琉球大学学術リポジトリ

食材性シロアリNasutitermes takasagoensisにおける組織複合域の生理学的機能に関するトランスクリプトーム解析

Doctoral Thesis of Philosophy

Transcriptomic analysis on the physiological roles of the mixed segment in the wood-feeding termite,

Nasutitermes takasagoensis

September 2016

by

Koku Hannadige Rohitha Pradeep Kumara

A dissertation submitted to the Graduate School of

Engineering and Science, University of the Ryukyus,
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Tropical Biology
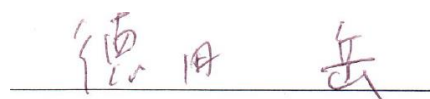
Marine and Environmental Sciences

Graduate School of Engineering and Science

University of the Ryukyus

Supervisor: Prof. Gaku Tokuda

We, the undersigned, hereby, declare that we have read this thesis and we have attended the thesis defense and evaluation meeting. Therefore, we certify that, to the best of our knowledge this thesis is satisfactory to the scope and quality as a thesis for the degree of Doctor of Philosophy in Tropical Biology under Marine and Environmental Sciences, Graduate School of Engineering and Science, University of the Ryukyus.

THESIS REVIEW & EVALUATION COMMITTEE MEMBERS

(**Chairman**) **Prof. Gaku Tokuda**

(**Committee**) **Prof. Joji Otaki**

(**Committee**) **Prof. Ryuuichi Itoh**

(**Committee**) **Prof. Naoya Shinzato**

ABSTRACT

Nasutitermes takasagoensis is a wood-feeding higher termite, which can digest cellulose using cellulases secreted from the midgut tissue. The adjacent gut region is called the mixed segment, which is a unique part of the gut present only in the most apical lineage of termites and consists of a complex of overlapping mesenteric and proctodeal epithelial tissues. It harbors numerous bacteria in the ectoperitrophic space. So far, physiological roles of the mixed segment have yet to be elucidated.

To understand the variety of functional enzymes and metabolic pathways acting in the mixed segment, I performed a transcriptome analysis with a pyrosequencer, 454 GS junior using mRNA extracted from the mixed segment. For better understanding of the mixed segment function, I compared transcriptome data of the midgut and the first proctodeal segment (P1) with those of the mixed segment. The first experiment sequenced three cDNA libraries from the midgut, the mixed segment and P1. I found that genes encoding vacuolar H^+-ATPase (V-ATPase) subunits were predominantly expressed in the mixed segment, which was confirmed by reverse-transcriptase quantitative polymerase chain reaction (RT-qPCR) analysis. The V-ATPase enzyme activity in these three tissues was in good agreement with the expression patterns, suggesting that V-ATPase is a prevalent enzyme in the mixed segment of the termites.

The second experiment sequenced an additional midgut and two mixed segment cDNA libraries and obtained transcripts were pooled with the three cDNA libraries obtained from the first experiment. Transcripts were assembled from the pool reads of six sequence data sets and the resulting transcripts were annotated using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database to assess metabolic functions occurring in the mixed segment when comparing midgut and P1. The predominant metabolic function detected in the mixed segment was energy production via oxidative phosphorylation. As KEGG orthologs involved in lysosomal and phagosomal systems were also characteristic of this segment, cellular components are probably recycled rather than newly produced, allowing higher energy production. The predominant ATP consumer in the mixed segment was V-ATPase, known to be involved in K^+ transport in many insects. The data showed that, V-ATPases appear to electrogenically drive NHE6 or NHE7 (NHE6/7) K^+/H^+ antiporters to stimulate K^+ secretion, increasing gut lumen pH. The mixed segment also expressed higher levels than did the

adjacent gut regions of carbonic anhydrase, another contributor to gut alkalization through HCO_3^- production. The dominant expressions of genes encoding NHE6/7 and carbonic anhydrase in the mixed segment were further confirmed by absolute quantitative RT-qPCR.

Finally, transcripts obtained from the first study were further annotated using different data bases (*i.e.,* Gene Ontology (GO), Pfam, CAZYmes, etc.) to understand the transcriptome profiles acting in the mixed segment. Annotations based on GO revealed the significant differences of functional variations in three tissues. Accordingly, Pfam annotation also revealed the differences in gene compositions in three tissues. Diverse genes encoding carbohydrate active enzymes were identified. These mRNAs encoding β-glucosidases and endo-β-1,4-glucanases have been mainly identified from the midgut transcriptome and almost absent in the mixed segment transcriptome.

In conclusion, unlike midgut, the mixed segment is not involved in cellulose digestion. The dominantly expressed V-ATPase in the mixed segment was presumed to be electrogenetically coupled with the NHE6/7 K^+/H^+ antiporter and HCO_3^- transport to increase the luminal pH. These mechanisms might promote the solubilization of lignocellulose during its passage to the hindgut and will lay the foundation for future gene expression and functional studies in termites to understand their lignocellulose digestion mechanism. These novel findings will allow to enhance our knowledge of how the mixed segment contribute to the digestive physiology of higher termites on plant biomass degradation in tropical ecosystems.

Author's publication list

Kumara, R.P., Saitoh, S., Aoyama, H., Shinzato, N. and Tokuda, G. (2015) Predominant expression and activity of vacuolar H^+-ATPases in the mixed segment of the wood-feeding termite, *Nasutitermes takasagoensis*. Journal of Insect Physiology, 78, 1-8, doi: 10.1016/j.jinsphys.2015.04.012.

Kumara, R.P., Saitoh, S., Aoyama, H., Shinzato, N. and Tokuda, G. (2016) Metabolic pathways in the mixed segment of the wood-feeding termite *Nasutitermes takasagoensis* (Blattodea (Isoptera): Termitidae). Applied Entomology and Zoology, 51, 429-440, doi: 10.1007/s13355-016-0417-4.

Acknowledgment

This dissertation would not have been possible without the guidance and the help of several individuals who in one way or another contributed and extended their valuable assistance in the preparation and completion of this study.

Contents

List of Figures

List of Tables

List of abbreviations

MD: Midgut

MS: Mixed segment

P1: First proctodeal segment

V-ATPases: Vacuolar type H^+-ATPases

ATP: Adenosine tri-phosphate

ADP: Adenosine di-phosphate

P_i: phosphate

cDNA: Complementary deoxyribonucleic acid

CAZymes: Carbohydrate-Active enzymes

GO: Gene ontology

mRNA: Messenger ribonucleic acid

rRNA: Ribosomal ribonucleic acid

mtGenome: Mitochondrial genome

DBs: Data bases

dbCAN: Database for automated Carbohydrate active enzyme annotation

KEGG: The Kyoto Encyclopedia of Genes and Genomes database

KAAS: KEGG Automatic Annotation Server

Pfam: Database of protein families

NCBI: National center for biotechnology information

Nr: Non-redundant protein sequence database

Nt: Nucleotide sequence database

gDNA: Genomic DNA

NADH-dh: Nicotinamide adenine dinucleotide (reduced form) – dehydrogenase gene

ORF: Open reading frame

PCA: Principal component analysis

TCA: Tricarboxylic acid

RPKM: Reads per kilobase of exon per million mapped sequence reads

ABC transporters: ATP-binding cassette transporters

P-ATPases: P-type ATPases

SLC: Solute carrier family

NHE: Sodium/hydrogen exchanger

HCO_3^-: Bicarbonate

Cl^-: Chloride

GH: Glycosyl hydrolase

GT: Glycosyl transferase

CBM: Carbohydrate-Binding Modules

CE: Carbohydrate esterases

JHBP: Haemolymph juvenile hormone binding protein

Tret1: Trehalose transporter

CHAPTER I

General Introduction

1.1 Termites

1.1.1 Ecological importance of termites

Termites are predominant terrestrial insects in the tropics and are major lignocellulose-decomposers. They have received attention from biologists largely because they inhabit nearly two-thirds of the Earth's land surface (Wood and Sands, 1978) and their enormous ecological impact. Of greatest importance is the role they play in recycling wood and other plant materials and it is thought to be responsible for a yearly atmospheric carbon flux of 1.9 billion tons (Cornwell et al. 2009). In addition, they are important food source for many other animals. Termite mounds are among extensive systems of tunnels and provide good ventilation system, shelter and breeding sites for many other creatures including many reptiles, birds and mammals. Apart from their ecological significance, termites, an economically significant pest group that can cause serious structural damage to buildings, crops or plantation forests and non-cellulosic materials such as electric cables comprise of more than 2500 recognized species, with an approximate 1000 species yet to be discovered (Eggleton 2011; Lo and Eggleton 2011). They are generally grouped according to their feeding behaviors, namely dead wood, grass-eating, leaf litter, micro-epiphytes and soil-feeding (Donovan et al. 2001).

1.1.2 Phylogenetic classification of termites

Termites comprise a group of eusocial insects usually classified as a part of the order Blattodea (Isoptera) (*i.e.,* cockroaches). They harbor a dense and diverse population of both eukaryotic and prokaryotic microorganisms in their gut since those microorganisms show strong mutualistic relationship with the host (Brune and Ohkuma 2011; Bignell 2000). Based on the presence or absence of hindgut protists, they can be divided into phylogenetically early branching lower termites and later branching higher termites. Lower termites consist of six families; Mastotermitidae, Kalotermitidae, Termopsidae, Hodotermitidae, Serritermitidae, Rhinotermitidae and higher termites belong to single family Termitidae, further divided into at least four subfamilies (*i.e.,* Macrotermitinae, Apicotermitinae, Nasutitermitinae, and Termitinae) that contains 70% of all termite species (Lo and Eggleton 2011).

1.1.3 General digestive mechanism of lower and higher termites

Termites have two independent cellulose-digestion systems: one in the midgut where cellulose digestion is accomplished by endogenous cellulases, and the other in the hindgut, which makes use of other cellulases from symbiotic flagellates and/or bacteria (Mikaelyan et al. 2014; Tokuda and Watanabe 2007; Warnecke et al. 2007; Nakashima et al. 2002). The digestive process in the gut of lower termites has been extensively studied to date (Brune 2014; Ni and Tokuda 2013). For cellulose digestion, these termites are mainly dependent on intestinal protists (Inoue et al. 2000), while higher termites do not possess symbiotic protists in their gut and they are purely prokaryotic (Brune 2014). Lower termites (such as *Neotermes koshunensis*, *Hodotermposis sjoestedti* and *Reticulitermes speratus*), show relatively lower cellulase activities in the midgut when compared to the salivary gland. On the other hand, an exception is the *Coptotermes formosanus,* which has relatively higher cellulase activity both in the salivary gland and the midgut (Lo et al. 2011). The midgut plays a central role in digestive physiology in higher termites. In the higher termite *Nasutitermes takasagoensis*, both cellulase genes encoding endogenous endo-β-1,4-glucanase and β-glucosidases are expressed in the midgut, which is the main site of cellulose digestion (Tokuda et al. 2012). In addition to the midgut, *Nasutitermes* termites show wood (or cellulosic) particle-associated cellulase activities in the bacterial community that resides in the enlarged part of the hindgut, which is called the paunch (Mikaelyan et al. 2014). Although higher termites have diverse feeding habits (Lee and Wood 1971) such as wood, soil, humus, lichen, and fungus etc., the digestive process except for the fungus-growers has been studied in limited groups of termites, such as those belonging to wood-feeding Nasutitermitinae and soil-feeding Termitinae (Ni and Tokuda 2013).

1.2 The termite gut; basic architecture, design and mechanism in (lower) and higher termites

Termite guts reveal highly diversified and very complicated structures. Generally, it is divided into three major parts: the foregut (FG), midgut (MD) and the hindgut (HG). The early branching termite families shows this general gut architectures, while later branching termites shows highly diversified gut architectures including FG, MD, mixed

segment and differentiated hindgut (first proctodeal segment (P1), paunch (P3), colon (P4), and rectum (P5)) (Holmgren 1909; Noirot 2001) (Figure 1). The highly compartmentalized digestive tract of the higher termites shows increased length, which may allow a larger transit time for food and sequential digestion (Bignell 1995 and Brauman et al. 2000). In addition, a characteristic elevation of the pH occurs in the mixed segment and anterior hindgut section (Brune 1998, 2014; Köhler et al. 2012). In general perspective, the foregut is an esophageal tract with the crop and gizzard that plays a role in mechanical grinding of ingested food materials (Tokuda et al. 2012). The next gut region is the midgut. It is believed to be chiefly a secretion site of digestive enzymes and nutrient absorptions. The hindgut is the largest compartment (Watanabe and Tokuda 2010). The P3 compartment (so-called the paunch) harbors numerous microorganisms (Köhler et al. 2012) and is generally acknowledged to be a site of microbial fermentation (Brune and Ohkuma 2011). The enteric valve (P2) seperates P1 and P3 (Watanabe and Tokuda 2010) while P4 is a narrow duct and is generally thought to be an absorption site of microbial end products (Bignell 2011). P5 is the terminal part of the gut that plays water reabsorption before voided as feaces (Watanabe and Tokuda 2010).
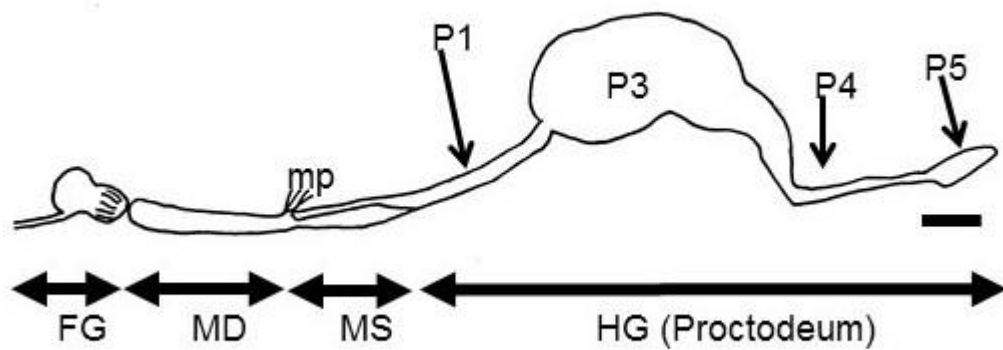
Figure 1: A schematic drawing of the gut of *N. takasagoensis*. FG, foregut; MD, midgut; MS, mixed segment; HG, hindgut that consists of the first proctodeal segment (P1), paunch (P3), colon (P4), and rectum (P5) (Holmgren 1909). The MS is a complex of epithelial tissues that extends from the MD to the P1; mp, Malpighian tubules. The scale bar indicates 1 mm.

1.3 The alkaline gut segments

1.3.1 General examples of alkaline gut segments in insects and mechanisms of gut alkalization

It was reported that various insects show different pH profiles in different part of the digestive tract (Moffet and Onken 2010; Onken et al. 2008; Boudko et al. 2001; Harrison 2001; Wieczorek et al. 1991, 2000; Clark et al. 1998; Bignell and Eggleton, 1995; Wieczorek 1992; Chamberlin 1990). It has been reported that an elevation of the pH (to pH 10 to 13) was observed in different insect taxa including Coleoptera, Diptera, Lepidoptera and Blattodea (Clark 1999). The mechanism of gut lumen alkalization was best studied using midguts of larval lepidopterans, generally called the tobacco hornworm, *Manduca sexta* (Wieczorek et al. 1991, 2000) and dipterans such as yellow fever mosquitoes (*Aedes aegypti*) (Boudko et al. 2001; Onken et al. 2008).

It is now well known that the vacuolar-type H^+-ATPases (V-ATPases) function to regulate the cellular pH homeostasis and ATP-driven active secondary ion and solute transport across divers biological membranes including insect epithelia (Wieczorek et al. 1992, Forgac 1989). In addition, this enzyme function to acdifys the intracellular compartments (such as lysosomes, etc.) (Forgac 1997) and serves as the breakdown of macromolecules. It has been suggested that V-ATPases are involved in active K^+ ion transport in many insect biological systems, which cause alkalinity (Wieczorek et al. 1991, Onken et al. 2008). On the other hand, it has already revealed that V-ATPase is abundant and active in cells rich in mitochondria (Brown et al. 1987, 1996; Harvey et al. 1983). For the *M. sexta* larvae, V-ATPases located in the apical membrane of midgut goblet cells associated with numerous mitochondria (Cioffi and Harvey 1981), generate a large membrane potential difference that drives K^+ ion into the midgut lumen via an electrogenic $K^+/2H^+$ antiporter system (Wieczorek et al. 1991; Azuma et al. 1995). In addition, these cells are highly abundant in carbonic anhydrase, the other agent of gut lumen alkalization (producing bicarbonate ions) (Ridgway and Moffett 1986). However, it is noteworthy to mention that there is a suspicion on the proposed $K^+/2H^+$ antiporter mechanism of alkalization in lepidopteran larval midgut (Moffet and Onken 2010). On the other hand, there is another way to transport K^+ ion through the apical membranes by using K^+-coupled amino acid transporters (Moffet and Onken 2010).

6

In contrast, yellow fever mosquitoes (*Aedes aegypti*) larvae show strong luminal alkalinity in the anterior midgut and appear to employ a different system than lepidopterans. In this larvae, the V-ATPases localized to the basolateral membrane of the anterior midgut (Zhuang et al. 1999) and associated with numerous mitochondria, contribute to luminal alkalization by energizing anion exchange across apical membrane (Boudko et al. 2001). A previous study reported that carbonic anhydrase activity seems to be less abundant in the anterior midgut of *A. aegypti* larvae (Corena et al. 2002). Nevertheless, this system still shows the strong luminal alkalization (Onken et al. 2008) and it was proposed that exchangers (such as Cl^-/HCO_3^-) in the luminal membrane are important for strong alkalization (Boudko et al. 2001b). On the other hand, electrogenic $K^+/2H^+$ antiporter system might be less important in this system (Onken et al. 2008).

1.3.2 Morphological features of the termite mixed segment in comparison to the alkaline gut segments of other insects

Most remarkable feature of the higher termite gut is the presence of so-called "mixed segment" between the midgut and hindgut (Figure 1). This unique complex structure mainly consists of the mesenteric and proctodeal epithelial tissues (Bignell et al. 1983; Tokuda et al. 2001), and is present only in the most derived group of higher termites (except for the basal lineages). Morphology of the epithelial tissue in the mixed segment of wood-feeding and soil-feeding higher termites was studied using *Nasutitermes takasagoensis* and *Cubitermes severus* (Tokuda et al. 2001; Bignell et al. 1983), respectively. The proctodeal epithelium consists of flat cuboidal cells covered with thick cuticular layers, while the mesenteric epithelium is composed of columnar cells (Tokuda et al. 2001). It was also found that basal invaginations of the mesenteric columnar cells in the mixed segment are extensive with well-developed mitochondria (Tokuda et al. 2001). It proposes the ATP-dependent active transport of ionic fluid across the mesenteric cells to the gut lumen (Bignell et al. 1983) (Figure 2). These morphological features are typical for ion transporting epithelia described in the previous section.

On the other hand, bacteria are observed in the ectoperitrophic space of the mixed segment and associated with degenerated microvilli (Bignell et al. 1983; Tokuda et al. 2000 and 2001), while they are distinct in morphology between wood-feeding

Nasutitermes and soil-feeding *Cubitermes* and *Procubitermes* and less diversified (Bignell et al. 1983; Czolij et al. 1985; Tokuda et al. 2000). In addition, bacteria who reside in the mixed segment and P1 well adapted to the alkaline environment (Brune 2014; Köhler et al. 2012; Thongaram 2003).

1.3.3 Physiological characteristics of the mixed segment in comparison to the adjacent gut segments (midgut and P1)

In contrast to the adjacent midgut that is slightly acidic (Brune et al. 1995; Köhler et al. 2012) and shows most of the cellulase activities, only a trace amount of cellulase activities in the mixed segment of the wood-feeding termite *N. takasagoensis* suggests that this organ is not involved in cellulose digestion (Tokuda et al. 1997, 2004, 2005). Although some physiological characteristics of the mixed segment have recently been clarified, yet the function has not been fully elucidated. One of the most significant features in mixed segment is the sharply elevated pH (Brune 2014; Köhler et al. 2012), since it is assumed to secrete alkaline fluid of K_2CO_3 to elevate the luminal pH that reaches over 10 in P1 (Bignell et al. 1983; Kappler and Brune 1999; Brune 2014). However, its role in digestive physiology is still unclear. In addition to the pH profiles, it was reported that mixed segment and P1 showed higher O_2 partial pressures, redox potentials and almost zero H_2 partial pressures in closely related *Nasutitermes corniger* (Köhler et al. 2012).
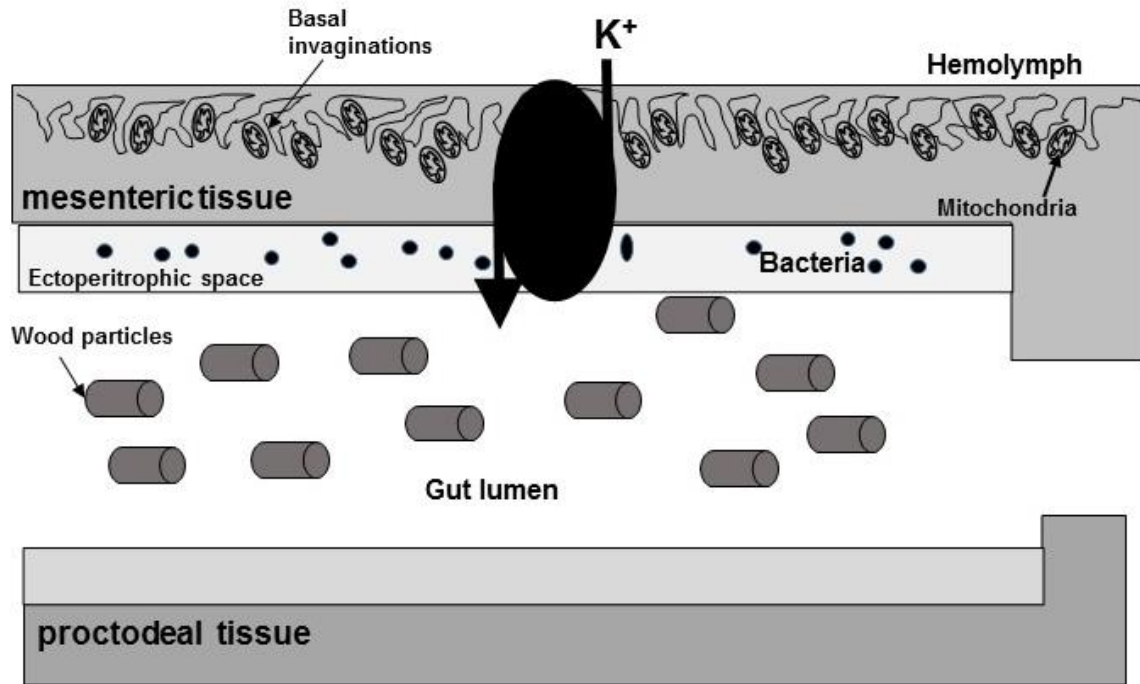
Figure 2: A putative schematic drawing of the mixed segment of *N. takasagoensis* with respective to the previous observations (Tokuda et al. 2001; Bignell et al. 1983). Black circle represents a channel molecule that involves K⁺ transport from hemolyph to the gut lumen. Not to scale.

1.4 Importance of the gut alkalization during lignocellulose digestion and relevant physiology

As explained above, all lower termite families share basic gut structure and lack any further significant compartmentalization. On the other hand, higher termites (such as advanced subfamilies of higher termites like the Termitinae and Nasutitermitinae) show highly compartmentalized hindgut (Noirot 2001; Holmgren 1909). It has been hypothesized that the expansion and elongation of hindgut during the evolution of higher termites may have required a sufficient irrigation system, which triggered to form the mixed segment that stimulates the flow rate of the gut fluid (Bignell 1994, 1983). As explained above (Chapter 1.3.2) gut alkalization is not limited to the termites, while this phenomenon was appeared in many insect biological systems and it has been assumed that it was an evolutionary adaptation to a diet rich in tannins or other polyphenolic constituents, since it enhances the solubility of dietary proteins and prevents precipitation of digestive enzymes (Brauman 2000; Martin et al. 1987; Felton and Duffey 1991). It was previously reported that this higher alkalinity increases the solubility of soil organic matter and autoxidation of humic substances in soil-feeding higher termites (Kappler and Brune 1999). On the other hand, its physiological basis in wood-feeding termites is still obscure, while an alkaline pH might promote autoxidation of polyphenolic compounds that loosen lignocellulosic structure and provide easy access of bacterial cellulolytic enzymes in the microbial rich hindgut paunch (Tokuda and Watanabe 2007).

1.5 Objectives of the study

If the function of the mixed segment is aimed to be analyzed, it should be compared with the midgut and P1 because the basic structure of the mixed segment is composed of the mesenteric and proctodeal epithelial tissues (Bignell et al. 1983; Tokuda et al. 2001). To understand the digestive physiology of the termites, tissue specific transcriptome analysis is powerful and attractive, because transcriptome data would provide essential information on the molecules (genes) and metabolic pathways working in the host tissues. To investigate the putative physiological functions of the mixed segment, in this study, I used transcriptome analysis for the first time, by using 454 GS junior (Roche) to identify genes and metabolic pathways specifically expressed in the mixed segment of the wood-feeding termite *N. takasagoensis.* For the better understanding of the mixed segment function, I also performed transcriptome analysis of the midgut and P1 (adjacent gut segments) and the resulted data were compared with the transcriptomic data obtained from the mixed segment.

The 454 pyrosequencing has been recognized as a sophisticated method for increasing sequencing depth and coverage while reducing time, labor, and cost (Ekblom and Galindo 2011). In the present study, I used the second generation of 454-based pyrosequencing (454 GS Junior, Roche) to define the transcriptomes of the midgut, mixed segment and P1 of the termite *N. takasagoensis*. I sequenced six independent cDNA libraries (three mixed segment, two midguts and one P1) and examined their functions based on lignocellulose digestion and the relevant physiology. This analysis is expected to dramatically increase in the number of known genes for this remarkable model of insect physiology and also to form an important source of cDNAs for eventual genome annotations.

In order to understand the contribution of the mixed segment of wood feeding higher termite *N. takasagoensis* to lignocellulose digestion and the relevant physiology, my doctoral study was structured with the following objectives:

1) To investigate the putative physiological functions of the mixed segment, using transcriptome analysis and comparison with those of midgut and P1. This analysis is enabled us to identify dominantly expressed genes in the mixed segment.

2) To investigate the metabolic network and functions characteristic in the mixed segment, using <u>K</u>yoto <u>E</u>ncyclopedia of <u>G</u>enes and <u>G</u>enomes (KEGG) database and comparison with those of midgut and P1. This analysis is expected to stimulate our understanding on the effective digestive system of higher termite *N. takasagoensis*.

3) To investigate the general pattern of transcriptome profiles observed in the mixed segment and comparison with the transcriptome profiles of midgut and P1 in the wood feeding termite *Nasutitermes takasagoensis* using different databases (such as Gene ontology, Pfam and CAZYmes). This analysis is expected to dramatically increase in the number of known genes and their precise functions in the mixed segment.

CHAPTER II

Predominant expression and activity of vacuolar H$^+$-ATPases in the mixed segment of the wood-feeding termite *Nasutitermes takasagoensis*

2.1 Introduction

Termites comprise a group of social insects usually classified within the order Blattodea (*i.e.,* cockroaches). They are ecologically important in terms of carbon recycling due to their ability to decompose lignocellulose. Termites of the most apical lineage are referred to as "higher termites" (family Termitidae) (Lo and Eggleton 2011).

A remarkable feature of the gut of higher termites is the presence of the so-called "mixed segment" between the midgut and the first proctodeal segment (P1) (Figure 1). The mixed segment consists of mesenteric and proctodeal epithelial tissues (Bignell et al. 1983; Tokuda et al. 2001). Although the exact function of the mixed segment is still ambiguous, some physiological characteristics have been reported. One of the features in the mixed segment is a sharp elevation in the pH (~ pH 10) compared to the adjacent midgut (pH 7) (Köhler et al. 2012; Brune 2014). It has been proposed that the mesenteric epithelium of the mixed segment secretes a potassium-rich fluid into the gut lumen (Bignell et al. 1983). The mesenteric columnar cells in the mixed segment have extensive basal membrane invaginations and well-developed mitochondria (Tokuda et al. 2001). The microbiology of the mixed segment has been investigated in the soil-feeding genera *Procubitermes* and *Cubitermes* (Bignell et al. 1983) and the wood-feeding *Nasutitermes* (Czolij et al. 1985; Tokuda et al. 2000). Despite differences in the morphology of the mixed segment, in both feeding groups, bacteria are located in the ectoperitrophic space of the mixed segment. In *Nasutitermes*, bacteria are rare in the midgut (Tokuda et al. 2001), though they are present in the two soil-feeders (Bignell et al. 1980). However, only a trace amount of cellulase activity has been detected in the mixed segment of the wood-feeding *Nasutitermes takasagoensis*, suggesting that this organ and the bacterial symbionts are not involved in cellulose digestion (Tokuda et al. 1997, 2004) in contrast to the adjacent midgut that is slightly acidic (Brune et al. 1995; Köhler et al. 2012) and secretes both endo-β-1,4-glucanases and β-glucosidases (Tokuda et al. 2012). These findings suggest that the mixed segment has a different function than the midgut, at least in the digestive process. Previous studies have shown that the mixed segment is involved in excretory processes (Bignell et al. 1983; Bignell 1994). It has been postulated that the structural elongation and specialization of the hindgut during the evolution of higher termites may have required an improved irrigation system, which triggered the elongation

of the mesenteric tissue to form the mixed segment that stimulates the flow rate of the gut fluid (Bignell 1994). To investigate the putative physiological functions of the mixed segment, in this study, I used transcriptome analysis to identify dominantly expressed genes in the mixed segment in the wood-feeding higher termite *N. takasagoensis*.

2.2 Materials and methods

2.2.1 Termites samples

Colonies of *N. takasagoensis* were collected on Iriomote Island, Okinawa, Japan, and were maintained in wood logs in the laboratory at 25 ± 3 °C until use.

2.2.2 Dissection and stabilization of termite gut tissues

Whole guts were dissected from mature worker termites with fine-tipped forceps, and the midgut, mixed segment, and the P1 tissues (Figure 1) were collected in separate microtubes. Filter-sterilized buffer (41.2 mM NaCl, 10.2 mM NaHCO$_3$, 5.7 mM sodium citrate, 14.5 mM KH$_2$PO$_4$, 50 mM EDTA) was used for the dissections. The gut samples were stabilized in an RNAlater RNA stabilization reagent (Qiagen, Hilden, Germany) according to the manufacturer's instructions.

2.2.3 RNA purification

Total RNA was extracted from the midgut and mixed segment samples of 750 worker termites, and from the P1 samples of 1750 workers; as the P1 tissue has thinner epithelium than the midgut and mixed segment mesenteric tissues (Tokuda et al. 2001), a large number of specimens were required for RNA extraction. The midgut and mixed segment samples were homogenized by using a plastic pestle, and total RNA was extracted using the RNeasy Mini Kit (Qiagen, Hilden, Germany) according to manufacturer's guidelines. P1 tissues were homogenized by using a rotor-stator homogenizer and total RNA was extracted using ISOGEN solution (Nippon Gene, Tokyo, Japan). The RNA quality was verified using an Agilent 2100 bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and quantification was done with the Qubit RNA Assay Kit for use with the Qubit 2.0 fluorometer (Invitrogen, Paisley, UK). Total RNA was pooled by tissue type and used for mRNA purification using the NucleoTrap mRNA

Mini Kit (Takara Bio, Otsu, Japan). The quality and quantity of the purified mRNAs were checked as described above. The purified mRNA was treated with 1 µl of RNase inhibitor (40 U µl^{-1}).

2.2.4 Sequencing library preparation

Double-stranded cDNA was synthesized from the purified mRNAs according to the cDNA rapid library preparation method from Roche (454 GS Junior Titanium Series, cDNA Rapid Library Preparation Method Manual, May 2010, Roche Applied Sciences, Indianapolis, IN). Sequencing of the cDNA libraries was performed in single-ended mode on the 454 GS Junior sequencer (Roche Diagnostics, Penzberg, Germany).

2.2.5 Sequence analyses of the 454 data

Raw sequencing data were cleaned by removing ribosomal RNA (rRNA) and mitogenome (mt) sequences using the Roche Reference Mapper software v. 2.7 (Figure 3). First, mt and rRNA sequences were assembled by mapping the raw reads onto sequences from closely related species (*i.e.*, 5.8S, 18S, and 28S rRNAs and the rRNA precursor genes from *Nasutitermes* sp. and *Zootermopsis angusticollis* [GenBank AY491151.1, AY491211.1, AY859614.1, and AY859615.1] and the mitochondrial genome from *Nasutitermes triodiae* [JX144940.1]). The assembled sequences were then used as references to filter out raw reads considered to be derived from the rRNA and mt genes.

Transcripts were assembled from the clean reads using the Newbler software v. 2.7 (Roche Diagnostics) with the default settings for cDNA mode. To eliminate bacterial sequences, each of the assembled sequences (isotigs) was subjected to a homology search (nucleotide BLAST against the NCBI Nucleotide collection database with a threshold *e*-value of 10^{-9}), then, a judgment of bacterial sequence was made based on whether the definition of the best hit included at least one of 'key words' or not, which are bacterial genus names or some other related words (such as "Bacterium", "bacterium", "Bacterial", "bacterial" "Uncultured", "uncultured", "cyanobacterium", "endosymbionts", "proteobacterium", and "16S"). Finally, isotigs showing significant similarity (*e*-value < 10^{-9}) to those from bacterial genes were excluded from further analysis.

16

Transcripts were quantified based on the output from Newbler by counting the number of reads aligned to the isotigs. For isogroups (*i.e.*, unigenes), quantification was based on the number of unique reads mapped to their family member isotigs. Numbers of total aligned reads and those of isotigs and isogroups were corrected by excluding reads aligned to bacterial isotigs (see above).

