Title
Expression of Plant Cell Wall Degrading Enzymes and Digestive Physiology in Phasmatodea

Author
Lin, Wei

Publication Date
2021

Peer reviewed|Thesis/dissertation

Expression of Plant Cell Wall Degrading Enzymes and Digestive Physiology in Phasmatodea

By

Wei Lin

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Entomology

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

Brian Johnson, Chair

Philip Ward

Geoffrey Attardo

Committee in Charge

2021

Table of Contents

ABSTRACT

Plant cell walls (PCWs) are the most abundant biomass on planet Earth. Unlike vertebrate animals, many invertebrates, especially insects, can digest PCWs, either to access the protoplasm or to break the wall itself into digestible, nutritive oligomers and monomers (Pauchet, Wilkinson et al. 2010, Calderon-Cortes, Quesada et al. 2012). This ability of many insects to break down recalcitrant PCWs has contributed greatly to their biodiversity and success around the world (Tokuda 2019). Traditionally, the breakdown of lignocellulosic and cellulosic material by insects was thought to be dependent on gastrointestinal microbes. Recent research with molecular and omics techniques (genomics, transcriptomics, and proteomics) has provided evidence to reveal PCW-degrading mechanisms in insects that use various arsenals of endogenous digestive enzymes. A few studies have reported on the endogenous PCW-degrading enzymes (PCWDEs) of Phasmatodea or stick insects. Much is still unknown with respect to how and where PCWDEs perform in a phasmid's digestive tract, in addition to the question of whether gut microbial PCWDEs are present to contribute. This dissertation applies RNA-Seq and metagenomic analysis to identify the repertoire of endogenous PCWDEs in *Aretaon asperrimus* and *Medauroidea extradentata*, to assess the expression profiles of these enzymes through the alimentary canal, and to investigate the potential digestive role that could be played by gut microbes. We identify abundant endogenous PCWDE repertoires in each species, including GH9 endoglucanases, GH1 cellobiases, and GH28 polygalacturonases. Most PCWDE families are highly expressed in the anterior section of the midgut, while beta-glucosidases show a global expression pattern in multiple tissues, including the head. We observed very low expressions of GH9 and GH28 in hindgut tissues (ileum and rectum). Hemicellulases from GH2, GH3, and GH30 also showed global expression profiles without high expression in the AMG. The compartmentalization of phasmid digestion is investigated by an annotation and enrichment test for differentially expressed genes and the 1,000 most highly expressed genes in each tissue. The metagenomic study indicates that the dominant bacterial phylum, Proteobacteria, in both species' gut communities, shows a great number of polygalacturonase sequences in

the posterior section, which infers a potential pectin degradation synergism between endogenous and bacterial pectinases in different sections of a phasmid's midgut. A lesser capability for PCW-degrading was observed in both species' midgut microbiota, with a lack of cellulolytic Fibrobacteres and Spirochaetes. Overall, the identified transcripts of various glycoside hydrolase families in this dissertation further support that phasmids have uncommonly rich endogenous repertoires of PCWDEs. All these sequences could benefit future evolutionary analyses of these gene families and, more broadly, the phylogenetic analyses in Phasmatodea. Our results from the expression profiles of identified PCWDEs take a step further to illustrate how these endogenous digestive enzymes perform differently among tissues. The discovery of many bacterial PCWDEs in phasmid's midgut in this dissertation indicates a possible synergism in the degradation of PCWs, which contradicts previous claims. All these findings contribute to the understanding of the digestive physiology in Phasmatodea and provide genomic data for future studies.

INTRODUCTION

The ability of herbivorous insects to consume plant material is one of the reasons behind the success of Insecta, which contains more than half of the living organisms on Earth (Wiens, Lapoint et al. 2015). Herbivorous insects are tremendously diverse in feeding habits and diets. Feeding habits of insect herbivores include chewing and sucking behaviors that depend largely on the mouthpart, which restricts the insect's food spectrum as well. Insects with chewing behaviors possess mandibles with sharp edges and serrated teeth, such as caterpillars in Lepidoptera, grasshoppers in Orthoptera, termites in Blattodea, bark beetles in Coleoptera, and stick insects in Phasmatodea. On the other hand, insect herbivores with piercing-sucking mouthparts mainly use their elongated stylets to secrete digestive enzymes into plant tissues and suck back the digested contents. Such insects include plant-hoppers in Hemiptera, and male mosquitoes in Diptera that feed on nectar or honeydews (Bernays 1998). Although insect herbivores evolved various structures that are adaptive to different feeding behaviors, the diet of herbivory is challenging due to the difficulty of the digestion of plant polysaccharides and the low protein levels of most plant tissues. The reasons behind the extraordinary diversity and abundance of insect herbivores remains largely unknown. Thus, herbivory in insects is a major scientific topic for entomologists.

Phasmatodea Systematics

Phasmatodeans, commonly called stick insects, are mainly nocturnal herbivores. Those few active in the day typically exhibit twig, bark, and leaf mimicry or display aposematic coloration and are toxic (Bradler, Robertson et al. 2014). More than 3200 species and about 570 genera of Phasmatodeans have been reported. Most live in the tropics while some are found in the temperate zone.

The 10 Polyneopteran orders are grasshoppers and crickets (Orthoptera), earwigs (Dermaptera), stoneflies (Plecoptera), mantids (Mantodea), cockroaches and termites (Blattodea), stick and leaf insects (Phasmatodea), heelwalkers (Mantophasmatodea), icecrawlers (Grylloblattodea), webspinners

(Embioptera) and ground lice (Zoraptera) (Wipfler, Letsch et al. 2019). The phylogenetic position of Phasmatodea among the lower Neoptera remains controversial. Almost every Orthopteroid order has been proposed to be a potential sister order to Phasmatodea (Kristensen 1975, Kjer 2004, Buckley, Attanayake et al. 2009, Friedemann, Wipfler et al. 2012). This challenging taxonomic situation comes mainly from overlap and complexity in various phenotypic traits. Increasing evidence from recent studies, however, are showing that the Embioptera may be the sister order to Phasmatodea (Terry and Whiting 2005, Misof, Liu et al. 2014, Wipfler, Letsch et al. 2019). The first evidence was reported by Rähle (1970) and was based on an additional para-glossae flexor muscle in both web-spinners and stick insects. Additional morphological characters that are shared by both Embioptera and Phasmatodea include the absence of a gonostyli in the male, a reduced ovipositor, eggs with a lid-like operculum, and nine longitudinal accessory fibers within spermatozoa (Buckley, Attanayake et al. 2009, Vallotto, Bresseel et al. 2016). Studies by Misof (2014), Shelomi (2020), and Wipfler (2019) have also presented molecular evidence supporting this sister taxonomic relationship.

The monophyly of Phasmatodea is well established with large amounts of apomorphic evidence. For instances, the presence of a pair of defensive glands on the prothorax (Kristensen 1975); the presence of distinctive appendages of unknown function on the posterior midgut (Shelomi and Kimsey 2014); the presence of a hook-bearing clasper on the ventral side of the male's abdomen (Tilgner, Kiselyova et al. 1999); the presence of unsegmented cerci, and the abdominal sternum VIII that covers a large portion of the female ovipositor (Kristensen 1975, Bradler and Buckley 2018).

With respect to division within the order: the presence and absence of a distinct triangular area located on the ventral apex of the tibiae, called the apicalis, traditionally grouped phasmids into two major categories: the Areolatae and Anareolatae. This division has not been supported by recent phylogenetic analysis, yet the presence or absence of the apicalis, is still a helpful feature to be used in phylogenetic studies (Zompro 2004). What is well-supported is a division between *Timema* (Timematodea) and all the

rest of stick insects (Euphasmatodea) (Whiting, Bradler et al. 2003, Gottardo, Mercati et al. 2012, Simon, Letsch et al. 2019). Thus, the elusive systematic questions nest within the Euphasmatodea where the relationships among the major clades are still poorly resolved. It has been suggested that a series of rapid speciation events occurred within a period of 10 million years which may explain this difficulty (Bradler and Buckley 2018). Moreover, the lack of congruency between Phasmatodea phylogeny and their geographical distributions poses challenges. In the Lanceocercata clade that contains phasmids from New Zealand, New Caledonia, Australia, and several adjacent Pacific islands, almost none of the traditional lineages are monophyletic (Buckley, Attanayake et al. 2009); however, the phylogeny of Lanceocercata reconstructed five lineages with little overlaps in geographical distribution (Buckley, Attanayake et al. 2010). These knowledge gaps, along with a lack of molecular data, greatly hinders the study of Phasmatodea (Brand, Lin et al. 2018).

The species selected in this dissertation belong to Heteropterygidae (Obriminae: *Aretaon asperrimus*) and Phasmatidae (Medaurinae: *Medauroidea extradentata*), which are both Old World species. Heteropterygidae is an areolatae group that inhabits the Oriental Region that contains more than 100 described species in approximately 30 genera. This group of phasmids have stout, robust, and thorny body shapes and are small (~4cm) to medium (~16cm) in size (Bradler and Buckley 2018). The Clitumnini phasmids have about 240 described species. They are wingless large insects (~30cm) from a range extending from India to Southeast Asia. The phylogenies of both Heteropterygidae and Clitumninae are still considered as unsolved questions (Zompro 2004, Buckley, Attanayake et al. 2009, Bradler, Cliquennois et al. 2015, Buscher, Buckley et al. 2018, Robertson, Bradler et al. 2018).

Given the knowledge gaps in our understanding of phasmid evolutionary history we could not choose phasmids based on phylogenetic placement for our studies. The difficulty of getting insects was exacerbated by most being tropical. Our focal species, *Aretaon asperrimus* (Thorny stick insect, abbreviation in this dissertation: ARE) and *Medauroidea extradentata* (Vietnamese stick insect,

abbreviation in this dissertation: VIET), were thus chosen due to their ease of attaining and our ability to house them in the lab.

Phasmid's Digestive Physiology

The main insect digestive tract includes the foregut, midgut, and hindgut. There are also the Malpighian tubules (the insect kidneys), and in some cases, gastric caecum, which play various roles in different taxa. The functions of these different gut compartments of insects vary, in fact, in general among species. However, the foregut is typically where food breakdown occurs (by the proventriculus); the midgut is a major center for digestion and initial absorption; and the hindgut's contribution is in excretion, and keeping electrolyte balance (Chapman 1998).

All life stages of phasmids (other than eggs) feed exclusively on leaves. Obligate folivory of this sort is relatively rare (Calderon-Cortes, Quesada et al. 2012) among insects. Most insect herbivores don't have a strict diet only on plant tissues, except for Lepidoptera and phytophagous Coleoptera, herbivores in Orthoptera like grasshoppers and crickets can even scavenge flesh from vertebrates and invertebrates. The phasmid digestion system reflects its diet. The alimentary canal is straight and narrow with no enlargements, diverticula or fermentation chambers (Cameron 1912, Clark 1976). The crop is an extremely elastic organ covered by a layer of cuticle and is usually stretched to store leaf pieces. The proventriculus is located between the foregut and the midgut in phasmids and has spine-like projections to grind leaf pieces. Such structures also occur in other herbivorous insects including many beetles, cockroaches, and crickets (Cheeseman and Pritchard 1984, Szinwelski, Rodrigues et al. 2009). The Phasmid midgut has two distinct compartments that have long been thought to have distinct functions. The anterior half of the midgut (AMG) is pleated and folded, whereas the posterior half (PMG) consists of a thin membrane that is embedded with small tubules of unknown function. These tubules are morphologically similar to the Malpighian tubules. Interestingly, these tubules are not found in any of the other Orthopteroid insects (Nation 1983, Klass,

Zompro et al. 2002). With respect to the functions of these compartments, it is thought that the phasmid's AMG is the major center for digestive enzyme secretion (Monteiro, Tamaki et al. 2014). As for the PMG, Azevedo described the special structure on PMG as "gastric caecae-like" and suggested its contribution to nutrient absorption (Azevedo, Fialho et al. 2013), but much is still unknown. The Phasmid's Malpighian tubules are positioned between the PMG and hindgut. The hindgut comprises the ileum and rectum. To this date, there is no previous report that specifically studies the hindgut of phasmids.

Plant Cell Wall Components and Plant Cell Wall Degrading Enzymes

The main components of primary PCWs, such as those in grasses and leaves, are cellulose, hemicellulose, and pectin. Cell walls become lignified, forming a complex structure called lignocellulose, during the formation of secondary PCWs (Albersheim, Darvill et al. 2011). The secondary structure of PCWs is not ubiquitous in plants. Only woody biomass contains a great amount of these reinforced cell walls, which provide extra mechanical support (Cosgrove 2005).

The primary structure of cellulose is linear glycosidic polymers connected by β-1,4-glucosidic bonds. These β-1,4-glucosic polymers assemble into parallel bundles of microfibril. These crystalline microfibrils are usually 3–5 nm wide and aligned with each other. This composition makes PCWs mechanically strong and extremely recalcitrant to enzymatic attack (Cosgrove 2005). Cellulase is a general term for enzymes that break down celluloses, and includes three major categories: *endo*-β-1,4-glucanases (EC 3.2.1.4), which attack polysaccharidic chains within the microfibril at random locations; *exo*-β-1,4-glucanases (i.e., cellobiohydrolases [EC 3.2.1.91], cellodexistrinases [EC 3.2.1.74]), which hydrolyze cellulose from the ends of microfibrils; and β-glucosidases (cellobiases) (EC 3.2.1.21), which cleave cello-oligosaccharides into glucose monomers (Cragg, Beckham et al. 2015).

Another major component of the PCW is hemicellulose, a general term for various polysaccharides with different compositions. Xyloglucans and glucuronoarabinoxylans are major hemicellulosic

components in primary cell walls. The backbone of xyloglucan consists of glucoses connected by β-1,4-glucosidic bonds, as in celluloses, but up to 75% of the glucose residues have xylose attached at the C_6 position. Xyloglucanases (also called xyloglucan-specific endo-β-1,4-glucanases; EC 3.2.1.151) primarily break bonds between glucose residues in the xyloglucan backbone, while α-xylosidases (EC 3.2.1.177) attack the bonds between C_6-xylose and glucose residues. Glucuronoarabinoxylans have xylan backbones with various saccharide substitutions including arabinose and glucuronic acids on the side chains (Ordaz-Ortiz and Saulnier 2005). The backbone of these complex polymers is hydrolyzed by xylanases (*endo*-β-1,4-xylanases; EC 3.2.1.8), whereas the various saccharidic side chains are attacked by other enzymes like α-L-arabinofranosidases (EC 3.2.1.55) and α-D-glucuronidases (EC 3.2.1.139). Other hemicelluloses like glucomannans and galactoglucomannans require a combination of mannanases (*endo*-β-1,4-mannanases; EC 3.2.1.78), *endo*-β-1,4-glucanase, and α-galactosidases (EC3.2.1.22) to break down. Furthermore, extra enzymes like β-xylosidases (EC 3.2.1.37), β-mannosidases (EC 3.2.1.25), and β-glucosidases are needed to fully degrade hemicellulases to monomeric sugars. Glucuronoxylans and glucomannans are major hemicellulosic components in angiosperms, whereas galactoglucomannans and glucuronoarabinoxylans are abundant in gymnosperms (Albersheim, Darvill et al. 2011).

Pectin is the third major component in primary PCWs, making up approximately 25–30% of their total polysaccharides. There are three major components in pectin. The first is a linear polymer of methyl esterified galacturonic acid residues, also called homogalacturonan. The second, rhamnogalacturonon-I (RG-I), consists of α-1,4-galacturonic acid and rhamnose residues. RG-I has arabinosyl and galactosyl side chains, and some galacturonic acid residues in the RG-I backbones are acetyl-esterified. The third component in pectin is the most complex polysaccharide in PCWs: rhamnogalacturonan-II (RG-II), which has oligogalacturonide backbones and side chains consisting of more than 20 different glycosidic bonds. Moreover, there is xylogalacturonan, which has a partly methyl esterified polygalacturonan backbone with xylose side chains (Albersheim, Darvill et al. 2011). Given these complex structure and components, there

is a general term for enzymes that cleave pectin: pectinase. Generally, there are *endo*-polygalacturonases (EC 3.2.1.15) and *exo*-polygalacturonases (EC 3.2.1.67) that hydrolyze unesterified homogalacturonan together. For the glycosyl hydrolysis of glycosidic bonds between D-galactosyluronic acid and α-1,2-L-rhamnosyl residues within RG-I, rhamnogalacturonan hydrolase (EC 3.2.1.171) plays a crucial role (Azadi, ONeill et al. 1995, McDonough, Kadirvelraj et al. 2004). As for the breakdown of RG-II that has various side chains, different enzymes are involved in this process, although there is no evidence thus far that insects use RG-II. Finally, *endo*-xylogalacturonan hydrolases (EC 3.2.1.-) exclusively hydrolyze polygalacturonic backbones that are substituted with xylose.

Insect Plant Cell Wall Degrading Mechanisms

The ability to degrade cell wall polysaccharides to obtain sugars was initially considered a rare trait in insects. Insects that are dependent on cellulase activity were thought to require symbionts to digest plant material (Breznak and Brune 1994). The discovery of endogenously produced cellulases in higher termites, however, led to a re-evaluation of endogenous PCWDEs in animals (Watanabe, Nakamura et al. 1997). Thus far, several PCW-degrading mechanisms performed by microbial and endogenous enzymes in insects have been suggested.

Hindgut Symbiotic Protists

Herbivorous insects can degrade PCW through hindgut symbiotic protists. Extensive studies have reported that termites possess a variety of symbiotic microorganisms, such as Archaea, bacteria, and protists (Ni and Tokuda 2013). In lower termites, the hindgut paunch harbors many flagellate protists that digest PCWs, which are assumed to have evolved from a set of these protists that were acquired by a common ancestor of termites and cockroaches (Brune 2014). Major PCWDEs expressed by symbiotic protists in lower termites and cockroaches (*Cryptocercus punctulatus*) are xylanases (hemicellulases), cellobiohydrolases (cellulases), and endoglucanases (cellulases) (Tartar, Wheeler et al. 2009, Todaka, Lopez et al. 2010). These

protists' enzymes contribute mainly to the digestion of celluloses and hemicelluloses and perform as a part of a termite's dual cellulolytic system. Lower termites can produce endogenous endoglucanases and β-glucosidases in the salivary gland and hindgut (Nakashima, Watanabe et al. 2002, Ni and Tokuda 2013). The symbiotic protist enzymes in termite hindguts allow further extraction of energy from the partially digested wood particles that are passed from the pharynx and foregut (Geib, Filley et al. 2008).

