Ahmad, 2009; Ngee and Lee, 2002).

Although *G. sulphureus* has a significant economic impact, its genomic data in public database are particularly scarce and only a few sequences have been published. The available sequences aremainly corresponded to mitochondrial (Bourguignon et al., 2017), ribosome fragments (Yeap et al., 2007), and endo-β-1,4-glucanase (Bujang et al., 2014). Even for other termites species, to date only three termite genome sequences are

* Corresponding author.
E-mail address: abdhafiz@usm.my (A.H. Ab Majid).

https://doi.org/10.1016/j.heliyon.2019.e02969
Received 28 June 2019; Received in revised form 4 November 2019; Accepted 28 November 2019

publicly available: two from the lower termite *Zootermopsis nevadensis* (Terrapon et al., 2014) and *Cryptotermes secundus* (Harrison et al., 2018) and one from the higher termite *Macrotermes natalensis* (Poulsen et al., 2014). The genome size of *Z. nevadensis* is 562 Mb smaller than *C. secundus* (1.30 Gb) and *M. natalensis* (1.31Gb) (Harrison et al., 2018) while the cockroach *Blattella germanica* genome (2.0 Gb) is considerably larger than all three termite genomes.

The actual genome size of *G. sulphureus* is unknown but a study done by Koshikawa et al. (2008) has mentioned that the Termitidae genome size based on C-value ranged from 1.48 pg to 1.90 pg. The rapid development of next-generation sequencing (NGS) technologies offers the opportunity to gather population genomic data in non-model organisms, in the absence of prior knowledge, at an affordable cost (Gayral et al., 2013). RNA sequencing (RNA-Seq) has been successful and increasingly used to define the transcriptome in termites (Dedeine et al., 2015; Geng et al., 2018; Huang et al., 2012; Rosenthal et al., 2011; Yuki et al., 2008). Genomes of higher termites are considered large, highly repetitive and, consequently, difficult to assemble (Korb et al., 2015). We therefore, decided to assemble its transcriptome. *De novo* transcriptomics (assembling transcriptomes without a reference genome) gives access to large numbers of genes at relatively low cost, with potential applications for SNP discovery and molecular markers (Gayral et al., 2013; Khizam and Ab Majid, 2019; Levy and Myers, 2016). *De novo* transcriptome has also been applied in other insect studies such as beetle (Sayadi et al., 2016), ants (Morandin et al., 2018), and cockroach (Kim et al., 2016).

We have previously recorded the metagenomic data from the worker and soldier gut of *G. sulphureus* (Hussin et al., 2018) and isolated their gut fauna using culture-dependent method (Hussin and Ab Majid, 2017). Lower and higher group termites were proven synergistically with their gut symbionts for lignocellulose digestion (Franco Cairo et al., 2016; Ni and Tokuda, 2013). Termites and their gut symbionts produced various cellulase enzymes in the salivary gland, midgut, and hindgut (Brune, 2014; Ni and Tokuda, 2013). These enzymes have demonstrated great potential for bioenergy applications (Brune, 2014; Scharf and Tartar, 2008; Singhania et al., 2013). Endo-β-1,4-glucanases and β-glucosidases are the key enzymes for cellulose digestion (Brune, 2014; Bujang et al., 2014; Lima et al., 2014; Ni and Tokuda, 2013).

In this present study, we used RNA-Seq approach to transcriptome profiling of worker heads from *G. sulphureus* using the Illumina HiSeq sequencing machine. The RNA was extracted from the head in order to minimize contamination by sequences from endosymbionts that inhabit the digestive tract of termites. We chose the worker because it is the dominant caste in the termite colony and plays a vital role in foraging and feeding other castes. We aimed to characterize the functional annotations of *G. sulphureus* head transcriptome. Furthermore, we validated the transcriptome data and compared the gene expression level of two cellulases genes between worker and soldier heads using real-time quantitative PCR. We hypothesized that the enzyme expression level would vary according to termite castes. Our results will provide an important resource for gene discovery and comparative genomics of termites and future termite pest management studies of the higher termite species.

2. Materials and methods

2.1. Termite materials for RNA-seq

The *G. sulphureus* colony was collected from a mound at the main campus of Universiti Sains Malaysia, Penang, Malaysia during August 2017. The termites together with the nest were transported to the laboratory using the black bucket and were used instantly. Only healthy workers (50 individuals) were selected from this single colony. The worker heads were separated from the bodies by using dissecting scissors. Then these 50 heads were pooled together and immediately stored in RNAlater RNA Stabilization Reagent (Qiagen, Germany) for further processing.

2.2. RNA extraction, cDNA library construction, and illumina sequencing

The head sample was shipped in dry ice to Major Biosystem Corporation, Taiwan for RNA extraction, cDNA library construction, and Illumina sequencing. Total RNA was extracted using TRIzol® reagent (Invitrogen, USA) according to the manufacturer's protocol. RNA purity was checked using a NanoDrop 2000 spectrophotometer. The RNA integrity was observed on 1% agarose gel electrophoresis and RNA integrity value was analyzed using Agilent 2100 Bioanalyzer. The sample was verified to have a RIN (RNA integrity number) value of 7.4. The RNA was used for the building of cDNA libraries using the Truseq™ RNA sample prep Kit (Illumina, San Diego, USA) as recommended by the manufacturer. The cDNA was amplified by PCR using Phusion DNA polymerase (NEB) for 15 PCR cycles and cDNA Library was obtained by gel extraction selected for target fragments on 2% Low Range Ultra Agarose (Certified Low Range Ultra Agarose, Bio-Rad). The library was then quantified using TBS380 (Picogreen, Invitrogen, Carlsbad, CA, USA) followed by bridge PCR amplification using the cBot Cluster Generation System (Truseq PE Cluster Kit v3-cBot-HS, Illumina, San Diego, USA). At last, the cDNA library was sequenced on the Illumina Hiseq X10 platform which generated 2 × 150 bp of paired-end raw reads.