Isotigs were annotated using KEGG (Kanehisa et al. 2010), Pfam (Finn et al. 2014), dbCAN (Yin et al. 2012), and NCBI non-redundant protein sequences (Nr) and nucleotide collection (Nt) databases by using online or local tools provided by the developers of the respective tools (Figure 4). Gene ontology (GO) categories were attributed by using GO terms collected from the Pfam and KEGG databases (pfam2go and kegg2go downloaded from http://www.geneontology.org/external2go). KEGG pathways were constructed using the online version of KEGG mapper (http//www.kegg.jp/kegg/tool/map_pathway/html).
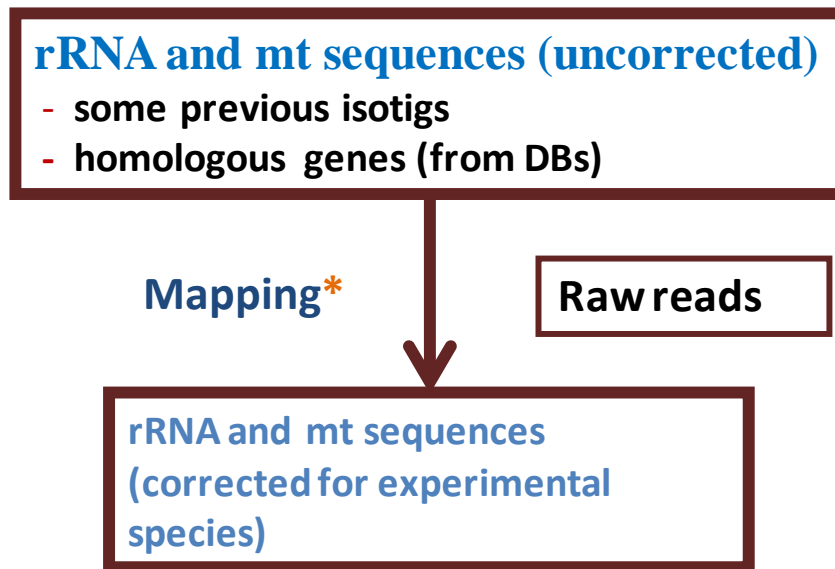
First step:

Constructing putative rRNA precursor gene and mitochondrial genome sequences

* using Roche Reference Mapper software
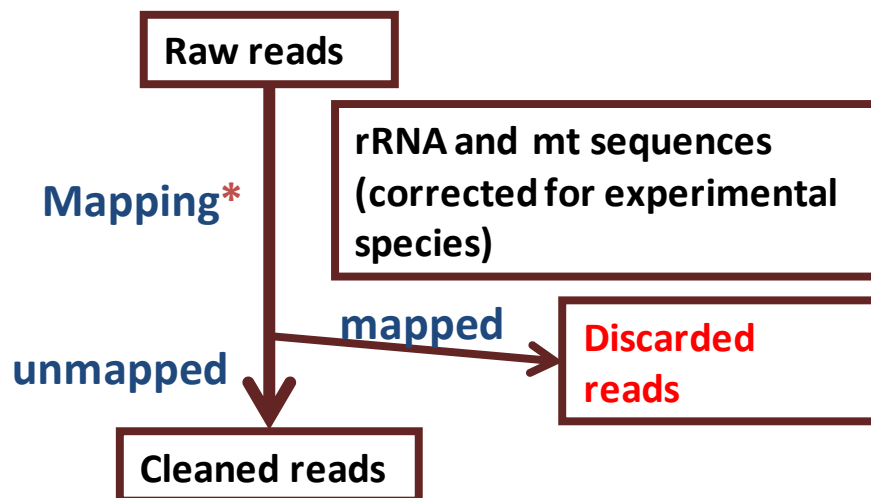
Secondary step:

Cleaning of data via mapping using putative rRNA and mt sequences as the references

Figure 3: Reference rRNA and mitogenome mapping process
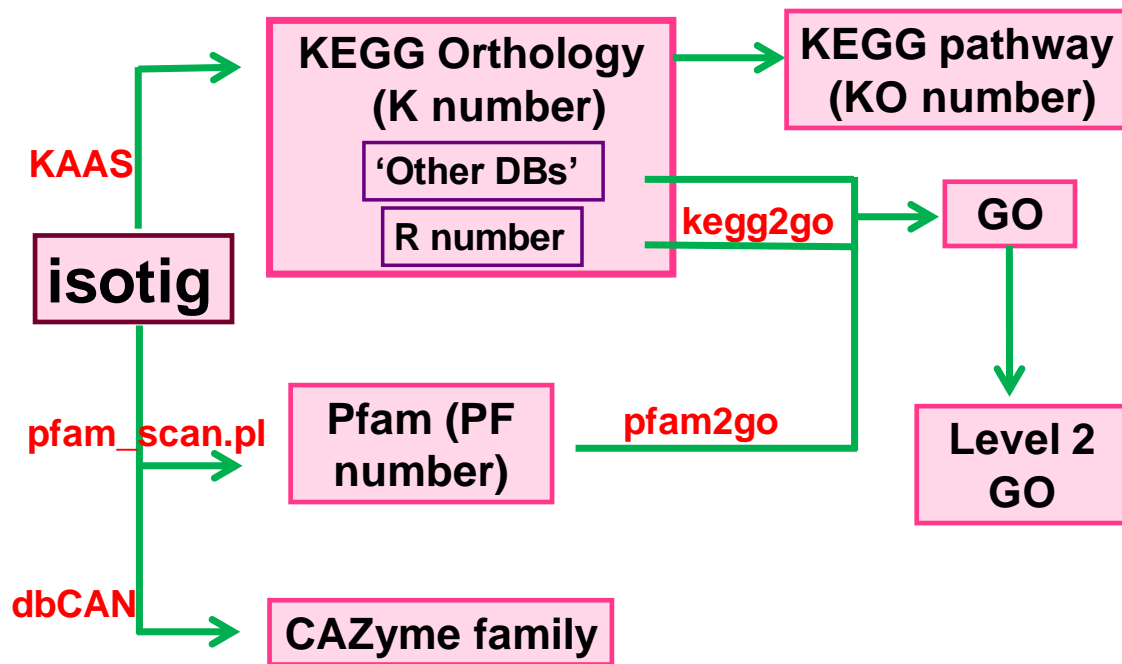
Figure 4: Reference data annotation process

2.2.6 Sanger sequencing

The contig sequences of vacuolar H^+-ATPase (V-ATPase) subunits were verified by Sanger sequencing. For this purpose, total RNA was extracted from the mixed segment samples as described above. First-strand cDNA was synthesized with a PrimeScript RT Reagent Kit with gDNA Eraser (Perfect Real Time) (Takara Bio, Shiga, Japan), according to the manufacturer's protocol. The sequences were aligned to identify highly conserved regions using Genetyx v.11 (Genetyx Corp., Tokyo, Japan) and gene-specific primers (Table 1) were designed using Primer3 (http://primer3.sourceforge.net/). The V-ATPase sequences were amplified by PCR with PrimeSTAR Max DNA Polymerase (Takara Bio) using the following thermal cycles: 35 cycles of 98 °C for 10 s, 55 °C for 15 s, and 72 °C for 5 s. The amplified cDNA fragments were purified using the Wizard SV Gel and PCR Clean-Up System (Promega, Madison, WI, USA). The purified DNA products were used as templates for direct sequencing using a BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA) on the ABI Prism 3130xl Genetic Analyzer (Applied Biosystems) according to the manufacturer's instructions. The sequences were subjected to homology search using nucleotide BLAST to confirm the identity of the genes.

Table 1: Primer sets used in this study for Sanger sequencing of the *N. takasagoensis* V-ATPase subunits

Gene	Primer sequence	Primer length (bp)
V0 subunit c (16kDa proteolipid subunit)	Forward: 5'-CCGCCTAGTTTAGGAGTGTGA-3' Reverse: 5'-GCCCACTGCCAACAAATTA-3'	21 19
V0 subunit d (d1-AC39 subunit)	Forward: 5'-GTAGGTGAGTGACTAGAAGAGAAGTG-3' Reverse: 5'-AGCAACACATTCCGAGATCC-3'	26 20
V1 subunit A	Forward: 5'-GGAGGTGTGGAGCACGTACA-3' Reverse: 5'-TCCGCACGATAGTTCTGACA-3'	20 20
V1 subunit B	Forward: 5'-ATGATGTCGTACGGTGATCG-3' Reverse:5'-TCTAACAGTCTGTGACTAAGTTCC-3'	20 24
V1 subunit C	Forward: 5'-GGCCTACGCGTTTAATATTGG-3' Reverse: 5'-GCAGCCAGTTCCTCTTCATT-3'	21 20
V1 subunit D	Forward: 5'-CTCGTGCATACAGAAGCGAA-3' Reverse: 5'-CCTCACCCAATTATATTTACAGGC-3'	20 24
V1 subunit E	Forward: 5'-TGGTGGTTAGTTGGGAGTGA-3' Reverse: 5'-AGTAAGTCCCACCCCACAAA-3'	20 20
V1 subunit F	Forward: 5'-CGAACTGTTGTAAGATAAATTTTGAGC-3' Reverse: 5'-GACACACACTTGTGTACATGGC-3'	27 22
V1 subunit G	Forward: 5'-GCTAGCCAGACGCAGGGA-3' Reverse: 5'-TTTAACAGAAGCATCACATATCTCTTC-3'	18 27
V1 subunit H	Forward: 5'-CCATGGGTATCGCTCTTGTT-3' Reverse: 5'-TCCTCCACTGCATTACCTCA-3'	20 20

2.2.7 Reverse-transcription quantitative polymerase chain reaction (RT-qPCR)

Next, the expression levels of the V-ATPase subunits in the midgut, mixed segment, and P1 samples were determined using RT-qPCR. Total RNA extraction and cDNA synthesis were carried out as described above. RT-qPCR was carried out using the Fast SYBR Green Master Mix (Applied Biosystems) on the Step One Plus Real-Time PCR System (Applied Biosystems), using the following thermal cycles: 95 °C for 20 s followed by 40 cycles of 95 °C for 3 s and 60 °C for 30 s. Melting curves to verify primer specificity were obtained as follows: 95 °C for 15 s, 60 °C for 60 s, followed by increasing the temperature with 0.3 °C every 15 s. The primers used in this study (Table 2) were designed using the Primer Express software (Applied Biosystems). I obtained sequences of four reference genes from Hojo et al. (2012) (18S ribosomal DNA [*Nt18S* - AB501106], elongation factor 1-alpha [*NtEF1* - AB501108], NADH dehydrogenase [*NtNADH-dh* - AB501109], and beta-actin [*Ntbeta-actin* - AB501107]) and evaluated their suitability as endogenous controls in the respective tissues. The RefFinder web-based program (http://www.leonxie.com/referencegene.php), in which geNorm (Vandesompele et al. 2002), Normfinder (Andersen et al. 2004), Best Keeper (Pfaffl et al. 2004), and the comparative delta-Ct method (Silver et al. 2006) are integrated, was used to compare and rank the tested candidate reference genes. The results revealed that the expression level of *NtNADH-dh* was the most stable; therefore, I used this gene as a reference gene in RT-qPCR analysis.

Relative quantification of gene expression in each of the three tissue types was accomplished using the standard curve method based on the expression levels of *NtNADH-dh*. RT-qPCR was carried out on cDNA from three biological replicates, and each reaction was run in duplicate.

Table 2: Primer sets used in this study for RT-qPCR analysis of the *N. takasagoensis* V-ATPase subunits

Gene	Primer sequence	Primer length (bp)
V0 subunit c (16-kDa proteolipid subunit)	Forward: 5'-TGAAATCGATCATTCCCGTAGTC-3' Reverse: 5'-CAGCCACAACCAGACCGTAGA-3'	23 21
V0 subunit d (d1-AC39 subunit)	Forward: 5'-TTCACCACCGACCCATCTCT-3' Reverse: 5'-GCTCAAAGCTGCCCAAAGG-3'	20 19
V1 subunit A	Forward: 5'-TCCGCACAGGTAAACCATTG-3' Reverse: 5'-CATCAAAAATGCTCCCCAAGA-3'	20 21
V1 subunit B	Forward: 5'-TTCGAGGAAAATGGCTCCAT-3' Reverse: 5'-TGGGATCGTTGGCCAAAT-3'	20 18
V1 subunit C	Forward: 5'-ATCAACGGCGTACAACAACCT-3' Reverse: 5'-CTGCCTGTTTGTTTCTTTTCCA-3'	21 22
V1 subunit D	Forward: 5'-GTTTCTTTGCCTATTTTTGAGAGCTA-3' Reverse: 5'-GCCAATCCGGCCAGTTC-3'	27 17
V1 subunit E	Forward: 5'-TGCCGAGATCCTGAATTCACT-3' Reverse: 5'-CCAGTACATTTGGCTCCAACAG-3'	21 22
V1 subunit F	Forward: 5'-GTCAATACAGTGGCGGATCATC-3' Reverse: 5'-CTTCAAGCGATTTGTGAAAAGAGA-3'	22 24
V1 subunit G	Forward: 5'-AAGGAAGATGTTGCAGCACGTA-3' Reverse: 5'-GCCTTCTGCATCTCATCTATCTTG-3'	22 24
V1 subunit H	Forward: 5'-TGCAAGTGGTTTTAATGCTGACA-3' Reverse: 5'-TCATCTGTGCCATGTAGCAACA-3'	23 22

2.2.8 ATPase activity measurement

Whole guts were dissected from 40 mature worker termites and midgut, mixed segment, and P1 tissue samples were collected in separate microtubes containing 150 μL of CelLytic B (Sigma, Saint Louis, MO, USA). The samples were sonicated and the crude enzyme extracts were recovered after centrifugation at $20,100 \times g$ for 20 min and were stored at –20 °C until use. Total ATPase activity was determined using the EnzChek Phosphate Assay Kit (Molecular Probes, Life Technologies, Eugene, OR, USA) to measure the production of inorganic phosphate (P_i) using ATP (5 mmol l^{-1}) as a substrate. V-ATPase activity was measured as the bafilomycin A1-sensitive portion of the total ATPase activity, as bafilomycin A1 is a specific V-ATPase inhibitor. The 1-ml reaction mixtures were prepared according to the manufacturer's instructions. The kinetics of P_i release were measured by determining the absorbance at 360 nm over time using a spectrophotometer (GeneQuant 1300 v. 1.6, GE healthcare, Little Chalfont, UK). Different concentrations of bafilomycin A1 were tested (10, 20, 30, 40, and 50 μmol l^{-1}) and I found that 30 μmol l^{-1} inhibited the maximum amount of V-ATPase, which was consistent with previous reports (Lin and Randall 1993). Therefore, I used this concentration for further experiments. All ATPase assays were carried out in triplicate. Standard curves were established according to the manufacturer's instructions. One unit (U) of ATPase activity was defined as the amount of enzyme that produces 1 nmol of phosphate (P_i) per min. To normalize V-ATPase activities to mm^3 of gut volume, the diameter and length of each gut segment were measured under a light microscope. Then, the average volume of each gut segment was calculated by assuming it a uniform cylinder.

2.2.9 Statistical analysis

For RT-qPCR, the relative expression values obtained for each gene in the different tissues were analyzed by using one-way analysis of variance (ANOVA). In case the P-value was significant ($P < 0.05$), pairwise comparisons using Student's t-test with Bonferroni correction were used to determine significant differences ($P_{adj} < 0.05$). For V-ATPase activity, Tukey's honestly significant difference (HSD) test was used for evaluating significant differences ($P < 0.05$) between different tissues.

2.3 Results

2.3.1 454 pyrosequencing and transcript assembly

An overview of the sequencing results is given in Table 3. I obtained 154714, 116228, and 114223 transcriptome sequencing reads for the midgut, mixed segment, and P1 tissues of *N. takasagoensis* workers, respectively, generated through three separate runs on the 454 GS Junior system. After removal of the rRNA, mt, bacterial, and low-quality sequences, 76197, 45283, and 24603 reads remained for the midgut, mixed segment, and P1 tissues, respectively. These sequences were assembled into 4813, 4563, and 3629 isotigs (transcripts) and further grouped into 3821, 3607, and 2823 isogroups (unigenes) for the midgut, mixed segment, and P1, respectively.

Table 3: Overview of *N. takasagoensis* transcriptome assembly features

Description	Midgut	Mixed segment	P1
rRNA reads[*]	25383	22370	60310
Mitogenomic reads[*]	15675	24015	2985
Bacterial reads[+]	971	3872	5024
Unassembled reads[†]	35698	20068	20825
Low-quality reads[†]	790	620	466
Putative host mRNA reads	76197	45283	24603
Total number of reads	154714	116228	114213
Isotigs (transcripts)	4813	4563	3629
Isogroups (genes)	3821	3607	2823

[*]These sequences were discarded prior to transcript assembly by using reference mapping.

[+]Bacterial sequences were discarded by using BLASTn after transcript assembly.

[†]These sequences were not used for further analysis.

P1: the first proctodeal segment.

2.3.2 Comparative analysis

Table 4 lists 10 genes identified in the mixed segment that were most abundant (by relative abundance of total mRNA reads) in the mixed segment compared to either the midgut or the P1, together with affiliations to specified isogroups. The most predominent gene in the mixed segment encoded a ricin-type (R-type) beta-trefoil lectin domain-containing protein, which was the sole family member of carbohydrate-active enzymes (CAZymes) among the most abundant genes and accounted for 1.72% of total reads. The second most abundant gene encoded a putative V-ATPase c-subunit. Notably, two genes encoding the subunits A and B were also found within the most abundant genes. The relative abundance of total V-ATPase subunit reads, which accounted for 5.98% of total reads in the mixed segment, was 3.5 times higher than that of the R-type beta-trefoil lectin domain-containing protein. In contrast, the V-ATPase subunits accounted for only 1.06% and 0.35% of the reads in the midgut and P1 samples, respectively (Figure 5). Based on these results, I focused further study on the validation of V-ATPase expression in the mixed segment.

Table 4: Relative abundance, isogroup affiliations, and functional annotations of 10 genes identified in the midgut, mixed segment, and P1

Isogroup	Relative abundance of reads (%)			No. of isotigs	Annotations					Level 2 GO		
	MD	MS	P1		BLAST	pfam	KEGG_orthology	KEGG pathway	CAZyme family	Molecular function	Biological process	Cellular component
00230	0.22	1.72	0.10	2	gb|AFZ78854.1|ricin-type beta-trefoil lectin domain-containing protein [*Coptotermes formosanus*]	PF00652: Ricin-type beta-trefoil lectin domain	N/A	N/A	CBM13	N/A	N/A	N/A
00829	**0.28**	**1.54**	**0.07**	**1**	**gb|ABV60304.1|putative V-ATPase c-subunit [*Lutzomyia longipalpis*]**	**PF00137: ATP synthase subunit c**	**K02155 V-type H$^+$-transporting ATPase 16kDa proteolipid subunit [EC: 3.6.3.14]**	**ko00190 Oxidative phosphorylation ko04142 Lysosome ko04145 Phagosome ko04721 Synaptic vesicle cycle**	**N/A**	**GO:0005215 GO:0003824**	**GO:0044699 GO:0009987 GO:0051234 GO:0051179**	**GO:0044425 GO:0032991 GO:0016020**
00011	0.04	1.46	0.01	15	ref|XP_001844560.1|serine protease 1 [*Costelytra zealandica*]	PF01400: Astacin (Peptidase family M12A)	K08606 meprin A, beta [EC: 3.4.24.18]	ko04974 Protein digestion and absorption	N/A	GO:0003824	GO:0008152	N/A
00636	0.42	1.23	0.02	1	ref|XP_001844554.1|4.1 G protein, putative [*Pediculus humanus corporis*]	N/A	N/A	N/A	N/A	N/A	N/A	N/A
00672	**0.20**	**1.16**	**0.03**	**1**	**ref|XM_001651408.1|*Aedes aegypti* ATP synthase subunit beta vacuolar partial mRNA**	**PF00006: ATP synthase alpha/beta family, nucleotide-binding domain**	**K02147 V-type H$^+$-transporting ATPase subunit B [EC: 3.6.3.14]**	**ko00190 Oxidative phosphorylation ko04145 Phagosome ko04721 Synaptic vesicle cycle**	**N/A**	**GO:0005215 GO:0003824 GO:0005488**	**GO:0044699 GO:0051234 GO:0009987 GO:0051179 GO:0008152**	**GO:0044425 GO:0032991 GO:0016020**
01106	0.11	0.95	0.24	1	ref|XP_003493661.1|mitochondrial ribosomal protein L2 [*Danaus plexippus*]	PF00153: Mitochondrial carrier protein	K15102 solute carrier family 25 (mitochondrial phosphate transporter), member 3	N/A	N/A	GO:0005215	N/A	N/A
00852	**0.12**	**0.92**	**0.07**	**1**	**ref|XP_976188.1|PREDICTED: V-type proton ATPase catalytic subunit A [*Tribolium castaneum*]**	**PF00006: ATP synthase alpha/beta family, nucleotide-binding domain**	**K02145 V-type H$^+$-transporting ATPase subunit A [EC: 3.6.3.14]**	**ko00190 Oxidative phosphorylation ko04145 Phagosome ko04721 Synaptic vesicle cycle**	**N/A**	**GO:0005215 GO:0003824 GO:0005488**	**GO:0044699 GO:0051234 GO:0009987 GO:0051179 GO:0008152**	**GO:0044425 GO:0032991 GO:0016020**
00638	0.09	0.88	0.004	1	ref|XP_001357737.1|aconitate hydratase, mitochondrial precursor [*Homo sapiens*]	PF01433: Peptidase family M1	K11140 aminopeptidase N [EC: 3.4.11.2]	ko00480 Glutathione metabolism	N/A	GO:0003824 GO:0005488	N/A	N/A
00105	0.22	0.71	0	3	gb|EKC29405.1|PREDICTED: tensin-3-like [*Takifugu rubripes*]	N/A	N/A	N/A	N/A	GO:0016491	N/A	N/A
00015	0.22	0.70	0.004	9	ref|XP_003706188.1|PREDICTED: pre-mRNA-splicing factor CWC22 homolog [*Megachile rotundata*]	N/A	K02184 formin 2	ko04320 Dorso-ventral axis formation	N/A	N/A	N/A	N/A
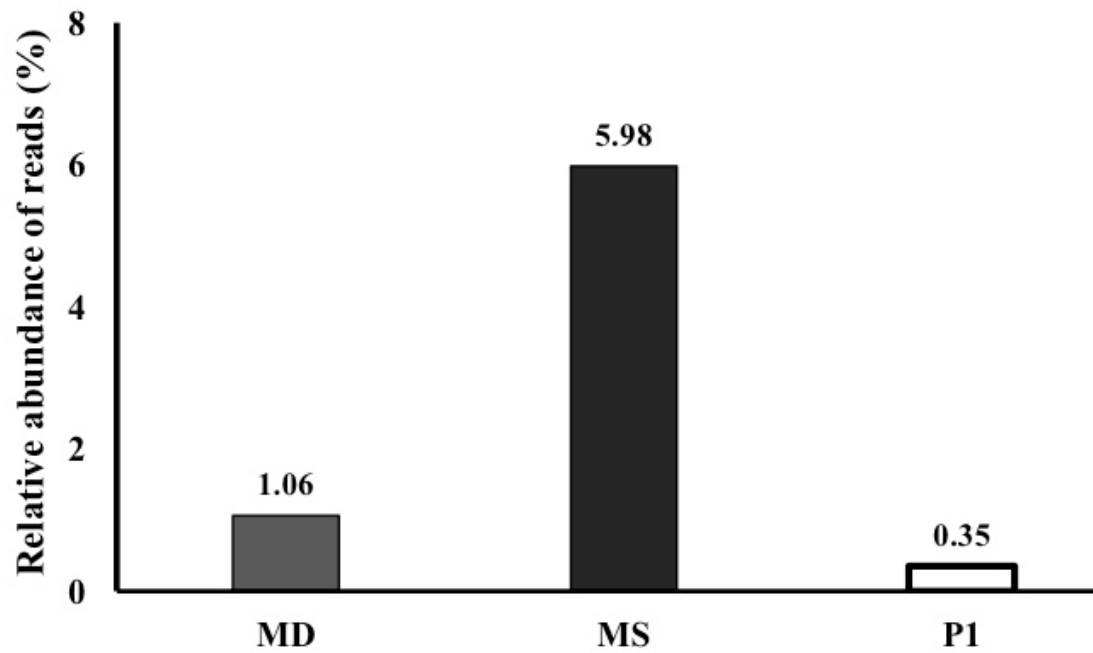
Figure 5: Expression of V-ATPase genes in the three gut segments of *N. takasagoensis*. Bar graphs represent the fraction (%) of reads affiliated with any of the V-ATPase subunits in the total mRNA reads from the indicated tissue. MD, midgut; MS, mixed segment; P1, first proctodeal segment.

2.3.3 Confirmation of the V-ATPase subunit sequences by Sanger sequencing

Eukaryotic V-ATPase consists of membrane-integral V_0 and cytoplasmic V_1 complexes (Baumann and Walz 2012). The V_0 complex of insect V-ATPase is composed of at least five heterologous subunits (a, c, c", d, and e), while the V_1 complex is composed of eight different subunits (A to H) (Figure 6). I identified contigs of two V_0 complex subunits (subunits c [16-kDa proteolipid] and d [d1-AC39]) and of each V_1 complex subunit in my transcriptome dataset. To verify that these sequences were no artifacts and encoded functional proteins, the sequences of the subunit ORFs were confirmed by Sanger sequencing.

V_1 V_1 is responsible for ATP hydrolysis

V_0 V_0 serves the translocation of protons from the cytosol to the extracellular space (ECS) or to the lumen of an organelle
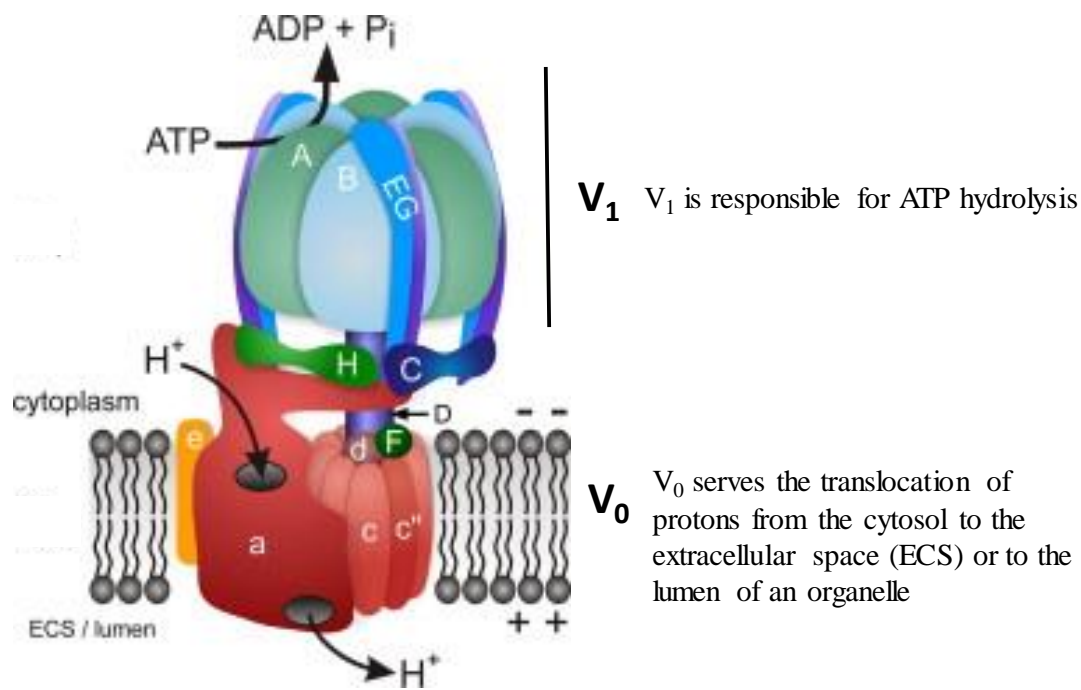
Figure 6: Schematic model of the eukaryotic V-ATPase. It is mainly consisting of membrane-integral V_0 and cytoplasmic V_1 complexes. The V_0 complex of insect V-ATPase is composed of at least five heterologous subunits (a, c, c", d, and e). The V_1 complex is composed of eight different subunits (A to H). The figure is taken from (Baumann and Walz 2012). Reprinted from Journal of Insect physiology, 58, Otto Baumann, Bernd Walz, The blowfly salivary gland – A model system for analyzing the regulation of plasma membrane V-ATPase, 450-458, 2012, with permission from Elsevier.

2.3.4 Relative expression of the V-ATPase subunits

The expression of each of the V-ATPase subunits in the midgut, mixed segment, and P1 samples were evaluated using RT-qPCR (Figure 7). A single peak was observed in the melting curve of each amplicon, suggesting selective amplification. All V-ATPase subunits displayed significantly higher mRNA expression in the mixed segment than in the midgut and P1 (Figure 7).

A V0 subunit c
B V0 subunit d
C V1 subunit A
D V1 subunit B
E V1 subunit C
F V1 subunit D
G V1 subunit E
H V1 subunit F
I V1 subunit G
J V1 subunit H
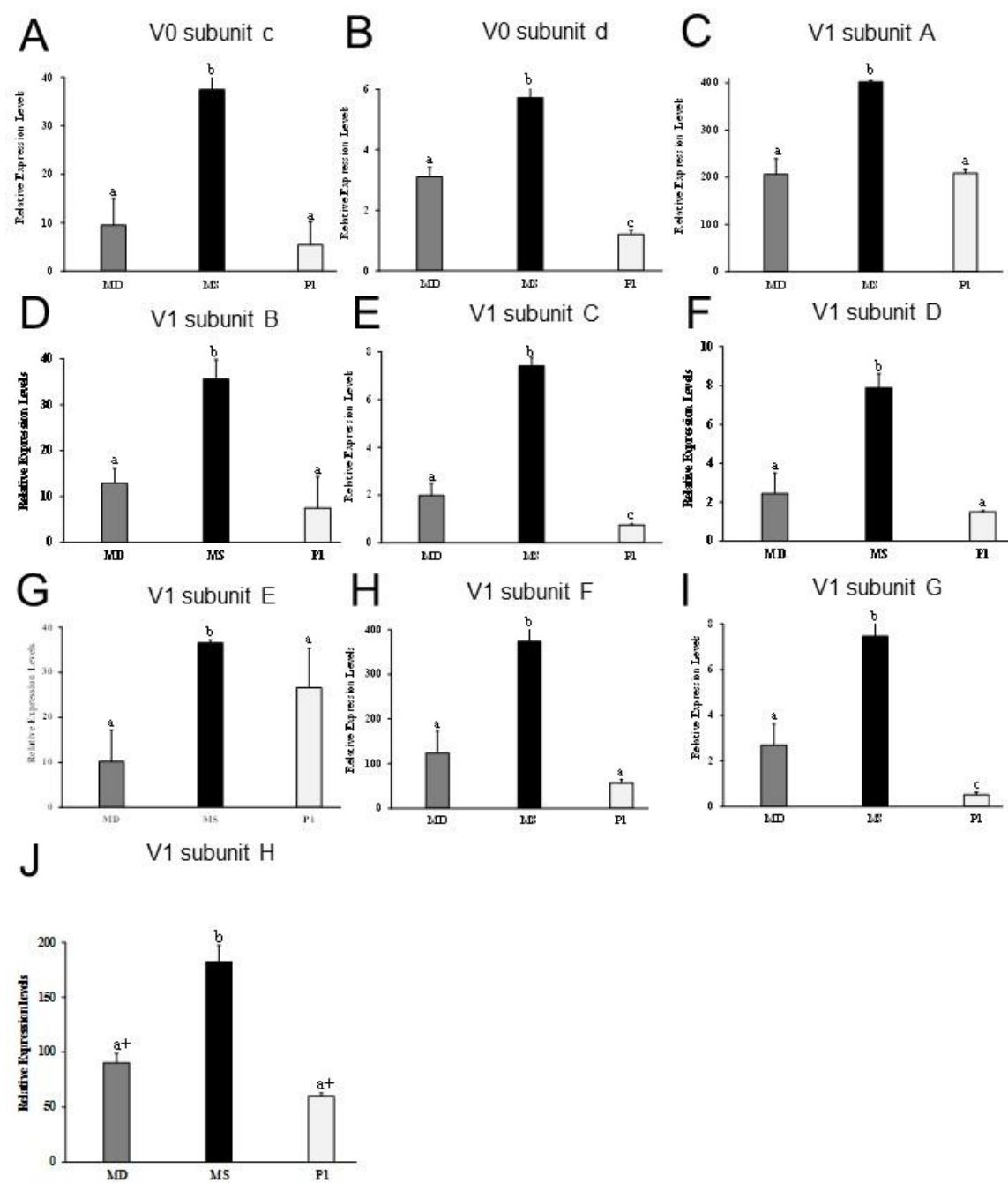
Figure 7: Relative mRNA levels of the genes encoding V-ATPase subunits (A–J) in the three gut segments of *N. takasagoensis*. Expression was quantified by RT-qPCR analysis with *NtNADH-dh* as a reference gene. A, c subunit (16-kDa proteolipid subunit; V_0); B, d subunit (d1-AC39 subunit; V_0); C, A subunit (V_1); D, B subunit (V_1); E, C subunit (V_1); F, D subunit (V_1); G, E subunit (V_1); H, F subunit (V_1); I, G subunit (V_1); J, H subunit (V_1). Error bars represent the standard deviation of three termite samples, each analyzed in duplicate. Different letters (a, b, and c) above the bars indicate significant differences (one-way ANOVA followed by pairwise tests using Student's *t*-test with Bonferroni correction ($P_{adj} < 0.05$). The symbol "+" denotes marginal significance ($0.05 \leq P_{adj} < 0.1$). MD; midgut, MS; mixed segment (MS), P1; first proctodeal segment.

2.3.5 V-ATPase activity

I determined V-ATPase activity in the midgut, mixed segment, and P1 tissues. The V-ATPase activity of the mixed segment was significantly higher than that in the midgut and P1 (Table 5). The V-ATPase activity in the mixed segment was 0.51 ± 0.06 U/gut, which was 58% of the total ATPase activity (0.88 ± 0.06 U/gut) (means $\pm$ SD; $n = 3$). To check if the significant difference was influenced by varying gut volumes, I determined the gut volumes of the midgut (0.089 ± 0.03 mm^3), the mixed segment (0.095 ± 0.02 mm^3), and P1 (0.086 ± 0.02 mm^3) (means $\pm$ SD; $n = 6$). The V-ATPase activity normalized to mm^3 gut volume was also significantly higher in the mixed segment than in the other two tissues (Table 5).

Table 5: V-ATPase activity in crude enzyme extracts of the three gut segments of
N. takasagoensis

	Midgut	Mixed segment	P1
V-ATPase activity (U/gut)	0.09 ± 0.01^a	0.51 ± 0.06^b	0.19 ± 0.03^a
V-ATPase activity (U/mm^3 gut volume)	1.06 ± 0.08^a	5.37 ± 0.64^b	2.16 ± 0.38^c

One unit of enzyme activity is defined as the amount of enzyme that produces 1 nmol of phosphate from ATP per min. Data are expressed the mean $\pm$ SD of three determinations. Different letters (a, b, and c) indicate significant differences (Tukey's HSD test, $P < 0.05$).

2.4 Discussion

In this study, I identified a functional enzyme that was dominantly expressed in the mixed segment. It has been postulated that the mixed segment complements the excretory function of the Malpighian tubules (Bignell et al. 1983; Bignell 1994), rather than the functions of the midgut, which is the site of cellulose digestion and absorption (Tokuda et al. 2012). My transcriptome data indicated that the abundant genes in the mixed segment were under represented in the adjacent tissues, and did not belong to the carbohydrate-metabolizing enzymes, except for one gene encoding a carbohydrate-binding module family protein (CBM13). These results suggested that the mixed segment plays physiologically distinct roles from those of the midgut and P1.

Although the present study was focused on V-ATPase, it should be noted that a unigene encoding an R-type beta-trefoil lectin domain protein was abundant in the mixed segment (Table 4). R-type lectins were originally identified as galactose and/or N-acetylgalactosamine-binding lectins (Cummings and Etzler 2009) and are thought to be located extracellularly or on the plasma membrane (Pace and Baum 2004). Given the fact that many clostridial cells are localized proximal to the mesenteric epithelial surface of the mixed segment in *N. takasagoensis* (Tokuda et al. 2000), in which the R-type lectins are presumably located, it can be postulated that these lectins are involved in interactions between the symbiotic bacteria and the host tissues. Indeed, a recent study in a coleopteran beetle suggested that increased expression of R-type lectins was elicited by microbial cell wall components such as β-1,3 glucans and peptidoglycans (Kim et al. 2013). A study in sandflies indicated that this type of lectins mediates the interaction between the host and pathogens in the midgut (Kamhawi et al. 2004). Further studies are expected to clarify the actual roles of R-type lectins in the mixed segment of termites.

The second most abundant gene in the transcriptome of the mixed segment was the V_1 c-subunit of V-ATPase. In addition, the relative abundance of two other V-ATPase subunits (V_0 subunits A and B) was most dominant in the mixed segment compared to the midgut and P1 segments, a result confirmed by RT-qPCR. Moreover, all V-ATPase subunits showed higher expression in the mixed segment than in the two other segments. It has been shown that the cytoplasmic V_1 complex hydrolyzes ATP and that the integral V_0 domain translocates protons across the membrane (Forgac 2007; Kane 2006).

34

Transcripts of all the subunits comprising the V_1 complex (*i.e.,* subunits A to H), and two of five subunits of the V_0 complex (*i.e.,* subunits c (16-kDa proteolipid subunit) and d (d1-AC39 subunit)) were identified in my transcriptome data. The fact that we did not detect the other V_0 complex subunits might be due to the relatively small output of GS Junior compared to other next-generation sequencers (Liu et al. 2012). However, the apparent V-ATPase activity indicates the presence of all functional subunits in these tissues.