Hindgut Symbiotic Bacteria

The second PCW-degrading mechanism is very well-known and emphasizes hindgut symbiotic bacteria. The higher termite is one of the most extensively studied taxonomic groups that have hindgut bacteria helping their digestion. Wood-feeding higher termites are assumed to have evolved independently in Termitidae from soil- and humus-feeders after losing their cellulolytic protists (Donovan, Eggleton et al. 2001). The assumption that higher termites' symbiotic bacteria replaced the flagellated protists in lower termites to degrade PCW was neglected until 2007. A metagenomic analysis in 2007 clarified the presence of more than 100 PCW-degrading enzymes in the hindgut microbiota of higher termites (*Nasutitermes* spp.) (Warnecke, Luginbuhl et al. 2007). Specifically, based on He, Ivanova et al. (2013), an abundant rumen bacteria phylum of higher termites, Fibrobacteres, plays a crucial role in the digestion of plant tissue. This result was supported by numerous cellulase homologs identified in *Fibrobacter succinogenes*. In general, higher termites that feed on grass or wood take advantage of symbiotic Fibrobacteres and Spirochaetes to digest food, whereas termites that prefer fungus, humus, or soil possess gut microbiomes dominated by Firmicutes, Bacteroidetes, and Proteobacteria (Mikaelyan, Kohler et al. 2015).

Hindgut bacterial PCW-degradation is also found in Orthoptera, Coleoptera, and Diptera (Egert, Stingl et al. 2005, Cook, Henriksen et al. 2007, Oh, Heo et al. 2008). Functionally similar to the hindgut paunch possessed by higher termites, a fermentation chamber located in the hindgut of Scarabaeidae (Coleoptera) and Tipulidae (Diptera) contains many symbiotic PCW-degrading microorganisms (Terra

1990). Similar microbe-rich tissue is found in Gryllidae and Gryllotalpidae (Orthoptera) as well, which present as projecting papillae in the hindguts (Nation 1983). These similar hindgut characteristics suggest that insects usually possess morphological projections in their hindguts to maintain microbial populations to digest PCWs. In Phasmatodea, there are no enlargements or projections on their straight and narrow alimentary canals, which restricts potential symbiosis with gut microorganisms; the only place similar to a fermentation chamber is a series of pyriform ampulles with long, thin filaments on the posterior end of their midgut, the function of which has yet to be determined by researchers. Thus far, a few studies have claimed that there is no significant contribution from symbiotic microbes to the digestion of PCW in phasmids (Cazemier, OpdenCamp et al. 1997, Shelomi, Lo et al. 2013).

Midgut Symbiotic Yeasts and Bacteria

Midgut symbiotic yeasts and bacteria comprise the third mechanism of PCW degradation in herbivorous insects (Martin 1983). The specialized cells that house symbiotic yeasts and bacteria in the midgut are called mycetocytes and are found in different tissues depending on the insect. In cockroaches, mycetocytes are located in haemocoel and the fat body; in cerambycid beetles, these specialized cells are located in the midgut caeca. Some hemipterans have even evolved specialized organs (mycetomes) formed by these cells(Douglas 2009). To date, midgut symbiotic yeast has been isolated from insects in seven orders: Orthoptera, Blattodea, Dermaptera, Hymenoptera, Neuroptera, Megaloptera, and Coleoptera (Suh, Nguyen et al. 2008). Among these insect orders, symbiotic yeasts provide significant PCW-degrading power to Coleoptera (Suh, Marshall et al. 2003, Berkov, Feinstein et al. 2007, Urbina, Schuster et al. 2013, Matos, Assuncao et al. 2017, Soto-Robles, Torres-Banda et al. 2019). Cellulolytic bacteria are found living in the midguts of Hemiptera, Coleoptera, Lepidoptera, and higher termites (Termitidae). However, compared to the hindgut PCW-degrading symbionts, midgut microorganisms are assumed to degrade polysaccharides as an initial process, as the enzymatic activities are relatively weak towards (hemi)cellulose and pectin (Huang, Zhang et al. 2010).

Biomass Exploitation from Fungal Symbiosis

Taking advantage of cultivated fungi to degrade plant material is another well-known mechanism in insects. Three taxonomic groups of insects – ambrosia beetles (Curculionidae: Scolytinae and Platypodinae), Formicidae ants (Myrmicinae: Attini), and Macrotermitidae termites – have independently evolved symbiotic associations with lignocellulolytic fungi cultivated for food. This process is known as fungiculture (Barcoto, Carlos-Shanley et al. 2020). Generally, insect hosts contribute to the supply of plant material and housekeeping, while the fungus symbionts convert plant material to digestible food for their hosts (Mueller and Gerardo 2002).

In ant-fungi symbiosis, leaf-cutting ants provide basidiomycete fungus (*Leucocoprinus gongylophorus*) with plant material to sustain their growth. In return, the fungus provides food for the ants(Moller, Licht et al. 2011). The foraged substrate is pretreated (licked by the hypopharynx and masticated into leaf fragments) by fungus-farming ants before integration into the fungus garden. The fungus garden then serves as an external digestive system that efficiently converts plant substrate into fungal biomass, in the form of hyphal swellings, also called gongylidia, that are rich in lipids and carbohydrates (Nygaard, Hu et al. 2016). A metaproteomic study further detected a series of PCWDEs in the ant fungus-garden, including a GH 9 endoglucanase, GH6 and GH7 cellobiohydrolases, and GH15 hemicellulases (Aylward, Burnum et al. 2012).

Termite-cultivated fungi all belong to the genus *Termitomyces* (Aanen, Eggleton et al. 2002). As with ant fungiculture, pretreatment occurs during the mastication and the passage of ingested plant materials and fungal nodules through the guts of young workers. This step reduces the cellulose crystallinity of the plant substrate by depolymerizing lignin, leaving the residues almost completely devoid of the various C-C- and C-O-bonded lignin units that are normally considered to be the most recalcitrant linkages. The pretreated plant material is embedded into a fungus garden where *Termitomyces* fungus and garden bacteria

are responsible for further degradation. The mature fungus garden, abundant in oligosaccharides and glucoses, is consumed by older workers, where endogenous and gut microbial enzymes fully break down the food. Metagenomics studies on *Odontotermes* and *Macrotermes* species found that the gut microbiomes of workers are enriched in enzymes such as GH92, GH43, and GH2, which are involved in the final digestion of oligosaccharides (Liu, Zhang et al. 2013, Poulsen, Hu et al. 2014, Hu, da Costa et al. 2019).

Ambrosia beetles do not forage plant substrate to feed their fungal cultivars, but instead transport their fungus to host trees. In the primitively eusocial ambrosia beetle *Xyleborinus saxesenii*, larvae pretreat woody tissue by digging galleries and smearing predigested feces. These predigested feces contain small woody particles that attach to gallery walls and serve as fungus beds for culturing. During pre-digestion in the gut, these woody particles are mixed with endo-β-1,4-xylanases that are active towards hemicelluloses (Licht and Biedermann 2012). The mechanisms of plant biomass degradation in ambrosia beetles remain largely unknown. However, it is suggested that a combination of endogenous enzymes, ectosymbiotic fungi, yeasts, and bacteria in beetle galleries all perform roles in the utilization of biomass (Li, Young et al. 2020).

Insects Endogenous Plant Cell Wall Degrading Enzymes

The last PCW-degrading mechanism in insects, which is the most neglected and understudied, is based on endogenous PCWDEs in insects. In insects, endogenous cellulases include *endo*-β-1,4-glucanases (endoglucanases) and β-glucosidases (cellobiases), but no cellobiohydrolases (Martin 1983, 1991). Cellobiohydrolases, regarding the digestion of PCWs, can be described as “wearing down” enzymes that act on the cellulosic microfibril ends. The reason that folivorous and xylophagous insects can thrive with a lack of these enzymes is the presence of mandibles and/or proventriculus that physically grind and crush plant material, allowing enzymes to access the substrates effectively (Nakashima, Watanabe et al. 2002, Fujita, Hojo et al. 2010). Additionally, a long alimentary tract (such as that of phasmids) enables a long food journey that gives cellulases more time to attack celluloses (Watanabe and Tokuda 2010). Plentiful

endoglucanases also compensate for the deficiency of cellulases against crystalline cellulose due to the lack of cellobiohydrolases. Molecular and biochemical evidence has shown that various endoglucanases are endogenously expressed in 16 insect orders having various feeding habits (Calderon-Cortes, Quesada et al. 2012). This suggests that the presence of PCWDEs in insects implies a potential phylogenetic signal, in addition to an adaptation to a feeding habit (Terra and Ferreira 1994, Boyd, Cohen et al. 2002). Among these 16 insect orders, Phasmatodea is the only order that is strictly folivorous while possessing a mysterious PCW-degrading mechanism that is poorly studied.

For endogenous insect hemicellulases, xylanases appear to be rare, and have only been reported in Coleoptera (Scrivener, Watanabe et al. 1997). Other hemicellulases, like β-1,3-glucanase, which hydrolyzes the glycosidic bonds between β-1,3-glucans, are also found endogenously expressed in insects as a part of the laminarinase complex (Terra and Ferreira 1994, Pauchet, Freitak et al. 2009). β-1,3-glucanases are presumed to be important for plant material digestion given their presence in various insect taxonomic groups with detritivorous and folivorous diets (Genta, Terra et al. 2003, Genta, Bragatto et al. 2009, Pauchet, Freitak et al. 2009, Shelomi, Jasper et al. 2014). Given that a major proportion of the linkages in hemicelluloses are β-1,4-glucosidic – the same as in celluloses – *endo*-β-1,4-glucanases and β-1,4-glucosidases can also degrade these polysaccharides, though with less effectiveness. Therefore, the rarity of endogenous hemicellulases in insects is assumed to be compensated for by various repertoires of cellulases.

Endogenous polygalacturonases, a major pectinase, have been extensively reported in Coleoptera and Hemiptera. In piercing-sucking hemipterans, salivary pectinases are involved in softening plant tissue before penetration and oviposition (Cherqui and Tjallingii 2000, Boyd, Cohen et al. 2002), whereas in beetles, pectinases are mostly applied to digest plant tissues for further breakdown (Vatanparast, Hosseininaveh et al. 2012). In recent years, studies on phasmids' endogenous pectinase have revealed some interesting results. While the genome of *Timema cristinae* possesses no pectinase genes (Soria-Carrasco,

Gompert et al. 2014, Brand, Lin et al. 2018), *Medauroidea extradentata*, *Clitarchus hookeri* and *Dryococelus australis* within the Timema's sister group Euphasmatodea contain 20 or more GH28 polygalacturonase genes that are expressed in the midgut (Shelomi, Jasper et al. 2014, Brand, Lin et al. 2018). Furthermore, these endogenous GH28 pectinases from phasmids have been shown to be homologous to those from *Proteobacteria*, instead of fungi or Bacteroidetes as in the case of pectinases in Coleoptera and Hemiptera (Kirsch, Gramzow et al. 2014). A horizontal GH28 gene transfer event is assumed to have occurred between a common ancestor of Euphasmatodea and γ-Proteobacteria 50 million years ago, and the transferred pectinase in Euphasmatodean evolved within the lineage with a series of expansions and duplications (Shelomi, Danchin et al. 2016).

Although a great deal of research has been conducted on endogenous PCWDEs in insects in recent years, the field is still in its early stages, and much is unknown. Major unresolved questions regarding PCW digestion in insects concern the structural features that render insect PCWDEs highly functionally stable in stringent digestive environments, synergism between endogenous and symbiotic PCWDEs, and the evolutionary history of PCWDEs in respect to dietary preferences. As for the structural and functional question, for instance, some insect herbivores have complex gut morphologies that facilitate the stepwise breakdown of PCW compounds. The nature of such compartments varies between insect orders, but they are well understood in termites, cockroaches, grasshoppers, and crickets. However, although phasmids are the only order of insects in which all species are obligatory leaf feeders, these compartments do not house symbionts and are thought instead to rely on the slow passage of food through a long gut to break down PCW compounds. It is possible that a wide variety of PCWDE enzymes may work together in a highly efficient manner to facilitate the breakdown of PCW compounds in this system. This hypothesis raises many interesting questions.

Whole Transcriptome Shotgun Sequencing

Next-generation sequencing (NGS) has been revolutionizing experimental design and providing high throughput data with nucleotide-level precision (Hrdlickova, Toloue et al. 2017). Nowadays, NGS technique is widely applied to study gene expression in almost every field of biological and medical research.

Traditional studies on gene expression mostly adopted the expressed sequence tag (EST) method of the early 90s (Adams, Kelley et al. 1991). The EST method investigates gene expressions by sequencing complementary DNA (cDNA) clones to reveal both the sequence and the abundance of corresponding RNA. Studies using expressed sequence tags contributed critically to the identification of novel genes in the early 90s. However, the high cost and semi-quantitative data of this method limited its application (Hrdlickova, Toloue et al. 2017). Meanwhile, the serial analysis of gene expression (SAGE) technique emerged. This techniques keeps costs much lower by sequencing a short tag region per cDNA (Velculescu, Zhang et al. 1995). While this technique can increase the data throughput and provide more precise expression levels, many of the short tags are challenging to align uniquely to references. A different type of technique that quantifies transcriptomes is based on the hybridization of fluorescently labeled DNA, for example, the DNA microarray method. However, this method has significant disadvantages, include a dependence upon *a priori* sequenced references and the complex statistical normalization that occurs when comparing expression levels across different conditions/treatments. The development of RNA sequencing (RNA-Seq) overcame these shortcomings with much higher throughput and accuracy. Briefly, the principle of RNA-Seq was to first construct a library of adaptors – attached cDNA fragments that are reverse-transcribed from total or partial sets of RNA (i.e., poly A+ mRNA). This library, with or without sequence amplification, is then sequenced from one end (single-end) or both ends (pair-end) to obtain raw reads. Depending on the platform on which the sequencing is performed, raw reads will be at different ranges of length. Three platforms that perform high-throughput sequencing include 454 GS Junior (Roche), MiSeq/HiSeq

(Illumina), and Ion Torrent PGM (Life Technologies). According to a study comparing these sequencing instruments, MiSeq/HiSeq generates the most accurate data with the highest throughput. The 454 sequencing data can generate the best assembly due to having the longest reads, and the Ion Torrent machine outputs the shortest reads with a medium throughput. Additionally, both 454 and Life Technologies platforms showed errors in homopolymers (Loman, Misra et al. 2012).

Generally, the purpose of RNA-Seq is to identify and quantify a set of transcripts expressed in a specific type of cells in an organism. Specifically, with transcriptomic data, scientists can determine the transcriptional structures of genes in terms of regulatory sites, introns/exons, and pro- and post-transcriptional modifications. Moreover, with deep coverage and advanced computational analysis, RNA-Seq is an extremely powerful way to study differential expressions of transcripts under certain physiological/environmental conditions or developmental stages.

Questions, Methods, and Goals

Plant-insect interactions are a prominent topic in many fields including evolution, ecology, phylogeny, and agriculture. Most studies investigate various arms races between the two. One well-studied context is based on the development of plant secondary compounds and the responses of insects, in myriad classes. Another active area of research explores the selective pressures stemming from the plant cell wall (PCW). The PCW is highly recalcitrant to digestion by design. This both serves structural reasons and it also makes digestion of it by herbivores extremely difficult (Hochuli 1996). The ability to utilize the energy in PCW, however, contributes greatly to insect biodiversity and ecological success, given its abundance in the biosphere.

This ability to digest the PCW was initially thought to be solely based on the presence of symbiotic microbes in the gut with PCW-degrading ability (Martin 1991). However, the report of an endogenous cellulase (GH9) from the termite *Reticulitermes speratus* challenged this widely accepted hypothesis

(Watanabe, Noda et al. 1998). Since then, more and more studies have reported insect endogenous PCW degrading enzymes (PCWDEs) (Watanabe and Tokuda 2010, Calderon-Cortes, Quesada et al. 2012). Endogenous PCWDEs have now been reported in most major insect orders. The various PCW-degrading capabilities and enzymatic repertoires in herbivores are thus drawing increasing attention from scientists.

A major obstacle to understanding the significance of PCWDEs in insects is the sampling bias towards Blattodea, Lepidoptera, Diptera, and Coleoptera (Davison and Blaxter 2005, Letsch and Simon 2013). To increase our understandings of insect PCWDEs, extensive studies are required for insect orders in Polyneoptera, such as grasshoppers and crickets (Orthoptera), stoneflies (Plecoptera), stick and leaf insects (Phasmatodea). These insects show a diversity of feeding strategies and their repertoires of PCWDEs might be more representative of the basal insect pattern than those found in model systems. Thus, more studies of the endogenous PCWDEs of these orders are necessary. The stick insects with their obligate folivary are prime candidates.

In addition to interesting scientists who study insect evolution and physiology, the biofuel industry is also interested in finding novel PCWDEs (Oppert, Klingeman et al. 2010). A major cost in the production of biofuels stems from the difficulty of degrading the PCW (Carroll and Somerville 2009). This step requires costly pre-treatments, and cellulolytic enzyme mixes. This reduces the competitiveness of biofuels compared to traditional fossil fuels. Consequently, the search for novel, more efficient, cellulases and other PCWDEs for the creation of efficient bioreactors is a priority for most biofuel industries (Li, McCorkle et al. 2009).

The goal of this dissertation is to use RNA-Seq to identify expressed PCWDEs in phasmids and document their expression profiles in distinct gut compartments. After studying endogenous PCWDEs in all tissues from both species, we will analyze differentially expressed genes between the two sections of the midgut and identify the most highly expressed genes in each tissue. We conclude with a metagenomic

study to investigate the taxonomical and functional composition of the microbiome of the phasmid's midgut. Previous work has suggested that stick insects have no symbionts, but this is unlikely.

This work will reduce a current knowledge gap regarding the diversity and function of PCWDEs in Phasmatodea. Specifically, we will expand our understanding of how phasmids use diverse structures and enzymes to optimize PCW digestion. Our data will shed light on the digestive physiology in phasmids based on tissue specific PCWDE expression. For biofuel industries, our results can contribute to the large community of researchers in the field mining biological systems for PCWDEs that are capable of efficient degradation of lignocellulosic biomass.