2.3. De novo transcriptome assembly

The raw reads (FASTQ data) were quality checked for base sequence quality, GC content, base N content, and adapter content using FastQC v0.11.5 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). These raw reads were then trimmed using Trimmomatic software v0.36 (Bolger et al., 2014) to remove adapters, low-quality reads, and to discard the reads with a minimum length of 36 bp. The resulted sequences were *de novo* assembled using Trinity software v2.5.1 with a default k-mer size of 25 (Grabherr et al., 2011; Haas et al., 2013) following default protocols. The outputs were saved in the FASTA format. The quality of the assembled transcripts has been verified using Bowtie2 v2.3.3.1 (Langmead and Salzberg, 2012).

2.4. Data deposition

All sequences acquired in this research were deposited under accession number (SRP147368) connected with BioProject PRJNA471675 in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) databases.

2.5. Transcriptome function annotation

All the assembled transcripts were functionally annotated using Trinotate pipeline v3.1.1 with a cutoff E-value of 10^{-5} (https://trinotate.github.io/). TransDecoder v5.0.2 (http://transdecoder.github.io) was used to predict coding regions within the generated transcripts and translate the longest open reading frame (ORF) (with a minimum at least 100 amino acids long) into peptide sequences. Trinity transcriptome file was searched for nucleotide sequences homology using BLASTX (NCBI-blast v2.7.1+) against several databases (SwissProt database, Kyoto Encyclopedia of Genes and Genomes (KEGG), Gene Ontology (GO), and evolutionary genealogy of genes: Non-supervised Orthologous Groups (EggNog). The transcripts were also searched for potential ribosomal RNA using RNAmmer (v1.2) (Lagesen et al., 2007). The transdecoder peptide sequence file for final candidate longest-ORF was searched for amino acid sequence homologies using BLASTP (NCBI-blast v2.7.1+), protein domain identification via a Pfam search using Hmmer v3.1b2 software (Finn et al., 2011), presence of signal peptide cleavage sites using SignalP v4.1 (Petersen et al., 2011), and prediction of transmembrane helices in proteins using TMHMM v2.0c (Krogh et al., 2001).

All the results above were loaded into the Trinotate SQLite database and reported in Excel format. The GO annotations were extracted from

the Trinotate output and were visualized using the Web Gene Ontology Annotation Plot (WEGO) 2.0 for plotting the biological process, molecular functions, and cellular components of GO terms (Ye et al., 2006). The Cluster of Orthologous Groups (COG) functional annotations were extracted from Trinotate output which derived from the EggNog database.

2.6. Validation and relative gene expression using real-time quantitative PCR

Total RNA was extracted from the heads of worker (30 heads) and soldier (30 heads) castes using a combination of Trizol reagent (Invitrogen, USA) and with RNeasy® MinElute® Cleanup Kit (Qiagen, Germany) following to manufacturer's protocol. The total RNA was treated with RNase-free DNase (Qiagen, Germany) to remove any trace of genomic DNA. The treated RNA samples were converted into cDNA using SuperScript™ III First-Strand Synthesis (Invitrogen, USA). The cDNA products were then used as a template in real-time qPCR. The RT-qPCR mixture was prepared using iTaqTM Universal SYBR® Green Supermix (Bio-Rad, USA). The reaction mixture (10 μL) contained 5 μL of 2X iTaq™ Universal SYBR® Green Supermix, 0.5 μLof each forward and reverse primers, 1 μL of cDNA template, and 3 μL of DEPC-treated water. The RT-qPCR was performed with the Bio-Rad® CFX96TM machine with the following conditions: polymerase activation at 95 °C for 30 s, followed by 40 cycles of DNA denaturation at 95 °C for 5 s and annealing/extension + plate read at 60 °C for 30 s, at last, melt-curve analysis at 65 °C - 95 °C with 0.5 °C increment/step. Two interesting genes were selected for validation: β-glucosidase and endo-β-1,4-glucanase. Primer sequences were listed in Table 1. Data generated from RT-qPCR were analyzed using the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001). All qPCR has been repeated in the three biological and three technical replications. Two genes expression was normalized against β-actin expression. Differences in the expression level of the two genes between workers and soldiers were tested using the t-test.

3. Results

3.1. Illumina sequencing and de novo assembly

In total, 93, 347, 116 raw sequencing reads were generated using the Illumina HiSeq X10 sequencing platform. Table 2 summarizes the statistics of the sequencing data and Trinity *de novo* assembly. After data filtering with Trimmomatic software, a total of 88, 639, 408 (95%) high-quality reads were obtained and used for assembly. *De novo* assembly using Trinity results in 243, 057 of transcripts and 193, 344 of genes with a GC percentage of 40.43%. The average length of transcripts was 813 bases and an N50 length was 1581. The quality of the assembled transcripts was checked using Bowtie2 and the result was 97% of the were correctly assembled.

3.2. Transcriptome annotation

Among 243, 057 Trinity transcripts, 27, 061 (11.13%) had top BLASTX hits in the SwissProt database. Meanwhile, only 51, 655 (21.25%) of all transcript sequences were effectively translated into peptide sequences using TransDecoder. From these numbers, 34, 893 (67.55%) of peptide sequences had top BLASTP hits in the SwissProt database and matched to 32, 221 (62.38%) protein entries in the Pfam database.

Table 1. Primers used in the relative gene expression experiment.

Primers	Sequence 5′-3′
β-glucosidase F	CCACCACGTTAACTCCATCTAT
β-glucosidase R	GACACCAACAGGGTACTTTACT
Endo-β-1,4-glucanase F	CAGGACTTGACAGGGGGATA
Endo-β-1,4-glucanase R	TAGTCCGTGGCCCATTTAAC
β-actin F	CTCTTCCAGCCTTCCTTCCT
β-actin R	CTTCTGCATCCTGTCAGCAA

Table 2. Summary of the head transcriptome of *G. sulphureus* from Trinity assembly.