The V-ATPase enzyme has been identified in a variety of intracellular compartments in eukaryotic cells that are rich in mitochondria (Brown et al. 1987; Brown and Breton 1996; Cidon and Nelson 1983; Dean et al. 1984; Forgac et al. 1983; Harikumar and Reeves, 1983; Harvey et al. 1983; Ohkuma et al. 1982; Xie et al. 1983) and plays a role in the acidification of vacuolar compartments (Kane and Stevens 1992; Bowman et al. 1992; Gluck 1992; Forgac 1992 and 1997). The mixed segment of *N. takasagoensis* harbors numerous mitochondria at the base of the mesenteric epithelial cells, which contain large autophagic vacuoles and are observed both in the midgut and the mixed segment (Tokuda et al. 2001). Although detailed studies on such autophagic vacuoles are scarce, a previous study in the midgut of a lower termite indicated strong acid phosphatase activity in the vacuole, similar to autolysosomes (Yamaoka and Nagatani 1980). Thus, it is probable that V-ATPase acidifies these vacuoles to promote degradation of cellular compartments for recycling.

Furthermore, it has been suggested that V-ATPases are involved in potassium ion transport in many insect biological systems, which causes alkalinity (Moffett and Onken 2010; Wieczorek et al. 1991, 1992, 2000); Onken et al. 2008). The physiological basis of alkalinity in the gut lumen of the mixed segment of higher termites has yet to be elucidated, but increased concentrations of potassium and carbonate ions have been found in the anterior hindgut of soil-feeding higher termites, suggesting the secretion of alkaline ions from the mesenteric tissue of the mixed segment (Bignell et al. 1983; Kappler and Brune 1999). Such alkalinity may help to access nutritional organic matters via solubilization of humic substances. The dominant expression and activity of V-ATPase in the present study further supported the idea of an ATP-driven secretion of potassium and carbonate ions (Kappler and Brune 1999). If the V-ATPase in the mixed segment is

coupled with the potassium secretion, it should be localized on the luminal surface of the mesenteric tissue. On the other hand, the morphological characteristics of the mixed segment, such as degenerated microvilli and the proximity of well-developed mitochondria to basal invaginations (Tokuda et al. 2001), render the possibility of the presence of V-ATPase in the basal membrane. V-ATPases localized to the basal membrane of the anterior midgut of mosquito larvae contribute to luminal alkalization by energizing anion exchange across the apical membrane (Boudko et al. 2001). Therefore, the localization of the V-ATPase in the mixed segment needs to be clarified to predict its actual functions.

To my knowledge, this is the first study to reveal the predominant expression and activity of V-ATPase in the mixed segment. My findings are a crucial first step towards the clarification of the physiological functions of the mixed segment, which have remained to be elucidated for more than half a century since the first description of this organ by Grassé and Noirot (1954).

CHAPTER III

Metabolic pathways in the mixed segment of the wood-feeding termite

Nasutitermes takasagoensis

3.1 Introduction

Termites are known worldwide for being significant wood-feeders and thought to be responsible for a yearly atmospheric carbon flux of 1.9 billion tons (Cornwell et al. 2009). This is attributable to termites' capability of efficiently digesting lignocellulosic materials.

Early branching families of termites, traditionally referred to as lower termites, harbor cellulolytic protists in the hindgut. Their digestive process has been extensively studied (Brune 2014; Ni and Tokuda 2013) and a dual cellulolytic system has been proposed for the termites and their protist symbionts (Nakashima et al. 2002). Termites of the most apical lineage (higher termites), which compose the single family Termitidae, have diverse feeding preferences, including wood, humus, soil, lichen, and fungus. With the exception of fungus-growers, the digestive process of Termitidae has been studied in limited groups of termites, such as the wood-feeding Nasutitermitinae and the soil-feeding Termitinae (Ni and Tokuda 2013). The foregut of these termites is generally involved in the fragmentation of ingested materials through a fine gizzard (Mikaelyan et al. 2014; Watanabe and Tokuda 2010), while the midgut is thought to secrete cellulolytic and proteolytic digestive enzymes (Bignell 2011; Ni and Tokuda 2013). Higher termites have lost the symbiotic protists in the hindgut, developing highly specialized hindgut compartments (P1 to P5) bearing several bacterial symbionts (Bignell 1994; Holmgren 1909) and generally acknowledged as a site of microbial fermentation. The physiological functions and microbial symbiosis within hindgut compartments have been described in detail elsewhere (Brune and Ohkuma 2011; Brune 2014; Brune and Dietrich 2015; Ohkuma and Brune 2011). In addition to these compartments, the gut of higher termites is equipped with a unique and complex structure of mesenteric and proctodeal epithelia, the mixed segment, which is situated between the midgut and P1 (Noirot 2001). Unlike the midgut, the mixed segment is not involved in cellulose digestion (Tokuda et al. 1997, 2004, 2005), instead being assumed to secrete a K_2CO_3 alkaline fluid to elevate the luminal pH that reaches over 10 in P1 (Bignell et al. 1983; Brune 2014; Kappler and Brune 1999). The morphology of the epithelial tissue lining the mixed segment of wood-feeding higher termites was studied in *Nasutitermes takasagoensis* (Shiraki) (Tokuda et al. 2001). The proctodeal epithelium consists of flat cuboidal cells covered with thick cuticular layers, while the mesenteric epithelium comprises columnar cells that have

38

degenerated microvilli associated with bacteria (Tokuda et al. 2000) and marked basal invaginations with well-developed mitochondria (Tokuda et al. 2001). These epithelial characteristics suggested an active transport of ionic fluids across the mesenteric cells and my previous study revealed predominant expression of vacuolar H^+-ATPases in *N. takasagoensis* mixed segment, also suggesting an active transport of ions (Chapter 2). These findings supported a proposed function for the mixed segment that complements excretion via the Malpighian tubules, improving irrigation through the specialized hindgut compartments (Bignell 1994; Bignell et al. 1983). However, the exact function of the mixed segment has yet to be clarified. Elucidation on the mixed segment's metabolic function is expected to improve our knowledge on the efficient digestive system of higher termites.

In the present study, transcriptome sequences from the mixed segment and adjacent gut regions (*i.e.,* the midgut and the first proctodeal segment) of the wood-feeding higher termite, *N. takasagoensis* were generated, in order to reveal the metabolic network and functions of this termite mixed segment. The metabolic analysis was performed using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa et al. 2014), which has been employed in many insect studies (*e.g.,* Hu et al. 2015; Lei et al. 2014; Luo et al. 2014; Nandety et al. 2013; Zhang et al. 2014).

3.2 Materials and methods

3.2.1 Termite samples, dissection, and stabilization of gut tissues

Colonies of *N. takasagoensis* were collected on Iriomote Island, Okinawa, Japan, and maintained up to a few weeks in the laboratory, with their own nest materials, at $25 \pm 3°C$ until use. Mature worker termites were used in the experiments. In addition to the cDNA libraries used in Chapter 2 analysis obtained from the midgut (MD2), mixed segment (MS2), and P1 tissues (P1), another midgut (MD1), and two mixed segment libraries (MS1 and MS3) were prepared as described below. Midgut and mixed segment tissue were collected from different colonies and stabilized as described in Chapter 2.

3.2.2 Purification of RNA and 454 pyrosequencing

Transcriptome libraries were prepared as described in Chapter 2. In brief, pooled tissues (750 tissues per sample) were homogenized using a plastic pestle and total RNA was isolated using the RNeasy mini kit (Qiagen, Hilden, Germany), according to manufacturer's instructions. Enrichment of the total RNA for mRNA species was performed with the NucleoTrap® mRNA Mini Kit (Takara Bio, Otsu, Japan) and verified using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA); RNA quantification was assessed with the Qubit® RNA Assay Kit in a Qubit® 2.0 fluorometer (Invitrogen, Paisley, UK). Double-strand cDNA synthesis was performed according to the cDNA Rapid Library Preparation method from Roche (Roche, May 2010). cDNA libraries were sequenced in a 454 GS Junior sequencer (Roche Diagnostics, Penzberg, Germany).

Reads obtained from the 454 pyrosequencing were preprocessed by removing ribosomal RNA (rRNA) and mitochondrial (mt) DNA sequences using Roche Reference Mapper 2.7 software (Roche Diagnostics) (see Chapter 2 for more details). Transcripts were assembled from the pooled reads of six sequence datasets (MD2, MS2, and P1 described in my previous chapter plus MD1, MS1, and MS3 obtained in the present study) using Newbler 2.7p (Roche Diagnostics) with the default settings for cDNA mode. This analysis enabled comparison of read abundances between samples based on a single assembly result. This large dataset resulted in a better assembly (*i.e.*, increased number of isotigs (transcripts) and reduced number of singleton reads) than assemblies described in Chapter 2 (MD2, MS2, and P1). Each assembled sequence (isotig) was then searched for homology against the National Center for Biotechnology Information (NCBI) non-redundant nucleotide (Nt) database, using the basic local alignment search tool (BLAST) with a threshold *e*-value of 10^{-9}, to identify bacterial and non-targeted eukaryote genes (*i.e.*, plant and vertebrate gene sequences). Isotigs showing significant similarity to those genes were discarded and the number of total reads aligned to contigs/isotigs was corrected by excluding reads aligned to these genes.

3.2.3 Functional annotations

The assembled sequences were assigned to metabolic pathways based on the KEGG database (Kanehisa et al. 2014). First, the nucleotide sequences of isotigs were queried on the KEGG Automatic Annotation Server (KAAS) using the Bi-directional best hit (BBH) assignment method and the insect reference genes dataset to assign KEGG orthologs' K-numbers. These K-numbers were then used to map KEGG reference pathways via the online version of the KEGG mapper (http//www.kegg.jp/keg/tool/map_pathway/html). The coding sequences of two termite genomes (Poulsen et al. 2014; Terrapon et al. 2014) were also mapped onto KEGG reference pathways, allowing comparisons with the transcriptome datasets obtained from *N. takasagoensis* gut libraries. After retrieving the possible KEGG pathways based on K-numbers, the KEGG pathways considered to be present in the focal termite species were manually selected to be included in the further analyses comparing termite genomic data, as KAAS does not determine whether a pathway is present or not from the input dataset. KEGG orthologs encoding transporters were detected and categorized based on the Transporter Classification Database (http://www.tcdb.org/) (Saier et al. 2014). Functions of the detected transporters were manually annotated.

3.2.4 Normalization of read counts

To exclude the bias caused by the different throughputs obtained between different libraries, read counts per isotig were normalized according to the upper quartile method (Bullard et al. 2010). Read abundance per pathway was then listed based on the normalized read counts.

3.2.5 Principal component analysis

Principal component analysis (PCA) was performed using the normalized read count data to identify the variables (KEGG pathways or orthologs) contributing the most to differential expression between tissues. This analysis was conducted in R 3.1.2 (Development Core Team, Vienna; http://www.R-project.org).

3.2.6 Reverse-transcription quantitative polymerase chain reaction (RT-qPCR)

The isotig sequences encoding carbonic anhydrase (isotig4607 and isotig7716), NHE6/7 and NHE8 antiporters were verified by Sanger sequencing as described in Chapter 2. The gene-specific primers (Table 6) were designed using Primer3 (http://primer3.sourceforge.net/). To determine the expression levels of these isotigs in the midgut, mixed segment, and P1 samples, RT-qPCR was performed as described in Chapter 2 with an NADH dehydrogenase (*NtNADH-dh*) gene as an endogenous control. The primers used in this study (Table 7) were designed using the Primer Express software (Applied Biosystems, Foster City, CA, USA). RT-qPCR was carried out from three biological replicates, and each reaction was run in duplicate. The relative expression values in these tissues were analyzed by using one-way analysis of variance (ANOVA). When the obtained p-value was significant ($p < 0.05$), pairwise comparisons using Student's t-test with Bonferroni correction were performed to determine significant differences ($p_{adj} < 0.05$).

Table 6: Primer sets used in this study for Sanger sequencing of the *N. takasagoensis*

Gene	Primer sequence	Primer length (bp)
Carbonic anhydrase (isotig4607)	Forward: 5'-ATTGCACTTACCGGTTCAGC-3' Reverse: 5'-TAACCGCTTCGCTACCAAGT-3'	20 20
Carbonic anhydrase (isotig7716)	Forward: 5'-TGCTACTCACCGCACTGTTA-3' Reverse: 5'-GCTTGGATCTGGTGTTGGTT-3'	20 20
NHE6/7	Forward: 5'-CGAGTTGTTTTATGTATTTGTGGC-3' Reverse: 5'-GTGATCCTTGAGAACCCCAA-3'	24 20
NHE8	Forward: 5'-ATCTGCTGGGATTGGAGTTG-3' Reverse:5'-CCTCCTTGTGAAGAAAGGGA-3'	20 24

Table 7: Primer sets used in this study for RT-qPCR analysis of the *N. takasagoensis*

Gene	Primer sequence	Primer length (bp)
Carbonic anhydrase (isotig4607)	Forward: 5'-GAGCTTCATTTGGTGCATTGG-3' Reverse: 5'-GGCTGCCTCGCCACAGT-3'	21 17
Carbonic anhydrase (isotig7716)	Forward: 5'-TCAGCAGTAGGCTCACCAATAATG-3' Reverse: 5'-GCCGGACTTGTCATGTTCATC-3'	24 21
NHE6/7	Forward: 5'-CTGCGACGCTAATAGAGGACAA-3' Reverse: 5'-CCCAAGTCATGCATCTCATCA-3'	22 21
NHE8	Forward: 5'-TGTAACCGCTTCCGTGAACA-3' Reverse:5'-ACCACTGAACCACATGATGAACA-3'	20 23

3.2.7 Absolute quantitative real-time PCR

For the absolute quantitative PCR, the concentration of the purified PCR product (obtained sequences from sanger sequencing primers) was measured with a spectrophotometer (Nanodrop ND-1000). Then, the average molecular weight of each product was calculated (carbonic anhydrase isotig4607 – 514154.26 Da, carbonic anhydrase isotig7716 – 363355.30 Da, NHE6/7 antiporter – 1566278.71 Da and NHE8 – 459710.46 Da; where Da = g/mol (1 mol = 6.02×10^{23} molecules). Finally, the initial number of copies per unit volume was calculated using the following equation.

$$\text{Number of copies/µl} = \frac{6.022 \times 10^{23} \text{ (molecules/moles)} \times \text{DNA concentrations (g/µl)}}{\text{Number of base pairs} \times 660 \text{ daltons}}$$

The volume of each PCR standards was adjusted to 10^{10} copies/µl. This stock solution was serially diluted to obtain a standard series from 10^9 to 10 copies/µl with each step differing by tenfold dilutions. The standard series of each gene was run under the same conditions (the conditions provided for qPCR analysis) and the copy numbers of genes were determined by reading off the standards series with the C_t values of the genes. Each standard was run in duplicate reactions.

3.2.8 Data deposition

The 454 reads of *N. takasagoensis* were deposited in the DNA Data Bank of Japan (DDBJ), under the DDBJ Sequence Read Archive (DRA) accession numbers DRA004254 (MD1 library), DRA004255 (MS1 library), and DRA004256 (MS3 library). The raw reads of MD2 (DRA003011), MS2 (DRA003012), and P1 (DRA003013) were obtained from Chapter 2. The verified sequences of the genes encoding NHE6/7 antiporter and carbonic anhydrase (isotig4607) were deposited in the DDBJ database under the accession numbers LC126081 and LC126080, respectively. The NHE8 and carbonic anhydrase (isotig7716) were provided as a supplement (File S1).

3.3 Results

3.3.1 Transcript assembly and detection of metabolic pathways in the mixed segment

As summarized in Table 8, 10,910 isotigs (transcripts) were obtained from the reads in the six libraries, with the isotigs from the mixed segment libraries MS1, MS2, and MS3 totalizing 7,018, 8,236, and 9,566, respectively. Using KAAS, K-numbers were assigned to 28.5% of the mixed segment isotigs, which were classified into 2,304 KEGG orthologs with different K-numbers (Table 9). Using the KEGG mapper, 1,852 of these orthologs were annotated onto 125 reference pathways identified for insects listed in the KEGG organism catalog (http://www.genome.jp/kegg/catalog/org_list.html) (Table 9). On the other hand, approximately 20% of KEGG orthologs were assigned to reference pathways that were related to human disease or that were not relevant to insects. Therefore, these orthologs were excluded from further analysis.

Table 8: Overview of *N. takasagoensis* transcriptome assembly features

Description	MD1	MD2	MS1	MS2	MS3	P1
Total number of reads	100490	154714	90793	116228	146705	114213
rRNA/ Mitogenomic reads[*]	72203	41848	57337	47005	45031	63761
Bacterial reads[+]	377	1408	996	3517	6007	4247
Non-targeted eukaryotic reads[§]	105	247	123	207	305	208
Unassembled reads[†]	6712	23204	8360	11950	22952	16185
Low-quality reads[†]	448	1077	594	1488	1738	1046
Putative host mRNA reads	20645	86930	23383	52061	70672	28766
Total number of isotigs (transcripts determined by total reads from all six libraries)			10910			
Isotigs (transcripts) assigned to each library	6195	9013	7018	8236	9566	6367

MD1 and MD2, libraries prepared from the midgut; MS1 to MS3, libraries from the mixed segment; P1, a library from the first proctodeal segment

*Ribosomal RNA (rRNA) and mitogenome sequences were removed using Roche Reference Mapper software, ver. 2.7 prior to transcript assembly.

+Bacterial sequences were discarded by using blastn after transcript assembly.

§Non-targeted eukaryotic sequences were discarded by using blastn after transcript assembly.

†These sequences were not used for further analysis.

Table 9: Frequency of distinct KEGG orthologs associated with the reference metabolic pathways of insects based on the gut libraries and published termite genomes

	Name of the reference pathway	*Z. nevadensis*	*M. natalensis*	MS libraries	MD libraries	P1 library
Metabolism						
Carbohydrate metabolism						
1	K00010 Glycolysis / Gluconeogenesis	28	28	25	24	21
2	K00020 Citrate cycle (TCA cycle)	22	19	20	20	17
3	K00030 Pentose phosphate pathway	18	18	14	14	11
4	K00040 Pentose and glucuronate interconversions	12	12	10	9	7
5	K00051 Fructose and mannose metabolism	15	15	13	12	9
6	K00052 Galactose metabolism	16	15	11	11	9
7	K00053 Ascorbate and aldarate metabolism	7	6	5	5	5
8	K00500 Starch and sucrose metabolism	19	19	17	16	12
9	K00520 Amino sugar and nucleotide sugar metabolism	28	27	20	20	11
10	K00620 Pyruvate metabolism	21	19	19	19	15
11	K00630 Glyoxylate and dicarboxylate metabolism	20	19	12	12	11
12	K00640 Propanoate metabolism	17	16	12	12	9
13	K00650 Butanoate metabolism	13	13	10	9	9
14	K00562 Inositol phosphate metabolism	34	30	18	17	11
Energy metabolism						
15	K00190 Oxidative phosphorylation	91	79	81	73	67
16	K00910 Nitrogen metabolism	5	5	4	4	4
17	K00920 Sulfur metabolism	8	6	4	4	4
Lipid metabolism						
18	K00061 Fatty acid biosynthesis	6	6	4	4	3
19	K00062 Fatty acid elongation	11	11	8	8	7
20	K00071 Fatty acid degradation	21	19	17	17	15
21	K00072 Synthesis and degradation of ketone bodies	5	5	4	3	4
22	K00100 Steroid biosynthesis	7	6	6	6	3
23	K00561 Glycerolipid metabolism	22	21	11	11	8

24	K00564 Glycerophospholipid metabolism	42	41	19	18	10
25	K00565 Ether lipid metabolism	13	11	7	7	5
26	K00600 Sphingolipid metabolism	20	18	11	11	8
27	K00590 Arachidonic acid metabolism	10	9	7	7	4
28	K00592 alpha-Linolenic acid metabolism	4	3	2	2	1
29	K01040 Biosynthesis of unsaturated fatty acids	6	6	5	5	5

Nucleotide metabolism

30	K00230 Purine metabolism	102	92	44	43	28
31	K00240 Pyrimidine metabolism	70	62	34	33	18

Amino acid metabolism

32	K00250 Alanine, aspartate and glutamate metabolism	22	21	16	15	10
33	K00260 Glycine, serine and threonine metabolism	25	26	13	12	9
34	K00270 Cysteine and methionine metabolism	24	21	13	13	9
35	K00280 Valine, leucine and isoleucine degradation	30	29	21	21	18
36	K00290 Valine, leucine and isoleucine biosynthesis	3	3	1	1	0
37	K00300 Lysine biosynthesis	1	1	1	1	1
38	K00310 Lysine degradation	30	25	12	12	8
39	K00220 Arginine biosynthesis	11	11	6	6	4
40	K00330 Arginine and proline metabolism	23	22	13	13	11
41	K00340 Histidine metabolism	10	10	4	4	3
42	K00350 Tyrosine metabolism	14	13	6	6	3
43	K00360 Phenylalanine metabolism	8	7	4	4	2
44	K00380 Tryptophan metabolism	16	14	8	8	8
45	K00400 Phenylalanine, tyrosine and tryptophan biosynthesis	4	4	2	2	1

Metabolism of other amino acids

46	K00410 beta-Alanine metabolism	16	16	14	14	9
47	K00430 Taurine and hypotaurine metabolism	5	4	4	4	2
48	K00450 Selenocompound metabolism	9	9	6	6	5
49	K00460 Cyanoamino acid metabolism	3	2	1	2	1
50	K00471 D-Glutamine and D-glutamate metabolism	2	2	1	1	1
51	K00472 D-Arginine and D-ornithine metabolism	1	1	1	1	1
52	K00480 Glutathione metabolism	22	21	16	16	12

Glycan biosynthesis and metabolism

53	K00510 N-Glycan biosynthesis	33	32	21	18	13
54	K00513 Various types of N-glycan biosynthesis	22	22	15	14	11
55	K00512 Mucin type O-Glycan biosynthesis	2	2	2	2	1
56	K00514 Other types of O-glycan biosynthesis	11	10	2	2	2
57	K00532 Glycosaminoglycan biosynthesis - chondroitin sulfate / dermatan sulfate	8	7	0	0	0
58	K00534 Glycosaminoglycan biosynthesis - heparan sulfate / heparin	13	12	2	2	2
59	K00533 Glycosaminoglycan biosynthesis - keratan sulfate	2	2	2	2	2
60	K00531 Glycosaminoglycan degradation	11	11	6	6	3
61	K00563 Glycosylphosphatidylinositol(GPI)-anchor biosynthesis	22	19	13	13	6
62	K00601 Glycosphingolipid biosynthesis - lacto and neolacto series	4	2	2	1	1
63	K00603 Glycosphingolipid biosynthesis - globo series	5	4	4	3	2
64	K00604 Glycosphingolipid biosynthesis - ganglio series	3	3	2	2	2
65	K00511 Other glycan degradation	10	10	8	8	5

Metabolism of cofactors and vitamins

66	K00730 Thiamine metabolism	3	3	1	1	1
67	K00740 Riboflavin metabolism	2	1	2	1	1
68	K00750 Vitamin B6 metabolism	3	4	1	1	0
69	K00760 Nicotinate and nicotinamide metabolism	8	8	6	6	4
70	K00770 Pantothenate and CoA biosynthesis	9	9	7	6	4
71	K00780 Biotin metabolism	3	3	1	1	1
72	K00785 Lipoic acid metabolism	3	2	2	1	2
73	K00790 Folate biosynthesis	11	10	7	6	6
74	K00670 One carbon pool by folate	13	12	2	1	0
75	K00830 Retinol metabolism	5	6	5	5	4
76	K00860 Porphyrin and chlorophyll metabolism	18	18	13	12	10
77	K00130 Ubiquinone and other terpenoid-quinone biosynthesis	9	8	3	2	1

Metabolism of terpenoids and polyketides

78	K00900 Terpenoid backbone biosynthesis	20	20	9	9	6
79	K00981 Insect hormone biosynthesis	11	11	3	2	3

Biosynthesis of other secondary metabolites

80	K00232 Caffeine metabolism	1	2	1	1	1

49

Xenobiotics biodegradation and metabolism

81	K00980 Metabolism of xenobiotics by cytochrome P450	5	5	6	6	4
82	K00982 Drug metabolism - cytochrome P450	4	4	4	4	3
83	K00983 Drug metabolism - other enzymes	15	14	10	10	6

Genetic Information Processing

Transcription

84	K03020 RNA polymerase	28	25	12	11	6
85	K03022 Basal transcription factors	30	27	9	9	5
86	K03040 Spliceosome	102	99	69	64	46

Translation

87	K03010 Ribosome	118	103	100	95	88
88	K00970 Aminoacyl-tRNA biosynthesis	28	28	19	18	12
89	K03013 RNA transport	115	105	77	74	62
90	K03015 mRNA surveillance pathway	54	54	42	36	36
91	K03008 Ribosome biogenesis in eukaryotes	63	54	34	30	21

Folding, sorting and degradation

92	K03060 Protein export	21	19	17	16	13
93	K04141 Protein processing in endoplasmic reticulum	106	98	82	81	67
94	K04130 SNARE interactions in vesicular transport	19	16	13	12	7
95	K04120 Ubiquitin mediated proteolysis	89	81	49	45	33
96	K04122 Sulfur relay system	8	6	4	4	2
97	K03050 Proteasome	34	33	33	33	29
98	K03018 RNA degradation	56	52	30	27	23

Replication and repair

99	K03030 DNA replication	31	28	9	9	7
100	K03410 Base excision repair	24	19	8	7	5
101	K03420 Nucleotide excision repair	35	31	17	17	13
102	K03430 Mismatch repair	17	16	7	7	5
103	K03440 Homologous recombination	22	19	6	6	3
104	K03450 Non-homologous end-joining	10	8	5	5	4
105	K03460 Fanconi anemia pathway	34	24	5	5	2

Environmental Information Processing

Membrane transport

106	K02010 ABC transporters		14	14	8	8	6

Signal transduction

107	K04013 MAPK signaling pathway - fly		16	14	8	8	7
108	K04310 Wnt signaling pathway		60	55	28	26	21
109	K04330 Notch signaling pathway		21	20	11	11	7
110	K04340 Hedgehog signaling pathway		25	23	7	7	7
111	K04350 TGF-beta signaling pathway		29	28	18	18	10
112	K04391 Hippo signaling pathway - fly		46	44	26	24	21
113	K04630 Jak-STAT signaling pathway		18	14	11	11	6
114	K04068 FoxO signaling pathway		50	50	24	23	23
115	K04070 Phosphatidylinositol signaling system		37	33	18	17	9
116	K04150 mTOR signaling pathway		29	26	17	16	11

Signaling molecules and interaction

117	K04080 Neuroactive ligand-receptor interaction		33	30	5	3	4
118	K04512 ECM-receptor interaction		10	12	3	3	2

Cellular Processes

Transport and catabolism

119	K04144 Endocytosis		113	98	76	68	53
120	K04145 Phagosome		41	37	36	36	34
121	K04142 Lysosome		64	55	47	48	33
122	K04146 Peroxisome		45	41	33	31	26
123	K04140 Regulation of autophagy		15	15	8	7	5

Organismal Systems

Sensory system

124	K04745 Phototransduction - fly		16	14	7	7	5

Development

125	K04320 Dorso-ventral axis formation		20	18	12	12	9

Environmental adaptation

126	K04711 Circadian rhythm - fly		5	7	3	3	3

	3032	2787	1852	1764	1367
Total number of KEGG orthologs					
% to the average number of KEGG orthologs from the genomes			63.7%	60.6%	47.0%

3.3.2 KEGG orthologs comprising KEGG reference pathways in the several MS
    libraries and termite genomes

The coding sequences (CDSs) of two available genome sequences from lower and higher termites (Poulsen et al. 2014; Terrapon et al. 2014) were mapped onto KEGG reference pathways and compared with the KEGG results from the MS libraries. In total, K-numbers were assigned to 33.8% of 15,860 CDSs from the lower termite *Zootermopsis nevadensis* Hagen, 28.3% of 16,310 CDSs of the higher termite *Macrotermes natalensis* (Haviland) (Table 10). These KEGG orthologs retrieved 126 KEGG reference pathways showing slight or no differences in their number of KEGG orthologs (Table 9). This considerable similarity between phylogenetically distinct termite species suggested a similar set of genes for *N. takasagoensis*.

Table 11 shows the frequency of KEGG orthologs in the three MS libraries and in the termite genomes with respect to each category of KEGG orthology. On average, the three MS libraries contained 55.4% of the KEGG orthologs to those found in the termite genomes. Although the proportion of KEGG orthologs detected in each category varied considerably, the pattern of these variations was fairly consistent among the three libraries (Table 11), suggesting they reflect selective gene expression patterns in the mixed segment. Compared to the average proportion (55.4% ± 6.6%), the percentage of detectable KEGG orthologs involved in 'Energy metabolism' in the mixed segment was markedly higher (81.8 ± 12.1%). On the other hand, the percentage of detectable KEGG orthologs involved in 'replication and repair' (30.0 ± 5.49%) as well as in 'Signaling molecules and interaction' (17.1 ± 1.34%), was considerably lower than the average percentage of KEGG orthologs detected in the mixed segment.

Table 10: Percentage of genes or transcripts assigned to KEGG orthologs

	Total number of genes or isotigs	Number of genes or isotigs assigned by K-numbers	Percentage of genes or isotigs with assigned K-numbers	Number of KEGG orthologs with different K-numbers
Z. nevadensis	15,860	5,263	33.2%	4,485
M. natalensis	16,310	4,615	28.3%	3,986
N. takasagoensis				
MS libraries	10,562	3,009	28.5%	2,304
MD libraries	9,620	2,812	29.2%	2,149
P1 library	6,367	2,074	32.6%	1,605

MS, mixed segment; MD, midgut; P1, first proctodeal segment

Table 11: Number (N) and percentage (%) of the several KEGG orthologs found for the mixed segment libraries according to their category

KEGG Category	Mean number of KEGG orthologs in termite genomes	KEGG orthologs in MS1		KEGG orthologs in MS2		KEGG orthologs in MS3		Mean % ± Standard deviation
		N	%	N	%	N	%	
Metabolism								
Energy metabolism	97	66	68.0	88	90.7	84	86.6	81.8±12.1
Xenobiotics biodegradation and metabolism	24	14	58.3	18	75.0	19	79.2	70.8±11.0
Carbohydrate metabolism	263	161	61.2	198	75.3	195	74.1	70.2±7.8
Metabolism of other amino acids	57	35	61.4	40	70.2	43	75.4	69.0±7.1
Lipid metabolism	162	81	50.0	89	54.9	100	61.7	55.6±5.9
Biosynthesis of other secondary metabolites	2	1	50.0	1	50.0	1	50.0	50.0±0.0
Amino acid metabolism	214	92	43.0	112	52.3	115	53.7	49.7±5.8
Glycan biosynthesis and metabolism	141	62	44.0	70	49.6	75	53.2	48.9±4.7
Metabolism of cofactors and vitamins	86	36	41.9	43	50.0	46	53.5	48.4±6.0
Nucleotide metabolism	163	47	28.8	65	39.9	66	40.5	36.4±6.6
Metabolism of terpenoids and polyketides	31	8	25.8	12	38.7	12	38.7	34.4±7.5
Subtotal	**1240**	**603**	**48.6**	**736**	**59.4**	**756**	**61.0**	**56.3±6.7**
Genetic information processing								
Folding, sorting and degradation	319	188	58.9	205	64.3	218	68.3	63.8±4.7
Translation	361	178	49.3	250	69.3	256	70.9	63.2±12.0
Transcription	156	56	35.9	80	51.3	80	51.3	46.2±8.9
Replication and repair	159	38	23.9	50	31.4	55	34.6	30.0±5.5
Subtotal	**995**	**460**	**46.2**	**585**	**58.8**	**609**	**61.2**	**55.4±8.0**
Environmental information processing								
Membrane transport	14	8	57.1	7	50.0	7	50.0	52.4±4.1
Signal transduction	319	135	42.3	150	47.0	159	49.8	46.4±3.8
Signaling molecules and interaction	43	7	16.3	7	16.3	8	18.6	17.1±1.3
Subtotal	**376**	**150**	**39.9**	**164**	**43.6**	**174**	**46.3**	**43.3±3.2**
Cellular processes								
Transport and catabolism	262	162	61.8	186	71.0	195	74.4	69.1±6.5
Subtotal	**262**	**162**	**61.8**	**186**	**71.0**	**195**	**74.4**	**69.1±6.5**
Organismal systems								
Development	19	10	52.6	11	57.9	12	63.2	57.9±5.2
Sensory system	15	6	40.0	7	46.7	6	40.0	42.2±3.9
Environmental adaptation	6	3	50.0	2	33.3	2	33.3	38.9±9.6
Subtotal	**40**	**19**	**47.5**	**20**	**50.0**	**20**	**50.0**	**49.2±1.4**
Total	2913	1394	**47.9**	1691	**58.1**	1754	**60.2**	**55.4±6.6**

3.3.3 Transcript assembly and detection of metabolic pathways in the midgut and first proctodeal segment

The isotigs present in two midguts and one P1 libraries (MD1, MD2, and P1) were selected from the total isotigs generated for the six libraries (Table 12) and compared to those of MS1, MS2, and MS3. K-numbers were assigned to 29.2 and 32.6% of the isotigs from MD and P1 libraries (Table 10), and retrieved 125 and 122 KEGG reference pathways, respectively (Table 9). The proportions found for each category of KEGG orthologs in the midgut and P1 revealed a similar tendency to those found for the mixed segment (Table 12).

Table 12: Frequency of KEGG orthologs among three gut regions to those of the termite genome (%)

KEGG Category	MD	MS	P1
Metabolism			
Energy metabolism	69.1±18.95	81.8±12.08	77.3
Xenobiotics biodegradation and metabolism	75.0±11.79	70.8±11.02	54.2
Carbohydrate metabolism	69.4±9.95	70.2±7.81	59.7
Metabolism of other amino acids	65.8±11.16	69.0±7.09	54.4
Lipid metabolism	55.2±3.06	55.6±5.89	45.1
Biosynthesis of other secondary metabolites	50.0±0.00	50.0±0.00	50.0
Amino acid metabolism	48.6±7.93	49.7±5.84	40.7
Glycan biosynthesis and metabolism	43.3±9.03	48.9±4.65	35.5
Metabolism of cofactors and vitamins	40.7±8.22	48.4±5.97	39.5
Nucleotide metabolism	36.5±10.85	36.4±6.56	28.2
Metabolism of terpenoids and polyketides	35.5±0.00	34.4±7.45	29.0
Subtotal	**53.7±9.12**	**56.3±6.71**	**46.5**
Genetic information processing			
Folding, sorting and degradation	57.5±10.86	63.8±4.72	54.5
Translation	57.2±13.91	63.2±12.02	60.7
Transcription	41.0±13.60	46.2±8.88	36.5
Replication and repair	23.9±13.34	30.0±5.49	24.5
Subtotal	**49.4±12.79**	**55.4±8.04**	**49.1**
Environmental information processing			
Membrane transport	53.6±5.05	52.4±4.12	42.9
Signal transduction	43.4±8.20	46.4±3.80	38.2
Signaling molecules and interaction	14.0±0.00	17.1±1.34	14.0
Subtotal	**40.4±7.15**	**43.3±3.21**	**35.6**
Cellular processes			
Transport and catabolism	66.8±7.29	69.1±6.51	57.6
Subtotal	**66.8±7.29**	**69.1±6.51**	**57.6**
Organismal systems			
Development	55.3±11.16	57.9±5.26	47.4
Sensory system	40.0±9.43	42.2±3.85	33.3
Environmental adaptation	50.0±0.00	38.9±9.62	50.0
Subtotal	**48.8±8.84**	**49.2±1.44**	**42.5**
Total	**51.6±10.0**	**55.4±6.6**	**46.9**

3.3.4 Differential expression of KEGG orthologs among the midgut, the mixed segment, and first proctodeal segment

To compare the expression of the KEGG orthologs comprising each pathway, read counts were normalized using the general quantile procedure, as it gives a less biased evaluation than common global normalization techniques such as the reads per kilobase of exon per million mapped sequence reads (*i.e.*, RPKM) (Bullard et al. 2010). Based on these normalized read counts comprising each pathway (Table S1), a PCA was applied to the six libraries (two MD, three MS, and one P1 libraries; Figure 8). In addition, to find the pathways that account for the differences of the gut libraries, another PCA was also applied for the 126 KEGG pathways based on the normalized read counts comprising each gut library. As the resulting biplots were heavily congested, a second PCA was performed, selecting only the 20 KEGG pathways with the largest coefficients of variation (Table S1). Although showing essentially the same patterns of separation as the analysis with all the 126 pathways, this second PCA had better resolution (Figure 9), separating MS from the other gut regions along the first principal component (PC1), which accounted for 85.4% of the observed variation. The separation of MS toward the negative side of PC1 related to the abundance of 'Phagosome' and 'Oxidative phosphorylation' reads, as evidenced by the positioning of these variables on the PCA biplot and by their high eigenvalues on PC1 (Table S1). MS was also separated from MD and P1 along the second principal component (PC2), which accounted for 10.1% of the observed variation. The observed separation towards the negative side of PC2 resulted from the abundance 'Lysosome' and 'Oxidative phosphorylation' reads, also distributed on this side of the diagram and presenting the highest eigenvalues for the mixed segment on PC2 (Table S1). These results suggested that energy production via oxidative phosphorylation, the turnover of cellular materials by phagocytosis, and lysosomal gene expression are more characteristic of the mixed segment than of the adjacent gut regions.