CHAPTER 1. ENDOGENOUS PLANT CELL WALL DEGRADING ENZYMES IN STICK INSECTS

Overview

Plant cell walls are the most abundant biomass on planet Earth. Many invertebrates, especially insects, can digest PCWs, either to access the protoplasm or to break the wall itself into digestible, nutritive oligomers and monomers (Pauchet, Wilkinson et al. 2010, Calderon-Cortes, Quesada et al. 2012). The ability of many insects to break down recalcitrant PCWs has contributed significantly to their biodiversity and success around the world (Tokuda 2019). Although scientists initially thought that the digestion of lignocellulosic material by insects was solely dependent on their gastrointestinal microbes, this thinking started to change when the first endogenous cellulase from glycoside hydrolase family 9 (GH9) was discovered in termites (Watanabe, Noda et al. 1998). Recent research with molecular and omics techniques (genomics, transcriptomics, and proteomics) has provided evidence to reveal PCW-degrading mechanisms that use various arsenals of endogenous digestive enzymes in insects.

Next-generation sequencing (NGS) has been revolutionizing experimental design and providing high throughput data with nucleotide-level precision (Hrdlickova, Toloue et al. 2017). The technique of RNA-Seq is widely applied to detect novel genes and to quantify transcripts nowadays. Specifically, with transcriptomic data, not only can scientists detect novel enzymes, but also investigate the expression profiles of the set of transcripts among between different tissues or different experimental conditions.

Phasmid's obligate folivory makes it an ideal order to study insect herbivory. It has been reported that many phasmids have various repertoires of endogenous PCWDEs (Shelomi, Watanabe et al. 2014, Shelomi, Heckel et al. 2016, Wu, Crowhurst et al. 2016, Brand, Lin et al. 2018), but there are still lots of unknowns. The midgut of phasmids has been proved to have high activities of cellulase and pectinase with multiple copies of GH9 endoglucanases, GH1 beta-glucosidases, and GH28 polygalacturonases (Shelomi,

Watanabe et al. 2014). However, with the distinctive structures between anterior and posterior midgut, it is unclear whether these two midgut compartments have different enzymatic activities.

The phasmid's hindgut does not have plausible tissues to house symbiotic microbes to help with the digestion of PCWs (Shelomi, Lo et al. 2013, Shelomi and Kimsey 2014); therefore, unlike most other insect herbivores that rely on hindgut symbionts to degrade PCWs, the digestion of PCWs in phasmids is claimed to depend on endogenous enzymes solely. It is reasonable to hypothesize that phasmids have endogenous PCWDEs expressed in the hindgut. However, most previous studies of phasmid's PCWDEs focused on the midgut, where endogenous PCWDEs showed high expression levels (Shelomi, Jasper et al. 2014, Wu, Crowhurst et al. 2016). The inclusion of the hindgut tissues is then necessary to recover a full inventory of PCWDEs in phasmids.

In this chapter, we assembled *de novo* transcriptomes from RNA samples that were extracted from six tissues of *A. asperrimus* and *M. extradentata*, including four tissues from the alimentary canal (anterior midgut, posterior midgut, ileum, rectum) and two from the torso (head and torso without the alimentary canal). The major experimental objective is to identify the endogenous PCWDEs of different GH families in in various tissues from *A. asperrimus* and *M. extradentata*. This study is a preliminary work for the following RNA-Seq analysis. More broadly, the chapter can provide transcriptomic data for future studies and help recovering potential industrial PCWDEs in biofuel production.

Materials and Methods

Insect Rearing and Dissection

The insects used were *Aretaon asperrimus* (Areolatae: Heteropterygidae) and *Medauroidea extradentata* (Anaerolatae: Phasmatidae), both cultured at room temperature in the Bohart Museum of Entomology and the Johnson Lab at the University of California, Davis. The phasmids were fed with cleansed leaves of *Rosa* sp.

Tissue for RNA extraction was dissected in 100% ethanol using sterilized razor blades with careful observation under microscopy. The head tissue was dissected first by cutting sample individuals' heads and removing the cuticle. Digestive tracts were dissected by slicing the ventral cuticle vertically and pulling out the whole alimentary canal. Anterior and posterior midguts, and the ileum and rectum, were then dissected according to the anatomy illustration in Azevedo (2013) with removal of the gut contents. All the rest of the tissue from the torso inside the opened cuticle was dissected and stored as whole-body tissue, which included the nervous system, thorax muscle, fat body, and so on. Every type of tissue was pooled from three adult individuals after a brief rinse in 100% ethanol, and then properly stored on dry ice until RNA isolation.

RNA Isolation and Library Preparation

Pooled tissue was ground in liquid nitrogen, followed by RNA extraction using the TRIzol RNA extraction reagent (Life Technologies) based on the manufacturer's protocol. Isolated RNA was cleaned up and collected using the RNeasy Mini Kit (Qiagen), with a DNase digestion step following the on-column protocol of PureLink™ DNase Kit (Invitrogen). Quality assessment of RNA samples was conducted using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) for purity, aQubit 3 Fluorometer for concentration, and an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA) for quality. cDNA Library construction was performed with TruSeq RNA Sample Prep Kit V2 (Illumina San Diego, CA, USA), guided by the manufacturer's protocol. Library quality was accessed by an Agilent 2100 Bioanalyzer, and library concentration was measured by a Qubit 3 Fluorometer. The procedure described above was repeated three times for each tissue in each species.

Deep Sequencing

Next, 150 base pair paired-end sequencing was performed on the HiSeq 4000 at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, and the raw data were uploaded to the NCBI SRA Database [*Aretaon asperrimus*: PRJNA601179; *Medauroidea extradentata*: PRJNA549703]. Raw read

quality was checked by FastQC prior to *de novo* assembly (Andrews 2010). Raw read numbers and SRA accession numbers for each biological replicate are shown in Table 1. Three library replicates were prepared for all tissues except for *A. asperrimus* anterior/posterior midgut and *M. extradentata* head and body due to a failure of library preparation and the unavailability of additional specimens.

de novo Transcriptome Assembly and Assessment

Before *de novo* assembly, we pooled forward/reverse reads from all replicates of corresponding tissue in each species. The Trinity assembler v2.11.0 (Grabherr, Haas et al. 2011) was applied to assemble *de novo* transcriptomes with the default setting for the two species using the pooled reads.

Benchmarking Universal Single-Copy Orthologs (BUSCO) version 4.1.2 (Simao, Waterhouse et al. 2015) with Insecta_odb10 database and transcriptome mode was applied to assess the assembled *de novo* transcriptomes for completeness and quality. Usually, BUSCO is applied to benchmark the quality of genome assembly based on evolutionarily informed expectation of single-copied gene contents. However, with the TBLASTN (Camacho, Coulouris et al. 2009), the HMMER (Eddy 2011), and the removal of alternative transcripts per gene (Seppey, Manni et al. 2019), BUSCO can generate a quality assessment of *de novo* transcriptomes, which is shown in Figure 1.

Transcriptome Annotation and PCWDE Identification

To annotate the two species' transcriptomes, we used the GO-Feat web server to compare each transcript to the NCBI nucleotide non-redundant (nr) database with the GO-Feat default settings. The GO-Feat web server then mapped the results to databases like UniProt, SEED, KEGG, InterPro, Pfam and Gene Ontology to fully annotate the transcriptomes (Araujo, Barh et al. 2018).

To identify as many potential transcripts for PCWDEs as possible, we downloaded all glycoside hydrolase (GH) nucleotide sequences from arthropods from the NCBI database. The query sequences from NCBI were blast-ed against the full transcriptomes with an expect value threshold of e^{-6}. Only transcripts

that aligned with at least 70% of the arthropods' GH genes were processed with following analysis. These potential PCWDE phasmid transcripts were converted into amino acid sequences by TransDecoder (Haas, Papanicolaou et al. 2013), which was further applied to predict the longest open reading frames (ORFs). The translated sequences were tested on SignalP 5.0 (http://www.cbs.dtu.dk/services/SignalP/) for the eukaryotic signal peptide cleavage sites (Armenteros, Tsirigos et al. 2019). The ORFs of these GH transcripts were then processed by HMMER 3.3.1 (http://hmmer.org/) to confirm functional domains and gene families. The numbers of transcripts and trinity genes within each GH family are listed in Table 2.

Amino Acid Alignment

To further verify the presence of conserved functional domains in these transcripts, we downloaded known PCWDE protein sequences from a variety of organisms including bacteria, fungi, plants, and invertebrates. These sequences, together with the longest isoform, were aligned with MUSCLE (Edgar 2004) in Unipro UGENE v36.0 (Okonechnikov, Golosova et al. 2012). The alignments were further manually curated in JalView (Waterhouse, Procter et al. 2009). We then searched for the active/catalytic sites in each enzyme type based on the Catalytic Site Atlas (Furnham, Holliday et al. 2014).

Results

de novo Transcriptome Assemblies

From the isolated RNA libraries of the pooled tissues of two species, we generated more than 50 million high quality, 150 bp, paired-end sequence reads from each tissue (Table 1) except for the head and body tissue of *M. extradentata*, for which we were only able to prepare two library replicates successfully. For the *A. asperrimus* transcriptome, Trinity assembly yielded 624,796 transcripts (N50 contig length=1,196). For the *M. extradentata* transcriptome, Trinity assembly yielded 765,325 transcripts with an 885 N50 contig length. All reads and the final transcriptome for *A. asperrimus* are available under BioProject accession PRJNA601179, and for *M. extradentata,* under PRJNA549703.

For assemblies with deep sequencing raw reads, an N50 value indicates the minimal length of the contigs or scaffolds that consist of the half the assembly, which only reflects the quality of sequencing and contiguity of the assembly. However, this metric is not very informative from an evolutionary perspective. In other words, it can be misleading without background information on the organism. BUSCO data sets were selected from the orthologous genes in OrthoDB (Kriventseva, Kuznetsov et al. 2019) that are present in at least 90% species of the interested taxonomical group – in our case, Insecta. By comparing the newly assembled transcriptome/genome to the corresponding BUSCO orthologous dataset, single-copy gene scores can be used to establish an evolutionarily informed expectation (Seppey, Manni et al. 2019). The BUSCO report scores of three categories of genes – C: complete [S: single-copy, D: duplicated], F: fragmented, and M: missing – indicate the absolute numbers and percentages of the total BUSCO genes found in the assembly. Non-model genomes commonly report complete single-copy scores ranging from 50% to 95%, while a transcriptomic assembly's score is expectedly lower, especially when the assembly is from few tissues or specific conditions, because not all BUSCO genes are necessarily expressed together. The BUSCO reports of our *de novo* assembled transcriptomes are shown in Figure 1.

Table 1. Total reads and trinity results for each transcriptomic library

de novo stick insect transcriptome	Tissue Replicates	Reads	Total Trinity Transcripts	Total Trinity Genes	Contig N50	SRA Accession
Aretaon asperrimus			624796	571318	1196	
	Head_1	22,149,099				SRR12075676
	Head_2	22,124,647				SRR12075675
	Head_3	27,072,432				SRR12075668
		Total: 71,346,178				
	Body_1 (without head and alimentary canal)	22,004,454				SRR12075667
	Body_2 (without head and alimentary canal)	34,614,366				SRR12075666
	Body_3 (without head and alimentary canal)	27,438,846				SRR12075665
		Total: 84,057,666				
	Anterior Midgut_1	26,499,034				SRR12075664
	Anterior Midgut_2	24,607,314				SRR12075663
		Total: 51,106,348				
	Posterior Midgut_1	26,351,839				SRR12075662
	Posterior Midgut_2	32,131,908				SRR12075661
		Total: 58,483,747				
	Ileum_1	32,394,686				SRR12075674
	Ileum_2	8,823,401				SRR12075673
	Ileum_3	27,011,423				SRR12075672
		Total: 68,229,510				
	Rectum_1	31,557,148				SRR12075671
	Rectum_2	32,775,943				SRR12075670
	Rectum_3	30,053,054				SRR12075669
		Total: 94,386,145				
Medauroidea extradentata			765325	696723	885	
	Head_1	26,227,592				SRR10437523
	Head_2	22,701,515				SRR10437522
		Total: 48,929,107				
	Body_1 (without head and alimentary canal)	21,825,971				SRR10437521
	Body_2 (without head and alimentary canal)	21,581,803				SRR10437520
		Total: 43,407,774				
	Anterior Midgut_1	50,887,853				SRR10437535
	Anterior Midgut_2	50,334,440				SRR10437534
	Anterior Midgut_3	29,559,179				SRR10437527
		Toal: 130,781,472				
	Posterior Midgut_1	44,930,507				SRR10437526
	Posterior Midgut_2	49,437,816				SRR10437525
	Posterior Midgut_3	36,574,523				SRR10437524
		Total: 130,942,846				
	Ileum_1	22,938,952				SRR10437533
	Ileum_2	20,136,171				SRR10437532
	Ileum_3	19,427,188				SRR10437531
		Total: 62,502,311				
	Rectum_1	17,727,052				SRR10437530
	Rectum_2	19,641,807				SRR10437529
	Rectum_3	22,638,613				SRR10437528
		Total: 60,007,472				

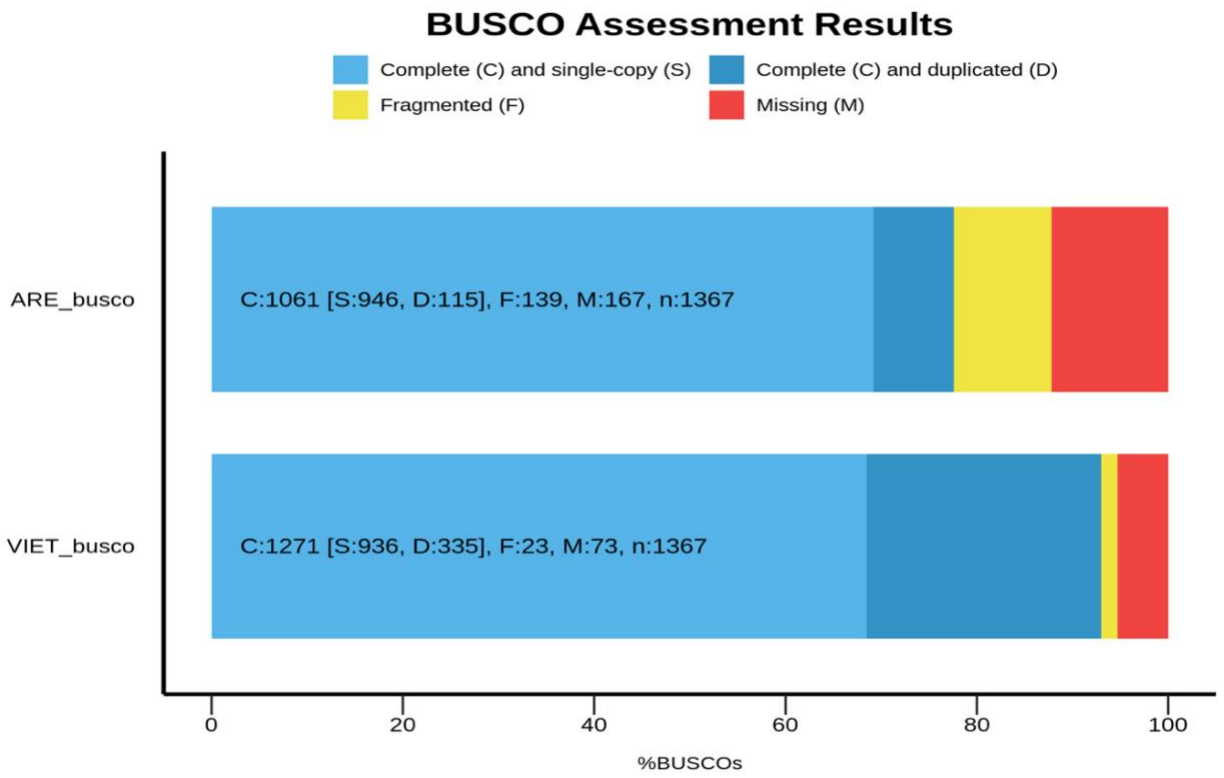


Figure 1. Benchmarking Universal Single-Copy Orthologs (BUSCO) of transcriptomes of *A. asperrimus* (are_busco) and *M. extradentata* (viet_busco). C: complete [S: single-copy, D: duplicated], F: fragmented, and M: missing indicate the absolute numbers and percentages of the total BUSCO genes found in the assembly.

The completeness of Insecta orthologous gene sets in *A. asperrimus* was approximately 77.6% (1,061/1,367), with 946 single-copy and 115 duplicated. In addition, there were 10.2% fragmented (139/1,367) and 12.2% missing (167/1,367) genes. For *M. extradentata*, the completeness of Insecta orthologous gene sets in its transcriptome was approximately 93.0% (1,271/1,367), with 936 single-copy and 335 duplicated. There were also 1.7% (23/1367) fragmented and 5.2% (73/1,367) missing genes. Both transcriptome assemblies had comparatively high completeness (77.7% and 93.0%), as most reported non-model insect transcriptomes scores are between to 62% and 85% (Tassone, Geib et al. 2016, Morandin, Pulliainen et al. 2018, Singh, Gupta et al. 2019). The *A. asperrimus* transcriptome had a lower percentage of complete Insecta orthologous gene sets (77.6%), compared to *M. extradentata* (93.0%). However, single-copy gene scores were close (*Aretaon*: 946 and *Medauroidea*: 936). For genomes, the complete genes found

with more than a single copy were designated as duplicated. These should be rare because the recovery of too many duplicates is indicative of erroneous assembly, as BUSCOs evolve under single-copy control (Waterhouse, Seppey et al. 2018). In our report (Figure 1), two-fold duplicated Insecta orthologous genes were aligned in *M. extradentata*, compared to *A. asperrimus*; on the other hand, more than three-fold of fragmented and missing genes were aligned in *A. asperrimus*, compared to *M. extradentata*. The reason behind this difference can be explained by read number differences from libraries. The number of total reads from *A. asperrimus* assembly was 427,609,594, while *M. extradentata* assemblies had 476,570,982 total reads (Table 1). From these reads, *M. extradentata* generated 140,000 more transcripts than *A. asperrimus*. Given the *de novo* method we applied to in this study, duplicated and fragmented gene scores likely reflected various transcripts of single unigenes instead of duplicated genes. BUSCO recovery tends to be higher when full organisms and/or multiple developmental stages are used to assembly transcriptomes. Our results showed that assembling from multiple tissues of full organisms can generate good representation of transcriptomes as well, given the high scores of complete single copy orthologs in both species.