Total Trinity 'genes'	193344
Total Trinity transcripts	243057
Percent GC	40.43
Stats based on All transcript contigs	
Contig N10	5272
Contig N20	3785
Contig N30	2850
Contig N40	2163
Contig N50	1581
Median contig length	384
Average contig length	812.90
Total assembled bases	197581041
Stats based on ONLY LONGEST ISOFORM per 'GENE'	
Contig N10	4285
Contig N20	2831
Contig N30	1942
Contig N40	1296
Contig N50	885
Median contig length	341
Average contig length	616.34
Total assembled bases	119166060

3.3. Functional classification by GO

In total, 17, 816 genes were assigned to 47, 598 GO terms. The GO terms were assigned to cellular compartment (16, 139, 33.91%), molecular function (15, 954, 33.52%), and biological process (15, 505, 32.57%). The terms were derived from 66 different functional groups (GO sub-categories level 2) (Figure 1). Under cellular compartment category, cell (14, 763 genes, 82.9%), cell part (14, 724 genes, 82.6%), and organelle (12, 020 genes, 67.5%) represented the majorities categories (Figure 1). For molecular function categories, binding (12, 855 genes, 72.2%) and catalytic activity (7, 738 genes, 43.4%) were depicted prominently. Meanwhile, under the biological process category, cellular process (13, 686 genes, 76.8%), and metabolic process (11, 301 genes, 63.4%) were the dominant subcategories.

3.4. Functional classification by COG

In total, 14, 223 of transcripts were assigned to multiple COG classification, resulting in 16, 211 of functional annotations. The results showed that these sequenced were annotated to 25 COGs functional categories (Figure 2). Among those categories, “General function prediction only” (2, 999, 18.5%) represented the largest group, followed by “Signal transduction mechanisms” (1, 560, 9.62%), “Posttranslational modification, protein turnover, chaperones” (1, 384, 8.54%), and “Translation, ribosomal structure and biogenesis” (1, 367, 8.43%). A few other categories, such as “Cell motility”, “Mobilome: prophages, transposons”, “Nuclear structure” were the smallest groups and “Extracellular structures” categories were absent.

3.5. Functional classification by KEGG

As a result, 12, 194 genes were matched in the KEGG pathway and were assigned to 392 KEGG pathways. The distribution of KEGG pathways was shown at the second hierarchical stage (Figure 3). The top five

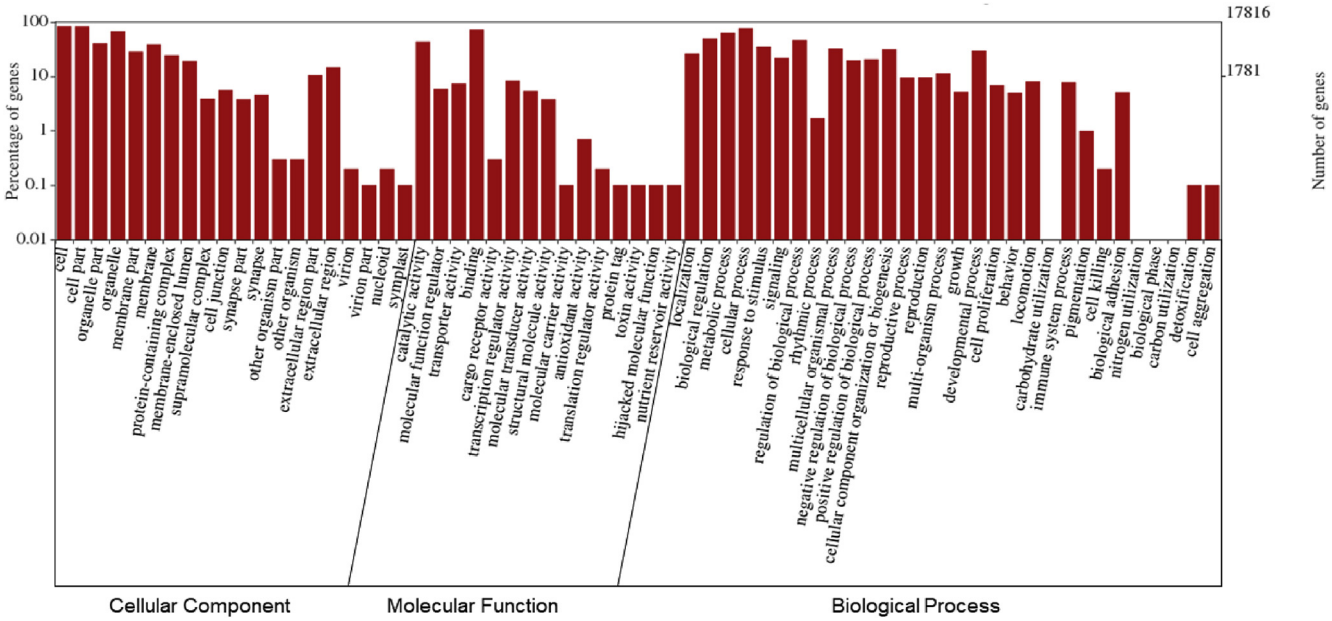

Figure 1. Histogram of Gene Ontology (GO) classification. The results were summarized into three categories: cellular component, molecular function, and biological process.