To address the hypothesis that the mixed segment plays a role in active ion transport, PCAs considering only the normalized read counts of 107 selected KEGG orthologs assumed to be involved in the transport of ions and other substances as well as the comprising six gut libraries were performed (Figure 10; Table S2 (and S3)). This analysis separated MS from MD and P1 along the PC1, which accounted for 89.2% of the

observed variation, reflecting the abundance of V-ATPases, F-type ATP synthases, and some transporters reads in MS libraries. Carbonic anhydrase, which catalyzes the hydration of carbon dioxide to produce protons and bicarbonate, was also characteristic of the mixed segment, while ABC transporters did not contribute to this separation. Along the PC2, which accounted for 5.8% of the observed variation, MS was only separated from P1, suggesting more functional similarities between the MS and MD than between these two regions and P1.

Finally, the functions of the transporters listed in Table S2 were manually annotated (Table S3) to find the molecules involved in potassium and bicarbonate secretion. The number of normalized read counts of P-ATPases, which are related to the active transport of potassium (Patrick et al. 2006), was relatively low in the mixed segment (Table S2). Thus, and being potassium secretion coupled with V-ATPase activity in lepidopteran insects (Wieczorek et al. 2000), only *SLC9A6_7* and *SLC9A8* orthologs could encode the potential K^+/H^+ antiporters known as NHE6 or 7 (NHE6/7) and NHE8 (Donowitz et al. 2013). When the normalized read counts of KEGG orthologs putatively involved in gut alkalization were compared (Figure 11), they showed different patterns in the three gut regions. Whereas V-ATPase (Figure 11a) and NHE6/7 (Figure 11b) revealed comparable expression patterns, NHE8 expression was much lower in all regions (Figure 11c); the dominant expression of carbonic anhydrase (two isotigs (4607 and 7716) belong to the same KEGG ortholog) in the mixed segment was also confirmed by the profile of the normalized read counts (Figure 11d), albeit its overall low number of normalized reads. The present dataset (Table S2) also contained an ortholog encoded by *SLC4*, a solute-carrier family gene, assumed to be potentially involved in HCO_3^-/Cl^- exchange (Mount and Romero 2004).
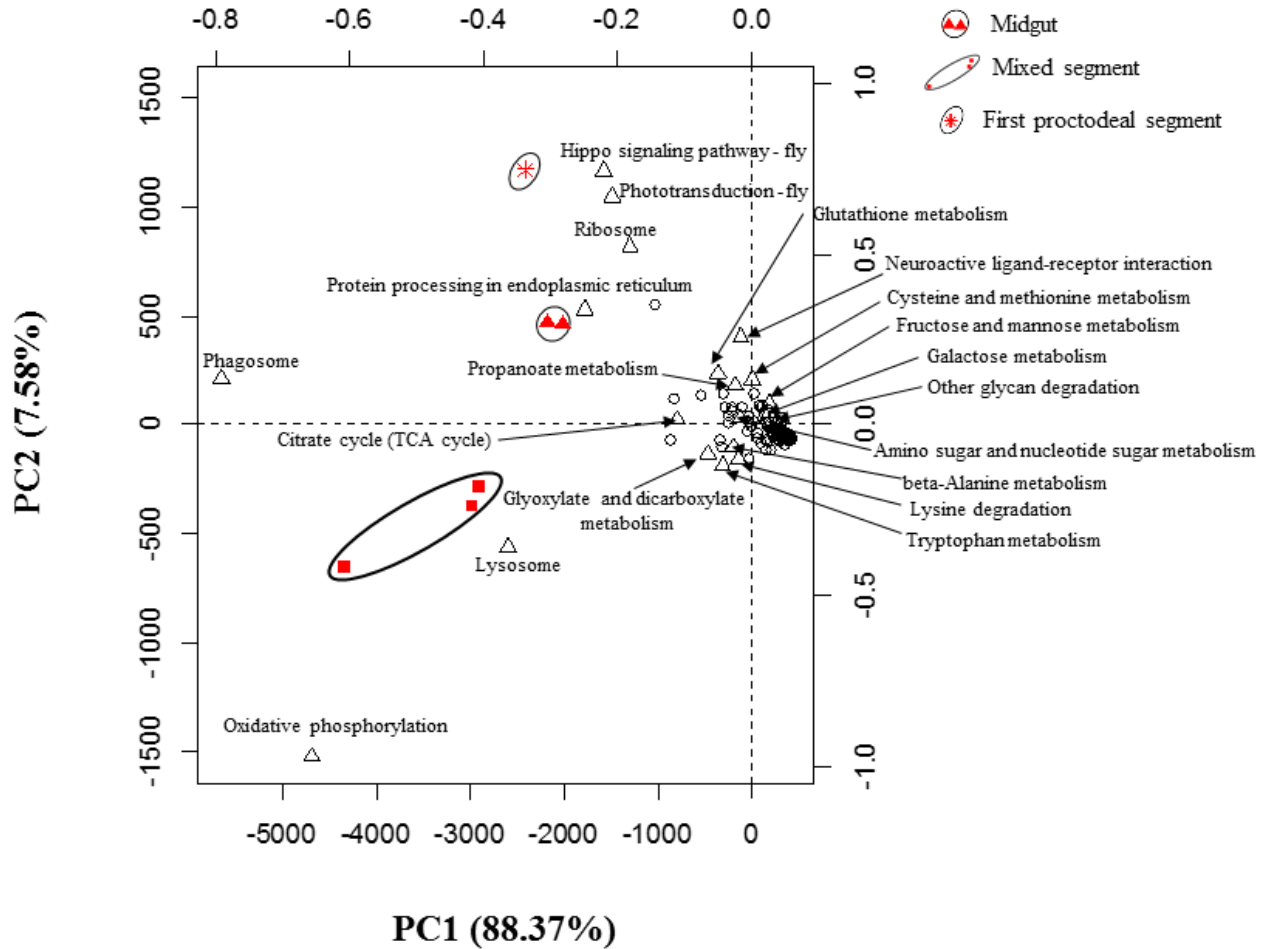
Figure 8: Principal components analysis biplot based on the normalized read counts of KEGG orthologs belonging to 126 reference pathways identified in the midgut (*filled triangles*), mixed segment (*filled squares*), and first proctodeal segment (*filled asterisk*) libraries. PC1 corresponds to the first principal component and PC2 to the second principal component. The percentage of variation explained by each PC is indicated between brackets. Open triangles (represent largest coefficients of variation) and open circles indicated pathways.
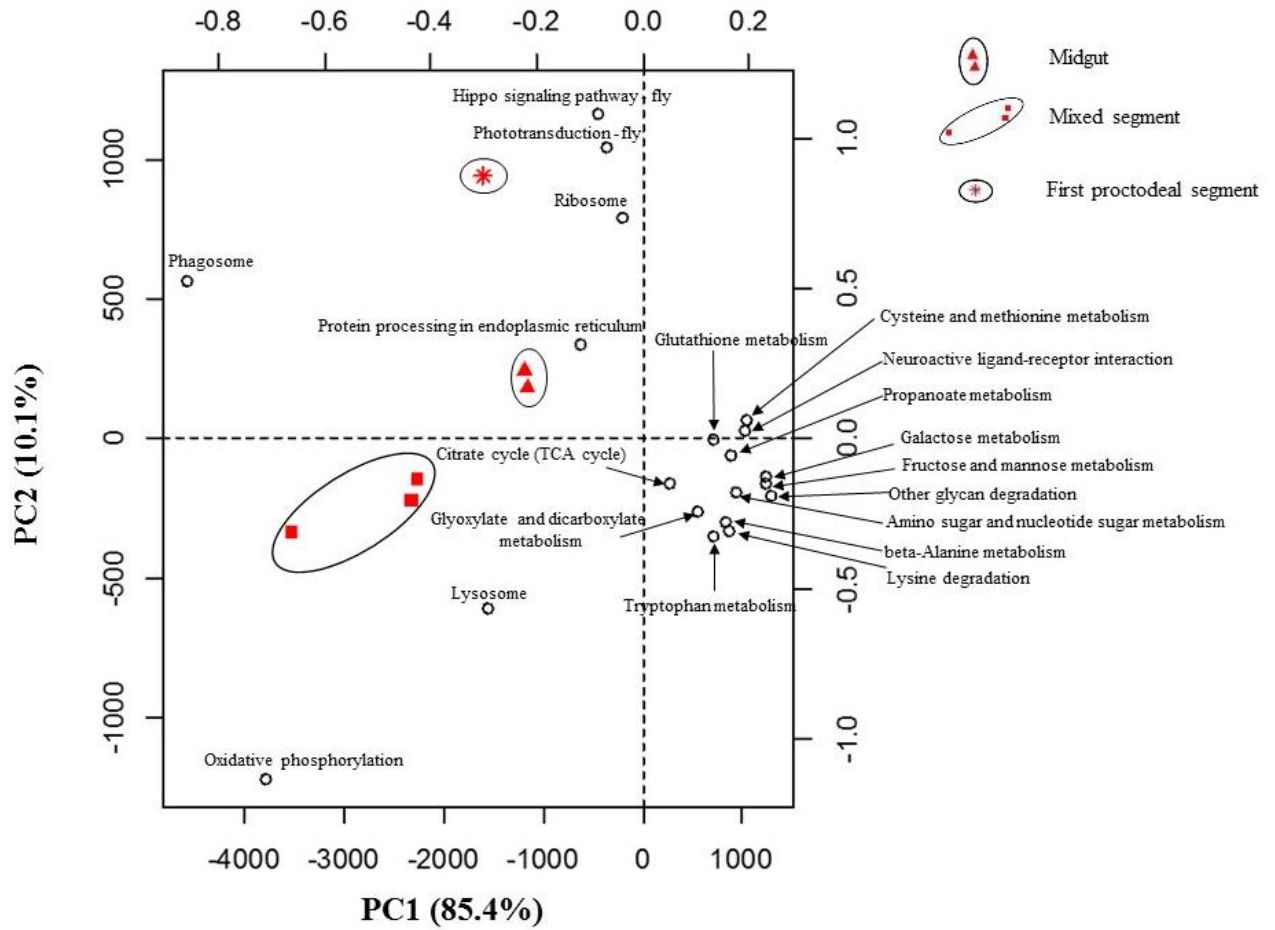
Figure 9: Principal components analysis biplot based on the normalized read counts of KEGG orthologs from the 20 reference pathways (*open circles*) having the largest coefficients of variation in the midgut (*filled triangles*), mixed segment (*filled squares*), and first proctodeal segment (*filled asterisk*) libraries. PC1 corresponds to the first principal component and PC2 to the second principal component. The percentage of variation explained by each PC is indicated between brackets.
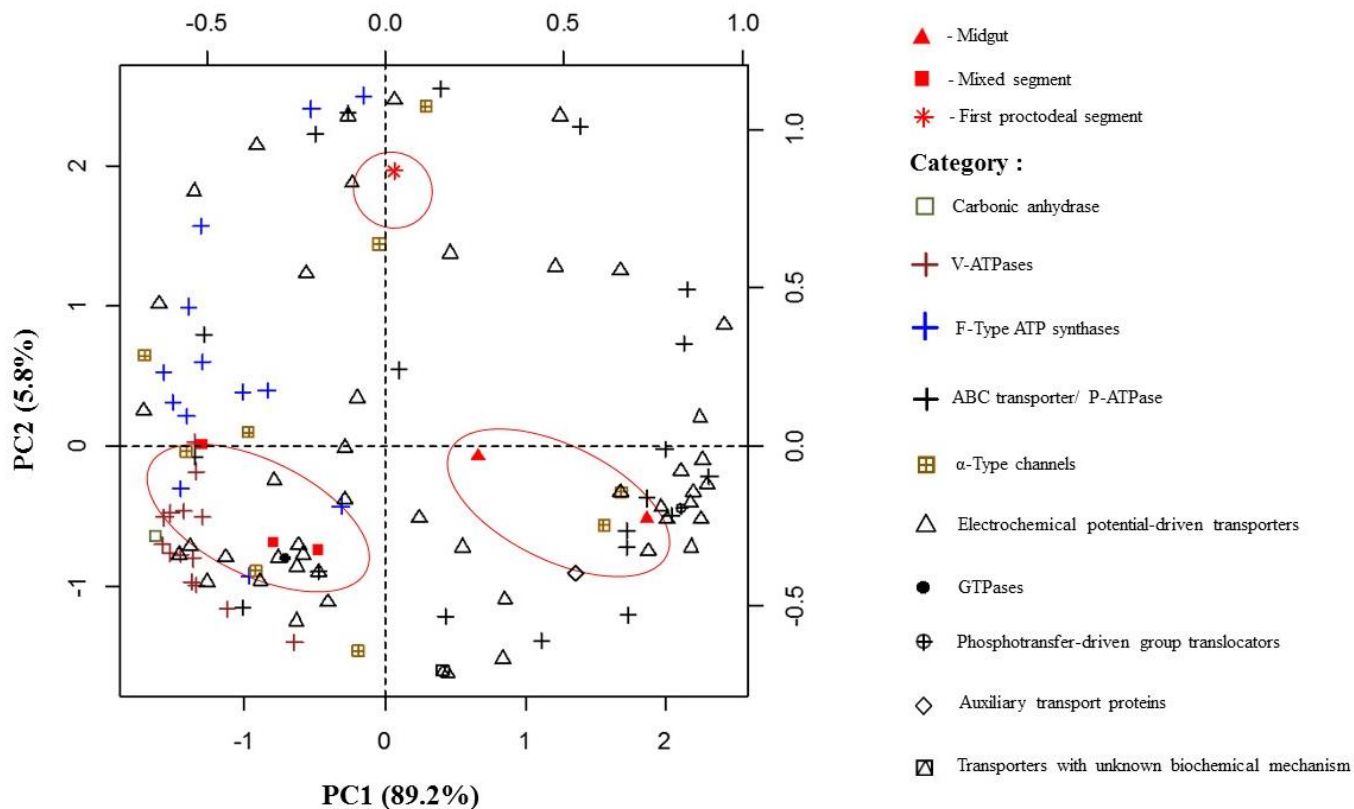
Figure 10: Principal components analysis biplot based on the normalized read counts of the 107 KEGG orthologs involved in the transport of ions and other substances, having the largest coefficients of variation in the midgut (*filled triangles*), mixed segment (*filled squares*), and first proctodeal segment (*filled asterisk*). PC1 corresponds to the first principal component and PC2 to the second principal component. The percentage of variation explained by each PC is indicated between brackets.

(a) V-ATPase

(b) NHE6/7

(c) NHE8

(d) Carbonic anhydrase
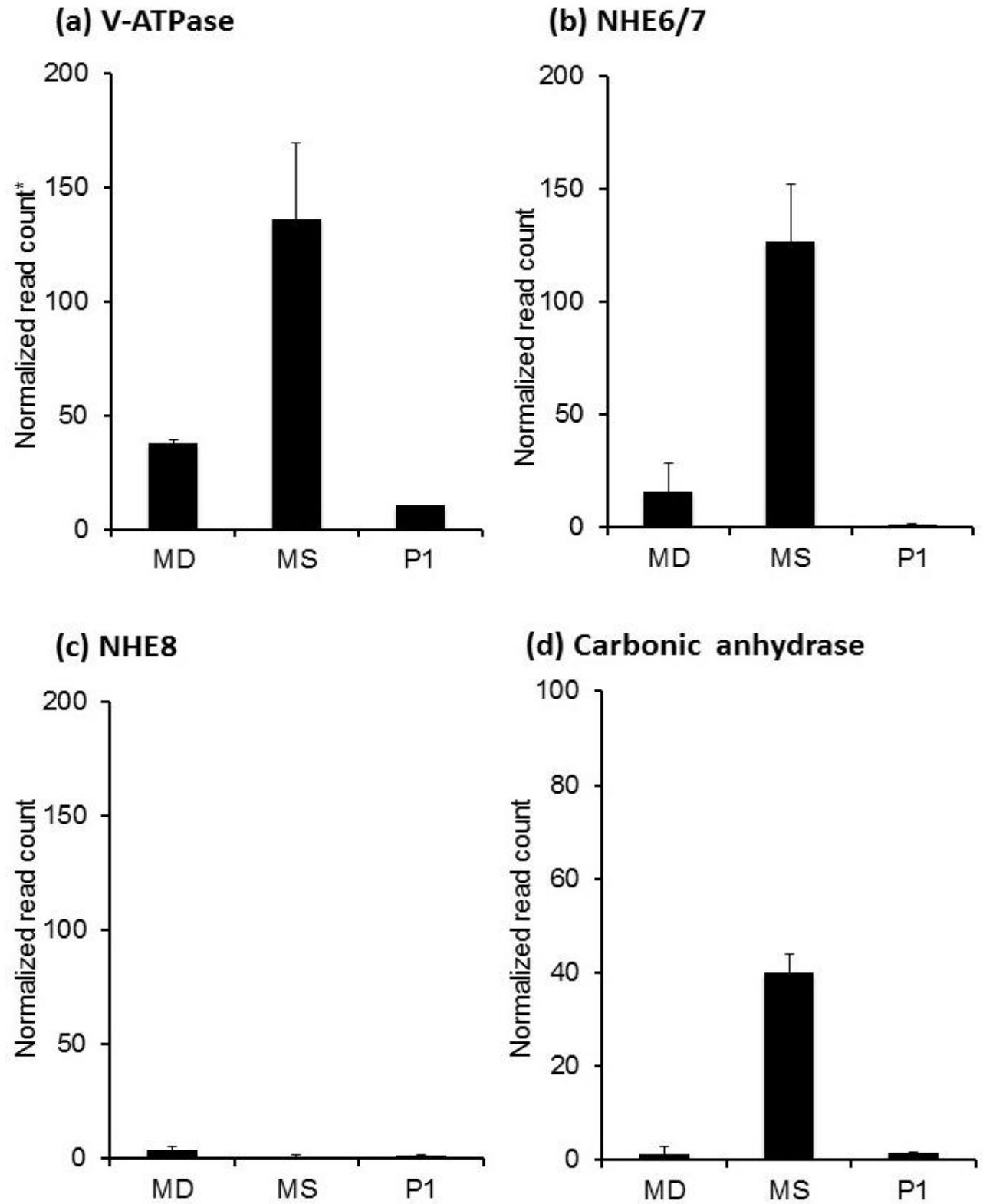
Figure 11: Average expression levels of (a) V-ATPase, (b) candidate K^+/H^+ antiporter (NHE6/7), (c) candidate K^+/H^+ antiporter NHE8, and (d) carbonic anhydrase in the midgut (MD), mixed segment (MS), and first proctodeal segment (P1), based on the normalized read counts from KEGG orthologs. Vertical lines correspond to the standard deviation. *Average normalized read counts of all subunits comprising V-ATPase.

3.3.5 Relative expression of the carbonic anhydrase orthologs, NHE6/7 and NHE8 (qRT-PCR)

In this study, I identified contigs encoding two orthologous carbonic anhydrases (isotig4607 and isotig7716), NHE6/7, and NHE8 in the transcriptome dataset. To verify that these sequences were not artifacts, the contig sequences were confirmed by Sanger sequencing. The expression of the four genes in the midgut, mixed segment, and P1 samples were evaluated using RT-qPCR (Figure 12). A single peak was observed in the melting curve of each amplicon, suggesting selective amplification. The genes encoding NHE6/7 and carbonic anhydrase (isotig4607) displayed significantly higher mRNA expression in the mixed segment than in the midgut and P1. On the other hand, gene encoding carbonic anhydrase (isotig7716) displayed significantly higher mRNA expression level in the P1 than in the midgut and mixed segment while gene encoding NHE8 was non-significant in all tissues.
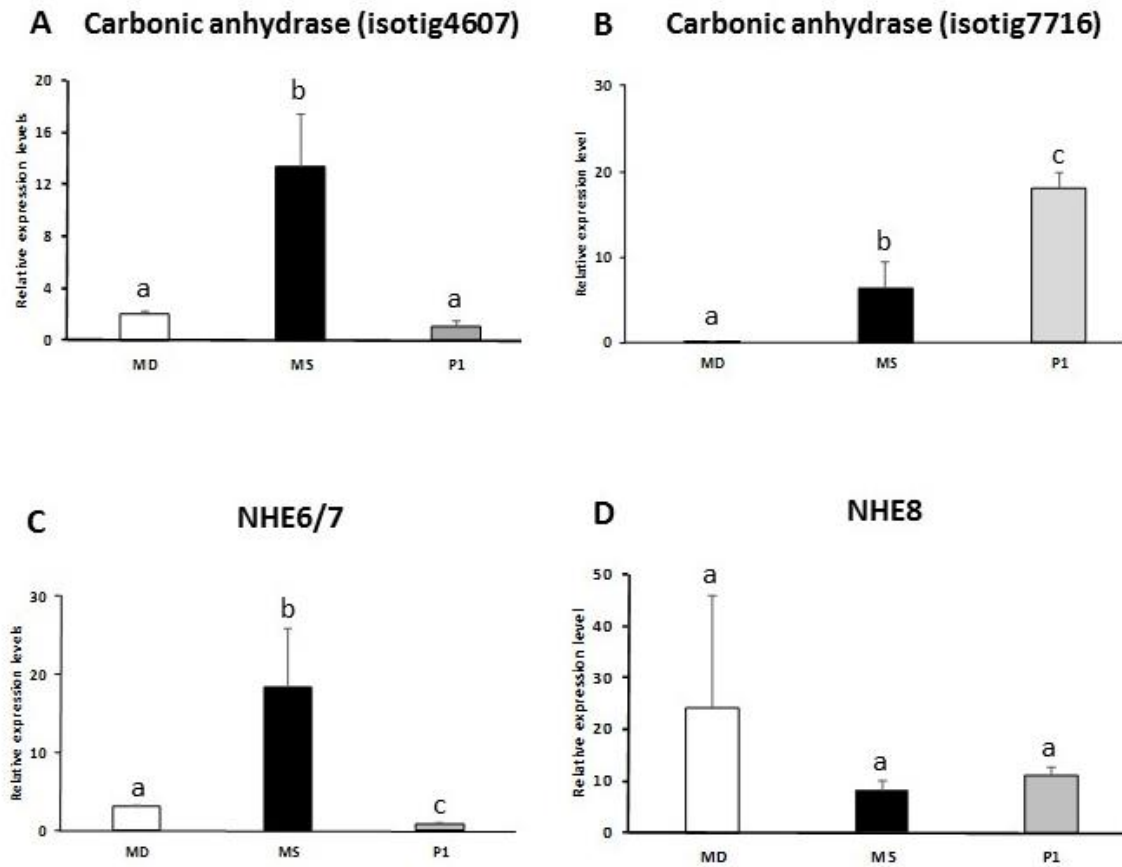
Figure 12: Relative mRNA levels of the genes encoding carbonic anhydrase (isotig4607) and isotig7716), NHE6/7 antiporter and NHE8 antiporter in the three gut segments of *N. takasagoensis*. Expression was quantified by RT-qPCR analysis with *NtNADH-dh* as a reference gene. (A) carbonic anhydrase (isotig4607); (B) carbonic anhydrase (isotig7716); (C) NHE6/7; (D) NHE8. Error bars represent the standard deviation of three biological replicates, each analyzed in duplicate. Different letters above the bars represent significant differences (one-way ANOVA followed by pairwise tests using Student's *t*-test with Benferroni correction ($p_{adj} < 0.05$). MD; midgut. MS; mixed segment, P1; first proctodeal segment.

3.3.6 Absolute quantification of the carbonic anhydrase orthologs, NHE6/7, and NHE8

The qPCR results confirmed only the relative expression profiles of the genes, but did not confirm the actual differences of expression levels between these genes. In addition, it did not demonstrate that mixed segment expresses carbonic anhydrase more dominantly than the adjacent tissues since one ortholog of carbonic anhydrase (isotig7716) was dominantly expressed in the P1 rather than the mixed segment. Based on these facts, it might give an impression that the NGS (transcriptome profiles) results (Figure 11) were not supported by qPCR. Therefore, to avoid the possible confusion, I determined the actual copy numbers of templates (each orthologs) using absolute quantitative real time PCR. The C_t values for standards were ranging from 10 to 10^9 copy numbers of NHE8, while these values were ranging from 10 to 10^8 copy numbers of other three orthologs. All the genes fell along a straight semi-log trendline with a correlation coefficient (R^2) value of 0.99 (Figure 13). The slope of the standards plot for all genes is indicated in each plot. Figure 14 summarizes the actual copy numbers obtained from genes encoding carbonic anhydrase (isotig4607, isotig7716 and sum of the both carbonic anhydrase), NHE6/7, and NHE8 in the three gut segments. It showed that the expression of NHE6/7 in the mixed segment was 15-fold higher levels than the NHE8. A combination of orthologs of carbonic anhydrase (isotig4607 and isotig7716) (*i.e.,* the sum of the copy numbers of two genes that correspond to the expression level of the carbonic anhydrase ortholog in transcriptomic dataset (Figure 11)) suggests that carbonic anhydrase was highly expressed in the mixed segment than the midgut and P1. Based on the absolute quantitative real time PCR data, the relationship of the expression pattern of all genes well supports the results obtained from the next generation sequencing.
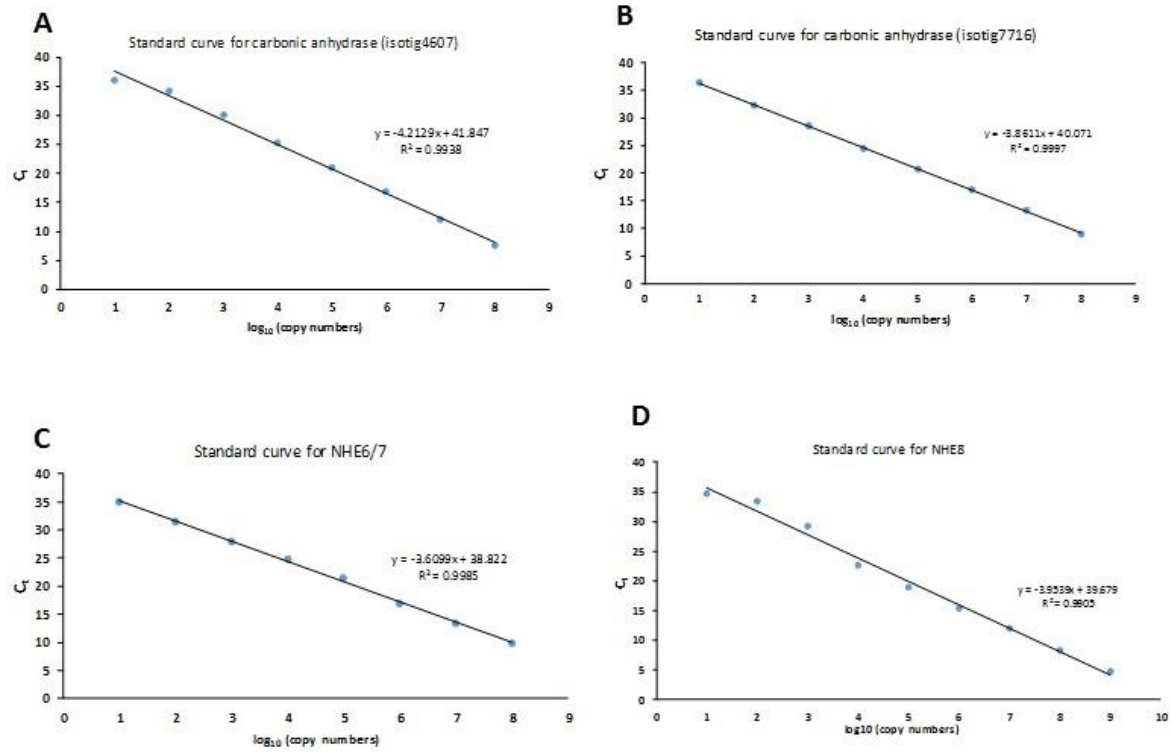
Figure 13: Standards (10-10^9) plot for genes encoding carbonic anhydrase (isotig4607 and isotig7716), NHE6/7 antiporter, and NHE8 antiporter with C_t values of standards (C_t – threshold cycle where amplification plot cuts the threshold level). (A) carbonic anhydrase (isotig4607); (B) carbonic anhydrase (isotig7716); (C) NHE6/7; (D) NHE8. R^2 value of 0.9938, 0.9997, 0.9985, and 0.9905 for carbonic anhydrase (isotig4607), carbonic anhydrase (isotig7716), NHE6/7, and NHE8, respectively. The corresponding slope is indicated in each plot.
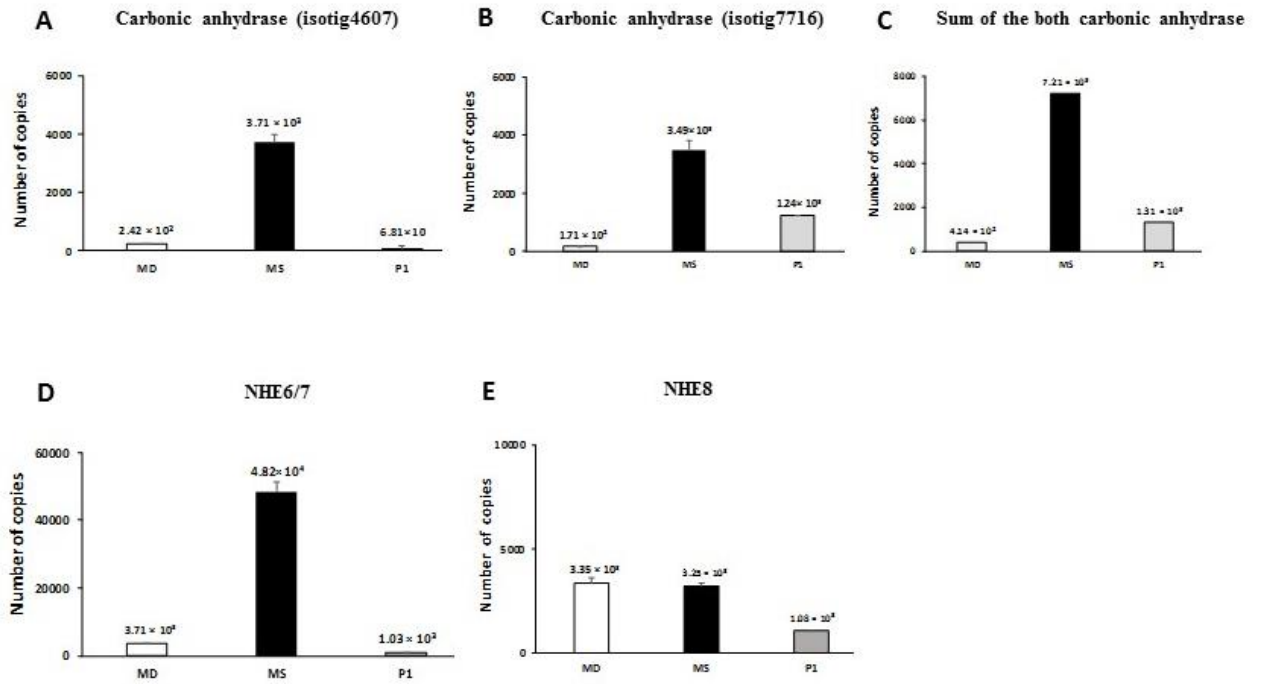
Figure 14: Summary of data obtained from absolute quantification RT-PCR. The five graphs present the actual copy numbers of the genes encoding carbonic anhydrase (isotig4607, isotig7716 and sum of the both carbonic anhydrase), NHE6/7, and NHE8 in the three gut segments of *N. takasagoensis*. (A) carbonic anhydrase (isotig4607); (B) carbonic anhydrase (isotig7716); (C) sum of copy numbers drived from the both carbonic anhydrase orthologs; (D) NHE6/7; (E) NHE8. Error bars represent the standard deviation of three biological replicates, each analyzed in duplicate. Copy numbers were normalized to the same number of tissues in each gut region. MD; midgut. MS; mixed segment, P1; first proctodeal segment.

3.4 Discussion

3.4.1 Proportion of KEGG orthologs in the mixed segment

Despite the prosperity of higher termites in the tropical forests, with their sophisticated gut structures including the mixed segment, the function of this structure remained unclear since its first description in *Apicotermes* (Grassé and Noirot 1954). This study attempted to characterize the metabolic functions occurring in the mixed segment. The transcripts detected here belonged to 125 of the 126 KEGG reference pathways identified in the CDSs of two termite species (Poulsen et al. 2014; Terrapon et al. 2014). Despite the relatively small output of the 454 GS Junior sequencer used in this study, resulting in a non-exhaustive dataset, most orthologs related to energy metabolism were detected in triplicate suggesting that the genes related to energy metabolism might be mostly expressed in the mixed segment. However, comparisons of KEGG orthologs proportions in each category between adjacent gut regions were unable to clearly discriminate the functions of the midgut, mixed segment, and first proctodeal segment.

3.4.2 Characteristics of the mixed segment based on comparisons of normalized read counts between the three gut portions

Reads' abundance distribution in the midgut, mixed segment, and first proctodeal segment libraries obtained in my previous study (Chapter 2) was in agreement with quantitative PCR results, despite the single library being sequenced from each gut region. The present study compared normalized read abundance among six libraries, including three independent libraries prepared from the mixed segment. The PCAs considering the differentially expressed KEGG orthologs in the several gut regions confirmed that energy production by oxidative phosphorylation is typical of the mixed segment, suggesting this region has a larger energy demand than the other gut regions. This is in agreement with the morphological characteristics of the mixed segment, where well-developed mitochondria are condensed in the base of the mesenteric cells (Tokuda et al. 2001). PCA results revealed that a few cellular processes, such as 'Lysosome' and 'Phagosome' activities, which are closely related to each other, are also characteristic of the mixed segment. A previous study revealed the presence of autophagic vacuoles (autophagosomes) in the mesenteric tissues of the mixed segment (Tokuda et al. 2001). It

has been suggested that lysosomes fuse with the autophagic vacuoles, with lysosomal enzymes being involved in the degradation of cytoplasmic components (Luzio et al. 2007). KEGG orthologs involved in amino acid biosynthesis were less abundant in the mixed segment (Table 12), meaning cellular components including amino acids are most likely recycled by autophagy rather than neogenesis. As carbon for amino acid biosynthesis derives from glycolysis and TCA cycle (Umbarger 1978), the predominant recycling of cellular components in the mixed segment might reduce the carbon available for amino acid biosynthesis, promoting energy production. This energy production via oxidative phosphorylation, in the presence of sufficient oxygen and sugar, is a typical behavior of normal animal cells (Warburg et al. 1927).