Table 2. Comparison of the Number of Potential PCWDE Genes Identified in ***A. asperrimus*** and ***M. medauroidea*** *Transcriptomes and Other Insects*

		Number of Genes (Number of transcripts in transcriptomic studies)							
		Cellulase				Pectinase	Hemicellulase		
Order	**Species (References)**	**GH9 endoglucanase**	**GH1 β-glucosidase**	**GH3 β-glucosidase**	**GH6 endoglucanase**	**GH28 polygalacturonase**	**GH30 (β-xylanase, β-glucosidase)**	**GH31 (glucosidase/mannosidase)**	**GH2 (β-mannosidase, β-xylosidase)**
Phasmatodea									
	Aretaon asperrimus	8 (11)	18 (27)	1 (3)	1 (3)	12 (26)	1 (3)	4 (5)	2 (3)
	Medauroidea extradentata	24 (31)	44 (46)	\	2 (2)	59 (92)	1 (1)	7 (7)	2 (3)
	Extatosoma tiaratum	4 (14)	16 (27)	\	\	18 (30)	\	\	\
	Ramulus artemis	5 (26)	17 (45)	\	\	17 (70)	\	\	\
	Peruphasma schultei	6 (8	4 (22)	\	\	7 (14)	\	\	\
	Sipyloidea sipylus	7 (11)	10 (22)	\	\	11 (36)	\	\	\
	Clitarchus hookeri (Chen 2016)	9	\	\	\	28	\	\	\
	Timema cristinae (Chen 2016)	5	\	\	\	0	\	\	\
Zygentoma									
	Thermobia domestica (Mallipeddi 2018)	85	19	\	\	\	\	73	39
	Ctenolepisma longicaudata (Mallipeddi 2018)	69	22	\	\	\	\	50	30
Blattodea									
	Coptotermes formosanus (Zhang 2012)	2	3	\	\	\	1	5	1
	Reticulitermes fla ipes (Zhang 2012)	4	2	\	\	\	3	\	2
	Blattodea germanica (Brand 2018)	3	\	\	\	32	\	\	\
	Zootermopsis nevadensis (Brand 2018)	3	\	\	\	0	\	\	\
Coleoptera									
	Dendroctonus ponderosae (Keeling 2013)	0	26	\	\	24	\	\	\
	Leptinotarsa decemlineata (Schoville 2018)	0	37	\	\	14	\	\	\
	Agrilus planipennis (BioProject: PRJNA343475)	1	30	\	\	5	\	\	\
	Anoplophora glabripennis (McKenna 2016)	5	65	\	\	20	\	\	\
	Tribolium castaneum (Tribolium_Genome_Sequencing_Consortium, 2008)	3	19	\	\	0	\	\	\

Note: Grey rows are species studied in this paper; "\" marks no data retrieved.

Identification of PCWDEs

Among the GHs that were found from blasting arthropods' GHs against two transcriptomes, we were able to further identify several groups of PCWDEs. This included cellulases in the GH9 and cellobiases in GH1, which together can digest cellulose polymers completely into carbohydrate monomers, and pectinase *endo*-polygalacturonases in GH28. All known PCWDEs we found from the transcriptomes are listed in Table 2 with the number of PCWDEs identified from other herbivorous insects.

Phasmids' expressed cellulases are found in families GH9 and GH1, in addition to a few in GH3, GH6, GH30. The *endo*-β-1,4-glucanases transcripts from GH9 were either included themselves or were successfully aligned to sequences including the known active sites in the GH9 family based on the references of *Thermobifida/Thermomonospora fusca* (PDB: 1js4) from the Catalytic Site Atlas (Sakon, Irwin et al. 1997, Zhou, Irwin et al. 2004, Furnham, Holliday et al. 2014). Specifically, in the GH9 hydrolyzing process, two conserved aspartic acids (D55, D58) help in forming hydrogen bonds to the attacking water: one tyrosine (Y206) activates the water, and the glutamic acid residue (E461) is a catalytic acid (Figure 3).

Pectinases in these two phasmids belong to GH28, *endo*- polygalacturonase, with five conserved catalytic residues based on protein reference from *Aspergillus niger* (PDB: 1czf) (van Santen, Benen et al. 1999). The hydrolytic water molecule is activated through hydrogen bonding to an aspartic acid (Asp, D180) where another aspartic acid (D202) attacks the glycosidic link in a single displacement reaction. The D201 aspartic acid then acts as a proton donor to the departing glycosidic oxygen. A conserved histidine (H223) participates in a proton relay with Asp (D180), ensuring a deprotonation before interacting with the hydrolytic water. The arginine (Arg, R256) and lysine (K258) modulate and decrease the *pKa* of Asp (D202) during the hydrolysis (Furnham, Holliday et al. 2014). The replacement of Arg (R256) residue with tyrosine

(Y267) was observed in our *A. asperrimus* pectinase transcripts, as described in Matan (2014). However, this replacement was not observed in all transcripts.

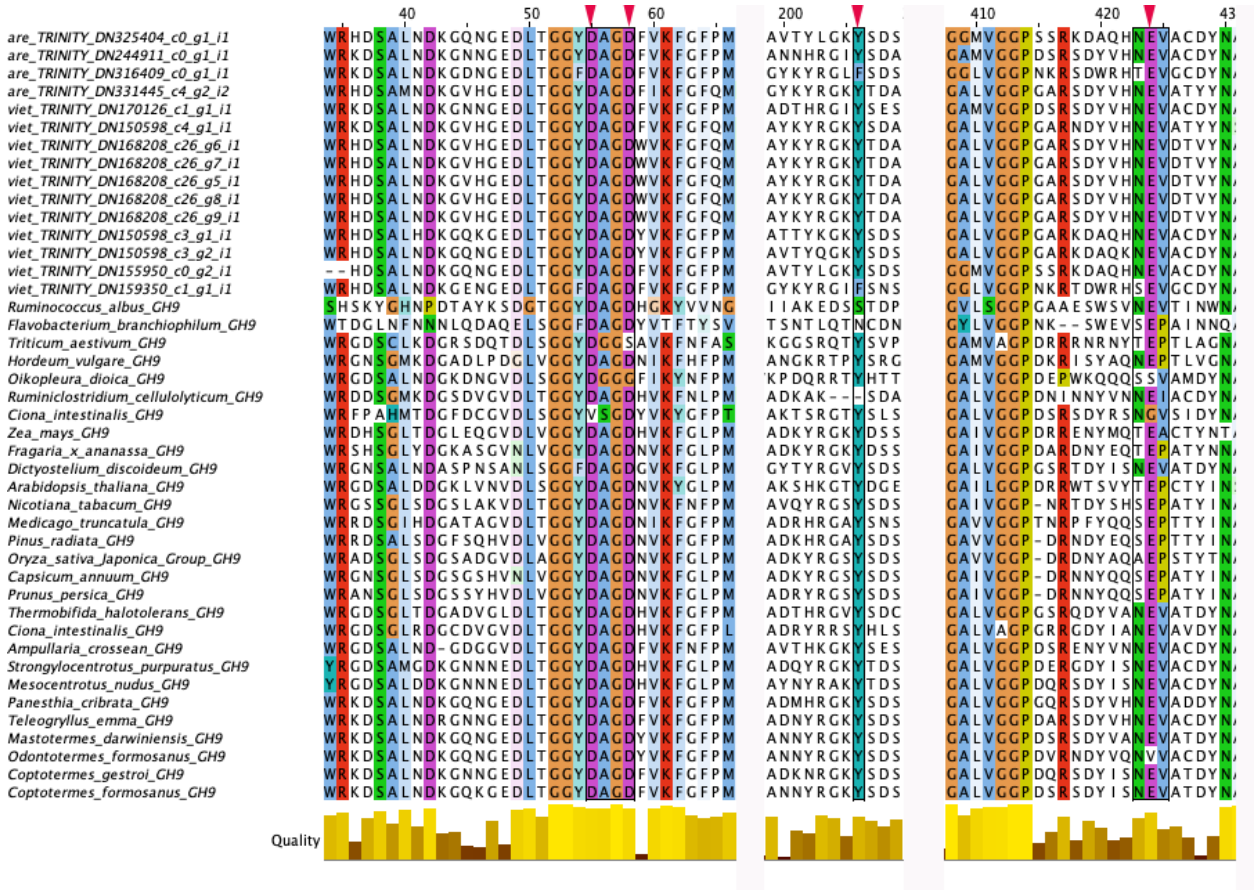


Figure 2. Sections of cellulase (endoglucanase, GH9) alignments with conserved functional residues. Catalytic sites are marked by the red arrows(Furnham, Holliday et al. 2014). Letters are colored by ClustalW algorithm based on conservation in JalView. Quality is the inverse likelihood of observing mutations based on the BLOSUM62 matrix(Waterhouse, Procter

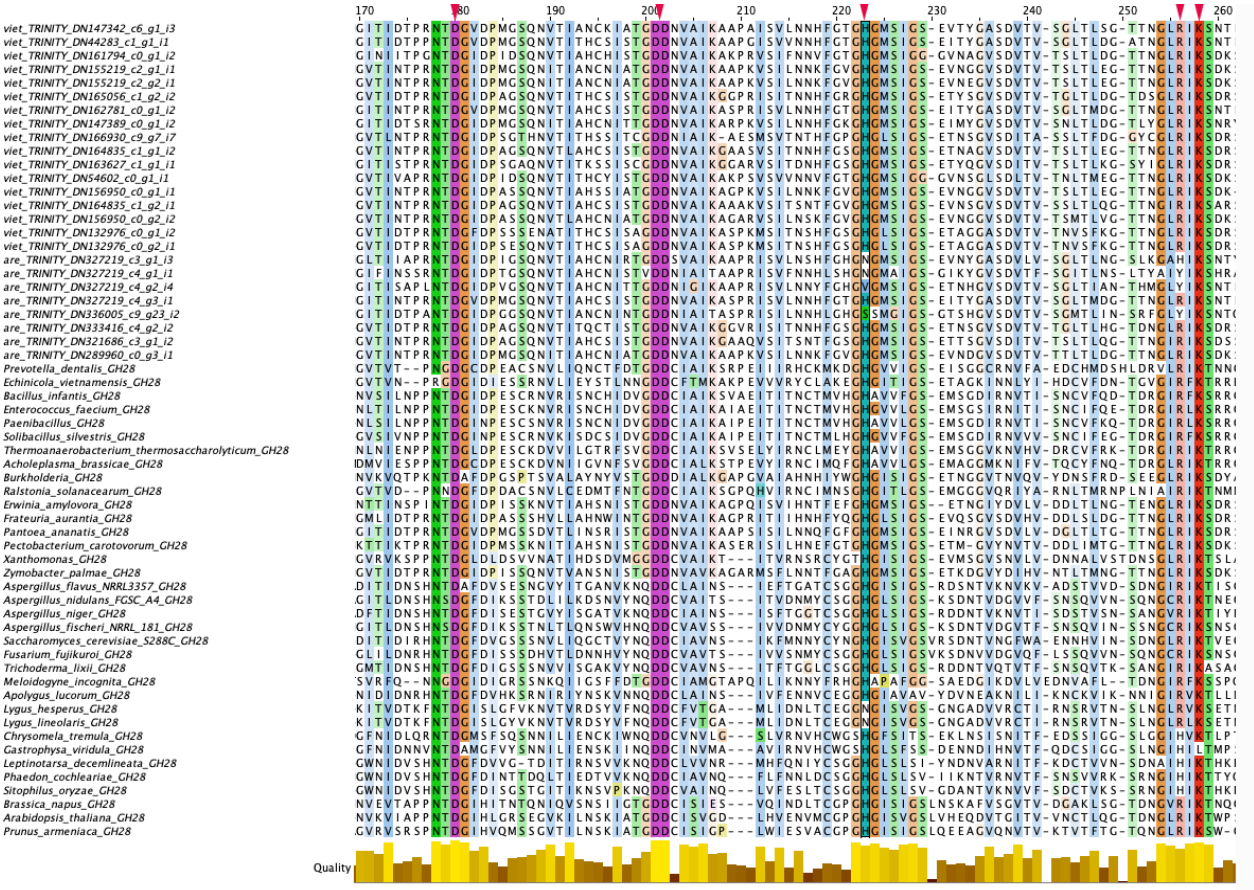


Figure 3. Section of pectinase (polygalacturonase, GH28) alignments with conserved functional residues. Catalytic residues marked by red arrows. Refer to the caption under Figure 2 for more information.

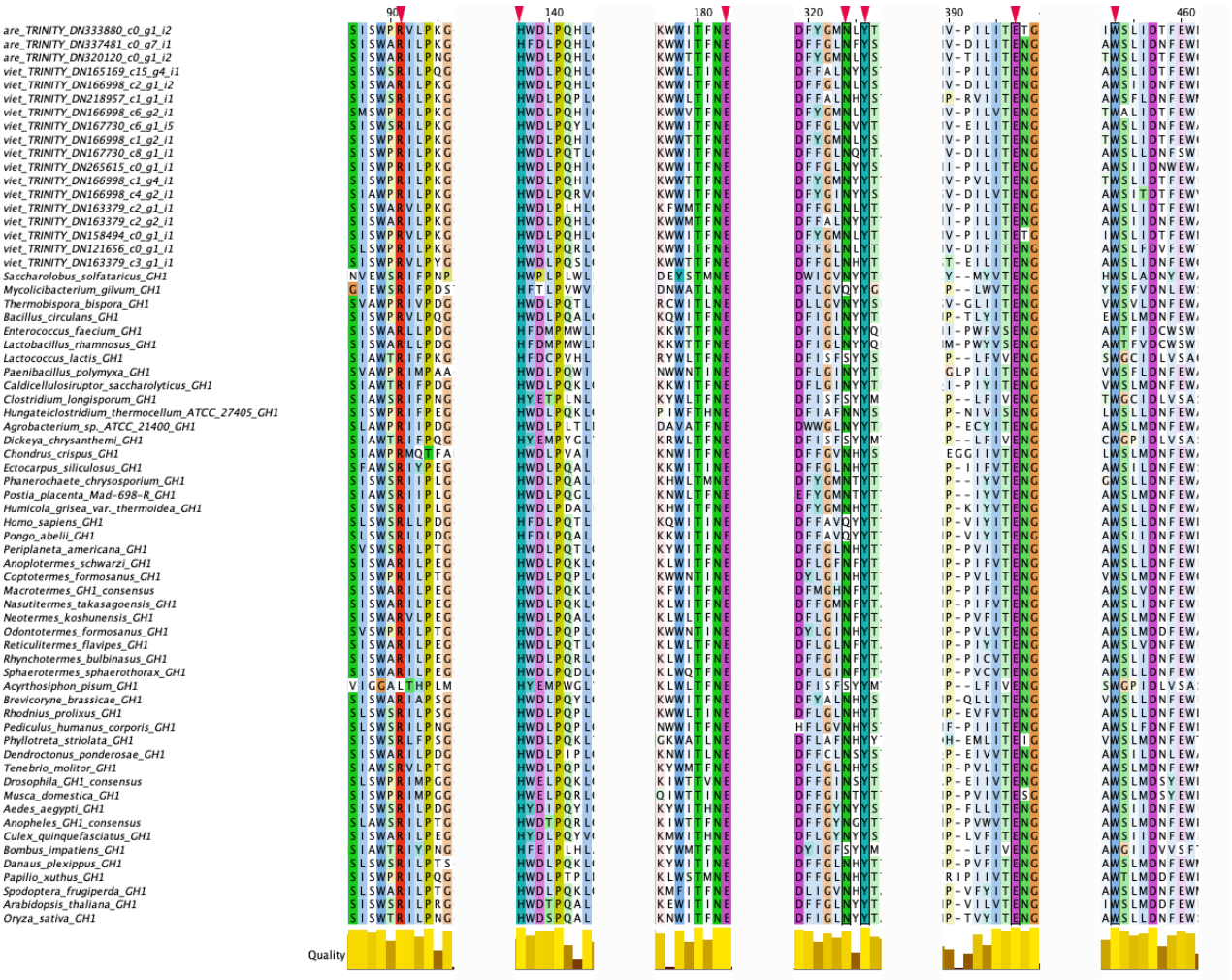


Figure 4. Sections of cellobiases (β-glucosidase, GH1) alignments with conserved functional residues. Catalytic residues marked by red arrows. Refer to the caption under Figure 2 for more information.

Most of the phasmid β-glucosidases/cellobiases contain all conserved functional residues present in the GH1 family, based on a study of white clover, *Trifolium repens* (PDB: 1cbg) (Barrett, Suresh et al. 1995): Arg (R91), His (H137), Glu (E183), Asn (N324), Tyr (Y326), Glu (E397) and Trp (W453) (Figure 5).

The detailed hydrolysis is first processed by the Glu residue (E397) performing a nucleophilic attack at the anomeric carbon in sugar, which results in the formation of a glucose-enzyme intermediate. Aglucone departure is facilitated by protonation of the glycosidic oxygen by the acid catalyst (Glu183). During the second catalytic step, a water molecule is activated by the catalytic base (Glu183) to serve as a nucleophile for hydrolysis of the glycosidic bond and release of the glucose. The rest of the catalytic residues are involved in modifying *pKa* and stabilizing the transition state (Furnham, Holliday et al. 2014). A few other PCWDE transcripts that we identified belonged to GH2, GH3, GH6, GH30, and GH31. GH2, GH3, GH6, and GH30 have β-glucosidase, *endo*-β-1,4-glucanase, and *endo*-β-1,4-xylanase functions, based on the CAZy database (http:/www.cazy.org), which are presumed to contribute to cellulose and hemicellulose degradation. However, these gene families have not been assumed to be present in insect genomes (Tokuda 2019). Additionally, it has been reported that GH31 has hydrolytic functions on the non-reducing terminals of α-1,4-linked glucose/mannose residues.

Discussion

Cellulases

Cellulose digestion is always accomplished by a mixture of enzymes with different specificities. The well-known cellulases to completely degrade polymers to monomers consists of *exo*-β-l,4-glucanase, *endo*-β-l,4-glucanase and β-l,4-glucosidase.