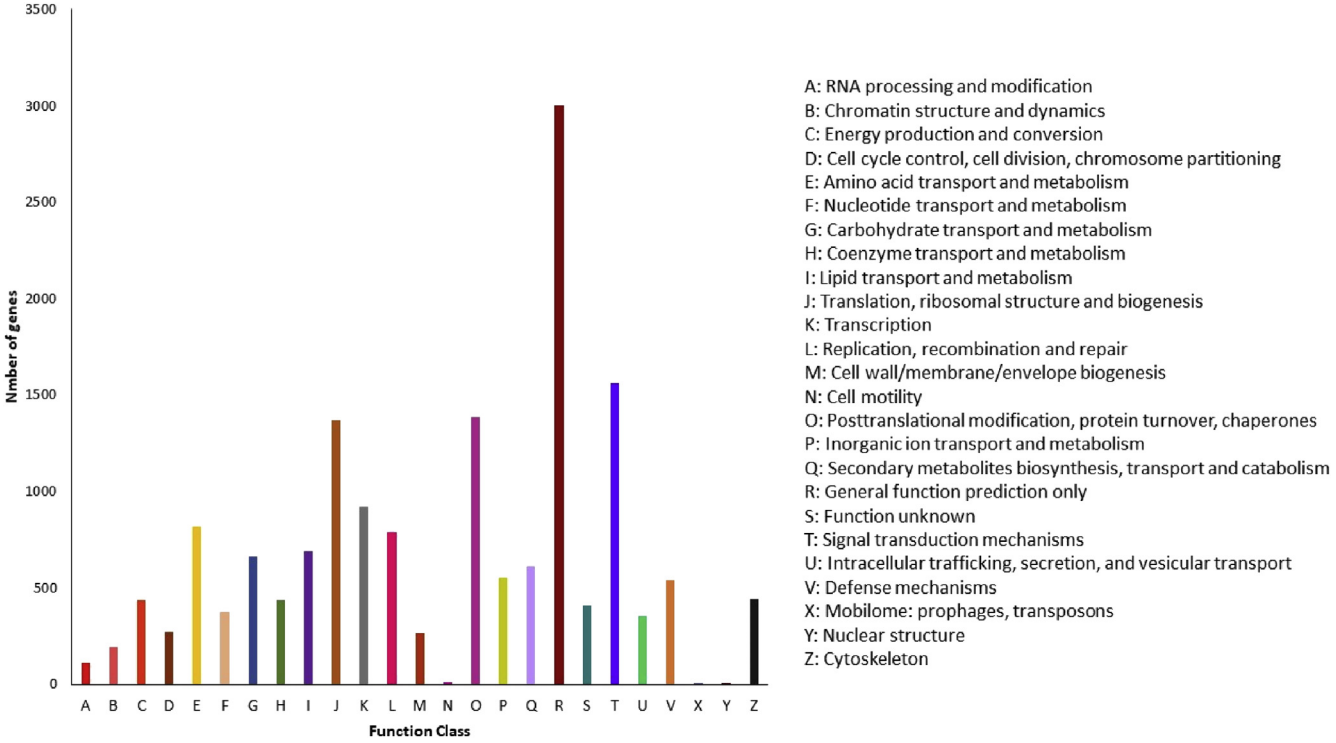

Figure 2. COG categories for the *G. sulphureus* workers' head transcriptome.

largest pathway groups were signal transduction (1439, 11.80%), global and overview maps (1346, 11.04%), infectious disease: viral (694, 5.69%), endocrine system (680, 5.58%), and cancers: specific types (526, 4.31%).

3.6. Gene expression analysis

To validate the RNA-seq outcomes of the worker head samples, two cellulase genes were selected, β-glucosidase and endo-β-1,4-glucanase. The relative gene expression experiment was performed using the RNA sample derived from the worker and soldier heads. Figure 4 shows the relative expression levels of both cellulases in both castes. The β-glucosidase was considerably greater in the head of the soldiers compared to the workers ($p < 0.05$). While the endo-β-1,4-glucanase expression level in worker heads was significantly higher than soldiers ($p < 0.05$).

4. Discussions

In the present study, we characterized the transcriptome data from the worker heads of the higher termite, *G. sulphureus* using the Illumina