3.4.3 Putative gut alkalization mechanisms in termites

The first proctodeal segment of several higher termites, including *Nasutitermes*, has been referred to as highly alkaline (pH > 10) (Brune 2014), with potassium and bicarbonate secretion in the mixed segment being responsible for the gut alkalization (Bignell 1994; Kappler and Brune 1999). Similarly, to that proposed for other insects (Wieczorek et al. 2000), a V-ATPase-coupled potassium ion transport would accomplish the luminal alkalization of the termite gut; this would require a K^+/H^+ antiporter to secrete potassium from the epithelial cells, which has yet to be identified. A previous study pointed out NHE8, a protein encoded by the solute carrier (SLC) family genes, to be involved in potassium secretion (Weihrauch et al. 2011) but a study considering the Malpighian tubules of the yellow fever mosquito contradicted the participation of NHE8 in the potassium secretion (Piermarini et al. 2009). The present study revealed NHE6/7 as a good K^+/H^+ antiporter candidate as its expression profiles were more similar to those of V-ATPase than NHE8. Although named after Na^+/H^+ exchangers, some NHEs can deal with K^+, Li^+, and NH_4^+ instead of Na^+ (Donowitz et al. 2013). In mammals, NHE6 and NHE7 are often localized in organelles, but some cells possess these antiporters in the plasma membrane, exchanging Na^+ or K^+ for H^+ to decrease the acidity generated by V-ATPase (Donowitz et al. 2013). Notably, this function seems conserved throughout the evolutionary history of eukaryotes. These antiporters share their evolutionary origin with Nhx1; localized in the vacuolar membrane of yeast, it exchanges the H^+ in the vacuole for

70

K^+ in the cytosol to regulate pH (Brett et al. 2005). Thus, the NHE6/7 found in termites' mixed segment seem to be electrogenically driven by V-ATPases to increase gut lumen pH.

The other agent of gut alkalization is bicarbonate and the abundance of carbonic anhydrase read found in the mixed segment suggested protons and bicarbonate are more actively produced here than in adjacent gut regions. An ortholog of a putative HCO_3^-/Cl^- exchanger, *SLC4* (Mount and Romero 2004) was expressed in the mixed segment. The SLC4 exchanger showed structural similarities to ClC chloride channels (Bonar et al. 2013), which were also found in the present study (Table S4), suggesting the possible involvement of these channels in HCO_3^-/Cl^- exchange (Jentsch et al. 2002).

3.4.4 Relative and absolute quantification of the carbonic anhydrase orthologs NHE6/7 and NHE8

The present study revealed NHE6/7 as a good K^+/H^+ antiporter candidate, based on the result confirmed by RT-qPCR. Moreover, higher expression of NHE6/7 in the mixed segment was further confirmed with the absolute RT-qPCR that showed the actual copy numbers present in each tissue. The expression profiles further support the transcriptome expression profiles of NHE antiporters. On the other hand, RT-qPCR expression profiles of two carbonic anhydrases appeared contradicting. However, their summation of actual copy numbers showed good agreement with the transcriptome profiles (two isotigs belong to the same KEGG ortholog), confirming the idea that carbonic anhydrase is more actively expressed in the mixed segment than the adjacent tissues.

3.4.5 Possible limitations of insect metabolic pathways annotations using the KEGG database

Although the present study provided some insights into the metabolic functions in the mixed segment, there are still some issues to be elucidated. The apical surface of the mesenteric tissue in the mixed segment is associated with dense bacterial community (Tokuda et al. 2000). Comprehensive expressions of KEGG orthologs in 'Xenobiotics biodegradation and metabolism' could be related to interactions with the intestinal bacteria, but actual molecules involved in the symbiosis are still not clear. In addition, the

present study indicated that only 30% of the obtained transcripts were assignable to KEGG orthologs that allow inference of the metabolic network, suggesting a limitation of annotations with KEGG database. This was probably due to the presence of several genes with unknown function in addition to unavailability of both KEGG reference pathways and data of KEGG orthologs that are optimized to insect gut metabolisms such as digestion, absorption, cuticular formation and sclerotization, formation of peritrophic membranes, and so on. This limitation of annotations with KEGG database probably prevented from discrimination of the metabolic network between the midgut and the mixed segment, which shared more than 90% of KEGG orthologs (Figure S1) in spite of structural differences between these tissues (Tokuda et al. 2001). Thus, development of more sophisticated method for constructions of metabolic pathways is awaited to annotate more precise functions in the mixed segment.

CHAPTER IV

Transcriptome analysis on the roles of the mixed segment in the wood-feeding termite *Nasutitermes takasagoensis*

4.1 Introduction

As described in Chapter 2, I identified V-ATPase as a functional enzyme that was dominantly expressed in the mixed segment. In order to reveal the metabolic network acting in the mixed segment, transcripts were annotated based on the KEGG database, as described in Chapter 3. It showed that the predominant function of the mixed segment appeared to be energy production via oxidative phosphorylation. In addition, results revealed that a few cellular processes, such as 'Lysosome' and 'Phagosome' activities, are also characteristic of the mixed segment. The dominantly expressed V-ATPase coupled with the highly expressed NHE6/7 K^+/H^+ antiporter and carbonic anhydrases most probably contribute to increase in the luminal pH by secretion of potassium and bicarbonate ions. Since approximately 70% of transcripts were not able to be annotated by KEGG database alone in Chapter 3, it was necessary to employ other databases to annotate the obtained transcripts more comprehensively.

In this chapter, I investigated the general pattern of transcriptome profiles expressed in the mixed segment when compared to the midgut and P1 in the wood feeding termite *Nasutitermes takasagoensis*. This chapter is expected to dramatically increase in the number of known genes and their precise functions with respective to the different databases (such as Gene ontology, Pfam and CAZYmes) in this termite. For the purpose, I further annotated the cDNA libraries used in the Chapter 2 with different databases. The results of this study further supported the results obtained in the previous chapters as well as previous findings.

4.2 Materials and methods

4.2.1 cDNA library preparation, cleaning of raw data, assembly, cleaning for bacterial genes and functional annotations

Prepapration of cDNA libraries, cleaning of raw data, assembly and cleaning for bacterial genes were performed as described in Chapter 2. Functional annotations of the assembled sequences (isotigs) were made against KEGG, Pfam, dbCAN, and BLAST Nonredundant protein sequences (Nr), and Nucleotide collection (Nt) databases (DBs) of NCBI as described in Chapter 2. Briefly, annotation for Pfam (Finn et al. 2014) and dbCAN (Yin et al. 2012) was made using the HMMER- v3 based scripts provided by the developer

(default settings were used for each database). For the BLAST search, Hi-per BLAST, a parallel BLAST solution was exploited. Gene ontology (GO) annotation was made by collecting GO terms via the Pfam and KEGG annotations using GO mapping files (pfam2go and kegg2go downloaded from http://www.geneontology.org/external2go) and 'Other DBs' information included in KEGG orthology entries (Figure 4).

4.2.2 Differential expression analysis

In this study, statistical tests were designed for 'cross-tabulation' of count data (the number of reads between the midgut, mixed segment and P1). Here, I calculated P values based on Pearson's Chi squared test or Fisher exact test. To identify the significance of each gene abundance among three tissues, P values were adjusted by using 'Bonferroni correction' ($P_{adj} < 0.05$).

4.3 Results

4.3.1 General overview of the annotation

Using an e-value threshold of 10^{-9}, 59.4% (midgut), 60.1% (mixed segment), and 60.7% (P1) transcripts returned a positive hit in BLASTX searches against the GenBank Nr protein database. Among the transcripts 32.1% (midgut), 33% (mixed segment), and 36% (P1) showed significant similarity to Genbank Nt database.

4.3.2 Comparative analysis

The distributions of obtained genes (isogroups) among three tissues were evaluated using a Venn diagram (Figure 15) based on annotations by BLAST searches. A total of 2347 genes were common to all three tissues. 1031, 237, and 178 genes were shared between midgut and mixed segment, between midgut and P1, and between mixed segment and P1 respectively. In addition, a total of 206, 51, and 61 genes were detected only in midgut, mixed segment, and P1, respectively. Unique genes detected in the mixed segment were those encoding astacin-like metalloendopeptidase, acetylcholinesterase, galactoside 2-L-fucosyltransferase 1/2, E3 ubiquitin-protein ligase SIAH1, mitochondrial inner membrane protein, MFS transporter (solute carrier family 22, organic cation transporter, member 3), solute carrier family 25 (mitochondrial phosphate transporter, member 3), in addition to

75

several unknown genes (Table 13).
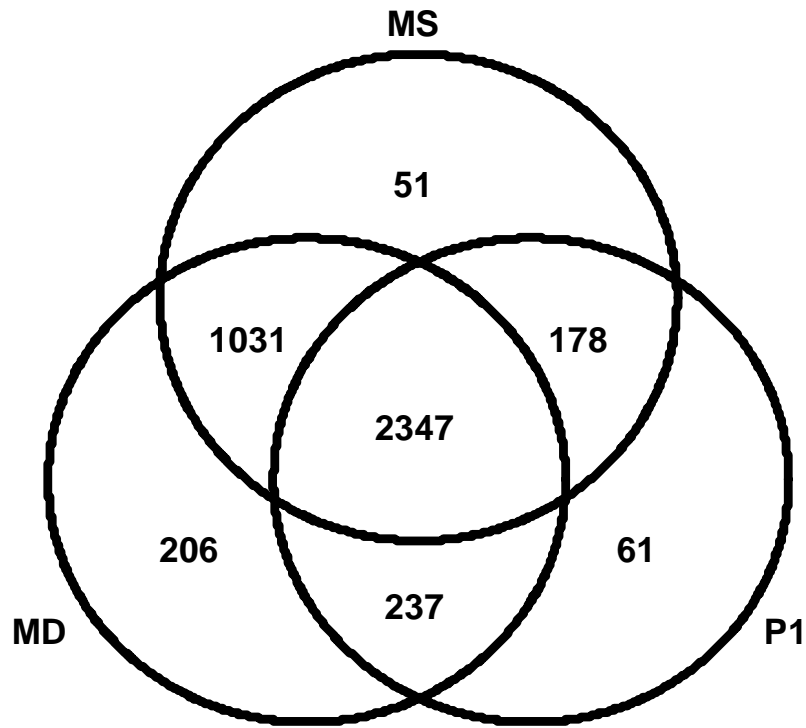
Figure 15: Venn diagram showing the comparisons of the unigenes (isogroups) among three tissues (MD, MS and P1) from *N. takasagoensis*. When at least 1 read derived from a library was assembled into an isogroup, the gene was registered as present in the library.

Table 13: List of genes identified only in the mixed segment

isogroup	KEGG_orthology	Pfam
isogroup00390	K08778 astacin-like metalloendopeptidase [EC:3.4.-.-]	PF01400: Astacin (Peptidase family M12A)
isogroup00407	K01049 acetylcholinesterase [EC:3.1.1.7]	PF00135: Carboxylesterase family
isogroup00465	NA	NA
isogroup00486	NA	NA
isogroup00524	K08778 astacin-like metalloendopeptidase [EC:3.4.-.-]	PF01400: Astacin (Peptidase family M12A)
isogroup00572	NA	NA
isogroup00613	NA	NA
isogroup01140	NA	NA
isogroup02162	K00718 galactoside 2-L-fucosyltransferase 1/2 [EC:2.4.1.69]	PF01531: Glycosyl transferase family 11
isogroup02726	NA	NA
isogroup02728	NA	NA
isogroup02743	NA	NA
isogroup02811	K07739 elongator complex protein 3 [EC:2.3.1.48]	PF04055: Radical SAM superfamily
isogroup02835	NA	NA
isogroup02937	NA	NA
isogroup03005	NA	NA
isogroup03009	NA	NA
isogroup03088	K15043 importin subunit alpha-2	PF00514: Armadillo/beta-catenin-like repeat
isogroup03095	NA	NA
isogroup03153	NA	PF08564: Cdc37 C terminal domain
isogroup03192	K00717 glycoprotein 6-alpha-L-fucosyltransferase [EC:2.4.1.68]	NA
isogroup03243	NA	NA
isogroup03328	K04506 E3 ubiquitin-protein ligase SIAH1 [EC:6.3.2.19]	NA
isogroup03498	NA	NA
isogroup03499	K03965 NADH dehydrogenase (ubiquinone) 1 beta subcomplex 9 [EC: 1.6.5.3 1.6.99.3]	PF05347: Complex 1 protein (LYR family)
isogroup03582	NA	PF09731: Mitochondrial inner membrane protein
isogroup03615	NA	NA
isogroup03654	K08778 astacin-like metalloendopeptidase [EC:3.4.-.-]	PF01400: Astacin (Peptidase family M12A)
isogroup03672	NA	NA
isogroup03685	NA	NA
isogroup03690	K08200 MFS transporter, OCT family, solute carrier family 22 (organic cation transporter), member 3	NA
isogroup03699	K14815 mRNA turnover protein 4	NA
isogroup03724	NA	NA
isogroup03825	NA	NA
isogroup03840	K00162 pyruvate dehydrogenase E1 component subunit beta [EC:1.2.4.1]	PF02779: Transketolase, pyrimidine binding domain
isogroup03842	K12845 U4/U6 small nuclear ribonucleoprotein SNU13	PF01248: Ribosomal protein L7Ae/L30e/S12e/Gadd45 family
isogroup03874	K05614 solute carrier family 1 (glial high affinity glutamate transporter), member 3	PF00375: Sodium:dicarboxylate symporter family
isogroup03932	NA	NA
isogroup03970	K04405 evolutionarily conserved signaling intermediate in Toll pathways	PF06239: Evolutionarily conserved signalling intermediate in Toll pathway

isogroup03974	NA	NA
isogroup03987	K06072 deoxyhypusine monooxygenase [EC:1.14.99.29]	PF13646: HEAT repeats
isogroup03996	NA	NA
isogroup04018	K16726 ankyrin repeat domain-containing protein 17	NA
isogroup04022	K08778 astacin-like metalloendopeptidase [EC:3.4.-.-]	PF01400: Astacin (Peptidase family M12A)
isogroup04032	NA	NA
isogroup04064	NA	NA
isogroup04081	NA	NA
isogroup04090	NA	NA
isogroup04101	NA	NA
isogroup04133	NA	NA
isogroup04138	K15102 solute carrier family 25 (mitochondrial phosphate transporter), member 3	NA

NA; not assigned

4.3.3 Gene Ontology analysis

Gene Ontology (GO) terms were assigned to a total of 2837, 2697, and 2163 genes for the midgut, mixed segment and P1 respectively. The GO terms were grouped into three main divisions: biological processes, molecular function and cellular components (Note that one gene can be associated with more than one GO term). The level 2 GO comparison between three tissues was based on the relative abundance of reads obtained from each division (Figure 16). Within the mixed segment, the GO term "catalytic activity" in the category of Molecular function was most overrepresented followed by "binding" in the same category and "metabolic process" in the category of Biological process. Among the biological process GO terms of three tissues, genes related to the metabolic process were highly abundant in the midgut. In addition, genes relevant to the cellular process, single organism process, and localization and establishment of localization were highly abundant in the mixed segment. Those for the biological adhesion, multicellular organismal process, and developmental process were highly abundant in P1. Among the cellular process GO terms, genes accounted for membrane and membrane part were highly abundant in the mixed segment, and those of cell, organelle, organelle part, and cell part were highly abundant in P1. In addition, genes relevant to the extracellular region, virion, and virion part were highly abundant in the midgut. Of the molecular function GO terms, genes involved in catalytic activity were highly abundant in the midgut, while those related to transporter activity were dominant in the mixed segment. Moreover, genes involved in binding were dominant in P1.
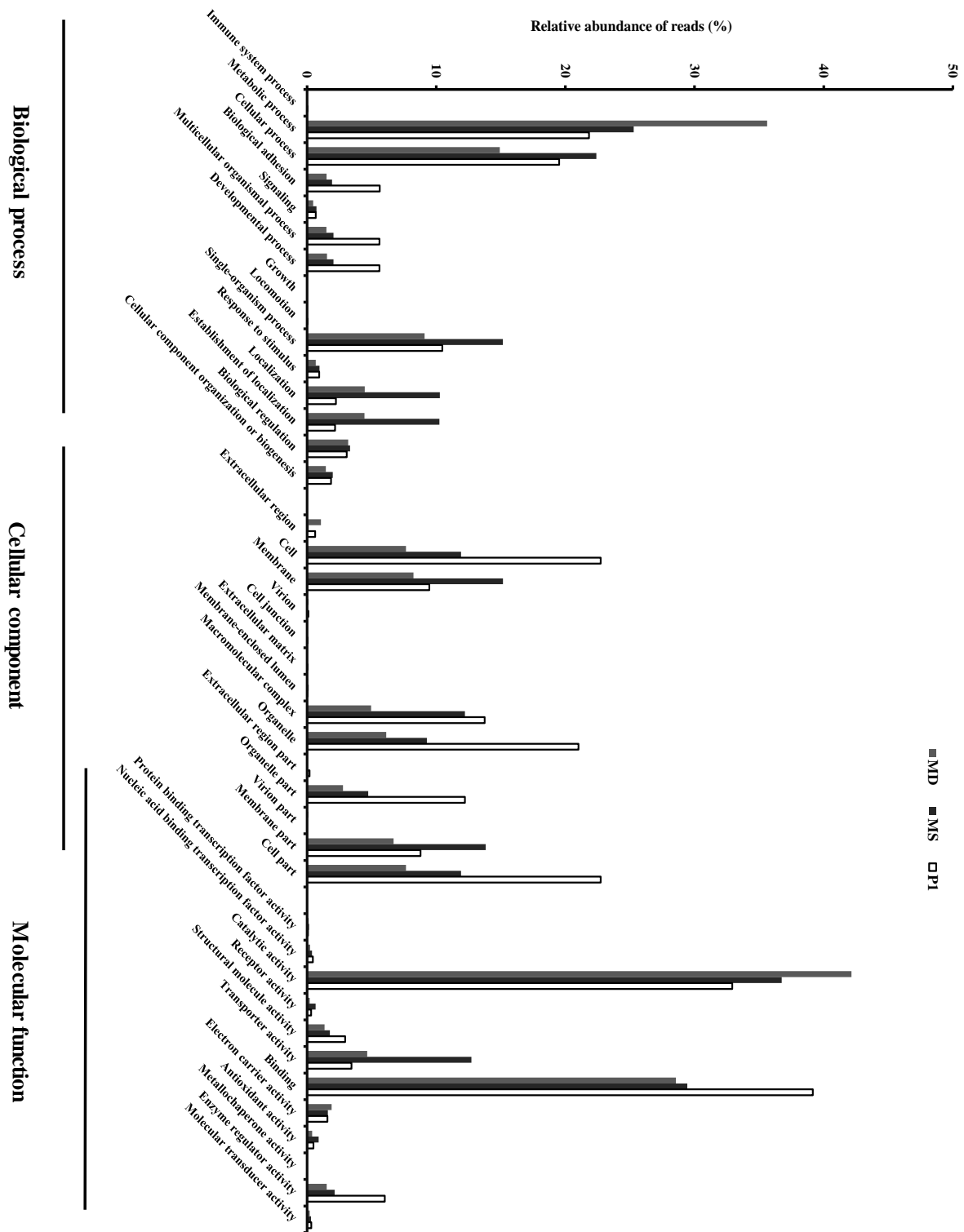
Figure 16. Gene Ontology (GO) classification of the *N. takasagoensis* transcriptomes. The results are summarized in three main GO categories: biological process, cellular component and molecular function. The level 2 GO comparison between three tissues was based on the relative abundance of reads (percentage of reads affiliated with each GO term in total number of mRNA reads). Note that one gene can be associated with more than one GO term. MD; midgut, MS; mixed segment, P1; first proctodeal segment.

4.3.4 Carbohydrate-active enzymes (CAZymes)

The assembled genes were annotated against database for automated carbohydrate-active enzyme annotation (dbCAN). As a result, 5.1% of the midgut reads were assigned to the CAZymes, while 3.1% of the mixed segment and 1.9% of the P1 reads were also affiliated with this kind of enzymes. These putative CAZymes include 12 families of glycoside hydrolases (GHs), 12 families of glycosyltransferases (GTs), one familly of carbohydrate esterases (CE), and three families of carbohydrate-binding modules (CBMs) (Table 14). The identified 12 GH families segregate into the four major functional categories of cellulases (GH1 and 9), hemicellulases (GH2, 16 and 27), chitinases (GH18 and 20) and α-carbohydrolases (GH13, 31, 37 and 38). The CAZymes were compared between three tissues were based on the relative abundance of reads. In our dataset, GHs 1, 9, 16 and 38 were highly abundant in the midgut. Approximately 0.3% of the midgut reads accounted for GH1 and 2.2% for GH9 but those were almost absent in the mixed segment (Figure 17). In addition, GH16 and GH38 were not detected from the mixed segment transcriptomes. On the other hand, 8 GHs families were represented in the mixed segment transcriptomes in low abundance (GH2, 13, 18, 20, 22, 27, 31 and 37). Of the identified CBMs, CBM13 was highly abundant in the mixed segment transcriptomes (2.1%) while CBM14 was highly abundant in the midgut (1%). In addition, the identified all CEs and GTs were present in mixed segment transcriptomes in low abundance (Table 14).
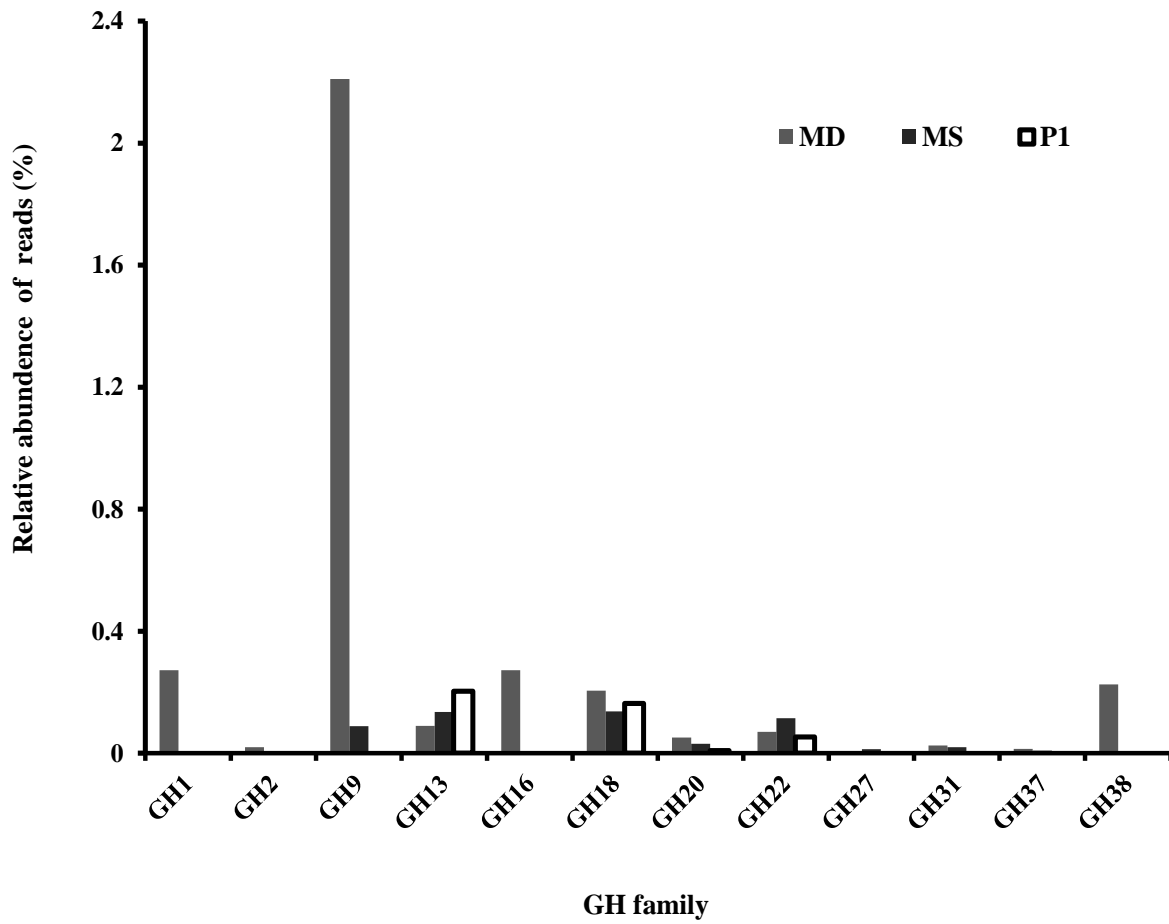
Figure 17: Comparison of relative abundance of reads (percentage of reads in total mRNA reads) affiliated with GH families among three tissues. MD; midgut, MS; mixed segment, P1; first proctodeal segment.

Table 14: Annotation of carbohydrate-active enzyme families and pfam domains in transcriptomes of *N. takasagoensis*

CAZY Familly	Pfam HMM name	Pfam description (pfam accession)	Known activities	*MD (%)	*MS (%)	*P1 (%)
CBM Family						
CBM13	Ricin_B_lectin	Ricin-type beta-trefoil lectin domain (PF00652)	Galactose-|mannose-|xylan-binding	0.30	2.10	0.20
CBM14	CBM_14	Chitin binding Peritrophin-A domain (PF01607), Sushi domain (SCR repeat(PF00084)	Chitin binding	1.00	0.10	0.40
CBM48	CBM_48	Carbohydrate binding module 48 (PF02922)	Glycogen-binding function, also found in the beta subunit (glycogen-binding) of AMP-activated protein kinases (AMPK)	0.003	0.01	0.01
GH Family						
GH1	Glyco_hydro_1	Glycosyl hydrolase family 1(PF00232)	β-Glucosidase ; β-galactosidase ; β-mannosidase, Others	0.27	0.002	0
GH2	Glyco_hydro_2_C	Glycosyl hydrolases family 2, TIM barrel domain (PF02836)	β-Galactosidase ; β-mannosidase; β-glucuronidase ; mannosylglycoprotein endo-β-mannosidase	0.02	0.01	0
GH9	Glyco_hydro_9	Glycosyl hydrolase family 9 (PF00759)	Endo-β-1,4-glucanase,endoglucanase; cellobiohydrolase ; β-glucosidase ; exo-β-glucosaminidase	2.21	0.09	0
GH13	Alpha-amylase	Alpha amylase, catalytic domain (PF00128)	α-Amylase; pullulanase ; cyclomaltodextrin glucanotransferase and related enzymes	0.09	0.13	0.20
GH16	Glyco_hydro_16	Glycosyl hydrolases family 16 (PF00722)	Xyloglucan:xyloglucosyltransferase ; keratan-sulfate endo-1,4-β-galactosidase; endo-1,3-β-glucanase, Others	0.27	0	0
GH18	Glyco_hydro_18	Glycosyl hydrolases family 18 (PF00704)	Chitinase ; endo-β-N-acetylglucosaminidase ; xylanase inhibitor; concanavalin B; narbonin	0.20	0.14	0.16
GH20	Glyco_hydro_20	Glycosyl hydrolase family 20, catalytic domain (PF00728)	β-Hexosaminidase ; lacto-N-biosidase	0.05	0.03	0.01
GH22	Lys	C-type lysozyme/alpha-lactalbumin family (PF00062)	Lysozyme type C ; lysozyme type i ; α-lactalbumin	0.07	0.11	0.05
GH27	Melibiase	Melibiase (PF02065)	α-Galactosidase ; α-N-acetylgalactosaminidase ; isomalto-dextranase ; β-L-arabinopyranosidase	0.004	0.01	0
GH31	Glyco_hydro_31	Glycosyl hydrolases family 31(PF01055)	α-Glucosidase; α-1,3-glucosidase ; sucrase-isomaltase ; α-xylosidase	0.02	0.02	0
GH37	Trehalase	Trehalase (PF01204)	α,α-Trehalase	0.01	0.01	0
GH38	Glyco_hydro_38	Glycosyl hydrolases family 38 (PF01074)	α-Mannosidase	0.23	0	0

CE Family

				MD	MS	P1
CE4	Polysacc_deac_1	Polysaccharide deacetylase (PF01522)	Chitin deacetylase	0.02	0.10	0.58

GT Family

				MD	MS	P1
GT1	UDPGT	UDP-glucoronosyl and UDP-glucosyl transferase (PF00201)	UDP-glucuronosyltransferase ; 2-hydroxyacylsphingosine 1-β-galactosyltransferase, Others	0.23	0.05	0.17
GT2	Glycos_transf_2	Glycosyl transferase family 2 (PF00535)	Cellulose synthase ; chitin synthase ; dolichyl-phosphate β-D-mannosyltransferase; Others	0.05	0.03	0.01
GT3	Glycogen_syn	Glycogen synthase (PF05693)	Glycogen synthase	0.01	0.004	0.02
GT4	DUF3542	Protein of unknown function (DUF3542) (PF12061)	Sucrose synthase ; sucrose-phosphate synthase ; α-glucosyltransferase; lipopolysaccharide N-acetylglucosaminyltransferase	0.003	0.01	0.01
GT7	Glyco_transf_7C	N-terminal domain of galactosyltransferase (PF02709)	Lactose synthase; β-N-acetylglucosaminyl-glycopeptide β-1,4-galactosyltransferase ; N-acetyllactosamine synthase	0.01	0.05	0.004
GT8	Glyco_transf_8	Glycosyl transferase family 8 (PF01501)	Lipopolysaccharide α-1,3-galactosyltransferase; UDP-Glc: (glucosyl)lipopolysaccharide α-1,2-glucosyltransferase	0.003	0.01	0.04
GT10	Glyco_transf_10	Glycosyl transferase family 10 (PF00852)	Galactoside α-1,3/1,4-L-fucosyltransferase	0.01	0.002	0
GT11	Glyco_transf_11	Glycosyl transferase family 11 (PF01531)	GDP-L-Fuc: galactoside α-1,2-L-fucosyltransferase	0	0.02	0
GT23	N/A	N/A	N-acetyl-β-D-glucosaminide, α-1,6-L-fucosyltransferase	0	0.01	0
GT32	Gly_transf_sug	Glycosyltransferase sugar-binding region containing DXD motif (PF04488)	α-1,6-Mannosyltransferase ; α-1,4-N-acetylglucosaminyltransferase	0.02	0.01	0.01
GT47	Adenylsucc_synt	Adenylosuccinate synthetase (PF00709)	GDP-Man: mannosyl-3-phosphoglycerate synthase	0.01	0.01	0.004
GT66	STT3	Oligosaccharyl transferase STT3 subunit (PF02516)	Dol-PP-α-oligosaccharide: protein β-oligosaccharyltransferase	0.02	0.02	0.03

*Relative abundance of reads affiliated with each CAZymes among three tissues is indicated. MD; Midgut, MS; Mixed segment, P1; First proctodeal segment. N/A, not available

4.3.5 Protein domains

A Pfam domain search yielded 2461 (51.1%), 2258 (49.5%) and 1782 (49.1%) transcripts in the midgut, mixed segment and P1 respectively. Based on the relative abundance of reads, I identified abundant Pfam domains in *N. takasagoensis* transcriptomes in each tissue (Table 15). Among the identified Pfam domains, haemolymph juvenile hormone binding protein (JHBP) (PF06585), insect allergen related repeat, nitrile-specifier detoxification (PF06757), glycosyl hydrolase family 9 (PF00759) and trypsin (PF00089) were abundant four in the midgut. Ricin-type beta-trefoil lectin domain (PF00652), ATP synthase subunit C (PF00137), astacin [Peptidase family M12A based on MEROPS (Rawlings et al. 2012)] (PF01400) and ATP synthase alpha/beta family, nucleotide-binding domain (PF00006, PF00306) were abundant four in the mixed segment. Reverse transcriptase (RNA-dependent DNA polymerase) (PF00078), actin (PF00022), myosin head (motor domain) and myosin tail (PF00063, PF01576) and myosin tail (PF01576) were abundant four in P1 (Table 15).

Table 15: Most abundant (5) Pfam domains identified from transcriptomes in the midgut, mixed segment and P1 of *N. takasagoensis*

Isogroup	Pfam description	Relative abundance of reads (%)		
		MD	MS	P1
MD				
isogroup02074	PF06585: Haemolymph juvenile hormone binding protein (JHBP)	4.09	0	0.004
isogroup00030	PF06757: Insect allergen related repeat, nitrile-specifier detoxification	3.14	0	0.004
isogroup00005	PF00759: Glycosyl hydrolase family 9	2.13	0.08	0
isogroup00027	PF00089: Trypsin	1.70	0	0
isogroup02056	PF00089: Trypsin	1.41	0	0
MS				
isogroup00230	PF00652: Ricin-type beta-trefoil lectin domain	0.22	1.72	0.10
isogroup00829[**]	PF00137: ATP synthase subunit C	0.28	1.54	0.07
isogroup00011	PF01400: Astacin (Peptidase family M12A)	0.04	1.46	0.01
isogroup00672[**]	PF00006: ATP synthase alpha/beta family, nucleotide-binding domain <> PF00306: ATP synthase alpha/beta chain, C terminal domain	0.20	1.16	0.03
isogroup01106	PF00153: Mitochondrial carrier protein	0.11	0.95	0.24
P1				
isogroup00001	PF00078: Reverse transcriptase (RNA-dependent DNA polymerase)	1.34	1.87	5.58
isogroup01077	PF00022: Actin	0.05	0.19	3.53
isogroup00012	PF00063: Myosin head (motor domain) <> PF01576: Myosin tail	0.03	0.16	2.86
isogroup00647	PF01576: Myosin tail	0.02	0.17	2.11
isogroup01081	PF13405: EF-hand domain	0.03	0.07	1.42

[**]These isogroups encodes V-ATPase subunits (see Table 4 for more information)

MD; Midgut, MS; Mixed segment, P1; First proctodeal segment

Names of gene that is considered chimeric are separated by "<>"

4.4 Discussion

This chapter represents the first report to elucidate the gene profiles acting in the mixed segment of wood feeding higher termite *N. takasagoensis*. The transcriptome analyses of the midgut, the mixed segment and P1 of *N. takasagoensis* have revealed potential differences in gene expressions and begun to provide insights into some fundamental physiological processes occurring in these tissues.

Comparative analysis showed that most genes were shared between the midgut and the mixed segment when compared with those between the midgut and P1, and between the mixed segment and P1. However, there were significant functional variations among the three tissues. On the other hand, there were also large variances in gene compositions in the three tissues. It is consisting with substantial differences of the mesenteric component showed in the mixed segment and midgut (Tokuda et al. 2001). Thus, I suggest that mixed segment is unique gut region present in the higher termites and it has a different function when compared with the midgut and P1.

GO analysis showed that genes involved in cellular process, single organism process, localization, establishment of localization, membrane, membrane part, and transporter activity were highly represented in the mixed segment. Those GO terms were represented mostly similar set of genes and KEGG pathways (Chapter 2; Table 4). Specifically, transcripts encode ATP synthases, V-ATPases, and many transporters were represented. The above gene expression pattern is also consisting with Pfam annotation. On the other hand, oxidative phosphorylation, phagosome, lysosome, protein processing in endoplasmic reticulum and endocytosis pathways were mostly represented in above mentioned GO categories. These categories supported the conclusion of the previous chapter that mixed segment is responsible for higher energy production, cellular renovation, and active ion transport when compared to the midgut and P1.