Within each of these three enzymatic categories, there are several glycoside hydrolase families showing the corresponding substrate specificity. For example, GH 9, GH5, and GH45 are well-known *endo*-β-l,4-glucanase families present among insects (Calderon-Cortes, Quesada et al. 2012). The distribution of GH5 among insects is limited in the Coleoptera order, and a horizontal gene transfer event has been supposed between bacterium and the common ancestor of Coleopterans. GH45 is a phylogenetically limited endoglucanase group in Polyphaga, Coleoptera. This family has been assumed to be horizontally acquired from fungi and diversified functionally in Polyphaga lineage (Busch, Danchin et al. 2019). On the other hand, GH9 has been proven to exist in the common ancestor of bilaterian animals (Lo, Watanabe et al. 2003). The first GH9 cellulase was found endogenously expressed in the salivary gland of *Reticulitermes speratus* (Watanabe, Noda et al. 1998), which was confirmed with a Southern blot analysis of DNA extracted from degutted samples. Since this ground-breaking study, there has been increasing interest in using termites and other insects to study endogenous PCWDEs. However, in both lower and higher termites, a variety of gut symbiotic microbes critically contribute to plant material breakdown. Thus, the digestive ability and influence of endogenous PCWDEs has been challenging to verify and quantify. While ancestral GH9 genes have been lost in most Coleoptera, Lepidoptera, and Diptera, most Polyneopteran orders and Zygentoma preserve a number of them. Considering the differences between diets and behaviors in Polyneoptera, repertoires of PCWDEs in this taxonomical group reflect evolutionary adaptations to feeding habits as well as phylogenetic signals. In Polyneoptera, only Blattodea (Cryptocercus and Isoptera) present

xylophagy, whereas other orders present various diets including detrivory (Orthoptera, Grylloblattodea, Dermaptera, Embioptera, Zoraptera, Plecoptera), herbivory (Isoptera, Phasmatodea, Orthoptera, Dermaptera, Plecoptera), and carnivory (Mantodea, Mantophasmatodea). With a lack of genomic data revealing Polyneopteran PCWDEs, the numbers of GH9 cellulases listed in Table 2 still indicate the unusual repertoires in phasmids. The ancestral GH9 genes in Metazoa have been proposed to be monophyletic and only present one or two copies (Davison and Blaxter 2005). Even though many recently sequenced insects have shown multiple GH9 sequences (Table 2), the numbers of GH9 we found in *A. asperrimus* and *M. extradentata* further support an ancient duplication or expansion in the GH9 family along the Phasmatodean lineage (Shelomi, Heckel et al. 2016).

Gene duplication and family expansion can lead to neofunctionalization, subfunctionalization, or loss of function (Taylor and Raes 2004). The stable presence of many copies of the GH9 family in Phasmatodea can be reasonably hypothesized as having been caused by a divergence of original functions. With the results from the expression analysis in Chapter 2, the various functions of these identified PCWDEs can be further extrapolated.

β-glucosidase is a general term that describes all enzymes that can attack β linkages between glucoses in polysaccharides, which include β-l,3, β-l,4, and β-l,6. We found abundant β-l,4-glucosidases in GH1, and other potential β-glucosidases in GH2, GH3, and GH30, in both phasmid species. β-glucosidases are commonly distributed in all insect orders, but not as abundantly as we found them in stick insects. The abundance of β-glucosidases in phasmids is shown in Table 2.

There have not been any reports that support the existence of *exo*-β-l,4-glucanase (cellobiohydrolase) in insects (Martin 1983, Calderon-Cortes, Quesada et al. 2012). We also did not find any. The lack of endogenous cellobiohydrolase makes it more intriguing that insect herbivores can success on such a disadvantageous and nutritionally poor diet.

We also identified a few PCWDEs including β-glucosidase, *endo*-β-1,4-glucanase, and *endo*-β-1,4-xylanase in GH2, GH3, GH6, and GH30. These families are not commonly reported in insects. However, the biochemical abilities of these families to degrade celluloses and hemicelluloses have been reported (Kirsch, Wielsch et al. 2012). So far, the only insect reported to express GH3 β-glucosidase is a lower termite, *Hodotermopsis sjostedti* (Yuki, Moriya et al. 2008). It would be interesting to know if other Polyneopterans, like Cryptocercidae and Embiopterans, possess GH3 in their genomes and have enzymatic functions. We assumed that the various families of glycoside hydrolases we found in phasmids' PCWDE repertoires were one of the reasons that they developed a unique and strict folivorous feeding habit. Further studies examining genomic data on chromosomal structures are required to support the assumption of cellulase gene expansion in Phasmatodean evolution.

β-1,3-glucanases from GH16 were not identified in our study. GH16 enzymes have been reported to be expressed in several phasmid species (Shelomi, Jasper et al. 2014) as well as some other detritivores (Calderon-Cortes, Quesada et al. 2012). We assumed that the reason behind this discordance could be that this family is still a recently described enzyme family with few recorded sequences in the literature or in the NCBI database.

We found rare hemicellulases such as xylosidase, xyloglucanase, and mannanase. We hypothesize that multiple functions of these PCWDEs regarding different glycosidic bonds existed in celluloses and hemicelluloses. In a previous study that analyzed the enzymatic functions of phasmids' cellulases, six GH9 cellulases from *M. extradentata* were found to be active towards glucomannan. Three of them showed strong activities towards xyloglucan, and one showed specificity towards xylan. However, none of the *A. asperrimus* cellulases showed (or were not tested for) enzymatic activities towards any component of hemicellulose in this study. Multi-functionalization was still very likely during the evolution of GH9 in phasmids due to the two strongly supported gene clusters of *Timema* GH9 and other Euphasmatodean GH9s, which are active respectively towards xyloglucan and xylan (Shelomi, Heckel et al. 2016). Nevertheless,

further biochemical and enzymatic analysis on more Euphasmatodeans is required to support this assumption and to reveal other novel functions in detoxification and immune-defense responses.

Overall, the number of Phasmatodean cellulases found in this study was consistent with the results from previous studies on phasmid digestive enzymes (Shelomi, Jasper et al. 2014, Shelomi, Heckel et al. 2016). Together with our results shown above, increasing evidence of insect endogenous PCWDEs further challenges the traditional claim that microbial PCWDEs in insects dominantly contribute to herbivorous digestion.

Pectinase

The polygalacturonases from GH28 enable herbivorous insects to break down the galacturonan backbones in pectin. This enzyme family is widely distributed among insects, especially Hemiptera, Coleoptera, and, recently, Phasmatodea. As shown in Table 2, all phasmid species except *T. cristinae* possess multiple GH28 pectinases in their transcriptomes.

The transcriptome of *M. extradentata* had 59 unigenes (92 transcripts) from GH28, whereas in *A. asperrimus*, we identified 12 unigenes with 26 transcripts from this family. Comparatively, other than *M. extradentata*, all the phasmids in Table 2 have similar number of GH28 transcripts (7–18). In a study that reported the genome of *M. extradentata*, 87 gene models (with nine pseudogenes) of polygalacturonase

Table 3. Feeding preferences of Stick Insect Species and Number of Pectinases Identified

				A. asperrimus	*M. extradentata*	*E.tiaratum*	*R. artemis*	*P. schultei*	*S. sipylus*	*C. hookeri*	*T. cristinae*
# of Pectinases				12 (26)	59 (92)	18 (30)	17 (70)	7 (14)	11 (36)	28	0
	Order	Genus	Common name								
Feeding preference (leaves)	Rosales	*Rosa*	Rose	✓	✓	✓	✓		✓	✓	
		Morus	Blackberry	✓	✓	✓	✓		✓		
		Rubus	Raspberry	✓	✓	✓	✓				
		Adenostoma	Chamise								✓
		Ceanothus	Redheart								✓
		Heteromeles	Christmas berry								✓
	Apiales	*Hedera*	Ivy	✓	✓						
	Fagales	*Quercus*	Oak	✓	✓	✓	✓		✓		✓
		Corylus	Hazel			✓	✓		✓		
		Betula	Birch						✓		
	Myrtales	*Eucalyptus*	Eucalypts		✓	✓					
		Leptospermum	White tea tree							✓	
		Metrosideros	White rata							✓	
	Sapindales	*Schinus*	Pepper tree					✓			
	Lamiales	*Syringa*	Lilac					✓			
		Ligustrum	Privet					✓			

References: Bedford (1978), Gunning (1987), Engel, Wang et al. (2016), Bradler and Buckley (2018)

were identified, whereas *Clitarchus hookeri* and *Dryococelus australis* have 28 and 20 gene models of the same family (Brand, Lin et al. 2018). The origin of the pectinase family GH28 in stick insects is assumed to be a gut gammaproteobacteria by way of a horizontal gene transfer event after the split of Timematodea and Euphasmatodea, which took place 60–100 million years ago (Shelomi, Danchin et al. 2016). Horizontally transferred pectinases then evolved within the Euphasmatodea lineage through a birth-death mechanism (Brand, Lin et al. 2018).

The food preference of *Timema cristinae* includes Californian and Oregonian native plants such as *Adenostoma*, *Ceanothus*, *Heteromeles.* Interestingly, there have been, so far, no reports about any Euphasmatodeans that can eat the leaves from these plants. The only food that *T. cristinae* shares with the

other stick insects in Table 3 is ivy (Apiales, *Hedera*). Thus, the transferred and expanded GH28 gene family in Euphasmatodea lineage potentially increased stick insects' feeding range. However, this assumption still needs support from further biochemical studies on novel enzymatic functions towards leaves on which *T. cristinae* does not feed. The feeding preference of *Peruphasma schultei* showed a drastic discordance with the other species in Table 3. The reason behind this difference is assumed to be their different habitat. *P. schultei* is a native species from South America, whereas *A. asperrimus*, *M. extradentata,* and *R. artemis* are native to Southeast Asia; *S. sipylus* inhabits Madagascar and South Africa; and *C. hookeri* and *E. tiaratum* are from New Zealand. The numbers of GH28 pectinase that stick insects have in their genomes are not significantly related to their preferences for food, except for *T. cristinae*. Instead, the development of food preferences seems to be more relevant to the geographical distribution of stick insects.

Conclusion

Overall, in this study, we assembled transcriptomes with RNA samples extracted from six tissues (head, torso, AMG, PMG, ileum, and rectum) of two species: *A. asperrimus* and *M. meaduroidea*. Both transcriptomes had high percentages of complete single-copied orthologs within the Insecta group, and thus both assemblies had good qualities for this and future studies. We successfully identified major endogenous PCWDEs including cellulases from GH3, GH6, and mainly GH1 and GH9; hemicellulases from GH2, GH30, and GH31; and pectinases from GH28. Among these genes, we confirmed the catalytic residues within endoglucanase, β-glucosidase, and polygalacturonase in GH9, GH1, and GH28. On the other hand, in the transcriptomic assembly from *M. extradentata*, GH9, GH1, and GH28 were also the three most abundant PCWDE families. Compared to other herbivorous insects and Polyneoterans, stick insects have an uncommonly high number of PCWDE genes. Ancient gene duplication and family expansion of cellulases and pectinases were supposed in previous studies (Shelomi, Danchin et al. 2016, Shelomi, Heckel et al. 2016). Digestive physiology and mechanism based on abundant endogenous PCWDEs are still under-

studied, and especially regarding the tissue- and sub-tissue specific enzymatic expressions of these families, there are still gaps of knowledge in this unique order of insect. The following chapter will further discuss this question based on tissue-specific expression analysis.

CHAPTER 2. PLANT CELL WALL DEGRADING ENZYMES' EXPRESSION IN PHASMATODEA'S TISSUES

Overview

The study of digestive physiology relating to herbivory in insects has mainly focused on orthopterans, termites, Lepidopterans, and beetles (Yuki, Moriya et al. 2008, Scharf, Karl et al. 2011, Pauchet, Kirsch et al. 2014, Liu, Song et al. 2015, Antony, Johny et al. 2017, Peterson and Scharf 2018, Gao, Liu et al. 2020). Initially it was thought that plant cell wall breakdown was only possible due to the presence of symbiotic microbes, but it is now known that most insect clades produce endogenous plant cell wall degrading enzymes (PCWDEs). Expression of these PCWDEs is thought to be associated with anatomical differences among tissues within the alimentary canal. For Phasmatodea, the mechanistic basis of plant cell wall breakdown in the alimentary canal is poorly understood.

Early work has assumed that the anterior midgut (AMG) of phasmids is the compartment where most PCW degradation happens, given the increased secretive surfaces resulting from its pleated structure. It also assumed that transcripts that are abundant in the posterior midgut (PMG) encode enzymes that break down the products of initial digestion in the AMG (dimers and monomers). It has been reported that, in phasmids, there is a 50% reduction in hydrolase gene expression in the PMG relative to the AMG (Shelomi, Jasper et al. 2014). Nevertheless, knowledge of tissue-specific expression of PCWDEs is still inadequate, and especially lacks deep sequencing data from both genomic and transcriptomic perspectives. Further, even when there is no enlargement or prominent projections attached, the hindgut (ileum and rectum) of phasmids is structurally distinctive to the anterior part of the alimentary canal and has significant length. If phasmids were degrading PCWs through a symbiont-independent scheme (Shelomi, Lo et al. 2013, Shelomi and Kimsey 2014), it is very likely that their hindguts would also express PCWDEs to contribute to PCW digestion. Chapter 1 described two phasmids' transcriptomes and the identification of endogenous

PCWDEs. This chapter will use RNASeq to catalog abundance and differential expression of PCWDEs across the relevant compartments and tissues.

Our goal is to explore the different expression patterns between the anterior midgut (AMG) and the posterior midgut (PMG), that are structurally distinctive in phasmids. Moreover, what are the putative functions of the two midgut sections, based one expression patterns? PCWDE genes (those with strong homology to known PCWDE genes and those without known functional domains but with expression patterns suggesting that they could be PCWDE genes) identified from Chapter 1 will be studied. Differential PCWDE expression patterns among different compartments will be documented, contributing to further studies. Moreover, associating these differences with the highest expressed genes in each tissue will assist future work exploring phasmid digestive physiology.

Methods

RNA isolation, Library construction, and Sequencing

Tissue for RNA extraction was dissected in 100% ethanol using sterilized razor blades under close observation through microscopy. Head tissue was dissected first by cutting individuals' heads with removal of the cuticle. Digestive tracts were dissected by slicing the ventral cuticle vertically and pulling out the whole alimentary canal. Anterior and posterior midguts, ileums, and rectums were then dissected according to the anatomy illustration in Azevedo (2013), with removal of the gut contents. All the rest of the tissue from the torso inside the opened cuticle was dissected and stored as whole-body tissue, which included the nervous system, thorax muscle, fat body, and so on. Every type of tissue was pooled from three adult individuals after a brief rinse in 100% ethanol, and then stored on dry ice until RNA isolation.

Pooled tissue was ground in liquid nitrogen, followed by RNA extraction using TRIzol RNA extraction reagent (Life Technologies) based on the manufacturer's protocol. Isolated RNA was cleaned up and collected by the RNeasy Mini Kit (Qiagen), with a DNase digestion step following the on-column

protocol of PureLink™ DNase Kit (Invitrogen). Quality assessment of RNA samples was conducted on a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) for purity, a Qubit 3 Fluorometer for concentration, and an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA) for quality. cDNA Library construction was performed with TruSeq RNA Sample Prep Kit V2 (Illumina San Diego, CA, USA), guided by the manufacturer's protocol. Library quality was accessed with an Agilent 2100 Bioanalyzer, and library concentration was measured with a Qubit 3 Fluorometer. The procedure described above was repeated three times for each tissue in each species.

Tissue-Specific Expression of PCWDEs

Kallisto v0.46.0 was used to pseudo-align reads to a full transcriptome and to quantify the abundance of transcripts in each tissue. This program was chosen over other programs as it does not rely on reference genomes. Moreover, Kallisto outperformed all the other RNA-Seq quantification tools in efficiency, and had an accuracy that was better than most (Bray, Pimentel et al. 2016). Normalized Transcripts Per Kilobase Million (TPM) matrices of all transcripts were generated by the abundance-to-matrix script from Trinity software package (Haas, Papanicolaou et al. 2013) with the estimation method specified by Kallisto. All PCWDE transcripts were retrieved through gene IDs of representative PCWDEs identified in Chapter 1. An expression heatmap was generated by heatmap2 within gplots R package.

Differentially Expressed Genes and Annotations

Differential expression analysis was conducted by Sleuth (Pimentel, Bray et al. 2017), an R package that compares transcript/gene expression levels across two or more biological conditions, in this case, AMG and PMG, using a Wald test model. Differentially expressed genes (DEGs) were those with an adjusted q-value <0.0001. The annotation of DEGs was conducted on the Blast2GO blastx program to compare each transcript to the SWISSPROT database with an expected value threshold of e^{-10}. Further annotation was identical to the annotation process of the top 1,000 expressed genes described below. We compared the

DEGs to a protein database because protein blast results are strong evidence in addition to the enriched Gene Ontology (GO) terms from enrichment tests.

Identification and Annotation of Top 1,000 Expressed Transcripts

The top 1,000 most expressed transcripts from each tissue were selected from Kallisto-sorted TPM values. Kallisto is optimal here because its psedoalignments can preserve the biological signals from raw reads without a shredding (Bray, Pimentel et al. 2016). Annotations of the top 1,000 most expressed genes were conducted using the Blast2GO tblastx program to compare each sequence to the NCBI-translated nucleotide collection (nr) database, with an expected value threshold of e^{-10}. These transcripts were then mapped to the GO database and annotated using Blast2GO with an expected value threshold of e^{-10}. InterPro annotations were performed using a Blast2GO remote connection to the InterProEBI server. Gene Ontology terms were modulated with ANNEX and GOSlim using the "generic" mapping (goslim_generic.obo) available in Blast2GO.

Enrichment Analysis for Gene Ontology

We performed functional annotations for both species' transcriptomes on the web server GO FEAT (Araujo, Barh et al. 2018). The annotation file was then parsed and edited according to the protocol of topGO 2.42.0 (Alexa and Rahnenfuhrer 2010), an R package that we used to perform an enrichment test.

Fisher's exact test was used to find enriched GO terms in the top 1,000 expressed genes in each tissue and in the DEGs between the AMG and the PMG (upregulated genes in AMGs and upregulated genes in PMGs) for both species.