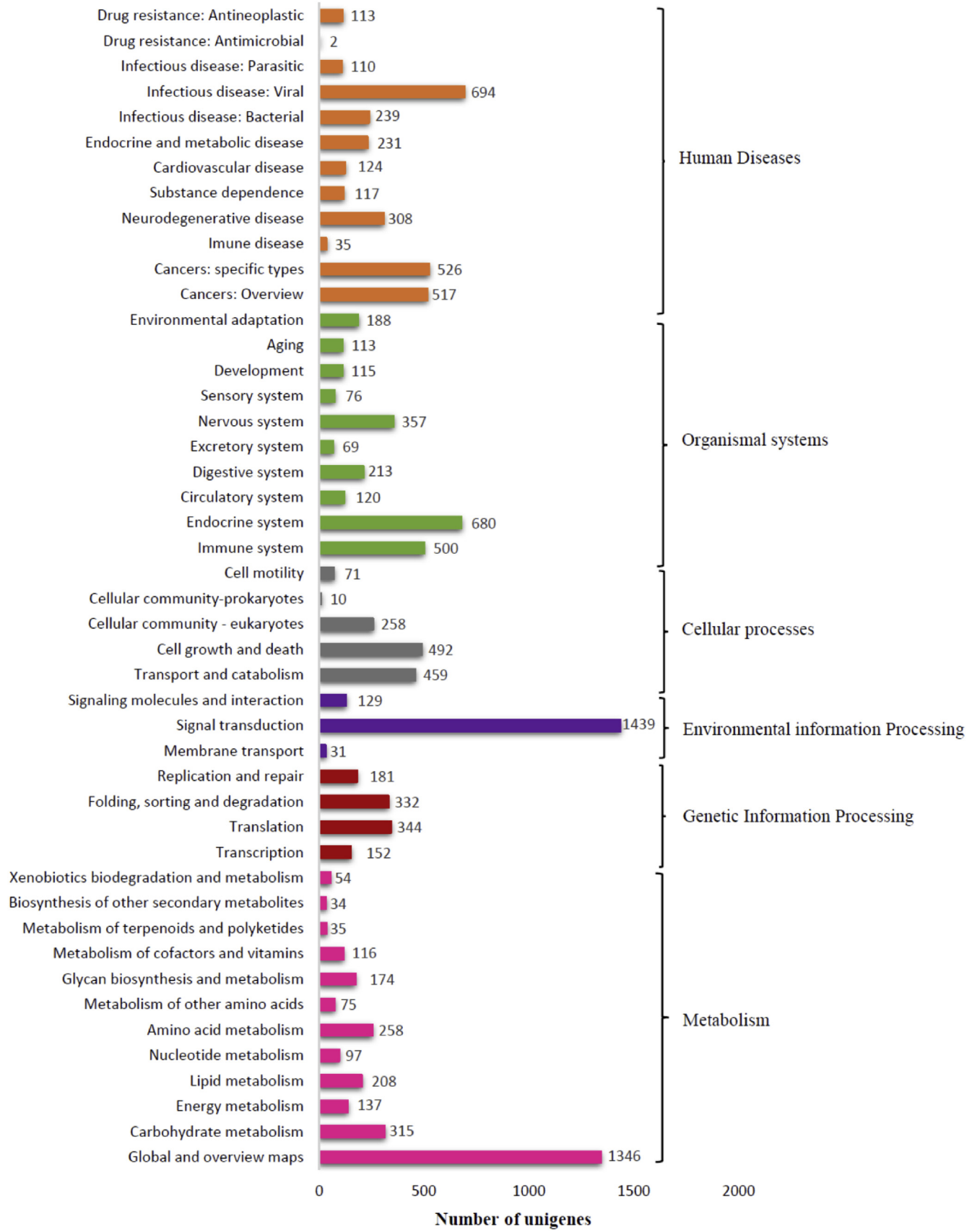

Figure 3. KEGG orthology (KO) classification. The second hierarchical stage of KEGG pathways were assigned into six categories: metabolism, genetic information processing, environmental information processing, cellular processes, organismal systems, and human diseases.

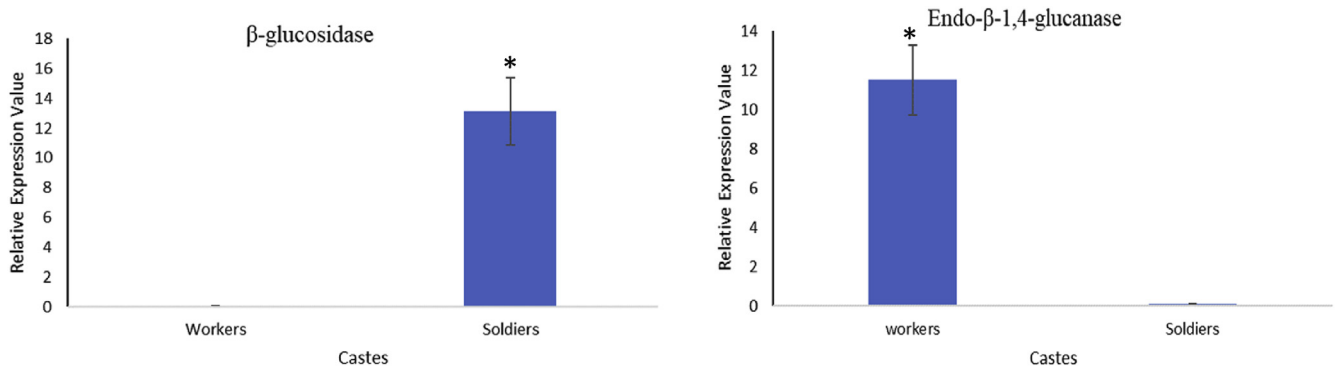

Figure 4. Relative expression levels of (left) β-glucosidase and (right) endo-β-1,4-glucanase in between workers' and soldiers' heads measured by real-time PCR. The y-axis indicates the relative expression value of genes and data are mean values of three independent replicates. Significant differences were identified by a t-test: *, $p < 0.05$.

HiSeq sequencing platform. Several transcriptome studies in termite species have been carried out by two of which have sequenced the worker heads from *Coptotermes gestroi* (Leonardo et al., 2011) and *Odontotermes formosanus* (Huang et al., 2012). While the rest of the studies have focused on the host termite, gut microbial symbionts, or combined the host and symbionts (Scharf, 2015). The transcriptome of *G. sulphureus* was not recorded before, so this is the first effort to examine the head transcriptome of this species.

The GO describes the functions of gene products in terms of their association with cellular components, molecular functions, and biological processes in a species-independent manner (Xie et al., 2012). In the cellular component category, most of our gene products were located in the cell and cellular part component. These genes were mainly molecular functions as binding and catalytic activity that were involved in the biological process category which was cellular processes and metabolic processes. These results represent most components are related to each other and important for maintaining the basic functions of genes in the cell and tissues of an organism. These components were also abundantly found in not only termites *O. formosanus* (Huang et al., 2012), *C. gestroi* (only in the biological process) (Leonardo et al., 2011), dry wood termite *Cryptotermes domesticus* (Wu et al., 2015), but also in other insects such as seed beetle *Callosobruchus maculatus* (Sayadi et al., 2016), oriental armyworm *Mythimna separate* (Bian et al., 2017), pear *Pyrus calleryana* Decne (Xu et al., 2015), and purple sweet potato *Impomoea batatas* L. (Xie et al., 2012).

The Cluster of Orthologous Groups (COG) is a database where the proteins are classified into 26 functional categories (A to Z) based on the orthology concept (Tatusov, 2000; Tatusov et al., 2003). From the results, the generated transcriptome profiles for the heads of *G. sulphureus* were not similar to *O. formosanus* except that both termites have the "general function prediction only" as the largest group in the COG categories (Huang et al., 2012). The "general function prediction only" is a member of a poorly characterized group where most of the proteins have similar biochemical activity but different in functions (R. L. Tatusov, 2000). The same goes for the KEGG pathways where both termites have a different pattern of pathways classification. The different diet of these termites (*O. formosanus*: fungus-feeding termite and *G. sulphureus*: wood-feeding termite) and distantly related in a phylogenetic relationship (Bujang et al., 2014) may contribute to the dissimilarity in transcriptome profiles.