A comparison of the cDNA abundance of each GH and other CAZymes detected in this study revealed an uneven distribution pattern among three tissues (Table 14). As elaborated in Figure 17, GH9 and GH1 were almost absent in the mixed segment. It is corresponding to the previously reported cellulase activities in the gut (Tokuda et al. 2004, 2005) and further indicated that mixed segment is not involved in the cellulose digestion. It is noteworthy that the mixed

segment lumen may have a small amount of sugar since my data suggested that a minute amount of different CAZymes could be able to produced by the mixed segment itself. Thus, it is possible that the mixed segment bacteria may utilize readily available sugars from the lumen very quickly to fulfill their nutrient requirements.

This study represents the first report to elucidate the gene expression profiles acting in the mixed segment of the wood-feeding termite *N. taksagoensis* and have begun to provide insights into some fundamental physiological processes occurring in these tissues.

CHAPTER V

General discussion

The prime objectives of this research were to obtain robust pools of host genes and metabolic pathways specific to the mixed segment of *Nasutitermes takasagoensis*. It is also essential to understand the roles of the midgut and P1 to analyze the functions of the mixed segment. My transcriptome analysis clearly suggests that genes related to energy metabolism and ion transport was mostly expressed in the mixed segment than the midgut and P1 and further confirmed the proposed role of the mixed segment as a transporting epithelium (Bignell et al. 1983). Here, I elaborate how my findings contribute to the cellulolytic and metabolic environment of such gut regions to enhance our understanding of the digestive physiology of termites. This section provides an overview of the major outcomes from the research described in the thesis. Specific results have been discussed in detail in their respective chapters.

5.1 Sequencing overview of the cDNA libraries

In this study, I compared transcriptomic data of the midgut, the mixed segment and P1 of *N. takasagoensis*. Six cDNA libraries (three mixed segments, two midguts and one P1) were created for these purposes. High-quality massive metatranscriptomic datasets were generated through 454 pyrosequencing.

The gut of *Nasutitermes* spps. revealed a highly structured microenvironment, with differences in metabolic activities and microbial community structure (Brune 2014, Köhler et al. 2012). In this context, complete elimination of prokaryotic RNA when enriched with poly-A$^+$ RNA is a critical obstacle since prokaryotic genes are recognized to be polyadenylated shortly (Sarkar 1997). In this investigation, I identified a strong prokaryotic homology for numerous genes (Chapter 2; Table 3 and Chapter 3; Table 8). In contrast, it is also possible that prokaryotic symbiotic genes have assimilated into host genomes (Hotopp et al. 2007). In this study, the prokaryotic sequences identified are removed by a homology search using blastn (*i.e.,* nucleotide search) against NCBI Nt database with a threshold *e*-value of 10^{-9}. However, assimilation of prokaryotic genes during poly-A$^+$ RNA purification also a major concern in future studies.

5.2 Vacuolar (H⁺)-ATPase (V-ATPase) functions in the mixed segment

It is noteworthy that V-ATPase subunits are amongst the most dominantly detected genes in the mixed segment and qPCR experiments confirmed this idea (Chapter 1; Figure 5 and Figure 7). In addition, the distribution of V-ATPase enzyme activity among the midgut, the mixed segment and P1 was also in good agreement with the expression pattern. This enzyme has been identified in a variety of intracellular compartments in eukaryotic cells, including clathrin-coated vesicles, endosomes, lysosomes, Golgi apparatus-derived vesicles, chromaffin granules, synaptic vesicles, multivesicular bodies, and the central vacuoles of yeast and plants (Cidon and Nelson 1983; Dean et al. 1984; Forgac et al. 1983; Harikumar and Reeves 1983; Ohkuma et al. 1982; Xie et al. 1983), and functions to acidify vacuolar compartments (Kane and Stevens 1992; Bowman et al. 1992; Gluck 1992; Forgac 1992 and 1997). This acidification plays an important role in a number of cellular processes including receptor-mediated endocytosis, intracellular targeting of newly synthesized lysosomal enzymes, macromolecular processing and degradation, and coupled transport of small molecules (Forgac 1989). Moreover, V-ATPases carried out several important functions in the plasma membrane of specialized cells (Brown et al. 1987), including steering ion transport and ionic balance (Wieczorek 1992).

The physiological basis of alkalinity in the gut lumen of the mixed segment of *N. takasagoensis* is not known, but likely involves an ATP-driven ion transport as documented in other insects (Moffett and Onken 2010; Wieczorek et al. 1991, 1992 and 2000; Boudko et al. 2001). Moreover, it has been suggested that V-ATPases were involved in K⁺ ion transport in many insect biological systems, which cause alkalinity (Wieczorek et al. 1991; Onken et al. 2008). On the other hand, it has already been revealed that V-ATPase is abundant in eukaryotic cells rich in mitochondria (Brown et al. 1987 and 1996; Harvey et al. 1983). The mixed segment of *N. takasagoensis* harbors numerous mitochondria in the mesenteric epithelial cells and those were localized in the basal part of the columnar cells (Tokuda et al. 2001). These morphological features are compatible with highly abundant ATP synthase genes detected in the mixed segment libraries. Further, our transcriptome comparison between six libraries using PCA showed that ATP synthase was clearly separated with mixed segment libraries (Chapter 3; Figure 10). In addition, typical justification for the mitochondria-richness and abundant ATP synthase genes of

these cells is that it provides abundant ATP for the dense arrays of V-ATPases. As described in Chapter 1, V-ATPases coupled with K^+ ion transport would accomplish the luminal alkalization of other insects; this would probably require a K^+/H^+ antiporter to secrete potassium from the epithelial cells (Wieczorek et al. 1991; Onken et al. 2008). Thus, the NHE6/7 found in the mixed segment of this termite seems to be electrogenically driven by V-ATPases to increase gut lumen pH and qPCR experiments confirmed the higher expression of NHE6/7 antiporter in the mixed segment than other two tissues.

On the other hand, the other agent of gut alkalization is bicarbonate. The higher read abundance of carbonic anhydrase found in the mixed segment, is further confirmed by qPCR experiments, suggesting that protons and bicarbonate are more actively produced here than in adjacent gut regions, where carbonic anhydrase catalyzes the reversible hydration of carbon dioxide to bicarbonate (Corena et al. 2004). It was reported that higher alkalinity in the P1 of higher termites arose from potassium and bicarbonate secretion in the mixed segment (Bignell 1994; Kappler and Brune 1999). As suggested by mosquito larvae, apical anion exchangers across apical membranes were energized by basally located V-ATPases, contributing to luminal alkalization (Boudko et al. 2001). There could be a similar mechanism acting in the mixed segment.

If the V-ATPase in the mixed segment is coupled with the potassium secretion, it should be localized on a luminal surface or basal membrane of the tissue. Based on the morphological features observed in the wood-feeding termite *N. takasagoensis*, it may render the possibility to localize the V-ATPases in the basal membrane like mosquito larvae. Further studies using immunohistochemistry are necessary to determine localization of V-ATPase enzymes in the mixed segment of this termite.

5.3 Metabolic pathways acting in the mixed segment

My study showed that genes responsible for energy production via oxidative phosphorylation pathway were highly abundant in the mixed segment. The results were further confirmed by comparison with phylogenetically distinct two genomic data published recently (Poulsen et al. 2014; Terrapon et al. 2014). Based on the frequency of KEGG orthologs in three MS libraries and in the termite genomes showed that average number of KEGG orthologs detected in the three MS libraries was approximately 55.4% of those in the termite genomes. Given that the reduced number of KEGG orthologs in MS libraries compared to the genomes was simply due to the small output of our DNA sequencer (*i.e.,* 454 GS junior), detection of a similar ratio of KEGG orthologs (*i.e.,* 55.4%±6.6%) was expected in every category. Compared to this average ratio (55.4%±6.6% derived from total KEGG orthologs detected in the mixed segment to the genomes), the percentage of detectable KEGG orthologs involved in 'Energy metabolism' in the mixed segment was much higher (*i.e.,* 81.8%±12.1%) (Chapter 3; Table 11), suggesting comprehensive expressions of genes that belong to this category.

It is well-known that mitochondria in the cells are responsible for energy production with ATP through TCA cycle and electron transport chain. It is consisting with the highly abundant and well developed mitochondria present in the mesenteric epithelium of the mixed segment (Tokuda et al. 2001). A combination of endogenous GH9 and GH1 cellulolytic activities ensures the production of glucose in the midgut lumen of *N. takasagoensis* (Tokuda et al. 2012), but either none or only trace amount of glucose might be present in the mixed segment lumen due to a trace of these transcripts in the mixed segment. On the other hand, it has previously been suggested that glucose molecules produced in the midgut are rapidly absorbed across the midgut wall to be catabolized by glycolysis in the host (Tokuda et al. 2005; Tokuda et al. 2012). Indeed, glucose metabolism via glycolysis in the termite midgut was demonstrated by a metabolomics analysis using *Hodotermpsis sjostedti* (Tokuda et al. 2014). This prompted us a few issues to be addressed such as contribution of the mixed segment to perform the glycolysis; participation of glucose and/or another nutrient source in the glycolysis pathway in the mixed segment. Although almost all genes of glycolysis and TCA cycle pathway were existed in the transcriptome dataset of the mixed segment, the absence of well-developed microvilli in the mesenteric tissue of this

region suggests less active absorption from the gut lumen than the midgut (Tokuda et al. 2001). I assume that glucose molecules enter to the mixed segment via insect haemolymphs and catabolized by glycolysis since insect haemolymphs play a part in the insect nutrition (Nelson 2003). On the other hand, it is possible that the tissue assimilates trehalose from the haemolymph to perform the glycolysis in the mixed segment since gene encoding Tret1, a homologue of Trehalose transporter (Kanamori et al. 2010; Kikuta et al. 2012) was detected in the mixed segment tissue. In addition, gene encoding α,α-trehalase was also detected in the mixed segment tissue which hydrolyze trehalose into glucose.

PCA analysis showed that genes responsible for phagosome and lysosome pathways were highly abundant in the mixed segment, while these pathways are closely related to each other. Highly represented phagosome pathway was consisting with the presence of large autophagic vacuoles observed in the epithelial cells of the mixed segment (Tokuda et al. 2001). The significance of these autophagic vacuoles in the mixed segment is unknown. However, the presence of autophagic vacuoles and the retention of regenerative nidi (Tokuda et al. 2001) are evidence of a cellular renovation. Moreover, autophagy is the major intracellular degradation system, and it delivered cytoplasmic materials to the lysosome and formed phagolysosomes (Luzio et al. 2007). In addition, autophagy produces new building blocks and energy for cellular renovation and homeostasis (Noboru et al. 2011). Collectively, the presence of large autophagic vacuoles suggests a quick turnover of the mesenteric cells in the mixed segment. Moreover, it will further supply energy and new building blocks to perform cellular functions in the mixed segment.

In contrast, KEGG orthologs related to 'Signaling molecules and interaction' and 'DNA replication and repair' in the mixed segment were scarce, suggesting that these categories do not represent the major functions in the mixed segment. In addition, since 16 of 125 pathways were mapped with only two or less KEGG orthologs (Chapter 3; Table 9), these pathways might not act in the mixed segment or play a relatively simple role like an interconversion between riboflavin and FAD as presumed in 'Riboflavin metabolism'.

Of the pathways identified in the three tissues, it was quite interesting to note that V-ATPase genes were shared among oxidative phosphorylation, phagosome and lysosome

pathways. Oxidative phosphorylation pathway produces ATPs in the mitochondria and it is characteristic in the mixed segment. On the other hand, ATP consumption should also characteristic in the mixed segment since V-ATPase is the most abundant ATPase identified in the mixed segment. A model proposed by Hochachka (2003) that shows ATP consumption stimulates ATP production in cells (Hochachka 2003). Thus, we suggest that highly abundant V-ATPase plays a key role in the mixed segment. It was found that ATP hydrolysis energies the formation of proton electrochemical gradients, which, in turn, energies ATP synthesis (Harold 1986). This proton-motive forces from H^+ V-ATPases generate transmembrane voltages and create membrane potential difference (Harvey and Wieczorek 1997) which further facilitates the secondary ion and solute transport across the membranes and other cellular processes in the mixed segment.

5.4 Carbohydrate-active enzymes in the mixed segment

Presence of diverse CAZymes in the transcriptome libraries is probably associated with the ability of the wood-feeding termite to thrive solely on a recalcitrant wood diet. As elaborated in Figure 17 (Chapter 4), GH9 and GH1 were highly abundant in the midgut. The termite *N. takasagoensis* can endogenously produce both endo-β-1,4-glucanases (GH9) and β-glucosidases (GH1) in the midgut. Moreover, it was previously described that more than 60% of the total cellulase activity is localized in the midgut (Tokuda et al. 2004, 2005 and 2012; Tokuda and Watanabe 2007). GH9 has been confirmed to actively digest cellulose to produce cellobiose and cellotriose (Zhang et al. 2009, 2011; Tokuda et al. 2012), which were further hydrolyzed to glucose by GH1 (Zhang et al. 2010; Tokuda et al. 2012). The mixed segment showed only a trace amount of cellulase activity (Tokuda et al. 1997, 2005), while P1 cellulase activity is negligible. The absence of cellulases in P1 is probably due to non-secretive characteristics of the epithelium such as its thin cuboidal tissue with a thick cuticle layer and denaturation by strong alkalinity (Tokuda et al. 2001). In consistent with the previous findings and observations, my transcriptomic data also showed similar expression patterns in GH9 and GH1. In addition to the presence of GH9 and GH1, I have identified putative hemicellulases, chitinases, and α-carbohydrolases, which could disrupt the compact structure of lignocelluloses and increase the

accessibility of cellulolytic enzymes. It might increase the digestibility of the lignocellulose substrate. These findings unveil the contribution of the midgut cellulolytic system to the higher termite *N. takasagoensis.* It has been reported that the presence of diverse cellulases and hemicellulases in the hindgut microbiota of *Nasutitermes* spp. (Warnecke et al. 2007; Mikaelyan et al. 2014). On the other hand, lower termites produce both GH9 and GH1 in the salivary glands. In addition, compared to the endogenous cellulases produced in lower termites, protistan symbionts produce more powerful cellulases such as β-glucosidases (GH3), endo-glucanases (GHF5, 7, 45), and cellobiohydrolases (GHF7). These protist symbionts also produce hemicellulolytic enzymes (Ni and Tokuda 2013).

5.5 Relative and absolute quantification of expression profiles of the V-ATPases, carbonic anhydrase orthologs, NHE6/7 and NHE8

Relative quantification of genes is commonly carried out to normalize the expression levels of the genes of interest with a suitable internal control gene. In this study, gene expression analysis was carried out using relative quantification RT-PCR. To obtain a better normalization for the relative quantification, in this study, I have evaluated four reference genes as described in Chapter 2. Based on the evaluation, the expression level of *NtNADH-dh* was identified as the most stable among three tissues and used as the internal control gene in Chapter 2 and Chapter 3. On the other hand, the absolute quantitative RT-PCR method relies on a standards constructed from known concentrations of serially diluted standards to measure the actual copy numbers of a particular target gene and corresponding levels of RT-PCR data without normalized to any internal control gene. Therefore, to obtain all biological information without any depreciation and without unnecessary arguments (Leong et al. 2007), measurement of a copy number of a target gene is considered to be more informative and reliable for comparisons. The relative expression profiles of genes encoding carbonic anhydrase (isotig4607 and isotig7716), NHE6/7 and NHE8 were the typical example for the above explanation. The relative expression profiles obtained from these genes (Chapter 3; Figure 12) only confirmed the expression profiles in each tissue since it did not confirm the expression profiles between genes. The direct quantification of

absolute copy number of a particular target gene in a sample, thus decreasing potential sources of misunderstandings normally occurred during gene comparisons.

5.6 Concluding remarks and future perspectives

Figure 18 summarizes the major metabolic network inferred for the mixed segment based on the data acquired so far. Considering that thick cuticular layers covering the thin cuboidal cells of the proctodeal tissue might prevent their interaction with luminal substances (Tokuda et al. 2001), the majority of the results obtained here might derive from the mesenteric tissue of the mixed segment. Compared to the adjacent gut regions, the predominant function of the mixed segment appeared to be energy production via oxidative phosphorylation, with cellular components being recycled by a lysophagosomal mechanism rather than by neogenesis. This allows a higher energy production using the available carbon. The dominantly expressed V-ATPase, an active proton pump, probably consumes a large energy amount and is electrogenetically coupled with the NHE6/7 K^+/H^+ antiporter to increase the luminal pH by secretion of potassium ions. The mixed segment was also characterized by a high expression of carbonic anhydrases that produce HCO_3^-, the other agent of the gut alkalization. These mechanisms might promote the solubilization of lignocellulose during its passage to the hindgut. It is generally thought that the solubilization of lignocellulose is beneficial to termites for further degradation (Bignell 2000), while its physiological basis in wood-feeding termites is still obscure. It is presumable that an alkaline pH might promote autoxidation of polyphenolic compounds that loosen lignocellulosic structure and enhance accessibility of bacterial cellulolytic enzymes in the hindgut (Tokuda and Watanabe 2007). Although the mixed segment has been assumed to be engaged in excretion, several transporters were more characteristic to the midgut or P1. Probably, there are distinct roles of transport between these tissues. Still, knowledge on the functions of the mixed segment are limited, and further attempts to elucidate the metabolic function occurring on this gut region will allow an understanding of how the mixed segment contributed to the successful prosperity of higher termites on plant biomass in tropical ecosystems (Lo and Eggleton 2011).

In conclusion, the present study showed the transcriptomic analyses of the mixed segment for the first time, which is unique in higher termites. In this thesis, I provided novel knowledge on the major enzymes, transporters, metabolic pathways as well as functional classifications of genes expressed in the mixed segment. I hope that my results presented in this thesis would greatly contribute to further understanding of relevant digestive mechanism and physiology in termites.
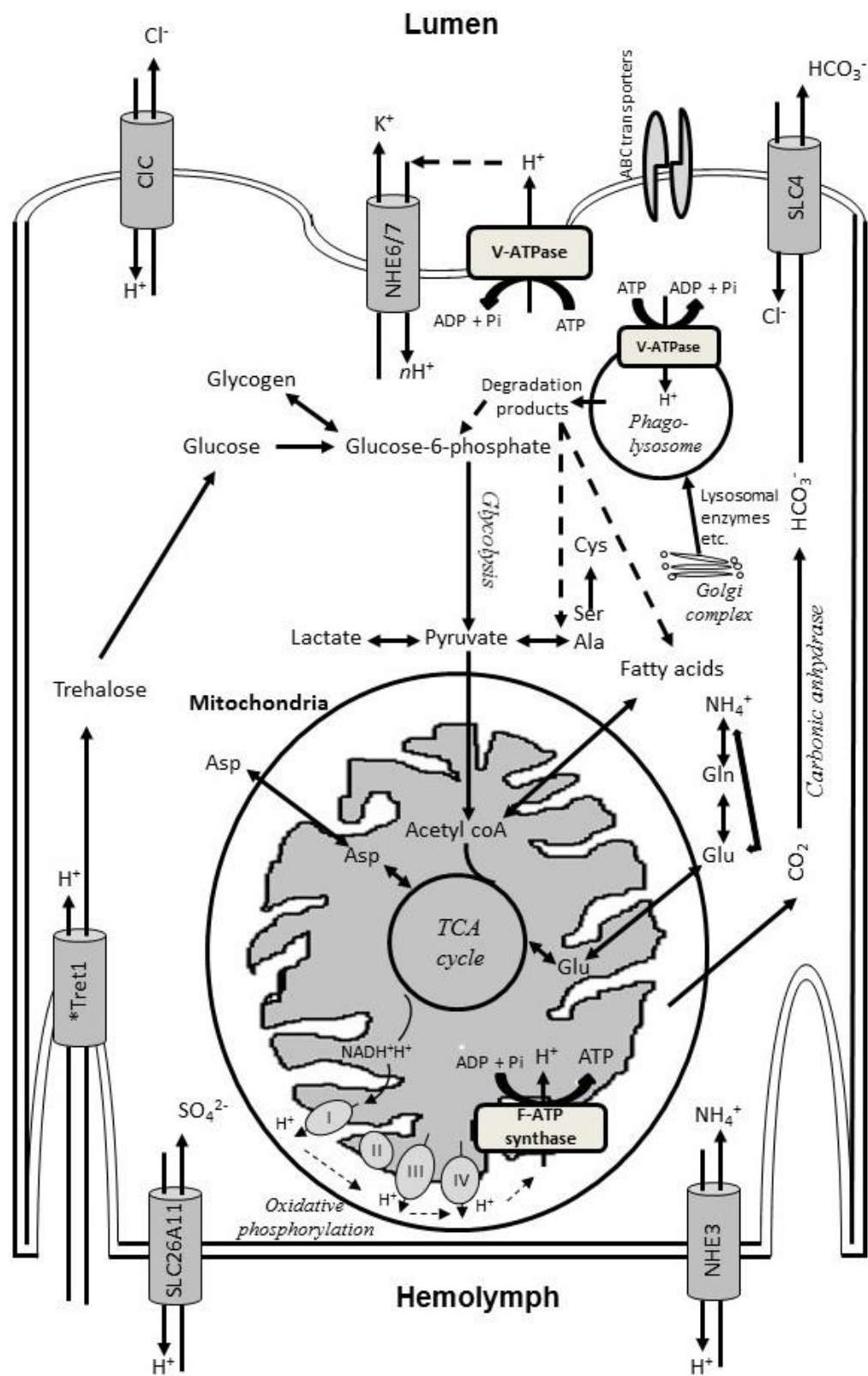
Figure 18: Putative major metabolic network in a mesenteric cell of the mixed segment. The winding of the basal membrane represents the observed invaginations in the mesenteric epithelium of the mixed segment in *N. takasagoensis* (Tokuda et al. 2001). *I, II, III,* and *IV* represent the four complexes of the respiratory chain in oxidative phosphorylation in the mitochondrion. *Tret1, a homologue of Trehalose transporter (Kanamori et al. 2010; Kikuta et al. 2012), was not assigned by a K-number, but the presence of a related isotig in mixed segment libraries was confirmed by a reciprocal BLAST search (the sequence of the isotig is available upon request).

References

Andersen, C.L., Jensen, J.L., Ørntoft, T.F., 2004. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res. 64, 5245–5250.

Azuma, M., Harvey, W.R., Wieczorek, H., 1995. Stoichiometry of K^+/H^+ antiport helps to explain extracellular pH 11 in a model epithelium. FEBS Letters 361, 153–156.

Baumann, O., Walz, B., 2012. The blowfly salivary gland – a model system for analyzing the regulation of plasma membrane V-ATPase. J. Insect. Physiol. 58, 450–458.

Bignell, D.E., 1994. Soil-feeding and gut morphology in higher termites. In: Hunt, J.H., Nalepa, C.A. (Eds.), Nourishment & Evolution in Insect Societies. Westview Press, Boulder, CO, pp. 131–159.

Bignell, D.E., 2000. Introduction to symbiosis. In: Abe, T., Bignell, D.E., Higashi, M. (Eds), Termites: evolution, sociality, symbioses, ecology. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 189-208

Bignell, D.E., 2011. Morphology, physiology, biochemistry and functional design of the termite gut: an evolutionary wonderland. In: Bignell, D.E., Roisin, Y., Lo, N. (Eds), Biology of termites: a modern synthesis. Springer, Dordrecht, The Netherlands, pp. 375-412.

Bignell, D.E., Eggleton, P., 1995. On the elevated intestinal pH of higher termites (Isoptera: Termitidae). Insect. Soc. 42, 57– 69.

Bignell, D.E., Oskarsson, H., Anderson, J.M., Ineson, P., 1983. Structure, microbial associations and function of the so-called "mixed segment" of the gut in two soil-feeding termites, *Procubitermes aburiensis* and *Cubitermes severus* (Termitidae, Termitinae). J. Zool. Lond. 201**,** 445–480.

Bonar, P., Schneider, H-P., Becker, H.M., Deitmer, J.W., Casey, J.R., 2013. Three-dimensional model for the human Cl^-/HCO_3^- exchanger, AE1, by homology to the *E. coli* ClC protein. J. Mol. Biol. 425, 2591-2608.

Boudko, D. Y., Moroz, L.L., Harvey, W. R., Linser, P. J., 2001a. Alkalinization by chloride/bicarbonate pathway in larval mosquito midgut. Proc. Natl. Acad. Sci. U.S.A. 98,

15354–15359.

Boudko, D.Y., Moroz, L.L., Linser, P.J., Trimarchi, J.R., Smith, P.J.S., Harvey, W.R., 2001b. In situ analysis of pH gradients in mosquito larvae using non-invasive, self-referencing, pH-sensitive microelectrodes. J. Exp. Biol. 204, 691–699.

Bowman, B.J., Vazquez-Laslop, N., Bowman, E.J., 1992. The vacuolar ATPase of *Neurospora crassa.* J. Bioenerg. Biomemb. 24, 361–370.

Brauman, A., Bignell, D.E., Tayasu, I., 2000. Soil-feeding termites: Biology, microbial associations and digestive mechanisms. In: Abe, T., Bignell, D.E., Higashi, M., (Eds.), Termites: evolution, sociality, symbiosis, ecology. Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 233–259.

Brett, C.L., Tukaye, D.N., Mukherjee, S., Rao, R., 2005. The yeast endosomal $Na^+(K^+)/H^+$ exchanger Nhx1 regulates cellular pH to control vesicle trafficking. Mol. Biol. Cell 16, 1396-1405.

Brown, D., Breton, S., 1996. Mitochondria-rich, proton secreting epithelial cells. J. Exp. Biol. 199, 2345–2358.

Brown, D., Gluck, S., Hartwig, J., 1987. Structure of the novel membrane coating material in proton-secreting epithelial cells and identification as an H^+ ATPase. J. Cell. Biol. 105, 1637–1648.

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

Brune, A., 1998. Termite guts: the world's smallest bioreacters. Trends Biotechnol. 16, 16–21.

Brune, A., Dietrich, C., 2015. The gut microbiota of termites: digesting the diversity in the light of ecology and evolution. Annu. Rev. Microbiol. 69,145-166.

Brune, A., Emerson, D., Breznak, J.A., 1995. The termite gut microflora as an oxygen sink: microelectrode determination of oxygen and pH gradients in guts of lower and higher termites. Appl. Environ. Microbiol. 61, 2681–2687.

Brune, A., Ohkuma, M., 2011. Role of the termite gut microbiota in symbiotic digestion. In: Bignell, D.E., Roisin, Y., Lo, N., (Eds), Biology of termites: a modern synthesis. Springer, Dordrecht, The Netherlands, pp. 439-475.

Bullard, J.H., Purdom, E., Hansen, K.D., Dudoi, S., 2010. Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics. 11, 94.

Chamberlin, M.E., 1990. Luminal alkalinization by the isolated midgut of the tobacco hornworm (Manduca sexta). J. Exp. Biol. 150, 467–71.

Cidon, S., Nelson, N., 1983. A novel ATPase in the chromaffin granule membrane. J. Biol. Chem. 258, 2892–2898.

Cioffi, M., Harvey, W. R., 1981. Comparison of potassium transport in three structurally distinct regions of the insect midgut. J. Exp. Biol. 91, 103–116.

Clark, T.M., 1999. Evolution and adaptive significance of larval midgut alkalinization in the insect superorder Mecopterida. J. Chem. Ecol. 25, 1945–1960.

Clark, T.M., Koch, A.R., Moffett, D.F., 1998. Alkalinization by *Manduca sexta* anterior midgut in vitro: requirements and characteristics. Comp. Biochem. Physiol. 121, 181–87.

Corena, M. P., Seron, T. J., Lehman, H. K., Ochrietor, J. D., Kohn, A., Tu, C. and Linser, P. J. 2002. Carbonic anhydrase in the midgut of larval *Aedes aegypti*: cloning, localization and inhibition. J. Exp. Biol. 205, 591-602.

Corena, M.P., Fiedler, M.M., VanEkeris, L., Linser, P.J., 2004. A comparative study of carbonic anhydrase in the midgut of mosquito larvae. Comp. Biochem. Physiol. 137, pp. 207–225.

Cornwell, W.K., Cornelissen, J.H.C., Allison, S.D., Bauhus, J., Eggleton, P., Preson, C.M., Scarff, F., Weedon, J.T., Wirth, C., Zanne, A.E., 2009. Plant traits and wood fates across the globe: rotted, burned, or consumed? Global Change Biol. 15, 2431-2449.

Cummings, R. D., Etzler, M. E. 2009. R-type lectins In: Varki, A, Cummings, R. D., Esko, J. D., Freeze, H. H., Stanley, P., Bertozzi, C. R., Hart, G. W., Etzler, M. E. (Eds.), Essentials of Glycobiology, 2nd Ed., Cold Spring Harbor, New York, pp. 403–414.

Czolij, R., Slaytor, M., O'Brien, R. W., 1985. Bacterial flora of the mixed segment and the hindgut of the higher termite *Nasutitermes exitiosus* Hill (Termitidae, Nasutitermitinae). Appl. Environ. Microbiol. 49, 1226–1236.

Dean, G. E., Fishkes, H., Nelson, P. J., Rudnick, G., 1984. The hydrogen ion-pumping adenosine triphosphatase of platelet dense granule membrane. Differences from F1F0 and phosphoenzyme-type ATPases. J. Biol. Chem. 259, 9569–9574.

Donovan, S.E., Eggleton, P., Bignell, D.E., 2001. Gut content analysis and a new feeding group classification of termites. Ecol. Entomol. 26, 356-366.

Donowitz, M., Tse, C.M., Fuster, D., 2013. SLC9/NHE gene family, a plasma membrane and organellar family of Na^+/H^+ exchangers. Mol. Asp. Med. 34, 236-251.

Eggleton, P., 2011. An introduction to termites: biology, taxonomy and functional morphology. In: Bignell, D.E., Roisin, Y., Lo, N., (Eds), Biology of Termites: A Modern Synthesis. Springer, Dordrecht, The Netherlands, pp. 1-26.

Ekblom, R., Galindo, J., 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity. 107, 1–15.

Felton, G.W., Duffey, S.S., 1991. Reassessment of the role of gut alkalinity and detergency in insect herbivory. J. Chem. Ecol. 17, 1821-1836.

Finn, R. D., Bateman, A., Clements, J., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Heger, A., Hetherington, K., Holm, L., Mistry, J., Sonnhammer, E. L. L., Tate, J., Punta, M., 2014. The Pfam protein families database. Nucl. Acids Res. 42, D222–D230.

Forgac, M., 1989. Structure and function of the vacuolar class of ATP-driven proton pumps. Physiol. Rev. 69, 765–796.

Forgac, M., 1992. Structure and properties of the coated vesicle H^+ ATPase. J. Bioenerg. Biomemb. 24, 341–350.

Forgac, M., 1997. Structure, function and regulation of the vacuolar (H^+)-ATPases. Annu. Rev. Cell. Dev. Biol. 13, 779–808.

Forgac, M., 2007. Vacuolar ATPases: rotary proton pumps in physiology and pathophysiology. Nat. Rev. Mol. Cell. Biol. 8, 917–929.

Forgac, M., Cantley, L., Wiedenmann, B., Altstiel, L., Branton, D., 1983. Clathrin-coated vesicles contain an ATP-dependent proton pump. Proc. Natl. Acad. Sci. USA. 80, 1300–1303.

Gluck, S.L., 1992. The structure and biochemistry of the vacuolar H^+ ATPase in proximal and

distal urinary acidification. J. Bioenerg. Biomemb. 24, 351–360.

Grassé, P.-P., Noirot, C., 1954. *Apicotermes arquieri* (Isoptère): Ses constructions, sa biologie. Considérations generals sur la sous-famille des *Apicotermitinae* nov. Ann. Sci. Nat. Zool. Biol. Animale [11] 16, 345–388.

Harikumar, P., Reeves, J. P., 1983. The lysosomal proton pump is electrogenic. J. Biol. Chem. 258, 10403–10410.

Harold, F.M., 1986. The Vital force: A study of bioenergetics. W.H. Freeman Co, New York, pp. 577.

Harrison, J. F., 2001. Insect acid-base physiology. Annual review of entomology. 46, 221-250.

Harvey, W.R., Cioffi, M., Wolfersberger, M. G., 1983. Potassium ion transport ATPase in insect epithelia. J. Exp. Biol. 106, 91–117.

Harvey, W.R., Wieczorek, H., 1997. Animal plasma membrane energization by chemiosmotic H^+ ATPases. J. Exp. Biol. 200, 203-216.

Hochachka, P.W., 2003. Intracellular convection, homeostasis and metabolic regulation. J. Exp. Biol. 206, 2001–2009.

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.

Holmgren, N., 1909. Termitenstudien. I. Anatomische Untersuchungen. Kungliga Svenska Vetensskapakademiens. Handlingar. 44, 1-215.

Hotopp, J.C., Clark, M.E., Oliveira, D.C., Foster, J.M., Fischer, P., Torres, M.C. *et al*., 2007. Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. Science 317, 1753-1756.

Hu, X., Zhang, X., Wang, J., Huang, M., Xue, R., Cao, G., Gong, C., 2015. Transcriptome analysis of BmN cells following over-expression of *BmSTAT*. Mol. Genet. Genomics 290, 2137-2146.

Inoue, T., Kitade, O., Yoshimura, T., Yamaoka, I., 2000. Symbiotic associations with protists. In: Abe, T., Bignell, D.E., Higashi, M., (Eds), Termites: Evolution, Sociality, Symbioses, Ecology. Dordrecht, The Netherlands, Kluwer Academic Publishers, pp. 275–288.

Jentsch, T.J., Stein, V., Weinreich, F., Zdebik, A.A., 2002. Molecular structure and physiological function of chloride channels. Physiol. Rev. 82, 503-568.

Kamhawi, S., Ramalho-Ortigao, M., Pham, V. M., Kumar, S., Lawyer, P. G., Turco, S. J., Barillas-Mury, C., Sacks, D. L., Valenzuela, J. G., (2004) A role for insect galectins in parasite survival. Cell 119, 329–4341.

Kanamori, Y., Saito, A., Hagiwara-Komoda, Y., Tanaka, D., Mitsumasu, K., Kikuta, S., Watanabe, M., Cornette, R., Kikawada, T., Okuda, T., 2010. The trehalose transporter 1 gene sequence is conserved in insects and encodes proteins with different kinetic properties involved in trehalose import into peripheral tissues. Insect. Biochem. Mol. Biol. 40, 30-37.

Kane, P.M., 2006. The where, when, and how of organelle acidification by the yeast vacuolar H^+-ATPase. Microbiol. Mol. Biol. Rev. 70, 177–191.

Kane, P.M., Stevens, T.H., 1992. Subunit composition, biosynthesis and assembly of the yeast V-ATPase. J. Bioenerg. Biomemb. 24, 383–394.

Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M., Hirakawa, M., 2010. KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucl. Acids Res. 38, D355–D360.

Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M., 2014. Data, information, knowledge and principle: back to metabolism in KEGG. Nucl. Acids Res. 42, D199-D205.

Kappler, A, Brune, A., 1999. Influence of gut alikalinity and oxygen status on mobilization and size-class distribution of humic acids in the hindgut of soil-feeding termites. Appl. Soil. Ecol. 13, 219–229.

Kikuta, S., Hagiwara-Komoda, Y., Noda, H., Kikawada, T., 2012. A novel member of the trehalose transporter family functions as an H^+-dependent trehalose transporter in the reabsorption of trehalose in malpighian tubules. Front. Physiol. 3, 290.

Kim, D. H., Patnaik, B. B., Seo, G. W., Kang, S. M., Lee, Y. S., Lee, B. L., Han, Y. S. (2013) Identification and expression analysis of a novel R-type lectin from the coleopteran beetle, *Tenebrio molitor*. J. Invertebr. Pathol. 114, 226–229.

Köhler, T., Dietrich, C., Scheffrahn, R.H., Brune, A., 2012. High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation of the gut in wood-feeding higher termites (*Nasutitermes* spp.). Appl. Environ. Microbiol. 78, 4691–4701.

Lee, K.E., Wood, T.G., 1971. Termites and soils. London, Academic Press, pp. 251.