Results

PCWDE Tissue-Specific Expression Patterns

The PCWDEs from most identified GH families are differentially expressed between the anterior and posterior midguts in both *A. asperrimus* and *M. extradentata* (Figure 5 and Figure 6). Moreover, the AMG is apparently the major compartment where most PCWDEs are highly expressed. The endoglucanases

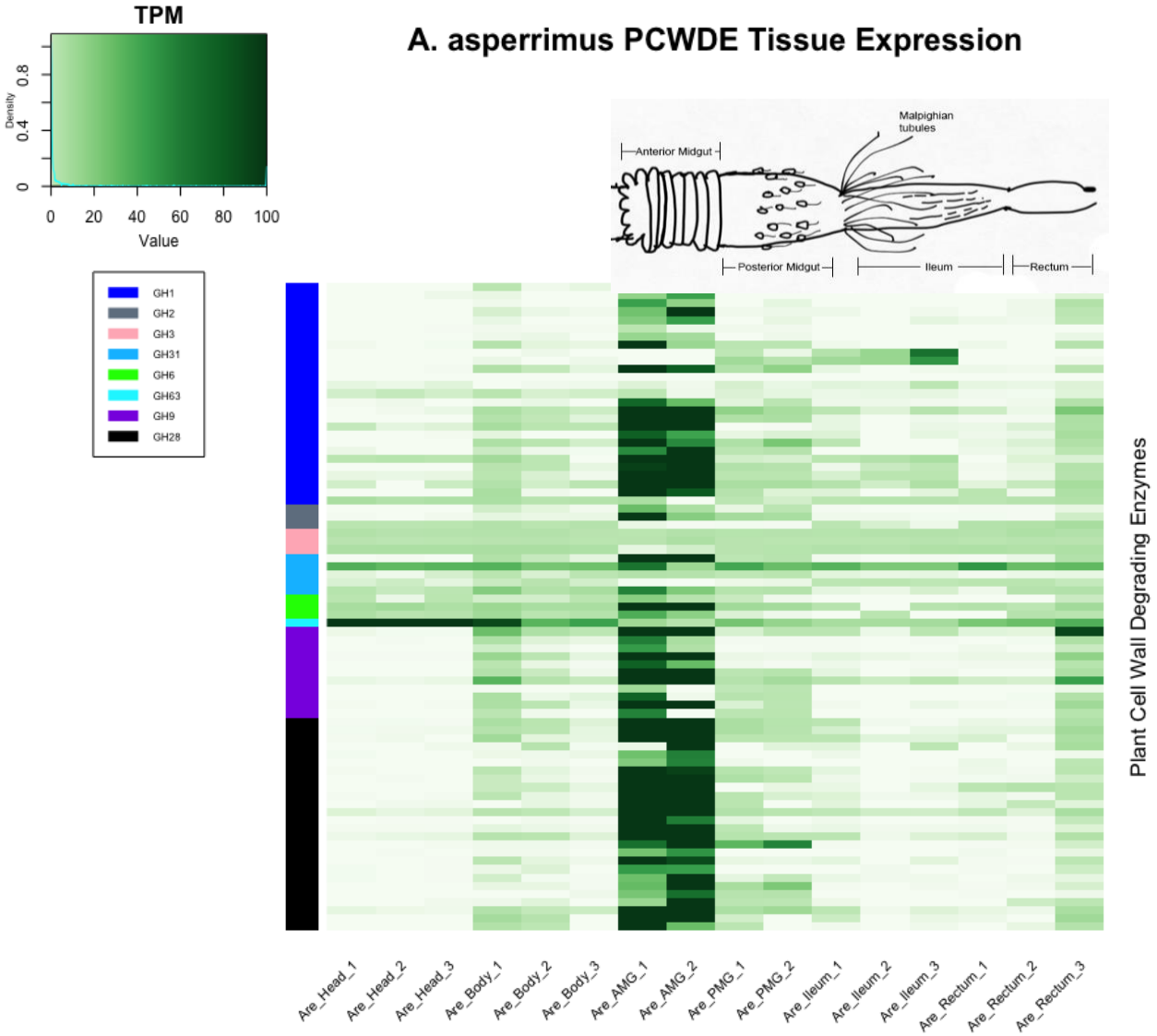


Figure 5. PCWDEs tissue-specific expressions of *A. asperrimus.* TPM: transcripts per million; a higher TPM represents a higher expression level within the tissue. Heatmap was generated by Heatmap2 in an R package "gplots". Generalized structure of midgut and hindgut of Phasmatodea was drawn based on the gut anatomy of *M. extradentata* by the author.

transcripts from GH9 were most highly expressed in the AMGs of both species. Interestingly, we observed light expressions of this family in one replicate sample of rectum from *A. asperrimus*, whereas all PMG replicates showed similarly low expression levels. Such a pattern is not found in *M. extradentata*, where GH9 is almost exclusively expressed in the midgut (both the AMG and the PMG).

The expression of GH9 genes was not expected to be observed in torso (body) samples from phasmids because RNA was extracted from phasmids with the alimentary canal and head removed. Thus, the torso (body) sample mainly consists of the fat body, nerve system, muscles, and circulatory system. However, as shown in Figures 5 and 6, there was sporadic expressions of GH9 cellulases in the torso samples of both species. It is assumed that the most likely places to express these transcripts would be the hemolymph and fat body, given that cellulases reportedly help herbivores avoid the poisonous secondary chemicals generated by PCW degradation (Peterson and Scharf 2016) and that the fat body and hemolymph are both major parts of the immune system in insects (Chapman 1998). These transcripts could also be the result of contamination during the dissection process.

Polygalacturonases from the GH28 family showed similar expression patterns to those from GH9. This gene family was most highly expressed in the anterior section of the midgut in both species. The PMG and the third replicate of the rectum in *A. asperrimus* expressed a few GH28 transcripts. On the other hand, in *M. extradentata*, fewer polygalacturonases were expressed in the PMG, but some had expression levels as high as those in the AMG. In both ileum and rectum tissues, no transcripts from GH28 showed TPMs higher than five, suggesting a weak pectinase activity in the hindgut of Vietnamese stick insects. Surprisingly, the expression of pectinases was shown in the head sample of both species at a low level (TPM < 5). These results suggest that the GH28 enzymes of phasmids may be produced in salivary glands and start pectin degradation from the mouthparts.

β-glucosidases from GH1 showed a more universally expressed pattern compared to GH9 and GH28 enzymes (Figures 5 and 6). In the ileum of the thorny stick insects (*A. asperrimus*), two GH1

transcripts in all replicates even showed higher expressions than in the AMG. The same situation was observed in Vietnamese stick insects. Moreover, in Vietnamese stick insects, one GH1 gene was expressed in all tissues, including the head, and two were expressed in every tissue from the alimentary canal. In general, it is assumed that β-glucosidases from GH1 are expressed throughout the whole alimentary canal (from mouthpart to rectum) in phasmids to continuously digest cellulosic oligomers. More specifically, similar to the case of Blattodea (Tokuda 2019), our results suggested that phasmids also express GH1 β-glucosidases in their salivary glands.

The minority of PCWDEs identified in *A. asperrimus* were in the families GH2, GH3, GH6, GH31, and GH63. Generally, the expressions of these enzyme families are not limited to the midgut nor to the alimentary canal but show global patterns across all tissues. Interestingly, the GH2 family showed two opposite patterns of expression: a midgut-limited expression and a "global" expression everywhere except the midgut (Figures 5 and 6; GH2: grey color-tagged) in both species. GH2 represents cellulolytic/hemicellulolytic specificities that include a variety of enzymes, such as β-mannosidase (EC 3.2.1.25), β-glucuronidase (EC 3.2.1.31), α-L-arabinofuranosidase (EC 3.2.1.55), endo-β-mannosidase (EC 3.2.1.152), exo-β-glucosaminidase (EC 3.2.1.165), α-L-arabinopyranosidase (EC 3.2.1.-), β-galacturonidase (EC 3.2.1.-), β-xylosidase (EC 3.2.1.37), and β-glucosidase (EC 3.2.1.21). On the other hand, all transcripts from GH3 showed expression in all selected tissues. According to CAZy, GH3 includes β-glucosidase (EC 3.2.1.21), β-1,4-xylosidase (EC 3.2.1.37), α-L-arabinofuranosidase (EC 3.2.1.55); β-1,4-glucosidase (EC 3.2.1.74), exo-1,3-1,4-glucanase (EC 3.2.1.-), β-1,3-glucosidase (EC 3.2.1.-), and xyloglucan-specific exo-β-1,4-glucanase/exo-xyloglucanase (EC 3.2.1.155). Considering representative hemicellulases like β-mannosidase, β-1,4-xylosidase, xyloglucanase, α-L-arabinofuranosidase, and β-1,3-glucosidase in these families, it is suggested that GH2 and GH3 have substantial digestive influences towards hemicellulosic material. Furthermore, as shown in Figure 5, *A. asperrimus* expresses hemicellulases that belong to GH2 and GH3 throughout the whole alimentary canal.

The GH6 family represents endoglucanase (EC 3.2.1.4), cellobiohydrolase (EC 3.2.1.91), and lichenase/endo-β-1,3-1,4-glucanase (EC 3.2.1.73). Cellobiohydrolase is a type of exo-β-1,4-glucanase that binds to the end of celluloses and keeps cleaving cellobioses until disassociation from the substrate (Fischer, Ostafe et al. 2013). Insects reportedly do not express endogenous cellobiohydrolases (Martin 1983, Martin 1991) but rather only take advantage of symbiotic microbes, such as the protozoan GH7 cellobiohydrolases in the hindgut of lower termites (Nakashima, Watanabe et al. 2002). However, sequences of cellobiohydrolases belonging to GH48 are present in Polyphaga beetles with unconfirmed functions. Given the fact that GH6 enzymes are expressed in the midguts of both phasmid species (Figures 5 and 6), it is plausible that cellulases from this family are involved in phasmid digestion.

For both species, two expression patterns of the enzymes from GH31 are shown in Figures 5 and 6: one is highly expressed in the AMG, and the other is globally expressed. GH31 is one of the major glycoside hydrolase families in eukaryotes, and mainly represents α-glucosidase (EC 3.2.1.20), α-mannosidase (EC 3.2.1.24), and α-xylosidase (EC 3.2.1.177). This enzyme family has been identified in protozoa, nematodes, insects, and all other eukaryotic lineages (Alam, Nakashima et al. 1996, Tibbot and Skadsen 1996, Sikora, Urinovska et al. 2010, Gabrisko 2013). The midgut-highlighted pattern suggested the involvement of this enzyme family in the digestion of plant cell starch. Another important role in metabolism played by this enzyme family is glycoprotein processing (e.g., ER glucosidase II) (Herscovics 1999).

Along with GH31, GH63 is another exo-acting α-glucosidase (processing α-glucosidase I [E.C. 3.2.1.106]) family that was expressed globally in all our studied tissues in *A. asperrimus*, with a significantly high expression level in the head – even higher than in the alimentary canal (Figure 5, GH63). According to the CAZy database (http://www.cazy.org/GH63.html), other enzymes belonging to the GH63 family include α-1,3-glucosidase (EC 3.2.1.84), mannosylglycerate α-mannosidase/mannosylglycerate hydrolase (EC 3.2.1.170), and glucosylglycerate hydrolase (EC 3.2.1.208). Eukaryotic glucosidase I

cleaves the terminal α-1,2-linked glucose residue of $Glc_3Man_9GlcNAc_2$, an oligo-saccharide precursor of N-linked glycoproteins. The following step is facilitated by α-glucosidase I from GH31 and involves removing two inner α-1,3-linked glucose residues. GH31 and GH63 together are critical for protein quality control in the endoplasmic reticulum (Herscovics 1999). α-glucosidase from GH63 is not likely to contribute to midgut digestion given its lower expression in the alimentary canal compared to the head.

With respect to *M. extradentata*, minor GH families showed expressions in all tissue samples as shown in Figure 6, including GH27 and GH30. GH30 is an enzyme family that represents major types of hemicellulase: endo-β-1,4-xylanase and β-xylosidase, according to the CAZy database (http://www.cazy.org/GH30.html) and Tokuda (2019). A globally high expression profile and midgut-specific expression were both observed for GH27, a gene family including α-galactosidase (EC 3.2.1.22), α-N-acetylgalactosaminidase (EC 3.2.1.49), isomalto-dextranase (EC 3.2.1.94), and β-L-arabinopyranosidase (EC 3.2.1.88). The tissue-specific expression pattern of GH27 suggested that its digestive role is played in the midgut, in addition to having an important biological function performed throughout the whole individual (see Figure 6, GH27). α-galactosidases cleave the terminal non-reducing α-D-galactose residues from α-D-galactosides, including galactose oligosaccharides, galactomannans, and galactolipids (Naumoff 2004). Galactomannan is a major component of hemicelluloses where the mannan backbone is decorated by α-D-galactose residues, which are especially abundant in the leaves from coniferous trees (Malgas, van Dyk et al. 2015). In fact, synergistic digestion performed by galactosidase and mannosidase can increase the efficiency of the hydrolysis of dominant galactomannan in softwood leaves (Varnai, Huikko et al. 2011). In Figure 6, we observed high expression of GH27 galactosidases and GH 31 mannosidases in the midgut of *M. extradentata*, which suggested additional digestive abilities of more hosts towards plant leaves.

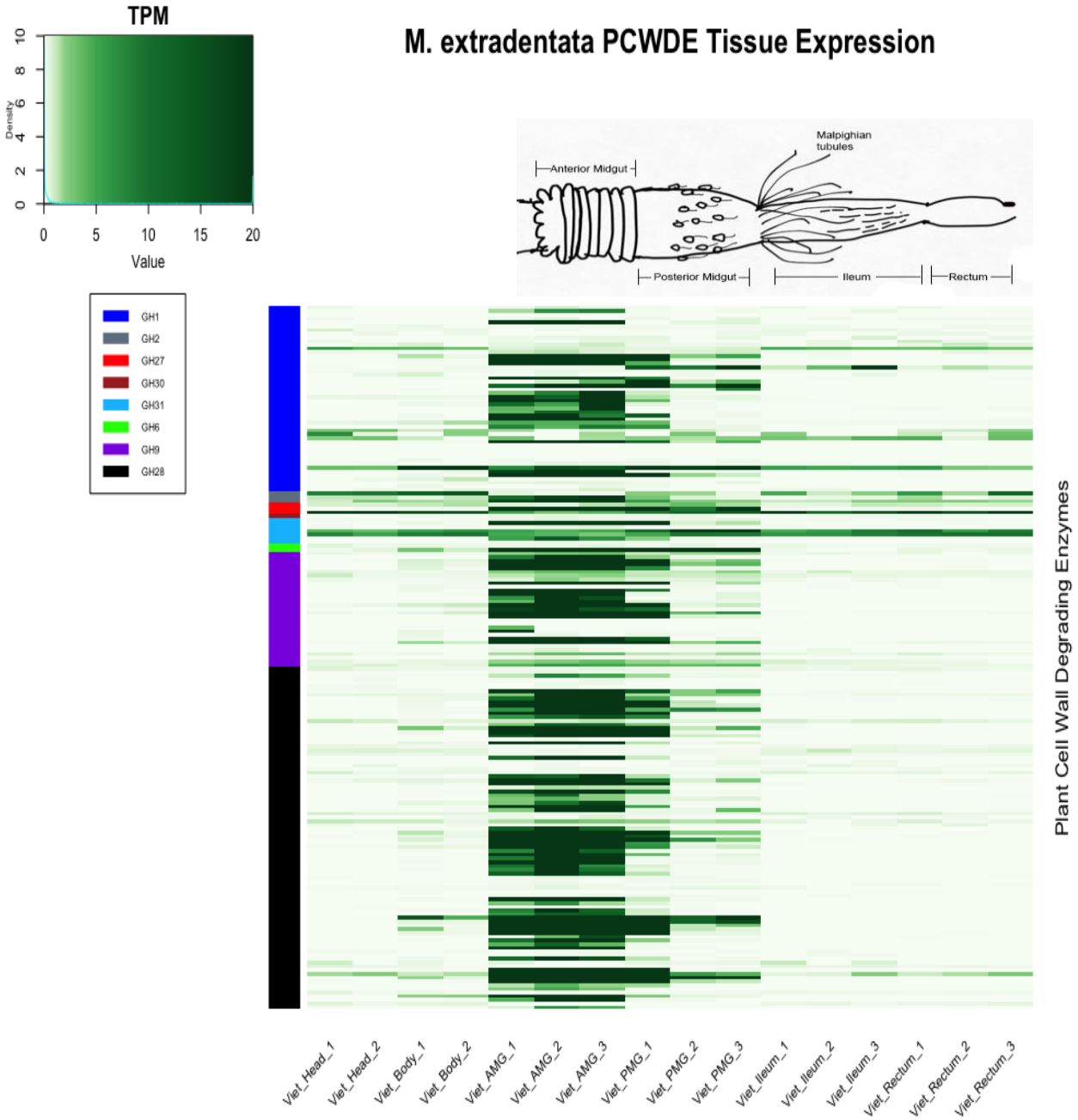


Figure 6. PCWDE tissue-specific expressions of *M. extradentata*. TPM: transcripts per million; a higher TPM represents a higher expression level within the tissue. Heatmap was generated by Heatmap2 in an R package "gplots". Generalized structure of midgut and hindgut of Phasmatodea was drawn based on the gut anatomy of *M. extradentata* by the author.