Two cellulase genes homologous to β-glucosidase and endo-β-1,4-glucanase were selected to validate the RNA-seq. Both genes also were selected to compare their expression levels between workers and soldiers. The expression of β-glucosidase was considerably greater in soldiers, while endo-β-1,4-glucanase was significantly higher in workers (Figure 4). These results were contradicted with the previous study in which *C. gestroi* workers' caste had a high-level expression of β-glucosidase compared to the caste of soldiers (Leonardo et al., 2011). A study of *Nasutitermes corniger* exhibited the greater level of exoglucanase and β-glucosidase activities in soldiers (Lima et al., 2014). Their relative expression analysis and enzyme activities were using the whole body or whole gut of the workers and soldiers, respectively. It is important to note that the relative expression level of genes does not correspond with its activity in the gut (Fujita et al., 2008). Further investigation concerning the digestive system of *G. sulphureus* should focus on enzyme activities for each host and gut compartment.

Since we prepared cDNA libraries from the worker and soldier heads, the cellulases we identified were likely expressed in the salivary gland. This finding was consistent with previous studies that β-glucosidase and endo-β-1,4-glucanase were found expressed in *C. gestroi* (Leonardo et al., 2011), *Crypototermes domesticus*, *Coptotermes formosanus*, *Ahmaditermes sichuanensis* (Li et al., 2012), and *Reticulitermes flavipes* (Tartar et al., 2009). From the nucleotide sequence homology using BLASTX, our β-glucosidases and endo-β-1,4-glucanase have been categorized into GH1 and GH9 (glycoside hydrolases) families. GH1 and GH9 are groups of enzyme families generated by the termite itself rather than the gut symbionts. This was consistent with the past studies of lower termites (Franco Cairo et al., 2016; Tartar et al., 2009) and higher termites (Bujang et al., 2014).

Cellulose digestion into glucose involves three different enzymes: endo-β-1,4-glucanase (EC.3.2.1.4), exoglucanase or cellobiohydrolases (EC.3.2.1.9.1), and β-glucosidases (EC.3.2.1.21) (Li et al., 2012). Endo-glucanases cut at random glycoside bonds within amorphous cellulose chains, while cellobiohydrolases cleave at the non-reducing ends of crystalline cellulose producing cellobiose, glucose, and cellotriose. The β-glucosidases (cellobiases) then hydrolyze the cellobiose to glucose (Lima et al., 2014; Terra and Ferreira, 1994). Workers and soldiers have different feeding habits, and the results obtained from this study may reflect the biological differences between castes. The worker caste forages and feeds other colony members. The soldier caste defends the colony from enemies and is fed by worker caste *via* trophallaxis. Thus, they receive the partially digested food from workers such as regurgitated food, saliva, or liquid excreta rich in symbionts (Nalepa, 2015). Lima et al. (2014) mentioned that the food originating from lignocellulose products requires the enzyme endo-β-1,4-glucanase for partial digestion. As a result, the expression of this enzyme is high in workers. This partially digested food is then supplied to the soldiers engaged in the enzymes exoglucanase and β-glucosidase for the final digestion of the cellulose component, resulting in a high-level expression of β-glucosidase in soldiers.

5. Conclusions

Our data transcriptome provided the functional annotations of genomic and molecular elements in the cells and tissues of *G. sulphureus*. These annotations can be used as a reference for comparative studies and genomic analysis of termites in the future. From our qPCR outcomes,

further in-depth molecular studies and enzyme operations of two cellulase genes should be performed to elucidate the metabolic pathways engaged in lignocellulase enzymes and their possible use in biotechnological procedures.

Declarations

Author contribution statement

Nurul Akmar Hussin: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

Mohd Nazalan Mohd Najimudin: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data.

Abdul Hafiz Ab Majid: Conceived and designed the experiments; Contributed reagents, materials, analysis tools or data; Wrote the paper.

Funding statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interest statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

References

Ab Majid, A.H., Ahmad, A.H., 2009. The status of subterranean termite infestation in Penang, Seberang Perai and Kedah, Malaysia. Malays. Appl. Biol. 38 (2), 37–48.

Ab Majid, A.H., Ahmad, A.H.A., 2011. Foraging population, territory, and control of *Globitermes sulphureus* (Isoptera: Termitidae) with Fipronil in Penang, Malaysia. Malays. Appl. Biol. 40, 61–65.

Bakaruddin, N.H., Dieng, H., Sulaiman, S.F., Ab Majid, A.H., 2018. Evaluation of the toxicity and repellency of tropical plant extract against subterranean termites, *Globitermes sulphureus* and *Coptotermes gestroi*. Inf. Process. Agric. 5, 298–307.

Bian, H.X., Ma, H.F., Zheng, X.X., Peng, M.H., Li, Y.P., Su, J.F., Wang, H., Li, Q., Xia, R., Liu, Y., Jiang, X.F., 2017. Characterization of the adult head transcriptome and identification of migration and olfaction genes in the oriental armyworm *Mythimna separate*. Sci. Rep. 7, 2324.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

Bourguignon, T., Lo, N., Sobotnik, J., Ho, S.Y.W., Iqbal, N., Coissac, E., Lee, M., Jendryka, M.M., Sillam-Dussès, D., Křížková, B., Roisin, Y., Evans, T.A., 2017. Mitochondrial phylogenomics resolves the global spread of higher termites, ecosystem engineers of the tropics. Mol. Biol. Evol. 34 (3), 589–597.

Brune, A., 2014. Symbiotic digestion of lignocellulose in termite guts. Nat. Rev. Microbiol. 12 (3), 168–180.

Bujang, N., Harrison, N., Su, N., 2014. A phylogenetic study of endo-beta-1,4-glucanase in higher termites. Insectes Sociaux 61, 29–40.

Dedeine, F., Weinert, L.A., Bigot, D., Josse, T., Ballenghien, M., Cahais, V., Galtier, N., Gayral, P., 2015. Comparative analysis of transcriptomes from secondary reproductives of three *Reticulitermes* termite species. PLoS One 10 (12), 1–18.

Finn, R.D., Clements, J., Eddy, S.R., 2011. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39, 29–37.