Lei, Y., Zhu, X., Xie, W., Wu, Q., Wang, S., Guo, Z., Xu, B., Li, X., Zhou, X., Zhang, Y., 2014. Midgut transcriptome response to a Cry toxin in the diamondback moth, *Plutella xylostella* (Lepidoptera; Plutellidae). Gene 533, 180-187.

Leong, D.T., Gupta, A., Bai, H.F., Wan, G., Yoong, L.F., Too, H.P., Chew, F.T., Hutmacher, D.W., 2007. Absolute quantification of gene expression in biomaterials research using real-time PCR. Biomaterials 28, 203–210.

Lin, H., Randall, D. J., 1993. Proton-ATPase activity in crude homogenates of fish gill tissue: Inhibitor sensitivity and environmental and hormonal regulation. J. Exp. Biol. 180, 163–174.

Liu, L., Li, Y., Li., S., Hu, N. *et al.*, 2012. Comparison of next-generation sequencing systems. J. Biomed. Biotechnol. Article ID: 251364.

Lo, N., Eggleton, P., 2011. Termite phylogenetics and co-cladogenesis with symbionts. In: Bignell, D.E., Roisin, Y., Lo, N., (Eds), Biology of termites: a modern synthesis. Springer, Dordrecht, The Netherlands, pp. 27-50.

Lo, N., Tokuda, G., Watanabe, H., 2011. Evolution and function of endogenous termite cellulases. In: Bignell, D.E., Roisin, Y., Lo, N. (Eds), Biology of termites: a modern synthesis. Springer, Dordrecht, The Netherlands, pp. 51-67.

Luo, J., Liu, X., Liu, L., Zhang, P., Chen, L., Gao, Q., Ma, W., Chen, L., Lei, C., 2014. *De novo* analysis of the *Adelphocoris suturalis* Jakovlev metathoractic scent glands transcriptome and expression patterns of pheromone biosynthesis-related genes. Gene 551, 271-278.

Luzio, J.P., Pryor, P.R., Bright, N.A., 2007. Lysosomes: fusion and function. Nat. Rev. Mol. Cell. Biol. 8, 622-632.

Martin, J.S., Martin, M.M., Bernays, E.A., 1987. Failure of tannic acid to inhibit digestion or reduce digestibility of plant protein in gut fluids of insect herbivores: implications for theories of plant defense. J. Chem. Ecol. 13, 605-622.

Mikaelyan, A., Strassert, J.F.H., Tokuda, G., Brune, A., 2014. The fiber-associated cellulolytic bacterial community in the hindgut of wood-feeding higher termites (*Nasutitermes* spp.). Environ. Microbiol. 16, 2711-2722.

Moffett, D.F., Onken, H., 2010. The cellular basis of extreme alkali secretion in insects: a tale of two tissues, in: Gerencser, G.A. (Eds), In Epithelial Transport Physiology. Humana Press, New York, pp. 91–112.

Mount, D.B., Romero, M.F., 2004. The SLC26 gene family of multifunctional anion exchangers. Pflügers. Arch/Eur. J. Physiol. 447, 710-721.

Nakashima, K., Watanabe, H., Saitoh, H., Tokuda, G., Azuma, J-I., 2002. Dual cellulose-digesting system of the wood-feeding termite, *Coptotermes formosanus* Shiraki. Insect Biochem. Mol. Biol. 32, 777-784.

Nandety, R.S., Kamita, S.G., Hammock, B.D., Falk, B.W., 2013. Sequencing and *de novo* assembly of the transcriptome of the glassy-winged sharpshooter (*Homalodisca vitripennis*). PLoS One 8, e81681.

Nelson, S.T., 2003. Trehalose - the insect 'blood' sugar. Adv. Insect Physiol. 31, 205–285.

Ni, J., Tokuda, G., 2013. Lignocellulose-degrading enzymes from termites and their symbiotic microbiota. Biotechnol. Adv. 31, 838-850.

Noboru, M., Masaaki, K., 2011. Autophagy: renovation of cells and tissues. Cell 147, 728–741.

Noirot, C., 2001. The gut of termites (Isoptera): comparative anatomy, systematics, phylogeny. II. Higher termites (Termitidae). Ann. Soc. Entomol. Fr (NS). 37, 431-471.

Ohkuma, M., Brune, A., 2011. Diversity, structure, and evolution of the termite gut microbial community. In: Bignell, D.E., Roisin, Y., Lo, N., (Eds), Biology of termites: a modern synthesis. Springer, Dordrecht, pp. 413-438.

Ohkuma, S., Moriyama, Y., Takano, K., 1982. Identification and characterization of a proton pump on lysosomes by fluorescein-isothiocyanate-dextran fluorescence. Proc. Nat. Acad. Sci. USA. 79, 2758-2762.

Onken, H., Moffett, S.B., Moffett, D.F., 2008. Alkalinization in the isolated and perfused anterior midgut of the larval mosquito, *Aedes aegypti*. J. Insect. Sci. 8, 1–20.

Pace, K. E. and Baum, L. G. 2004. Insect galectins: roles in immunity and development. Glycoconj. J. 19, 607–614.

Patrick, M.L., Aimanova, K., Sanders, H.R., Gill, S.S., 2006. P-type Na^+/K^+-ATPase and V-type H^+-ATPase expression patterns in the osmoregulatory organs of larval and adult mosquito *Aedes aegypti*. J. Exp. Biol. 209, 4638-4651.

Pfaffl, M.W., Tichopad, A., Prgomet, C., Neuvians, T.P., 2004. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper – excel-based tool using pair-wise correlations. Biotechnol. Lett. 26, 509–515.

Piermarini, P.M., Weihrauch, D., Meyer, H., Huss, M., Beyenbach, K.W., 2009. NHE8 is an intracellular cation/H^+ exchanger in renal tubules of the yellow fever mosquito *Aedes aegypti*. Am. J. Physiol. Renal. Physiol. 296, F730-F750.

Poulsen, M., Hub, H., Li, C., Chen, Z., Xu, L., Otani, S., Nygaard, S., Nobre, T., Klaubauf, S., Schindler, P.M., Hauser, F., Pan, H., Yang, Z., Sonnenberg, A.S.M., Beer, Z.Wd., Zhang, Y., Wingfield, M.J., Grimmelikhuijzen, C.J.P., Vries, R.Pd., Korb, J., Aanen, D.K., Wang, J., Boomsma, J.J., Zhang, G., 2014. Complementary symbiont contributions to plant decomposition in a fungus-farming termite. Proc. Natl. Acad. Sci. USA. 111, 14500-14505.

Rawlings, N.D., Barrett, A.J., Bateman, A., 2012. MEROPS: the database of proteolytic enzymes, their substrate and inhibitors. Nucl. Acids Res. 40, 343–350.

Ridgway, R.L., Moffett, D.F., 1986. Regional differences in the histochemical localization of carbonic anhydrase in the midgut of tobacco hornworm (*Manduca sexta*). J. Exp. Zool. 237, 407–412.

Saier, M.H., Reddy, V.S., Tamang, D.G., Västermark, Å., 2014. The transporter classification database. Nucl. Acids Res. 42, D251-D258.

Sarkar, N., 1997. Polyadenylation of mRNA in prokaryotes. Annu. Rev. Biochem. 66, 173-197.

Silver, N., Best, S., Jiang, J., Thein, S.L., 2006. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. BMC Mol. Biol. 7, 33.

Terrapon, N., Li, C., Robertson, H.M., Ji, L., Meng, X., Booth, W., Chen, Z., Childers, C.P., Glastad, K.M., Gokhale, K., Gowin, J., Gronenberg, W., Hermansen, R.A., Hu, H., Hunt, B.G., Huylmans, A.K., Khalil, S.M.S., Mitchell, R.D., Munoz-Torres, M.C., Mustard, J.A., Pan, H., Reese, J.T., Scharf, M.E., Sun, F., Vogel, H., Xiao, J., Yang, W., Yang, Z., Yang, Z., Zhou, J., Zhu, J., Brent, C.S., Elsik, C.G., Goodisman, M.A.D., Liberles, D.A., Roe, R.M., Vargo, E.L., Vilcinskas, A., Wang, J., Bornberg-Bauer, E., Korb, J., Zhang, G., Liebig, J., 2014. Molecular traces of alternative social organization in a termite genome. Nat. Commun. 5, 3636.

Thongaram, T., Kosono, S., Ohkuma, M., Hongoh, Y., Kitada, M., Yoshinaka T., *et al*. 2003. Gut of higher termites as a niche for alkaphiles as shown by culture-based and culture-independent studies. Microbes. Environ. 18, 152-159.

Tokuda, G., Lo, N., Watanabe, H., 2005. Marked variations in patterns of cellulase activity against crystalline- vs. carboxymethyl-cellulose in the digestive systems of diverse, wood-feeding termites. Physiol. Entomol. 30, 372-380

Tokuda, G., Lo, N., Watanabe, H., Arakawa, G., Matsumoto, T., Noda, H., 2004. Major alteration of the expression site of endogenous cellulases in members of an apical termite lineage. Mol. Ecol. 13, 3219–3228.

Tokuda, G., Nakamura T., Murakami, R. Yamaoka, I., 2001. Morphology of the digestive system in the wood-feeding termite *Nasutitermes takasagoensis* (Shiraki) [Isoptera: Termitidae]. Zool. Sci. 18, 869–877.

Tokuda, G., Tsuboi, Y., Kihara, K., Saitoh, S., Moriya, S., Lo, N., Kikuchi, J., 2014. Metabolic profiling of ^{13}C-labelled cellulose digestion in a lower termite: insights into gut symbiont function. Proc. R. Soc. B 281, 2014.0990.

Tokuda, G., Watanabe, H., 2007. Hidden cellulases in termites: revision of an old hypothesis. Biol. Lett. 3:336-339

Tokuda, G., Watanabe, H., Hojo, M., Fujita, A., Makiya, H., Miyagi, M. *et al*., 2012. Cellulolytic environment in the midgut of the wood-feeding higher termite *Nasutitermes*

takasagoensis. J. Insect. Physiol. 58, 147–154.

Tokuda, G., Watanabe, H., Matsumoto, T., Noda, H., 1997. Cellulose digestion in the wood-eating higher termite, *Nasutitermes takasagoensis* (Shiraki): distribution of cellulases and properties of endo-β-1,4-glucanase. Zool. Sci. 14, 83–93.

Tokuda, G., Yamaoka, I., Noda, H., 2000. Localization of symbiotic clostridia in the mixed segment of the termite *Nasutitermes takasagoensis*. Appl. Environ. Microbiol. 66, 2199–2207.

Umbarger, H.E., 1978. Amino acid biosynthesis and its regulation. Annu. Rev. Biochem. 47, 533-606.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. *et al*., 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 3, RESEARCH0034.

Warburg, O., Wind, F., Negelein, E., 1927. The metabolism of tumors in the body. J. Gen. Physiol. 8, 519-530.

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

Watanabe, H., Tokuda, G., 2010. Cellulolytic systems in insects. Annu. Rev. Entomol. 55, 609-632.

Weihrauch, D., Donini, A., O'Donnell, M.J., 2011. Ammonia transport by terrestrial and aquatic insects. J. Insect Physiol. 58, 473-487.

Wieczorek, H., 1992. The insect V-ATPase, a plasma membrane proton pump energizing secondary active transport: Molecular analysis of electrogenic potassium transport in the Tobacco hornworm midgut. J. Exp. Biol. 172, 335–343.

Wieczorek, H., Gruber, G., Harvey, W., Huss, M., Merzendorfer, H., Zeiske, W., 2000. Structure and regulation of insect plasma membrane H^+ V-ATPase. J. Exp. Biol. 203, 127–135.

Wieczorek, H., Putzenlechner, M., Zeiske, W., Klein, U., 1991. A vacuolar-type proton pump energizes K^+/H^+ antiport in an animal plasma membrane. J. Biol. Chem. 266, 15340–15347.

Wood, T.G., Sands, W.A., 1978. The role of termites in ecosystems. In: Brian, M.V., (Eds), Production ecology of ants and termites, Cambridge University Press, Cambridge, pp. 245–292.

Xie, X. S., Stone, D. K., Racker, E., 1983. Determinants of clathrin-coated vesicle acidification. J. Biol. Chem. 258, 14834–14838.

Yamaoka, I., Nagatani, Y., 1980. Phagocytic cells in the midgut epithelium of the termite, *Reticulitermes speratus* (Kolbe). Zool. Mag. 89, 308–311.

Yin, Y., Mao, X., Yang, J., Chen, X., Mao, F., Xu, Y., 2012. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucl. Acids Res. 40, W445–W451.

Zhang, D., Lax, A.R., Raina, A.K., Bland, J.M., 2009. Differential cellulolytic activity of native-form and C-terminal tagged-form cellulase derived from *Coptotermes formosanus* and expressed in *E. coli*. Insect Biochem. Mol. Biol. 39, 516–522.

Zhang, D., Lax, A.R., Bland, J.M., Allen, A.B., 2011. Characterization of a new endogenous endo-β-1,4-glucanase of Formosan subterranean termite (*Coptotermes formosanus*). Insect Biochem. Mol. Biol. 41, 211–218.

Zhang, D., Lax, A.R., Bland, J.M., Yu, J., Fedorova, N., Nierman, W.C., 2010. Hydrolysis of filter-paper cellulose to glucose by two recombinant endogenous glycosyl hydrolases of *Coptotermes formosanus*. Insect Sci. 17, 245–252.

Zhuang, Z., Linser, P.J., Harvey, W.R., 1999. Antibody to H^+ V-ATPase subunit E colocalizes with portasomes in alkaline larval midgut of a freshwater mosquito (*Aedes aegypti*). J. Exp. Biol. 202, 2449–2460.

Zhang, D., Lax, A.R., Henrissat, B., Coutinho, P., Katiya, N., Nierman, W.C., Fedorova, N., 2012. Carbohydrate-active enzymes revealed in *Coptotermes formosanus* (Isoptera: Rhinotermitidae) transcriptome. Insect Mol. Biol. 21, 235–245.

Zhang, W., Song, W., Zhang, Z., Wang, H., Yang, M., Guo, R., Li, M., 2014. Transcriptome analysis of *Dastarcus helophoroides* (Coleoptera: Bothrideridae) using Illumina HiSeq Sequencing. PLoS One 9, e100673.

Supplementary Materials

File S1

Carbonic anhydrase (isotig7716)

TTGGGCAGATGATTATCCGGCGTGCGCCGGAAAGTTCCAGTCTCCAGTTTGACATAGAGGAACATCTCGTGCGAC
CAGTGCGCTTGCCACCCTTGGTATTTAATGGCTTCAACAGACTGCCACTCAGCAGTAGGCTCACCAATAATGGCCA
TACCGTGATGCTGCAGATGAACATGACAAGTCCGGCATTGCTGTCTGGAGGGCCACTCAGCGACAACTACGTGTT
CTATCAGCTCCATTTCCACTGGGGCTCCAACGATTCTGTGGGAAGTGAGGACACACTTAACAACCGATCGTTTCCA
TTGGAGATGCACATGGTCACATACAAGCACATGTATGGAAACTTTGACAACGCTACAAATTATTCAGACGGTCTAG
CAGTGATTGCTGTCTTCTTTGAGATCTATGGGGACGGTAACCCATACTACTCAGAGATTACACAGGTGCTGCCACA
AGTGACAGAGCCAGACTCACAGGTCAAACTAAAGCACGCGTTAACGCTGTTCTCACTGCTACCAGAGACAAAGCA
CCTCTATTTCACATACCAAGGATCACTCACTACACCACCCTGCTTGGAGGTGGTAACCT

NHE8

ATGTATTGCGTTGGCAAGCGCATTGTTGTTAAAACATGTGGACTTGAGAAAAAATCCATCCCTTGAGTTTGGGATA
ATGCTGGTATTCACATATGCACCTTATGCTCTTGCTGAAGGTATTCATTTATCAGGTATCATGGCAATTCTGTTCTGT
GGCATTGTCATGTCACATTATACTCACTTCAACCTCTCCACAGTGACACAAATCACCATGCAGCAAACAATGCGTAC
TCTGGCATTTATCGCTGAGACTTGTGTCTTTGCGTATCTTGGTTTGGCGTTGTTCAGTTTTCAGCACCGTGTGGAGC
CTGCATTGGTTGTGTGGAGCATTGTGCTGTGTCTTATTGGCCGGGCTTGCAATATCTTTCCACTTGCTATATTGTGT
AACCGCTTCCGTGAACACCAGATCACCTACAAGATGATGTTCATCATGTGGTTCAGTGGTTTGCGAGGTGCTATTG
CCTATGCTCTCTCTCTTCATCTGCAGTTCAGTGATGAAACACGGCATGTCATTGTGACCACGACACTCATCATAGTA
CTCTTCACAACTCTGGTGTTTGGTGGATCTACAATGCCTATGATGAAGCTGCTGCTGGCAAACAAGAAGGGAAGC
AGTCATCGGAGACGGGAAAAGGAGATAACCCTCAGCAAGACCAAAGAATGGGGTCAGGCAATAGATTCAGAACA
TCTGTCAGAATATACAGAAGAGGAGATGGAGGTGAATTTTATCCAGTCTCGTATCAA

Figure S1
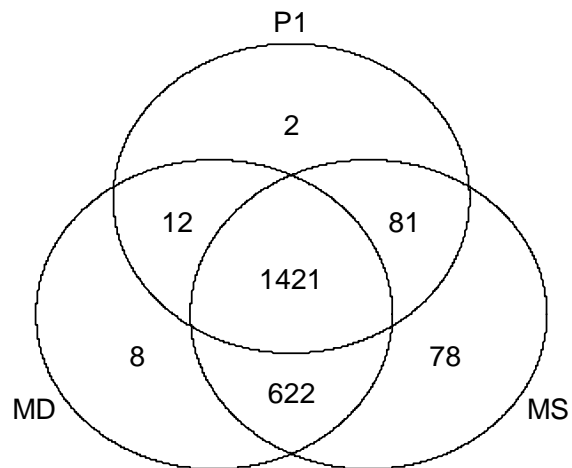
Figure S1: Venn diagram based on the K-numbers among mixed segment (MS1, MS2, and MS3), MD (MD1 and MD2), and P1 libraries. When at least 1 read derived from a library was assembled into an isotig, the k-number was registered as present in the library. Based on the present/absent information per isotig for libraries, a Venn diagram was drawn by k-number.

Table S1: Normalized read counts and coefficient of variation in each KEGG reference pathway

| ko-number | Definition | Normalized read count in each pathway | | | | | | | Eigenvalues | |
		MD1	MD2	MS1	MS2	MS3	P1	Coefficient of variation (CV)	PC1	PC2
ko00190	ko00190 Oxidative phosphorylation	860	1101.25	2150	3910.83	2435.63	745	796.9484351	-4685.1	-1521.3
ko04080	ko04080 Neuroactive ligand-receptor interaction	462.5	1167.5	35	24.1667	46.25	8.33333	739.2362605	-111.33	404.967
ko04391	ko04391 Hippo signaling pathway - fly	802.5	775.625	641.667	757.5	518.75	1760	226.8281822	-1572.7	1166.92
ko04142	ko04142 Lysosome	1050	998.125	1518.33	1733.33	1497.5	328.333	218.3254648	-2588.5	-560
ko04745	ko04745 Phototransduction - fly	730	723.75	598.333	790	478.75	1646.67	209.4059723	-1483.6	1048.9
ko04145	ko04145 Phagosome	1680	1764.38	2528.33	3611.67	2446.25	2403.33	200.2762613	-5655.6	214.207
ko03010	ko03010 Ribosome	545	684.375	343.333	810	590	1413.33	185.7044225	-1295.1	819.194
ko00640	ko00640 Propanoate metabolism	187.5	671.25	173.333	208.333	185	198.333	142.8647036	-173.74	180.629
ko00052	ko00052 Galactose metabolism	247.5	260.625	41.6667	55.8333	38.75	65	94.5500367	187.321	97.2186
ko00480	ko00480 Glutathione metabolism	552.5	597.5	193.333	294.167	263.75	233.333	84.761354	-359.54	234.502
ko00270	ko00270 Cysteine and methionine metabolism	130	125	103.333	176.667	120.625	423.333	82.4686635	7.4925	207.879
ko04141	ko04141 Protein processing in endoplasmic reticulum	1362.5	1175	765	920.833	794.375	731.667	68.3216971	-1769.8	529.731
ko00020	ko00020 Citrate cycle (TCA cycle)	337.5	785.625	573.333	570.833	483.75	280	66.070947	-787.81	25.8937
ko00520	ko00520 Amino sugar and nucleotide sugar metabolism	432.5	302.5	211.667	193.333	235.625	66.6667	61.5734809	-113.79	31.0132
ko00380	ko00380 Tryptophan metabolism	215	290.625	318.333	479.167	323.125	81.6667	60.8043832	-308.3	-185.37
ko00511	ko00511 Other glycan degradation	145	168.75	41.6667	48.3333	41.25	26.6667	48.1057126	260.693	13.0335
ko00051	ko00051 Fructose and mannose metabolism	200	200.625	48.3333	67.5	66.875	70	46.6058142	195.695	56.0025
ko00310	ko00310 Lysine degradation	170	226.25	253.333	360.833	243.75	58.3333	45.8742857	-136.23	-155.5
ko00410	ko00410 beta-Alanine metabolism	242.5	330	258.333	352.5	250.625	55	44.3561226	-186.98	-104.24
ko00630	ko00630 Glyoxylate and dicarboxylate metabolism	255	223.75	470	491.667	416.875	235	44.2314257	-459.85	-134.08
ko00280	ko00280 Valine, leucine and isoleucine degradation	252.5	376.875	281.667	450	293.75	118.333	43.2616467	-312.49	-95.063
ko04070	ko04070 Phosphatidylinositol signaling system	167.5	116.25	211.667	290.833	210	50	40.2778773	-25.418	-150.86
ko03013	ko03013 RNA transport	637.5	668.125	453.333	600	520	915	40.141739	-1035.3	554.261
ko00500	ko00500 Starch and sucrose metabolism	280	286.875	161.667	118.333	98.75	183.333	33.7737974	17.9399	141.082
ko00982	ko00982 Drug metabolism - cytochrome P450	180	260	53.3333	125.833	111.25	138.333	33.7082734	103.994	83.1735
ko00980	ko00980 Metabolism of xenobiotics by cytochrome P450	187.5	271.25	58.3333	136.667	120.625	145	33.301354	83.6313	86.1103

ko00983	ko00983 Drug metabolism - other enzymes	222.5	235	110	72.5	98.75	123.333	32.2737635	119.284	92.0718
ko00061	ko00061 Fatty acid biosynthesis	35	18.75	60	50.8333	59.375	146.667	32.0645728	280.216	16.0865
ko00071	ko00071 Fatty acid degradation	257.5	378.75	301.667	439.167	311.875	153.333	31.7421925	-335.45	-70.273
ko00340	ko00340 Histidine metabolism	80	131.875	108.333	190.833	111.875	15	31.720553	150.558	-114.45
ko04144	ko04144 Endocytosis	502.5	416.25	600	672.5	591.875	341.667	29.8355363	-860.88	-63.623
ko00903	ko00903 Limonene and pinene degradation	65	99.375	83.3333	159.167	89.375	6.66667	29.2462027	206.427	-111.83
ko00561	ko00561 Glycerolipid metabolism	165	214.375	118.333	205.833	137.5	38.3333	28.7168621	69.6281	-61.1
ko00830	ko00830 Retinol metabolism	75	113.125	13.3333	21.6667	30	78.3333	28.2201184	312.437	33.0005
ko00760	ko00760 Nicotinate and nicotinamide metabolism	120	93.75	33.3333	25	29.375	41.6667	27.6344111	308.098	8.26745
ko00040	ko00040 Pentose and glucuronate interconversions	230	306.25	116.667	219.167	150.625	130	27.5518179	-20.609	45.9334
ko00510	ko00510 N-Glycan biosynthesis	195	167.5	76.6667	108.333	99.375	65	22.477868	158.82	10.8886
ko00790	ko00790 Folate biosynthesis	57.5	112.5	26.6667	40.8333	45.625	18.3333	22.34154	314.187	-31.298
ko00900	ko00900 Terpenoid backbone biosynthesis	102.5	149.375	63.3333	113.333	68.75	25	21.9681724	219.166	-44.271
ko00531	ko00531 Glycosaminoglycan degradation	52.5	51.25	8.33333	15.8333	6.875	5	21.6173001	377.686	-38.252
ko00250	ko00250 Alanine, aspartate and glutamate metabolism	212.5	169.375	80	113.333	85	105	21.5595449	141.39	47.2276
ko00010	ko00010 Glycolysis / Gluconeogenesis	355	324.375	225	353.333	224.375	188.333	19.7401264	-238	38.8525
ko02010	ko02010 ABC transporters	127.5	93.125	141.667	96.6667	134.375	23.3333	18.6107545	180.516	-77.836
ko00860	ko00860 Porphyrin and chlorophyll metabolism	105	141.875	40	60	65	100	16.1973443	238.996	37.5928
ko00620	ko00620 Pyruvate metabolism	250	253.75	213.333	399.167	248.75	253.333	15.7851871	-242.23	12.0932
ko00670	ko00670 One carbon pool by folate	20	22.5	30	63.3333	38.125	0	15.3643214	346.419	-91.484
ko00062	ko00062 Fatty acid elongation	102.5	143.75	143.333	185.833	138.75	56.6667	15.0706306	104.019	-83.7
ko00230	ko00230 Purine metabolism	315	299.375	196.667	217.5	192.5	176.667	15.027755	-107.45	80.2086
ko00604	ko00604 Glycosphingolipid biosynthesis - ganglio series	35	28.125	5	10	1.25	3.33333	14.7058144	397.496	-47.285
ko00240	ko00240 Pyrimidine metabolism	230	241.25	173.333	200.833	168.125	98.3333	14.4309303	-9.4029	-4.0388
ko03060	ko03060 Protein export	180	163.125	90	115	93.125	81.6667	14.2341018	151.852	13.857
ko00981	ko00981 Insect hormone biosynthesis	92.5	154.375	66.6667	94.1667	70	56.6667	13.9633041	219.558	-14.972
ko00513	ko00513 Various types of N-glycan biosynthesis	152.5	152.5	81.6667	105	92.5	63.3333	12.7844273	178.805	-5.3556
ko00053	ko00053 Ascorbate and aldarate metabolism	125	188.75	118.333	209.167	138.75	101.667	12.29994644	64.7431	-35.1
ko04146	ko04146 Peroxisome	492.5	491.875	436.667	669.167	503.75	496.667	12.1944641	-827.89	123.745
ko00650	ko00650 Butanoate metabolism	82.5	134.375	113.333	150.833	118.125	53.3333	11.5799808	155.859	-70.002

ko00232	ko00232 Caffeine metabolism	107.5	74.375	45	40	48.75	45	11.4447574	294.382	-11.775
ko04310	ko04310 Wnt signaling pathway	247.5	109.375	173.333	177.5	170.625	161.667	11.252504	17.5652	17.7945
ko00603	ko00603 Glycosphingolipid biosynthesis - globo series	47.5	30.625	11.6667	20.8333	12.5	8.33333	10.1302826	377.485	-47.497
ko00565	ko00565 Ether lipid metabolism	50	15	20	14.1667	14.375	11.6667	10.1168192	380.286	-48.145
ko00330	ko00330 Arginine and proline metabolism	357.5	373.75	233.333	323.333	274.375	310	8.7292988	-300.9	144.479
ko00910	ko00910 Nitrogen metabolism	35	40	80	88.3333	51.875	46.6667	8.4831658	279.668	-69.657
ko03450	ko03450 Non-homologous end-joining	5	6.25	10	14.1667	9.375	33.3333	8.3943333	394.494	-44.958
ko00920	ko00920 Sulfur metabolism	37.5	61.875	33.3333	77.5	39.375	80	8.0040034	291.925	-21.512
ko04350	ko04350 TGF-beta signaling pathway	70	40.625	66.6667	99.1667	60	103.333	7.8324056	245.885	-20.843
ko03410	ko03410 Base excision repair	32.5	44.375	48.3333	63.3333	74.375	23.3333	7.5131004	305.694	-75.679
ko00970	ko00970 Aminoacyl-tRNA biosynthesis	107.5	80.625	80	87.5	91.875	35	7.4060881	234.57	-53.772
ko04068	ko04068 FoxO signaling pathway	185	159.375	185	259.167	205.625	160	7.1559266	-48.649	-31.449
ko00564	ko00564 Glycerophospholipid metabolism	87.5	53.75	70	51.6667	63.75	26.6667	7.0727201	288.904	-50.565
ko00770	ko00770 Pantothenate and CoA biosynthesis	57.5	48.75	26.6667	25	26.25	21.6667	6.5074224	350.221	-36.825
ko04140	ko04140 Regulation of autophagy	12.5	35.625	21.6667	50.8333	27.5	38.3333	5.857216	347.046	-51.877
ko00072	ko00072 Synthesis and degradation of ketone bodies	5	9.375	3.33333	19.1667	12.5	25	5.6579132	394.728	-51.866
ko03040	ko03040 Spliceosome	355	370	380	466.667	381.875	451.667	5.3697885	-539.21	136.446
ko00471	ko00471 D-Glutamine and D-glutamate metabolism	15	15.625	30	33.3333	12.5	8.33333	5.3288717	376.112	-70.191
ko04340	ko04340 Hedgehog signaling pathway	70	25.625	40	40.8333	37.5	48.3333	5.0335253	324.913	-31.887
ko00592	ko00592 alpha-Linolenic acid metabolism	27.5	43.75	21.6667	13.3333	15	23.3333	5.0040826	373.503	-37.53
ko00601	ko00601 Glycosphingolipid biosynthesis - lacto and neolacto series	0	0	6.66667	9.16667	7.5	0	4.833333	414.272	-70.073
ko00300	ko00300 Lysine biosynthesis	0	0.625	6.66667	10	1.25	1.66667	4.8165808	415.624	-67.504
ko00030	ko00030 Pentose phosphate pathway	127.5	88.125	86.6667	107.5	83.75	70	4.4159735	203.293	-27.084
ko00563	ko00563 Glycosylphosphatidylinositol(GPI)-anchor biosynthesis	45	30.625	35	29.1667	30.625	10	4.3329484	355.436	-58.994
ko04512	ko04512 ECM-receptor interaction	20	12.5	3.33333	7.5	5.625	18.3333	4.2204592	401.787	-44.943
ko00360	ko00360 Phenylalanine metabolism	17.5	34.375	13.3333	22.5	24.375	8.33333	4.1735871	378.144	-59.078
ko00780	ko00780 Biotin metabolism	20	13.75	3.33333	10	4.375	6.66667	4.1429211	404.35	-54.023
ko03030	ko03030 DNA replication	7.5	15	10	20.8333	23.75	30	4.1261349	382.175	-50.267
ko01212	ko01212 Fatty acid metabolism	215	286.875	250	305.833	249.375	283.333	4.1129596	-210.92	66.3623
ko04330	ko04330 Notch signaling pathway	52.5	15.625	35	35	38.75	31.6667	4.0651432	343.605	-49.491

117

ko03430	ko03430 Mismatch repair	7.5	10.625	8.33333	19.1667	21.25	26.6667	3.9804195	387.377	-52.484
ko00350	ko00350 Tyrosine metabolism	20	40	16.6667	29.1667	33.75	15	3.9579964	364.125	-57.278
ko03050	ko03050 Proteasome	130	138.125	96.6667	166.667	125	140	3.9194895	106.29	4.66459
ko04630	ko04630 Jak-STAT signaling pathway	25	16.25	31.6667	25.8333	38.125	11.6667	3.8081463	365.147	-67.803
ko04711	ko04711 Circadian rhythm - fly	37.5	13.75	25	15.8333	15.625	20	3.7295405	377.794	-47.902
ko00260	ko00260 Glycine, serine and threonine metabolism	42.5	31.25	40	63.3333	43.75	28.3333	3.6767001	321.391	-64.163
ko00562	ko00562 Inositol phosphate metabolism	75	51.25	78.3333	75	73.75	41.6667	3.5959916	266.308	-57.46
ko03440	ko03440 Homologous recombination	5	10.625	10	14.1667	12.5	25	3.4803459	394.673	-50.601
ko03420	ko03420 Nucleotide excision repair	35	29.375	28.3333	42.5	40	60	3.451026	332.404	-31.607
ko00785	ko00785 Lipoic acid metabolism	7.5	13.125	23.3333	15.8333	29.375	23.3333	3.4305556	380.25	-57.462
ko01040	ko01040 Biosynthesis of unsaturated fatty acids	62.5	85.625	55	60	55	43.3333	3.2877281	285.325	-34.614
ko00600	ko00600 Sphingolipid metabolism	62.5	53.125	51.6667	39.1667	37.5	30	3.1948352	321.108	-40.906
ko00590	ko00590 Arachidonic acid metabolism	55	69.375	35	59.1667	51.25	40	3.0794945	303.76	-39.381
ko00512	ko00512 Mucin type O-Glycan biosynthesis	20	6.25	13.3333	13.3333	11.875	3.33333	3.0516055	398.507	-63.754
ko00400	ko00400 Phenylalanine, tyrosine and tryptophan biosynthesis	15	11.875	8.33333	6.66667	10	0	3.0396586	406.223	-61.893
ko03460	ko03460 Fanconi anemia pathway	5	9.375	8.33333	12.5	11.25	21.6667	2.8375382	398.4	-52.172
ko04130	ko04130 SNARE interactions in vesicular transport	70	43.75	41.6667	51.6667	50	36.6667	2.7882979	310.827	-42.927
ko00130	ko00130 Ubiquinone and other terpenoid-quinone biosynthesis	0	5.625	3.33333	9.16667	5	1.66667	2.5227591	414.403	-65.936
ko00740	ko00740 Riboflavin metabolism	0	0.625	0	3.33333	5	1.66667	2.3112745	420.843	-64.405
ko04013	ko04013 MAPK signaling pathway - fly	45	30.625	50	37.5	30.625	26.6667	2.2914304	338.594	-52.741
ko00460	ko00460 Cyanoamino acid metabolism	7.5	9.375	3.33333	2.5	3.125	1.66667	2.094697	415.994	-59.349
ko00100	ko00100 Steroid biosynthesis	20	11.25	13.3333	13.3333	13.75	26.6667	2.0833333	388.311	-45.272
ko03008	ko03008 Ribosome biogenesis in eukaryotes	72.5	81.875	68.3333	99.1667	83.75	98.3333	1.9471717	223.7	-17.331
ko03015	ko03015 mRNA surveillance pathway	292.5	256.25	235	287.5	251.875	271.667	1.8449652	-203.26	81.1609
ko03020	ko03020 RNA polymerase	17.5	16.875	13.3333	21.6667	23.125	30	1.6845238	377.175	-47.538
ko00730	ko00730 Thiamine metabolism	2.5	1.25	5	9.16667	6.875	5	1.6587995	412.438	-64.773
ko03022	ko03022 Basal transcription factors	27.5	13.75	16.6667	23.3333	20.625	13.3333	1.6513713	379.609	-58.663
ko00514	ko00514 Other types of O-glycan biosynthesis	0	1.875	1.66667	0.83333	1.875	5	1.537037	421.486	-60.053
ko04122	ko04122 Sulfur relay system	5	16.875	10	14.1667	14.375	15	1.4981584	395.368	-56.626
ko00430	ko00430 Taurine and hypotaurine metabolism	10	11.25	6.66667	4.16667	5.625	3.33333	1.4731387	410.732	-58.69