Differentially Expressed Genes between the AMG and PMG

We identified 1,120 differentially expressed genes (DEGs) between the AMG and the PMG in *A. asperrimus*. In *M. extradentata*, with more reads generated from the two compartments of the midgut, we were able to identify 2,431 differentially expressed genes. Among these DEGs, we identified highly expressed genes in the AMG and PMG respectively according to the beta value (b) in the Sleuth results. According to the beta (b) value from the Sleuth results (Pimentel, Bray et al. 2017), we were able to

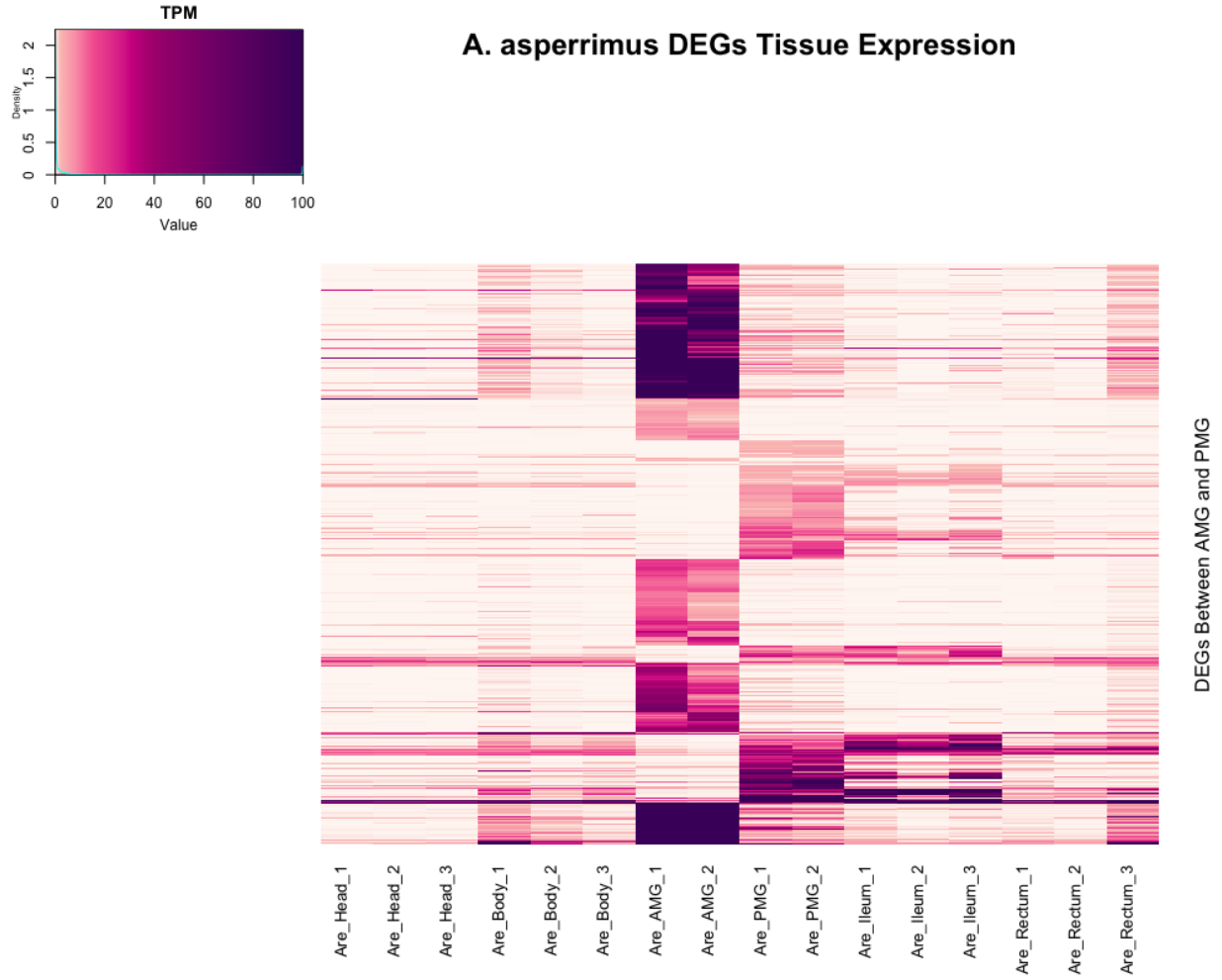


Figure 7. Tissue-specific expression heatmap of all differentially expressed genes between the AMG and PMG of *A. asperrimus*. The expression level was indicated by the TPM value generated from Kallisto abundance analysis. The heatmap was generated by Heatmap2 in an R package "gplots".

identified highly expressed genes in the AMG and PMG. The compartmentation of digestive functions along phasmid alimentary canals was highly plausible (Azevedo, Fialho et al. 2013, Shelomi, Jasper et al.

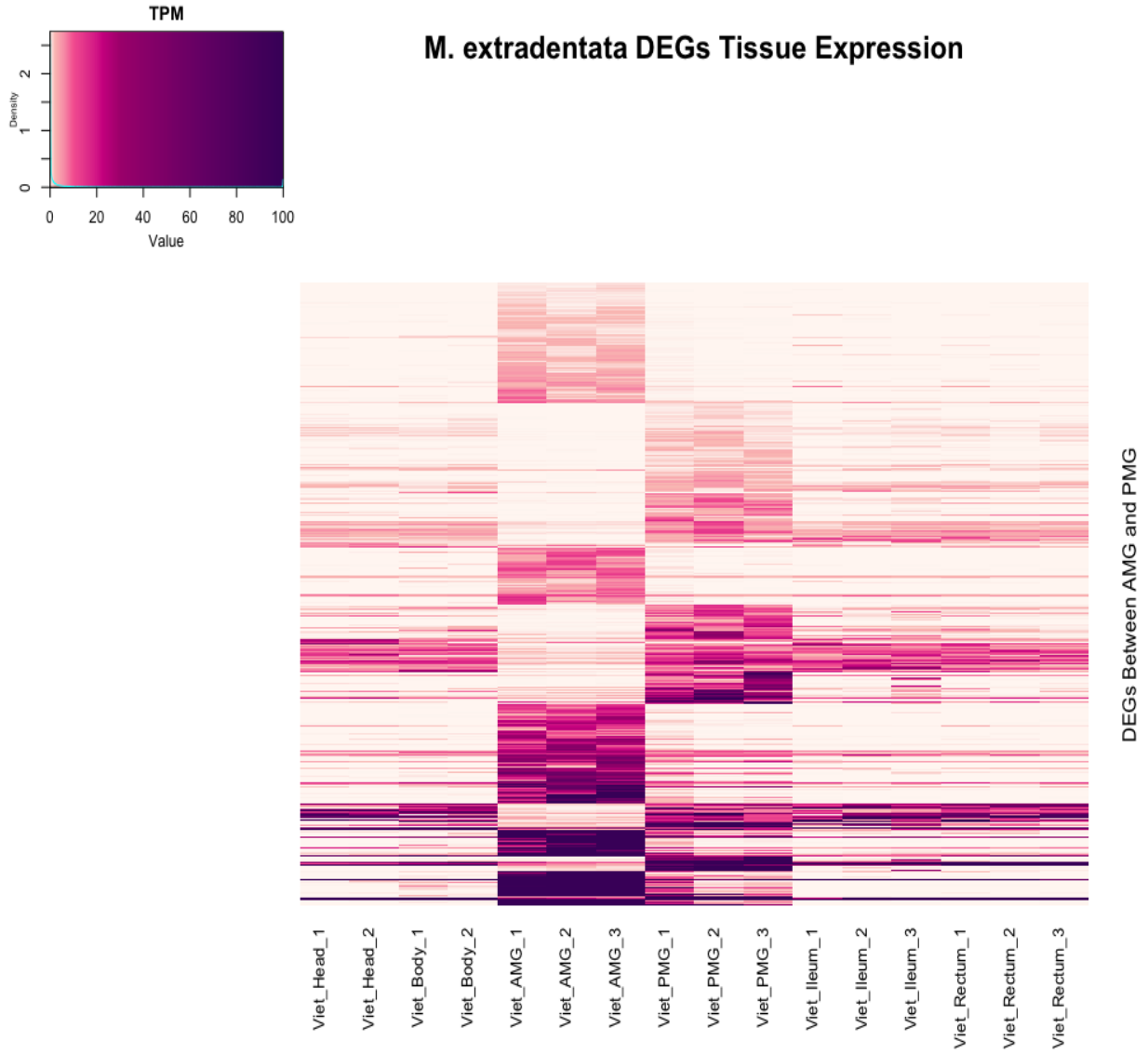


Figure 8. Tissue-specific expression heatmap of all differentially expressed genes between AMG and PMG of *M. extradentata*. Expression level was indicated by TPM value generated from Kallisto abundance analysis. Heatmap was generated by Heatmap2 in an R package "gplots".

2014). We mapped expression levels of the DEGs identified between the AMG and PMG to further illustrate the different roles played by phasmid digestive tissues.

We extracted tissue-specific TPM values of these DEGs (between AMG and PMG) and generated heatmaps for both species, as shown in Figure 7 and Figure 8. All highly expressed genes in the AMGs of both species had low TPMs in other alimentary and non-alimentary tissues. This pattern suggested the unique physiological, and presumably digestive, role played by the AMGs of phasmids.

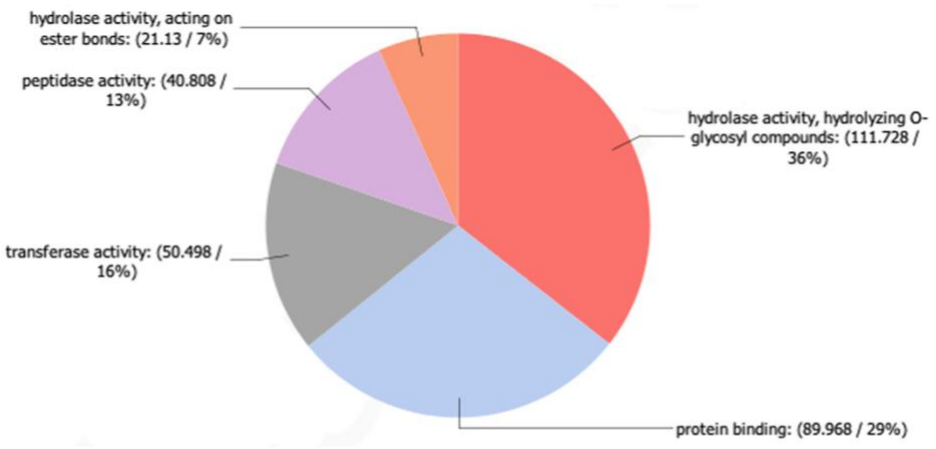
Interestingly, we observed that a proportion of up-regulated genes in PMGs also expressed highly in the ileum and rectum. Similar expression levels of these genes were present in the head and torso tissues as well. In general, we assumed a transitionally digestive role played by the PMG given that the following digestive tissues share weakened but similar expression patterns of up-regulated genes in the PMG.

To learn more about compartmental digestion between the AMG and PMG, we conducted gene ontology annotation and enrichment analysis for these DEGs. The results are shown in Figures 9–12. Figure 9 shows distributions of molecular function GO terms annotated to top nodes assigned for more than 25 sequences. In both species, hydrolase activities involving hydrolyzing oxygen-glycosyl compounds were one of the top three GO terms.

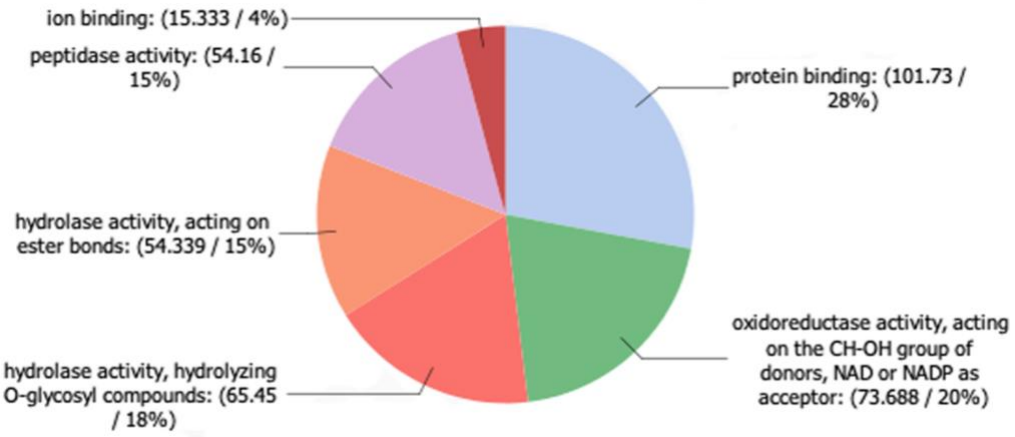
We also observed that protein binding and peptidase activity were comparatively up-regulated in the AMG, which suggested that the AMG is not only a major sub-tissue that hydrolyzes carbohydrates, but also digestively active towards proteins. However, as shown in Figure 10, protein binding and peptidase activity are the two dominant molecular functions up-regulated in the PMGs of both species. Therefore, our results suggested that, compared to the AMG, the PMG is highly active in protein digestion.

Most enriched GO terms in these DEGs were based on annotations of overall transcriptomes of two phasmids species, as shown in Figures 11 and 12. Our results from enrichment analysis of DEGs between midgut sections in both species showed a strong confidence in up-regulated enzymatic functions in carbohydrate degradation in the AMG, especially polygalacturonase activity, which was the most enriched GO term in the AMGs of both species (Figures 11 and 12). This result was also supported by PCWDE-

tissue specific expression heatmaps in Figure 5 and Figure 6, where GH28 polygalacturonases showed significantly higher expressions in the AMG then the PMG. Moreover, glucosylceramidase, β-glucosidase, and serine-type endopeptidase activities were also up-regulated in the AMG of phasmids. Among up-regulated genes in the PMGs of both species (Figures 11 and 12), our results indicated that most enriched GO terms included metallopeptidase, aminopeptidase, transmembrane transporter, and metallocarboxypeptidase activities. Therefore, we have confidence in inferring that the PMG is critical to degrading various protein oligomers after the up-regulated endopeptidase activity in the AMG that dissolves the polypeptide chains. The enriched transmembrane transporter and symporter activities in the PMG additionally indicated that in the posterior section of phasmid's midgut, ion and molecule transportation is an important function besides further food breakdown.

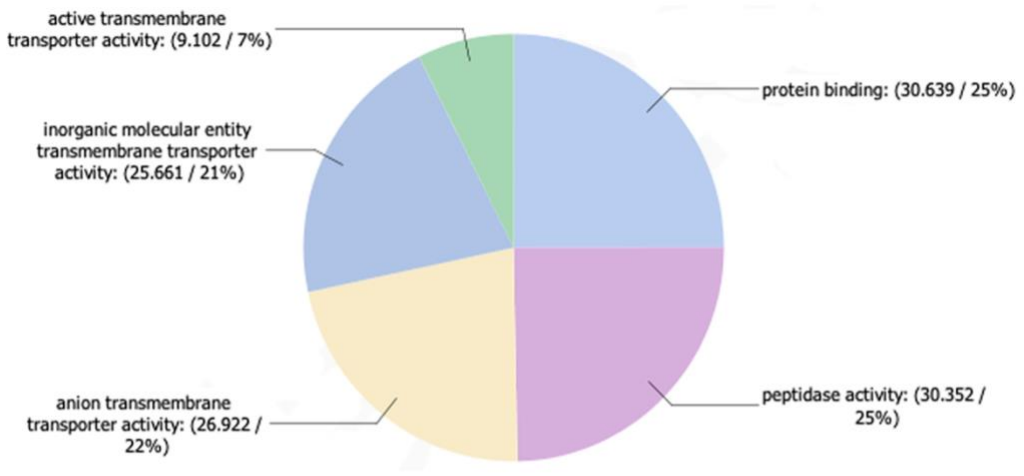


A. Top GO terms of up-regulated genes in AMG of *Aretaon asperrimus*

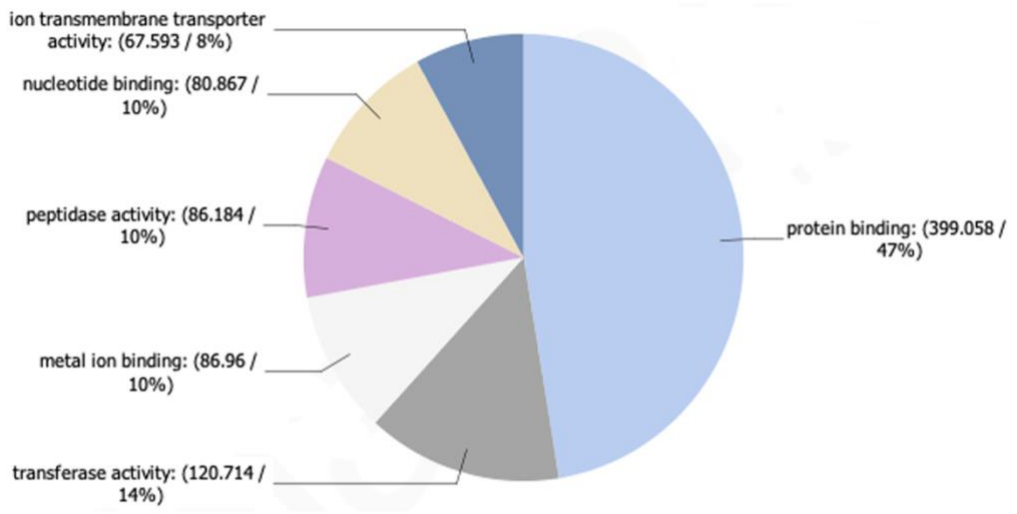


B. Top GO terms of up-regulated genes in AMG of *Medauroidea extradentata*

Figure 9. Multi-level pie chart of molecular functions annotated for up-regulated DEGs in the AMG of *A. asperrimus* (A) and *M. extradentata* (B) with node score and percentage. Node score was the sum of sequences directly or indirectly associated with a given GO term weighted by the distance of the term to the term of "direct annotation" (Conesa, Gotz et al. 2005). The minimal number of sequences a filtered GO node has assigned is 25.



A. Top GO terms of up-regulated genes in PMG of *Aretaon asperrimus*



B. Top GO terms of up-regulated genes in PMG of *Medauroidea extradentata*

Figure 10. Multi-level pie chart of molecular functions annotated for up-regulated DEGs in PMG of *A. asperrimus* (A) and *M. extradentata* (B) with node scores and percentages. Node scores are the sum of sequences directly or indirectly associated to a given GO term weighted by the distance of the term to the term of "direct annotation" (Conesa, Gotz et al. 2005). The minimal number of sequences a filtered GO node has assigned is 25.

Top 1,000 Most Highly Expressed Genes Annotation

To further explore differences between selected alimentary and body tissues from phasmids, we further annotated the top 1,000 highly expressed genes in each tissue from *A. asperrimus* and *M. extradentata* and conducted enrichment tests based on the overall transcriptomes. The results are shown in Figures 13–24.