Franco Cairo, J.P.L., Carazzolle, M.F., Leonardo, F.C., Mofatto, L.S., Brenelli, L.B., Gonçalves, T.A., Uchima, C.A., Domingues, R.R., Alvarez, T.M., Tramontina, R., Vidal, R.O., Costa, F.F., Costa-Leonardo, A.M., Leme, A.F.P., Pereira, G.A.G., Squina, F.M., 2016. Expanding the knowledge on lignocellulolytic and redox enzymes of worker and soldier castes from the lower termite *Coptotermes gestroi*. Front. Microbiol. 7 (OCT).

Freymann, B.P., Buitenwerf, R., Desouza, O., Olff, H., 2008. The importance of termites (Isoptera) for the recycling of herbivore dung in tropical ecosystems: a review. Eur. J. Entomol. 105 (2), 165–173.

Fujita, A., Miura, T., Matsumoto, T., 2008. Differences in cellulose digestive systems among castes in two termite lineages. Physiol. Entomol. 33 (1), 73–82.

Gayral, P., Melo-Ferreira, J., Glémin, 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 (4), 1–15.

Geng, A., Cheng, Y., Wang, Y., Zhu, D., Le, Y., Wu, J., Xie, R., Yuan, J.S., Sun, J., 2018. Transcriptome analysis of the digestive system of a wood-feeding termite (*Coptotermes formosanus*) revealed a unique mechanism for effective biomass degradation. Biotechnol. Biofuels 11 (1), 1–14.

Govorushko, S., 2018. Economic and ecological importance of termites: a global review. Entomol. Sci. 1–15.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29 (7), 644–652.

Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Lieber, M., MacManes, M.D., Ott, M., Orvis, J., Pochet, N., Strozzi, F., Weeks, N., Westerman, R., William, T., Dewey, C.N., Henschel, R., LeDuc, R.D., Friedman, N., Regev, A., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8 (8), 1494–1512.

Harrison, M.C., Jongepier, E., Robertson, H.M., Arning, N., Bitard-Feildel, T., Chao, H., Bornberg-Bauer, E., 2018. Hemimetabolous genomes reveal molecular basis of termite eusociality. Nature Ecol. Evol. 2 (3), 557–566.

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 (11), 1–11.

Hussin, N.A., Ab Majid, A.H., 2017. Inter and intra termites colonies comparisons of gut microbial diversity from worker and soldier caste of *Globitermes sulphureus* (Blattodea: Termitidae) using 16S rRNA gene. Malays. J. Microbiol. 13 (3), 228–234.

Hussin, N.A., Zarkasi, K.Z., Ab Majid, A.H., 2018. Characterization of gut bacterial community associated with worker and soldier castes of *Globitermes sulphureus* Haviland (Blattodea: Termitidae) using 16S rRNA metagenomic. J. Asia Pac. Entomol. 21 (4), 1268–1274.

Khizam, N.A.N., Ab Majid, A.H., 2019. Development and annotation of species-specific microsatellite markers from transcriptome sequencing for a higher group termite, *Globitermes sulphureus* Haviland (Blattodea: Termitidae). Meta Gene 20, 1–6.

Kim, I.W., Lee, J.H., Subramaniyam, S., Yun, E.Y., Kim, I., Park, J., Hwang, J.S., 2016. De novo transcriptome analysis and detection of antimicrobial peptides of the American cockroach *Periplaneta americana* (Linnaeus). PLoS One 11 (5), 1–16.

Korb, J., Poulsen, M., Hu, H., Li, C., Boomsma, J.J., Zhang, G., Liebig, J., 2015. A genomic comparison of two termites with different social complexity. Front. Genet. 6 (9), 1–12.

Koshikawa, S., Miyazaki, S., Cornette, R., Matsumoto, T., Miura, T., 2008. Genome size of termites (Insecta, Dictyoptera, Isoptera) and wood roaches (Insecta, Dictyoptera, Cryptocercidae). Naturwissenschaften 95, 859–867.

Krishna, K., Grimaldi, D.A., Krishna, V., Engel, M.S., 2013. Treatise on the Isoptera of the world. Bull. Am. Mus. Nat. Hist. 377 (1).

Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L., 2001. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. J. Mol. Biol. 305 (3), 567–580.

Kuswanto, E., Ahmad, I., Dungani, R., 2015. Threat of subterranean termites attack in the asian countries and their control: a review. Asian J. Appl. Sci. 8 (4), 227–239.

Lagesen, K., Hallin, P., Rødland, E.A., Stærfeldt, H.H., Rognes, T., Ussery, D.W., 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 35 (9), 3100–3108.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9 (4), 357–359.

Lee, C.Y., 2002. Control of foraging colonies of subterranean termites, *Coptotermes travians* (Isoptera: Rhinotermitidae) in Malaysia using hexaflumuron baits. Sociobiology 39 (3), 411–416.

Lee, C.Y., Vongkaluang, C., Lenz, M., 2007. Challenges to subterranean termite management of multi-genera faunas in Southeast Asia and Australia. Sociobiology 50 (1), 213–221.

Lenz, M., Lee, C.Y., Lacey, M.J., Yoshimura, T., Tsunoda, K., 2011. The potential and limits of termites (Isoptera) as decomposers of waste paper products. J. Econ. Entomol. 104 (1), 232–242.

Leonardo, F.C., da Cunha, A.F., da Silva, M.J., Carazzolle, M.F., Costa-Leonardo, A.M., Costa, F.F., Pereira, G.A., 2011. Analysis of the workers head transcriptome of the Asian subterranean termite, *Coptotermes gestroi*. Bull. Entomol. Res. 101 (4), 383–391.

Levy, S.E., Myers, R.M., 2016. Advancements in next-generation sequencing. Annu. Rev. Genom. Hum. Genet. 17 (1), 95–115.

Li, Z.Q., Zeng, W.H., Zhong, J.H., Liu, B.R., Li, Q.J., Xiao, W.L., 2012. Cellulase activity in higher and lower wood-feeding termites. Sociobiology 59 (4), 1157–1166.