118

ko00220	ko00220 Arginine biosynthesis	52.5	46.25	63.3333	65	52.5	43.3333	1.4353763	294.726	-52.459
ko00533	ko00533 Glycosaminoglycan biosynthesis - keratan sulfate	2.5	0	5	3.33333	0.625	3.33333	1.411385	419.552	-62.471
ko04120	ko04120 Ubiquitin mediated proteolysis	202.5	163.125	198.333	200.833	195	176.667	1.3367935	-25.803	11.8169
ko00472	ko00472 D-Arginine and D-ornithine metabolism	0	3.125	0	3.33333	1.875	3.33333	1.3184524	420.825	-61.664
ko00450	ko00450 Selenocompound metabolism	30	29.375	30	34.1667	30.625	16.6667	1.2778455	356.788	-60.029
ko04150	ko04150 mTOR signaling pathway	55	44.375	43.3333	40.8333	61.25	50	1.2503239	311.578	-35.676
ko00750	ko00750 Vitamin B6 metabolism	0	1.25	0	0	1.25	0	1	424.817	-63.177
ko03018	ko03018 RNA degradation	300	285	295	323.333	271.875	295	0.9914332	-276.74	78.9232
ko04320	ko04320 Dorso-ventral axis formation	95	78.125	91.6667	91.6667	71.875	90	0.9766211	221.321	-16.266
ko00534	ko00534 Glycosaminoglycan biosynthesis - heparan sulfate / heparin	7.5	6.25	6.66667	5.83333	6.875	3.33333	0.3497619	411.42	-61.934

Table S2: List of KEGG orthologs (transporter genes) and normalized read counts employed for PCA in Figure 10

| K-number | Gene product | Normalized read counts in each gut library | | | | | | Eigenvalues | |
		MD1	MD2	MS1	MS2	MS3	P1	PC1	PC2
K01672	Carbonic anhydrase	0	2.5	40	44.16667	36.25	1.667	-1.63815	-0.63577
K01551	Arsenite-transporting ATPase	2.5	3.75	0	5	1.875	1.667	0.096531	0.548632
K01530	Phospholipid-translocating ATPase	35	13.75	5	10	6.875	23.333	2.148722	1.123722
K01537	Ca^{2+}-transporting ATPase	7.5	1.25	1.666667	2.5	1.25	0.000	1.719533	-0.59583
K05850	Ca^{2+}-transporting ATPase, plasma membrane	2.5	6.875	0	3.333333	4.375	28.333	0.393742	2.557798
K05853	Ca^{2+} transporting ATPase, sarcoplasmic/endoplasmic reticulum	15	10	11.66667	5	4.375	11.667	2.130849	0.734075
K01539	Sodium/potassium-transporting ATPase subunit alpha	27.5	21.875	11.66667	6.666667	3.125	53.333	1.387571	2.281414
K01540	Sodium/potassium-transporting ATPase subunit beta	20	9.375	10	12.5	24.375	3.333	0.429381	-1.2132
K17686	Cu^{2+}-exporting ATPase	7.5	1.25	1.666667	0.833333	0.625	0.000	2.037441	-0.49173
K13525	Transitional endoplasmic reticulum ATPase	27.5	13.125	8.333333	9.166667	17.5	13.333	1.993142	-0.0142
K02127	F-type H^+-transporting ATPase subunit b	5	6.875	16.66667	42.5	16.25	11.667	-1.51333	0.318277
K02128	F-type H^+-transporting ATPase subunit c	42.5	76.25	68.33333	167.5	109.375	81.667	-1.57985	0.530704
K02129	F-type H^+-transporting ATPase subunit e	0	0.625	0	2.5	0	0.000	-0.84368	0.40403
K02130	F-type H^+-transporting ATPase subunit f	0	0	1.666667	4.166667	1.25	15.000	-0.15396	2.504564
K02131	F-type H^+-transporting ATPase subunit 6	0	1.875	10	9.166667	3.75	1.667	-1.45442	-0.29323
K02132	F-type H^+-transporting ATPase subunit alpha	40	65.625	71.66667	139.1667	55	65.000	-1.30379	0.605638
K02133	F-type H^+-transporting ATPase subunit beta	52.5	80	98.33333	215	95	71.667	-1.41691	0.22066
K02134	F-type H^+-transporting ATPase subunit delta	2.5	13.125	8.333333	9.166667	5	1.667	-0.31397	-0.42494
K02135	F-type H^+-transporting ATPase subunit epsilon	0	1.875	3.333333	1.666667	1.875	0.000	-0.96909	-0.92006
K02136	F-type H^+-transporting ATPase subunit gamma	27.5	24.375	25	75.83333	30	26.667	-1.01148	0.391364
K02137	F-type H^+-transporting ATPase subunit O	0	8.125	5	40	15	16.667	-1.40046	0.998945
K02138	F-type H^+-transporting ATPase subunit d	2.5	5	5	21.66667	11.25	16.667	-1.31284	1.578713
K02140	F-type H^+-transporting ATPase subunit g	0	0.625	1.666667	6.666667	1.25	11.667	-0.52393	2.409266

K02144	V-type H$^+$-transporting ATPase subunit H	27.5	30.625	78.33333	110	115.625	18.333	-1.53662	-0.75528
K02145	V-type H$^+$-transporting ATPase subunit A	22.5	55.625	100	344.1667	103.75	26.667	-1.35613	0.033817
K02146	V-type H$^+$-transporting ATPase subunit d	32.5	44.375	70	103.3333	87.5	10.000	-1.37635	-0.9619
K02147	V-type H$^+$-transporting ATPase subunit B	45	94.375	265	438.3333	381.25	13.333	-1.58747	-0.69514
K02148	V-type H$^+$-transporting ATPase subunit C	12.5	10.625	56.66667	111.6667	63.75	3.333	-1.52857	-0.47347
K02149	V-type H$^+$-transporting ATPase subunit D	30	15	120	102.5	116.875	8.333	-1.34587	-0.98137
K02150	V-type H$^+$-transporting ATPase subunit E	17.5	20.625	73.33333	150	106.875	8.333	-1.57288	-0.50534
K02151	V-type H$^+$-transporting ATPase subunit F	2.5	4.375	3.333333	26.66667	28.125	3.333	-1.34263	-0.17301
K02152	V-type H$^+$-transporting ATPase subunit G	20	26.25	46.66667	130.8333	85.625	5.000	-1.43658	-0.45631
K02153	V-type H$^+$-transporting ATPase subunit e	10	16.25	15	44.16667	30.625	3.333	-1.29932	-0.50441
K02154	V-type H$^+$-transporting ATPase subunit a	52.5	29.375	143.3333	117.5	111.875	8.333	-1.12544	-1.15342
K02155	V-type H$^+$-transporting ATPase 16kDa proteolipid subunit	150	132.5	421.6667	582.5	350	28.333	-1.36913	-0.79567
K03661	V-type H$^+$-transporting ATPase 21kDa proteolipid subunit	17.5	28.125	48.33333	80.83333	58.125	5.000	-1.45836	-0.77021
K03662	V-type H$^+$-transporting ATPase S1 subunit	67.5	37.5	123.3333	94.16667	82.5	6.667	-0.64555	-1.39237
K05643	ATP-binding cassette, subfamily A (ABC1), member 3	20	8.75	83.33333	45	66.875	0.000	-1.01352	-1.14807
K05658	ATP-binding cassette, subfamily B (MDR/TAP), member 1	40	36.25	20	21.66667	36.875	1.667	1.111135	-1.38113
K05662	ATP-binding cassette, subfamily B (MDR/TAP), member 7	0	1.875	5	3.333333	1.875	3.333	-1.28805	0.79732
K05665	ATP-binding cassette, subfamily C (CFTR/MRP), member 1	2.5	3.125	5	5	6.25	11.667	-0.50029	2.231987
K05668	ATP-binding cassette, subfamily C (CFTR/MRP), member 5	35	27.5	20	15	17.5	3.333	1.732806	-1.20369
K05673	ATP-binding cassette, subfamily C (CFTR/MRP), member 4	12.5	10.625	1.666667	5.833333	4.375	1.667	1.867491	-0.36554
K05674	ATP-binding cassette, subfamily C (CFTR/MRP), member 10	5	1.25	3.333333	0.833333	0	0.000	1.725873	-0.71597
K05676	ATP-binding cassette, subfamily D (ALD), member 2	12.5	3.75	3.333333	0	0.625	1.667	2.312	-0.20631
K06158	ATP-binding cassette, subfamily F, member 3	0	0.625	1.666667	0	1.25	0.000	-0.4773	-0.89109
K06174	ATP-binding cassette, sub-family E, member 1	2.5	8.125	5	6.666667	5	11.667	-0.26563	2.382532
K06184	ATP-binding cassette, subfamily F, member 1	0	0	0	2.5	1.875	0.000	-1.3565	-0.07406
K05011	Chloride channel 2	2.5	1.25	5	3.333333	7.5	1.667	-0.92202	-0.8785
K05012	Chloride channel 3/4/5	12.5	6.875	20	9.166667	17.5	3.333	-0.19209	-1.45454
K05016	Chloride channel 7	0	0.625	1.666667	4.166667	1.875	1.667	-1.71535	0.653085
K05019	Chloride channel, nucleotide-sensitive, 1A	0	2.5	0	5	1.875	0.000	-0.97412	0.105822
K04978	Transient receptor potential cation channel subfamily M member 3	0	0	5	0	0	5.000	-0.04308	1.44541

K05315	Voltage-dependent calcium channel alpha 1, invertebrate	0	0.625	1.666667	0	0	8.333	0.29186	2.42758
K14686	Solute carrier family 31 (copper transporter), member 1	27.5	16.875	8.333333	17.5	5	5.000	1.681354	-0.321
K15122	Solute carrier family 41	15	1.875	13.33333	0	0.625	1.667	1.559602	-0.55829
K15040	Voltage-dependent anion channel protein 2	20	61.875	46.66667	107.5	63.125	30.000	-1.41693	-0.03068
K07232	Cation transport protein ChaC (glutathion transport)	5	16.875	35	5.833333	23.75	0.000	-0.40481	-1.1096
K14688	Solute carrier family 30 (zinc transporter), member 1	27.5	16.875	8.333333	17.5	5	5.000	1.681354	-0.321
K14692	Solute carrier family 30 (zinc transporter), member 5/7	5	7.5	3.333333	6.666667	2.5	0.000	0.549333	-0.71581
K14711	Solute carrier family 39 (zinc transporter), member 5	2.5	2.5	0	0	1.25	0.000	1.969689	-0.42692
K14713	Solute carrier family 39 (zinc transporter), member 7	15	4.375	5	4.166667	2.5	0.000	1.876985	-0.74105
K14714	Solute carrier family 39 (zinc transporter), member 8	2.5	0	1.666667	0	2.5	0.000	0.854824	-1.08623
K14715	Solute carrier family 39 (zinc transporter), member 9	0	2.5	1.666667	0.833333	3.75	0.000	-0.58302	-0.76977
K14719	Solute carrier family 39 (zinc transporter), member 13	42.5	21.25	40	58.33333	53.125	8.333	-0.63394	-1.24323
K14708	Solute carrier family 26 (sodium-independent sulfate anion transporter), member 11	0	1.875	86.66667	62.5	46.875	0.000	-1.38998	-0.71048
K13855	Solute carrier family 4 (anion exchanger), member 2	0	0.625	0	0.833333	4.375	3.333	-0.56085	1.235903
K12040	Solute carrier family 9 (sodium/hydrogen exchanger), member 3	0	0.625	13.33333	1.666667	4.375	0.000	-0.61418	-0.69804
K12041	Solute carrier family 9 (sodium/hydrogen exchanger), member 6/7	25	7.5	155	107.5	117.5	1.667	-1.26743	-0.96173
K14724	Solute carrier family 9 (sodium/hydrogen exchanger), member 8	5	3.125	0	0.833333	1.875	1.667	2.244894	0.211436
K14427	Solute carrier family 12 (potassium/chloride transporter), member 4/5/6	25	5.625	3.333333	4.166667	4.375	0.000	2.012078	-0.51364
K14428	Solute carrier family 12 (potassium/chloride transporters), member 8	20	13.75	8.333333	1.666667	2.5	1.667	2.254246	-0.51362
K05038	Solute carrier family 6 (neurotransmitter transporter, amino acid) member 5/7/9/14	45	25	11.66667	3.333333	12.5	0.000	2.181029	-0.71514
K05336	Solute carrier family 6 (neurotransmitter transporter), invertebrate	2.5	4.375	3.333333	9.166667	8.125	0.000	-1.13977	-0.78741
K13868	Solute carrier family 7 (L-type amino acid transporter), member 9	0	0	5	4.166667	1.875	35.000	0.068818	2.477038
K14209	Solute carrier family 36 (proton-coupled amino acid transporter)	37.5	26.875	18.33333	13.33333	11.25	26.667	2.420639	0.873429
K08193	MFS transporter, ACS family, solute carrier family 17 (sodium-dependent inorganic phosphate cotransporter), other	10	11.875	1.666667	0	0	0.000	2.105053	-0.17367
K12303	MFS transporter, ACS family, solute carrier family 17 (sodium-dependent inorganic phosphate cotransporter), member 9	0	0.625	3.333333	0.833333	1.25	0.000	-0.76581	-0.7889
K08202	MFS transporter, OCT family, solute carrier family 22 (organic cation transporter), member 4/5	5	3.125	0	0	0.625	0.000	2.292194	-0.2687
K14206	Solute carrier family 15 (oligopeptide transporter), member 1	52.5	46.25	1.666667	0	0	1.667	2.266271	-0.09162
K14343	Solute carrier family 10 (sodium/bile acid cotransporter), member 3/5	2.5	1.875	3.333333	2.5	6.25	1.667	-0.62627	-0.85391

K07299	MFS transporter, SP family, solute carrier family 2 (facilitated glucose transporter), member 1	0	3.125	0	0.833333	0.625	1.667	0.465347	1.383962
K13783	MFS transporter, OPA family, solute carrier family 37 (glycerol-3-phosphate transporter), member 1/2	2.5	3.125	3.333333	1.666667	1.25	3.333	1.21554	1.287578
K15272	Solute carrier family 35 (UDP-sugar transporter), member A1/2/3	0	1.25	1.666667	0.833333	1.25	0.000	-0.88953	-0.95692
K15277	Solute carrier family 35 (adenosine 3'-phospho 5'-phosphosulfate transporter), member B3	2.5	5.625	1.666667	7.5	1.25	1.667	-0.20082	0.35214
K15278	Solute carrier family 35 (UDP-xylose/UDP-N-acetylglucosamine transporter), member B4	10	2.5	0	5	0.625	6.667	1.67432	1.259844
K15279	Solute carrier family 35 (GDP-fucose transporter), member C1	0	1.875	1.666667	0	0.625	0.000	0.244177	-0.50198
K15280	Solute carrier family 35, member C2	2.5	3.125	1.666667	5	3.75	1.667	-0.78883	-0.23857
K15281	Solute carrier family 35	20	3.75	1.666667	0	1.25	0.000	2.197536	-0.32074
K15283	Solute carrier family 35, member E1	0	2.5	0	1.666667	1.25	0.000	-0.28717	-0.00527
K15284	Solute carrier family 35, member E2	10	1.875	1.666667	0	0.625	0.000	2.178994	-0.39981
K15377	Solute carrier family 44 (choline transporter-like protein), member 2/4/5	32.5	42.5	35	29.16667	34.375	1.667	0.447301	-1.61019
K13354	Solute carrier family 25 (peroxisomal adenine nucleotide transporter), member 17	2.5	1.25	3.333333	2.5	4.375	6.667	-0.23611	1.882954
K15119	Solute carrier family 25, member 39/40	0	1.875	0	1.666667	1.25	3.333	-0.26199	2.36076
K05863	Solute carrier family 25 (mitochondrial adenine nucleotide translocator), member 4/5/6/31	80	108.125	165	255.8333	122.5	306.667	-0.9137	2.149374
K14684	Solute carrier family 25 (mitochondrial phosphate transporter), member 23/24/25/41	0	1.25	10	10	11.875	11.667	-1.61091	1.022883
K15101	Solute carrier family 25 (mitochondrial ornithine transporter) member 2/15	0	1.875	3.333333	6.666667	5.625	8.333	-1.36163	1.824993
K15102	Solute carrier family 25 (mitochondrial phosphate transporter), member 3	7.5	51.875	150	364.1667	181.25	98.333	-1.71786	0.258543
K15105	Solute carrier family 25 (mitochondrial aspartate/glutamate transporter), member 12/13	2.5	3.75	0	0.833333	0	6.667	1.249847	2.360935
K15108	Solute carrier family 25 (mitochondrial thiamine pyrophosphate transporter), member 19	0	0	3.333333	2.5	2.5	0.000	-1.46409	-0.77013
K15109	Solute carrier family 25 (mitochondrial carnitine/acylcarnitine transporter), member 20/29	2.5	1.875	3.333333	0.833333	1.875	0.000	0.838452	-1.5091
K15111	Solute carrier family 25 (mitochondrial S-adenosylmethionine transporter), member 26	0	0.625	1.666667	0	1.25	0.000	-0.4773	-0.89109
K15113	Solute carrier family 25 (mitochondrial iron transporter), member 28/37	0	0.625	0	0	1.875	0.000	-0.28995	-0.3737
K01528	Dynamin GTPase	0	0	3.333333	0	4.375	0.000	-0.71952	-0.79441
K08745	Solute carrier family 27 (fatty acid transporter), member 1/4	32.5	28.75	5	5.833333	9.375	1.667	2.111232	-0.43912
K06519	Solute carrier family 3, member 2 (accessory protein facilitates transport of amino acids and polyamines)	7.5	1.875	5	3.333333	3.125	1.667	1.358898	-0.89411
K15382	Solute carrier family 50 (sugar transporter)	40	44.375	43.33333	25	56.25	8.333	0.409028	-1.59577

Table S3: List of KEGG orthologs encoding transporters and categorized based on the Transporter Classification Database (http://www.tcdb.org/) (Saier *et al.* 2014). Functions of the detected transporters were manually annotated.

K-number	Gene or Family	Category	Description	Reference
K01672	**CA**	Carbonic anhydrase	Catalyse the reversible hydration of carbon dioxide to bicarbonate	Corena *et al.* (2004) Comp. Biochem. Physiol. 137C, 207-298
K01551	**ATPat**	P-ATPase	Active arsenite-exporting transporter	Clark *et al.* (2007) Nature. 450, 203-218
K01530	**Flippase**	P-ATPase	Active phospholipids transporter	Clark *et al.* (2007) Nature. 450, 203-218
K01537	**PMCA**	P-ATPase	Active Ca^{2+} transporter	Krebs and Guerini. (1996) Biomembr. 5, 101-131
K05850	**PMCA**	P-ATPase	Active Ca^{2+} transporter	Krebs and Guerini. (1996) Biomembr. 5, 101-131
K05853	**SERCA**	P-ATPase	Active Ca^{2+} transporter	Krebs and Guerini. (1996) Biomembr. 5, 101-131
K01539	**ATPNa$^+$/K$^+$ alpha**	P-ATPase	Active Na^+/K^+ transporter	Khodabandeh S. (2006) Zoological Studies. 45, 510–516
K01540	**ATPNa$^+$/K$^+$ beta**	P-ATPase	Active Na^+/K^+ transporter	Khodabandeh S. (2006) Zoological Studies. 45, 510–516
K17686	**ATPCu^{2+}**	P-ATPase	Active Cu^{2+}-exporting transporter	Gupta A. (2012) Curr. Genomics. 13, 124-133
K13525	**VCP**	P-ATPase	Active ER retrotranslocon transporter	Bolte *et al.* (2011) Bioessays, 33, 368-376
K02127	**ATPeF0B**	F-Type ATP synthases	Active proton transporter and ATP synthase	Müller and Grüber. (2003) Cell. Mol. Life Sci. 60, 474–494
K02128	**ATPF0C**	F-Type ATP synthases	Active proton transporter and ATP synthase	Müller and Grüber. (2003) Cell. Mol. Life Sci. 60, 474–494
K02129	**ATPeF0E**	F-Type ATP synthases	Active proton transporter and ATP synthase	Müller and Grüber. (2003) Cell. Mol. Life Sci. 60, 474–494
K02130	**ATPeF0F**	F-Type ATP synthases	Active proton transporter and ATP synthase	Müller and Grüber. (2003) Cell. Mol. Life Sci. 60, 474–494
K02131	**ATPeF0F6**	F-Type ATP synthases	Active proton transporter and ATP synthase	Müller and Grüber. (2003) Cell. Mol. Life Sci. 60, 474–494
K02132	**ATPeF1A**	F-Type ATP synthases	Active proton transporter and ATP synthase	Müller and Grüber. (2003) Cell. Mol. Life Sci. 60, 474–494
K02133	**ATPeF1B**	F-Type ATP synthases	Active proton transporter and ATP synthase	Müller and Grüber. (2003) Cell. Mol. Life Sci. 60, 474–494
K02134	**ATPeF1D**	F-Type ATP synthases	Active proton transporter and ATP synthase	Müller and Grüber. (2003) Cell. Mol. Life Sci. 60, 474–494
K02135	**ATPeF1E**	F-Type ATP synthases	Active proton transporter and ATP synthase	Müller and Grüber. (2003) Cell. Mol. Life Sci. 60, 474–494
K02136	**ATPeF1G**	F-Type ATP synthases	Active proton transporter and ATP synthase	Müller and Grüber. (2003) Cell. Mol. Life Sci. 60, 474–494
K02137	**ATPeF0O**	F-Type ATP synthases	Active proton transporter and ATP synthase	Müller and Grüber. (2003) Cell. Mol. Life Sci. 60, 474–494
K02138	**ATPeFD**	F-Type ATP synthases	Active proton transporter and ATP synthase	Müller and Grüber. (2003) Cell. Mol. Life Sci. 60, 474–494
K02140	**ATPeFG**	F-Type ATP synthases	Active proton transporter and ATP synthase	Müller and Grüber. (2003) Cell. Mol. Life Sci. 60, 474–494
K02144	**ATPeV1H**	V-ATPases	Active proton transporter	Wieczorek et al. (2003) J Bioenerg Biomembr. 35, 359-366
K02145	**ATPeV1A**	V-ATPases	Active proton transporter	Wieczorek et al. (2003) J Bioenerg Biomembr. 35, 359-366

K02146	**ATPeV0D**	V-ATPases	Active proton transporter	Wieczorek et al. (2003) J Bioenerg Biomembr. 35, 359-366
K02147	**ATPeV1B**	V-ATPases	Active proton transporter	Wieczorek et al. (2003) J Bioenerg Biomembr. 35, 359-366
K02148	**ATPeV1C**	V-ATPases	Active proton transporter	Wieczorek et al. (2003) J Bioenerg Biomembr. 35, 359-366
K02149	**ATPeV1D**	V-ATPases	Active proton transporter	Wieczorek et al. (2003) J Bioenerg Biomembr. 35, 359-366
K02150	**ATPeV1E**	V-ATPases	Active proton transporter	Wieczorek et al. (2003) J Bioenerg Biomembr. 35, 359-366
K02151	**ATPeV1F**	V-ATPases	Active proton transporter	Wieczorek et al. (2003) J Bioenerg Biomembr. 35, 359-366
K02152	**ATPeV1G**	V-ATPases	Active proton transporter	Wieczorek et al. (2003) J Bioenerg Biomembr. 35, 359-366
K02153	**ATPeV0E**	V-ATPases	Active proton transporter	Wieczorek et al. (2003) J Bioenerg Biomembr. 35, 359-366
K02154	**ATPeV0A**	V-ATPases	Active proton transporter	Wieczorek et al. (2003) J Bioenerg Biomembr. 35, 359-366
K02155	**ATPeV0C**	V-ATPases	Active proton transporter	Wieczorek et al. (2003) J Bioenerg Biomembr. 35, 359-366
K03661	**ATPeV0B**	V-ATPases	Active proton transporter	Wieczorek et al. (2003) J Bioenerg Biomembr. 35, 359-366
K03662	**ATPeVS1**	V-ATPases	Active proton transporter	Wieczorek et al. (2003) J Bioenerg Biomembr. 35, 359-366
K05643	**ABCA3**	ABC transporter	Active substrate transporter	Broehan et al. (2013) BMC Genomics. 14:6
K05658	**ABCB1**	ABC transporter	Active substrate transporter	Broehan et al. (2013) BMC Genomics. 14:6
K05662	**ABCB7**	ABC transporter	Active substrate transporter	Broehan et al. (2013) BMC Genomics. 14:6
K05665	**ABCC1**	ABC transporter	Active substrate transporter	Broehan et al. (2013) BMC Genomics. 14:6
K05668	**ABCC5**	ABC transporter	Active substrate transporter	Broehan et al. (2013) BMC Genomics. 14:6
K05673	**ABCC4**	ABC transporter	Active substrate transporter	Broehan et al. (2013) BMC Genomics. 14:6
K05674	**ABCC10**	ABC transporter	Active substrate transporter	Broehan et al. (2013) BMC Genomics. 14:6
K05676	**ABCD2**	ABC transporter	Active substrate transporter	Broehan et al. (2013) BMC Genomics. 14:6
K06158	**ABCF3**	ABC transporter	Non-transporter	Dermauw and Leeuwen (2014) Insect Biochem. Mol. Biol. 45, 89-110
K06174	**ABCE1**	ABC transporter	Non-transporter	Dermauw and Leeuwen (2014) Insect Biochem. Mol. Biol. 45, 89-110
K06184	**ABCF1**	ABC transporter	Non-transporter	Dermauw and Leeuwen (2014) Insect Biochem. Mol. Biol. 45, 89-110
K05011	**CLCN2**	α-Type channels	Transmembrane Cl⁻channels	Moffett D. F. (1994) Adv. Comp. Environ. Physiol. 19, 155-172
K05012	**CLCN3_4_5**	α-Type channels	Transmembrane Cl⁻channels	Moffett D. F. (1994) Adv. Comp. Environ. Physiol. 19, 155-172
K05016	**CLCN7**	α-Type channels	Transmembrane Cl⁻channels	Moffett D. F. (1994) Adv. Comp. Environ. Physiol. 19, 155-172
K05019	**CLNS1A**	α-Type channels	Transmembrane Cl⁻channels	Moffett D. F. (1994) Adv. Comp. Environ. Physiol. 19, 155-172
K04978	**TRPM**	α-Type channels	Transmembrane Ca^{2+} permeable cation channels	Ranganathan *et al.* (1995) Annu. Rev. Neurosci. 18, 283–317
K05315	**VDCC**	α-Type channels	Transmembrane Ca^{2+} permeable cation channels	Gielow *et al.* (1995) J Neurosci. 15, 6085– 6093

K14686	**SLC31A1**	α-Type channels	Transmembrane Cu^{2+} ion transporter	Petris M. J. (2004) Pflügers Archiv. 447, 752-755.
K15122	**SLC41**	α-Type channels	Transmembrane Mg^{2+} ion transporter	Sahni and Scharenberg. (2013) Molecular aspects of medicine. 34, 620 - 628.
K15040	**VDAC2**	α-Type channels	Transmembrane macro and small molecules transport channels (mitochondrial)	Ryerse *et al.* (1997) Biochim. Biophys. Acta. 1327, 193–203
K07232	**ChaC**	Electrochemical potential-driven transporters	Glutathione transporter	Quresh *et al.* (2013) Journal of Asia-Pacific Entomology. 16, 301-307
K14688	**SLC30A1**	Electrochemical potential-driven transporters	Zinc transporter	Montanini *et al.* (2007) BMC Genomics. 8, 107
K14692	**SLC30A5_7**	Electrochemical potential-driven transporters	Zinc transporter	Montanini *et al.* (2007) BMC Genomics. 8, 107
K14711	**SLC39A5**	Electrochemical potential-driven transporters	Zinc transporter	Eide D.J. (2004) Pflugers Arch - Eur. J. Physiol. 447, 796–800
K14713	**SLC39A7**	Electrochemical potential-driven transporters	Zinc transporter	Eide D.J. (2004) Pflugers Arch - Eur. J. Physiol. 447, 796–800
K14714	**SLC39A8**	Electrochemical potential-driven transporters	Zinc transporter	Eide D.J. (2004) Pflugers Arch - Eur. J. Physiol. 447, 796–800
K14715	**SLC39A9**	Electrochemical potential-driven transporters	Zinc transporter	Eide D.J. (2004) Pflugers Arch - Eur. J. Physiol. 447, 796–800
K14719	**SLC39A13**	Electrochemical potential-driven transporters	Zinc transporter	Eide D.J. (2004) Pflugers Arch - Eur. J. Physiol. 447, 796–800
K14708	**SLC26A11**	Electrochemical potential-driven transporters	Sodium-independent sulfate transporter	Vincourt et al. (2003) FASEB J. 17, 890-892
K13855	**SLC4A2**	Electrochemical potential-driven transporters	Cl^-/HCO_3^- exchange transporter	Piermarini *et al.* (2010) Am. J. Physiol. Regul. Integr. Comp. Physiol. 298, R642–R660.
K12040	**SLC9A3**	Electrochemical potential-driven transporters	Cation:proton antiporter	Donowitz *et al.* (2013) Mol Aspects Med. 34, 236–251
K12041	**SLC9A6_7**	Electrochemical potential-driven transporters	Cation:proton antiporter	Donowitz *et al.* (2013) Mol Aspects Med. 34, 236–251
K14724	**SLC9A8**	Electrochemical potential-driven transporters	Cation:proton antiporter	Piermarini *et al.* (2009) Am. J. Physiol. Renal. Physiol. 296, F730-F750
K14427	**SLC12A4_5_6**	Electrochemical potential-driven transporters	Cation:chloride cotransporter	Pullikuth *et al.* (2003) J. Exp. Biol. 206, 3857-3868
K14428	**SLC12A8**	Electrochemical potential-driven transporters	Cation:chloride cotransporter	Pullikuth *et al.* (2003) J. Exp. Biol. 206, 3857-3868
K05038	**SLC6A5_7_9_14**	Electrochemical potential-driven transporters	Amino acid transporter	Boudko D.Y. (2012) J. Insect Physiol. 58, 433-449
K05336	**SLC6A**	Electrochemical potential-driven transporters	Amino acid transporter	Boudko D.Y. (2012) J. Insect Physiol. 58, 433-449
K13868	**SLC7A9**	Electrochemical potential-driven transporters	Amino acid transporter	Boudko D.Y. (2012) J. Insect Physiol. 58, 433-449
K14209	**SLC36**	Electrochemical potential-driven	Proton-dependent small amino acid transporter	Boudko D.Y. (2012) J. Insect. Physiol. 58, 433–449

transporters

K08193	**SLC17A**	Electrochemical potential-driven transporters	Sodium-dependent inorganic phosphate cotransporter	Sreedharan *et al.* (2010) BMC Genomics. 11, 17
K12303	**SLC17A9**	Electrochemical potential-driven transporters	Sodium-dependent inorganic phosphate cotransporter	Sreedharan *et al.* (2010) BMC Genomics. 11, 17
K08202	**SLC22A4_5**	Electrochemical potential-driven transporters	Organic cation/carnitine transporter	Eraly *et al.* (2004) Physiol. Genomics. 18, 12-24
K14206	**SLC15A1**	Electrochemical potential-driven transporters	Oligopeptide transporter	Smith *et al.* (2013) Mol. Aspects. Med. 34, 323–336
K14343	**SLC10A3_5**	Electrochemical potential-driven transporters	Sodium/bile acid cotransporter	da Silva *et al.* (2013) Mol. Aspects. Med. 34, 252–269
K07299	**SLC2A1**	Electrochemical potential-driven transporters	Facilitated glucose transporter	Stock *et al.* (2013) PLoS ONE. 8, e84461
K13783	**SLC37A1_2**	Electrochemical potential-driven transporters	Glycerol-3-phosphate transporters	Reddy *et al.* (2012) FEBS. J. 279, 2022–2035
K15272	**SLC35A1_2_3**	Electrochemical potential-driven transporters	UDP-sugar transporter	Reddy *et al.* (2012) FEBS. J. 279, 2022–2035
K15277	**SLC35B3**	Electrochemical potential-driven transporters	Adenosine 3'-phospho 5'-phosphosulfate transporter	Reddy *et al.* (2012) FEBS. J. 279, 2022–2035
K15278	**SLC35B4**	Electrochemical potential-driven transporters	UDP-xylose/UDP-N-acetylglucosamine transporter	Reddy *et al.* (2012) FEBS. J. 279, 2022–2035
K15279	**SLC35C1**	Electrochemical potential-driven transporters	GDP-fucose transporter	Reddy *et al.* (2012) FEBS. J. 279, 2022–2035
K15280	**SLC35C2**	Electrochemical potential-driven transporters	Triose-phosphate transporter	Reddy *et al.* (2012) FEBS. J. 279, 2022–2035
K15281	**SLC35**	Electrochemical potential-driven transporters	Triose-phosphate transporter	Reddy *et al.* (2012) FEBS. J. 279, 2022–2035
K15283	**SLC35E1**	Electrochemical potential-driven transporters	Triose-phosphate transporter	Reddy *et al.* (2012) FEBS. J. 279, 2022–2035
K15284	**SLC35E2**	Electrochemical potential-driven transporters	Triose-phosphate transporter	Reddy *et al.* (2012) FEBS. J. 279, 2022–2035
K15377	**SLC44A2_4_5**	Electrochemical potential-driven transporters	Choline transporter	Saier M. H. (2000) Microbiol. Mol. Biol. Rev. 64, 354–411
K13354	**SLC25A17**	Electrochemical potential-driven transporters	Mitochondrial adenine nucleotide transporter	Walker and Runswick. (1994) . In Molecular Biology of Mitochondrial Transport Systems. 41-53
K15119	**SLC25A39_40**	Electrochemical potential-driven transporters	Mitochondrial transporter	Haitina *et al.* (2006) Genomics. 88, 779–790
K05863	**SLC25A4_5_6_31**	Electrochemical potential-driven transporters	Mitochondrial adenine nucleotide transporter	Walker and Runswick. (1994) In Molecular Biology of Mitochondrial Transport Systems. 41-53
K14684	**SLC25A23_24_25_41**	Electrochemical potential-driven transporters	Mitochondrial phosphate transporter	Walker and Runswick. (1994) In Molecular Biology of Mitochondrial Transport Systems. 41-53
K15101	**SLC25A2_15**	Electrochemical potential-driven transporters	Mitochondrial ornithine transporter	Walker and Runswick. (1994) In Molecular Biology of Mitochondrial Transport Systems. 41-53

K15102	**SLC25A3**	Electrochemical potential-driven transporters	Mitochondrial phosphate transporter	Walker and Runswick. (1994) In Molecular Biology of Mitochondrial Transport Systems. 41-53
K15105	**SLC25A12_13**	Electrochemical potential-driven transporters	Mitochondrial aspartate/glutamate transporter	Arco and Satrustegui. (2005) Cell. Mol. Life Sci. 62, 2204–2227
K15108	**SLC25A19**	Electrochemical potential-driven transporters	Mitochondrial thiamine pyrophosphate transporter	Arco and Satrustegui. (2005) Cell. Mol. Life Sci. 62, 2204–2227
K15109	**SLC25A20_29**	Electrochemical potential-driven transporters	Mitochondrial carnitine/acylcarnitine transporter	Arco and Satrustegui. (2005) Cell. Mol. Life Sci. 62, 2204–2227
K15111	**SLC25A26**	Electrochemical potential-driven transporters	Mitochondrial S-adenosylmethionine transporter	Arco and Satrustegui. (2005) Cell. Mol. Life Sci. 62, 2204–2227
K15113	**SLC25A28_37**	Electrochemical potential-driven transporters	Mitochondrial iron transporter	Arco and Satrustegui. (2005) Cell. Mol. Life Sci. 62, 2204–2227
K01528	**DNM**	GTPases	Formation of coated vesicles facilitator	Chen *et al*. (1991) Nature. 351, 583 - 586
K08745	**SLC27A1_4**	Phosphotransfer-driven group translocators	Fatty acid transporter	Stahl A. (2004) Pflug. Arch. 447, 722–727
K06519	**SLC3A2**	Auxiliary transport proteins	Amino acids and polyamines transporter	Palacin and Kanai. (2004) Pflug. Arch. 447, 490-494
K15382	**SLC50**	Transporters with unknown biochemical mechanism	Sugar transporter	Wright E. M. (2013) Mol. Aspects Med. 34, 183–196