Lima, T. de A., Pontual, E.V., Dornelles, L.P., Amorim, P.K., Sá, R.A., Coelho, L.C.B.B., Napoleão, T.H., Paiva, P.M.G., 2014. Digestive enzymes from workers and soldiers of termite *Nasutitermes corniger*. Comp. Biochem. Physiol., B 176, 1–8.

Livak, K.J., Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C}_{T}$ method. Methods 25, 402–408.

Morandin, C., Pulliainen, U., Bos, N., Schultner, E., 2018. *De novo* transcriptome assembly and its annotation for the black ant *Formica fusca* at the larval stage. Sci. Data 5, 1–8.

Nalepa, C.A., 2015. Origin of termite eusociality: trophallaxis integrates the social, nutritional, and microbial environments. Ecol. Entomol. 40 (4), 323–335.

Neoh, K.B., Jalaludin, N.A., Lee, C.Y., 2011. Elimination of field colonies of a mound-building termite *Globitermes sulphureus* (Isoptera: Termitidae) by bistrifluron bait. J. Econ. Entomol. 104 (2), 607–613.

Ngee, P.S., Lee, C.Y., 2002. Colony characterization of a mound-building subterranean termite, *Globitermes sulphureus* (Isoptera: Termitidae) using modified single-mark recapture technique. Sociobiology 40 (3), 525–532.

Ni, J., Tokuda, G., 2013. Lignocellulose-degrading enzymes from termites and their symbiotic microbiota. Biotechnol. Adv.

Petersen, T.N., Brunak, S., Von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8 (10), 785–786.
Poulsen, M., Hu, H., Li, C., Chen, Z., Xu, L., Otani, S., Zhang, G., 2014. Complementary symbiont contributions to plant decomposition in a fungus-farming termite. Proc. Nat. Acad. Sci. 111 (40), 14500–14505.
Rosenthal, A.Z., Matson, E.G., Eldar, A., Leadbetter, J.R., 2011. RNA-seq reveals cooperative metabolic interactions between two termite-gut spirochete species in co-culture. ISME J. 5 (7), 1133–1142.
Rust, M.K., Su, N.Y., 2012. Managing social insects of urban importance. Annu. Rev. Entomol. 57 (1), 355–375.
Sayadi, A., Immonen, E., Bayram, H., Arnqvist, G., 2016. The *de novo* transcriptome and its functional annotation in the seed beetle *Callosobruchus maculatus*. PLoS One 11 (7), e0158565.
Scharf, M.E., 2015. Omic research in termites: an overview and a roadmap. Front. Genet. 6 (MAR).
Scharf, M.E., Tartar, A., 2008. Termite digestomes as sources for novel lignocellulases. Biofuels Bioprod. Biorefining.
Singhania, R.R., Patel, A.K., Sukumaran, R.K., Larroche, C., Pandey, A., 2013. Role and significance of beta-glucosidases in the hydrolysis of cellulose for bioethanol production. Bioresour. Technol. 127, 500–507.
Tartar, A., Wheeler, M.M., Zhou, X., Coy, M.R., Boucias, D.G., Scharf, M.E., 2009. Parallel metatranscriptome analyses of host and symbiont gene expression in the gut of the termite *Reticulitermes flavipes*. Biotechnol. Biofuels 2 (25), 19.
Tatusov, R.L., 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. Nucleic Acids Res. 28 (1), 33–36.
Tatusov, Roman L., Fedorova, N.D., Jackson, J.D., Jacobs, A.R., Kiryutin, B., Koonin, E.V., Krylov, D.M., Mazumder, R., Mekhedov, S.L., Nikolskaya, A.N., Rao, B.S., Smirnov, S., Sverdlov, A.V., Vasudevan, S., Wolf, Y.I., Yin, J.J., Natale, D.A., 2003. The COG database : an updated version includes eukaryotes. BMC Bioinf. 4, 41.
Terra, W.R., Ferreira, C., 1994. Insect digestive enzymes: properties, compartmentalization and function. Comp. Biochem. Physiol. 109B (1), 1–62.
Terrapon, N., Li, C., Robertson, H.M., Ji, L., Meng, X., Booth, W., Liebig, J., 2014. Molecular traces of alternative social organization in a termite genome. Nature Commun. 5 (3636), 1–12.
Wu, W., Huang, Z., Li, Z., Zhang, S., Liu, X., Gu, D., 2015. De novo transcriptome sequencing of *Cryptotermes domesticus* and comparative analysis of gene expression in response to different wood species. Gene 575, 655–666, 2016.
Xie, F., Burklew, C.E., Yang, Y., Liu, M., Xiao, P., Zhang, B., Qiu, D., 2012. De novo sequencing and a comprehensive analysis of purple sweet potato (*Impomoea batatas* L.) transcriptome. Planta 236 (1), 101–113.
Xu, Y., Li, X., Lin, J., Wang, Z., Yang, Q., Chang, Y., 2015. Transcriptome sequencing and analysis of major genes involved in calcium signaling pathways in pear plants (*Pyrus calleryana* Decne.). BMC Genomics.
Ye, J., Fang, L., Zheng, H., Zhang, Y., Chen, J., Zhang, Z., Wang, J., Li, S., Li, R., Bolund, L., Wang, J., 2006. WEGO: a web tool for plotting GO annotations. Nucleic Acids Res. 34, W293–W297. WEB. SERV. ISS.
Yeap, B., Othman, A.S., Lee, V.S., Lee, C., 2007. Genetic Relationship between *Coptotermes gestroi* and *Coptotermes vastator* (Isoptera : Rhinotermitidae). Kirton 2005), pp. 467–474.
Yuki, M., Moriya, S., Inoue, T., Kudo, T., 2008. Transcriptome analysis of the digestive organs of *Hodotermopsis sjostedti*, a lower termite that hosts mutualistic microorganisms in its Hindgut. Zool. Sci. 25 (4), 401–406.