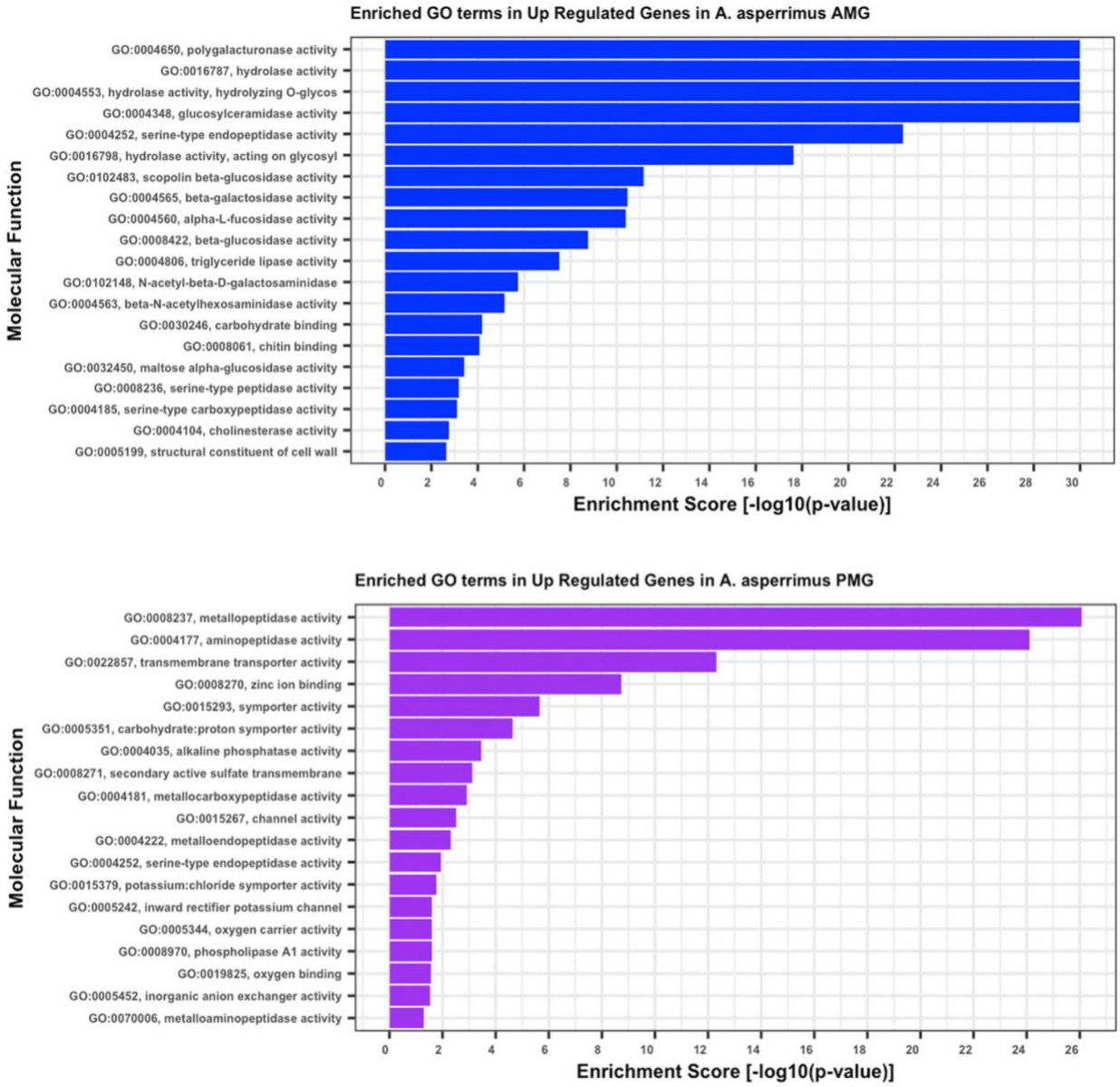


Figure 11. Top molecular function GO categories enriched for the most differentially expressed genes in the anterior and posterior midguts of *A. asperrimus*. Enrichment scores were derived from p-values generated by a weighted Fisher's exact test, which was conducted based on *A. asperrimus* transcriptomes assembled from Chapter 1.

The result of our analysis of the top 1,000 highly expressed sequences in each tissue further suggested the compartmentalization of digestion in phasmids. All species showed highly similar GO distributions in each tissue (Figure 13). In the top 1,000 expressed genes in the PMGs, we did not see

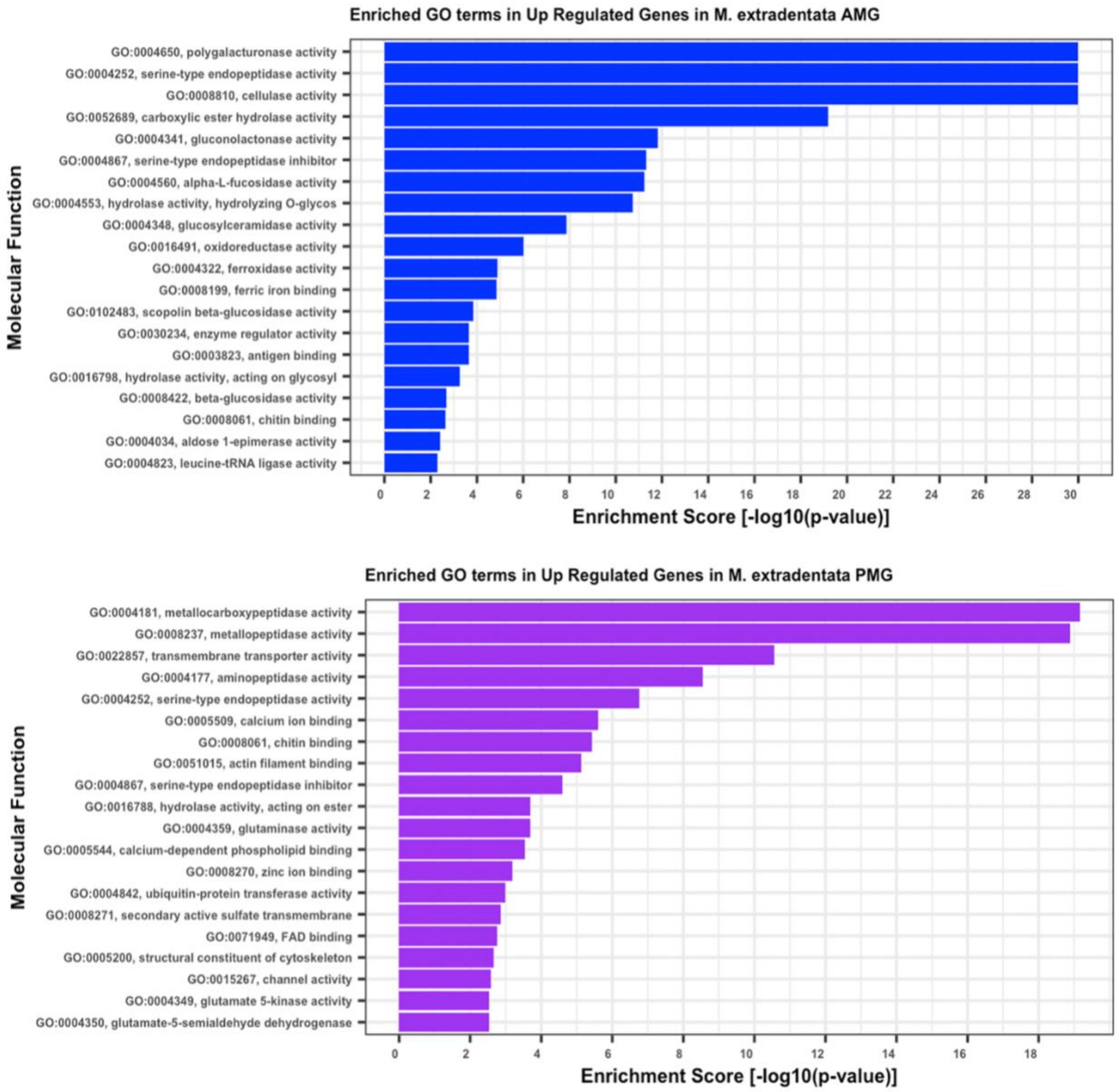


Figure 12. Top molecular function GO categories enriched for the most differentially expressed genes in the anterior and posterior midguts of *M. extradentata*. Enrichment scores were derived from p values generated by a weighted Fisher's exact test, which was conducted based on *M. extradentata* transcriptomes assembled from Chapter 1.

hydrolase activity as a major molecular function. This can be compared to the AMGs, where sequences annotated to hydrolase activity were more than a quarter of the function.

Surprisingly, hydrolase activity was not one of the top GO terms in the PMGs of both species. Instead, oxidoreductase activity and metal ion binding activity showed high proportions of GO terms in the PMG, as well as in hindgut tissues (ileum and rectum). In the hindguts of both species, most highly expressed enzymes included hydrolases, oxidoreductases, metal ion binding, transferases, and transmembrane transporters. Enrichment tests were also conducted, and the results showed us more information about compartmentalization of digestion in phasmid alimentary canals with overall transcriptome contexts. Most enriched molecular function GO terms in each tissue from two phasmids species are shown in Figures 14–17.

Besides the structural constituent of the ribosome being the most enriched GO term in every gut tissue of both species, the results suggested enriched polygalacturonase activity and heme binding in the AMG (Figure 14). Heme binding proteins (HBPs) are a group of metalloproteins that perform in electron transfer, diatomic gas transportation/storage, chemical catalysis, transcriptional regulation, ion channel chemosensing, circadian clock control, and microRNA processing (Liou, Charoenkwan et al. 2014). In Figure 15, which shows enriched GO terms in the most highly expressed genes in the PMG, our results suggested that enzymes active towards peptides are more enriched. These enzymes include unfolded protein binding proteins, glutathione transferases, and protein disulfide oxidoreductases. Moreover, highly expressed enzymes in phasmid PMGs showed enriched activities in sugar processing, including adenylosuccinate synthases, phosphopyruvate hydrolases, and NADH dehydrogenases. For the hindgut in phasmids, the results from the ileum and rectum are shown in Figures 16 and 17. We observed more molecular functions in ATPase and GTPase activities in the hindgut, as well as activities from transmembrane transporters. Overall, the enriched GO terms for the most highly expressed genes in phasmid hindguts have similar profiles as those in the PMG.

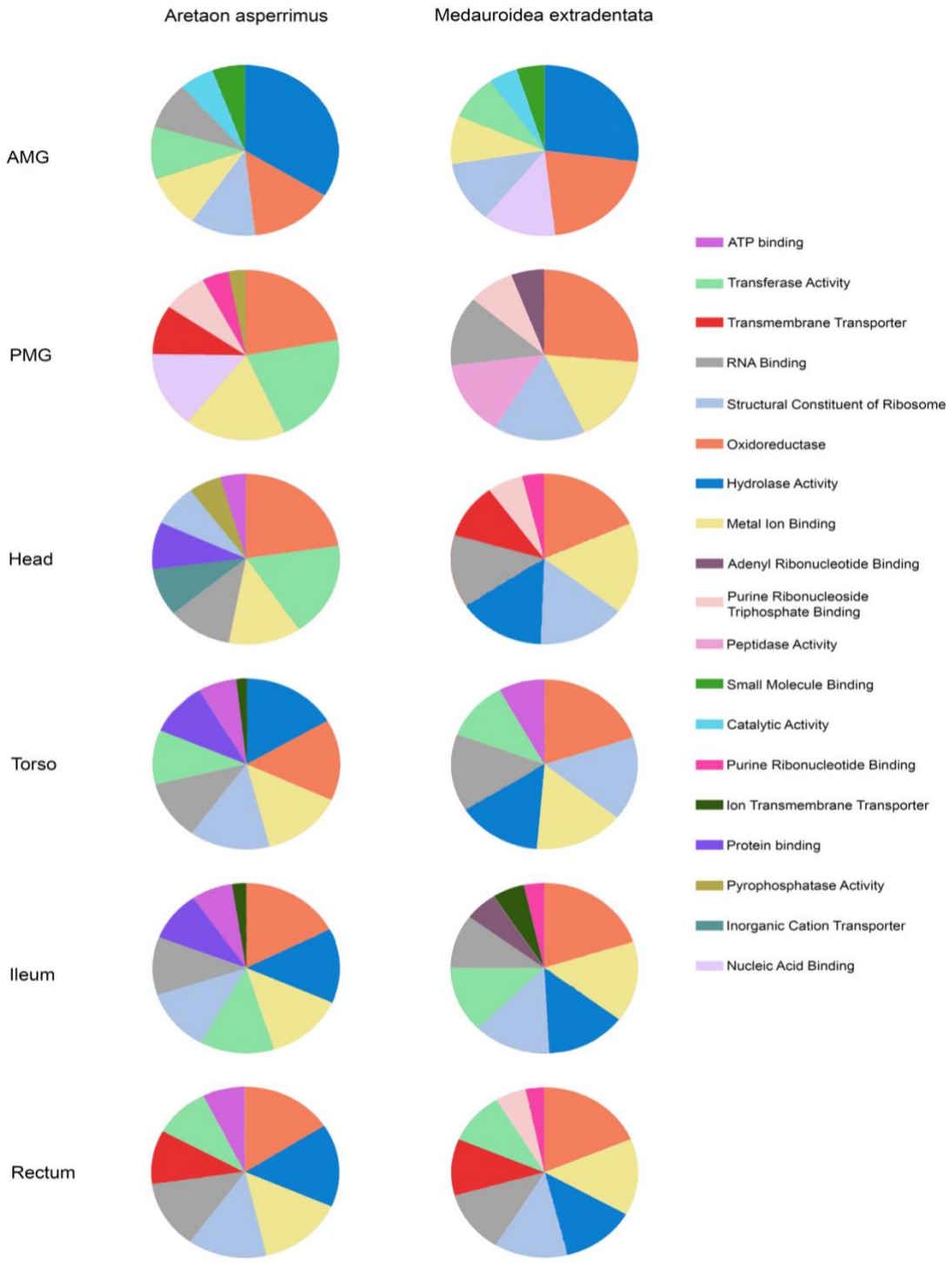


Figure 13. Multi-level pie charts of molecular functions annotated for the top 1,000 highly expressed genes in the AMG, PMG, head, torso, ileum, and rectum of *A. asperrimus* and *M. extradentata*. The distribution of top GO terms was plotted according to node scores from Blast2GO annotation (Conesa, Gotz et al. 2005). Details of parameters are shown in captions under Figure 9 and 10.

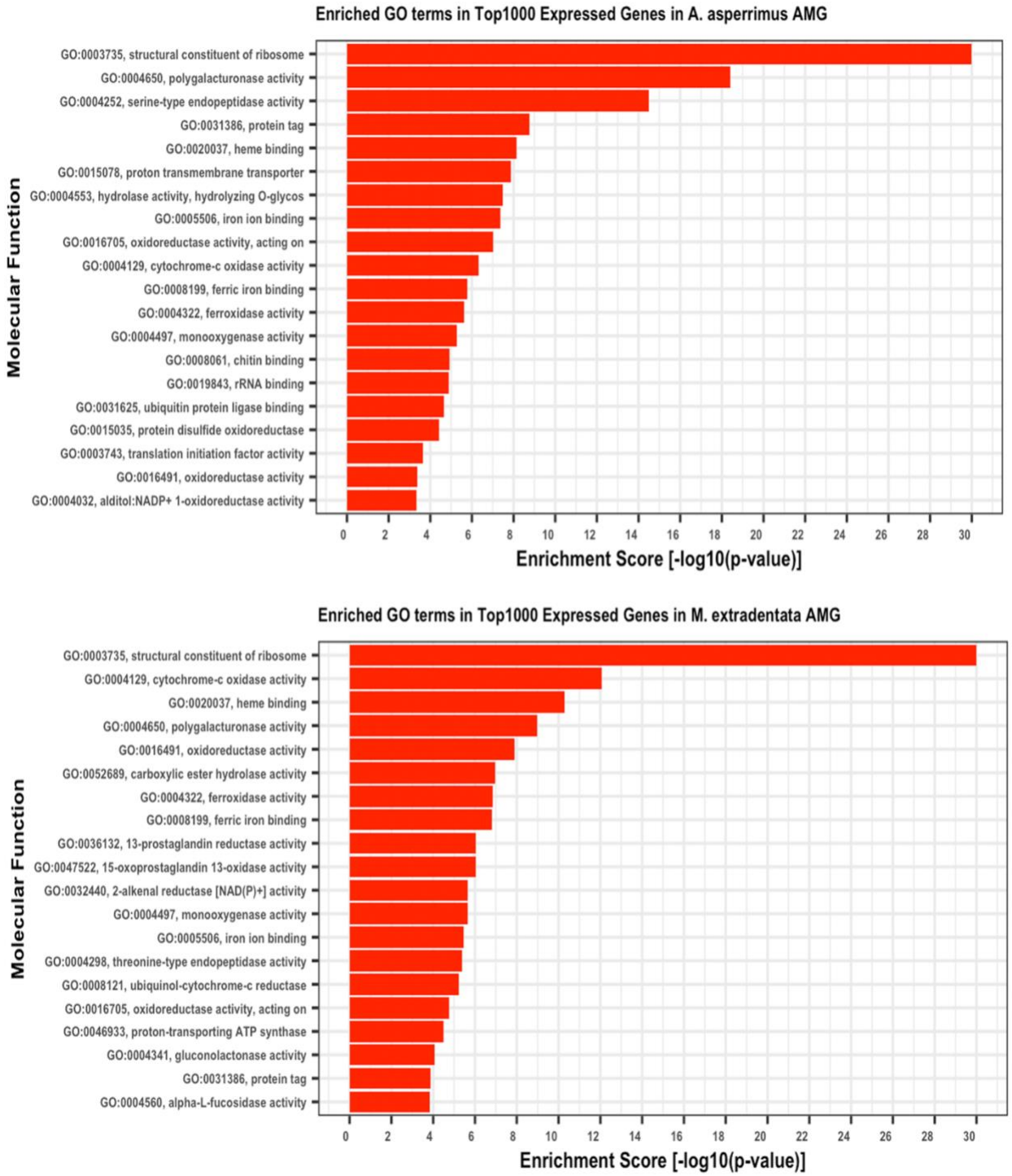


Figure 14. Top molecular function GO categories enriched for the 1000 most highly expressed genes in **anterior midgut** of *A. asperrimus* and *M. extradentata*. Enrichment score was derived from p values generated by weighted Fisher's Exact test, which was conducted based on species transcriptomes assembled from Chapter 1.

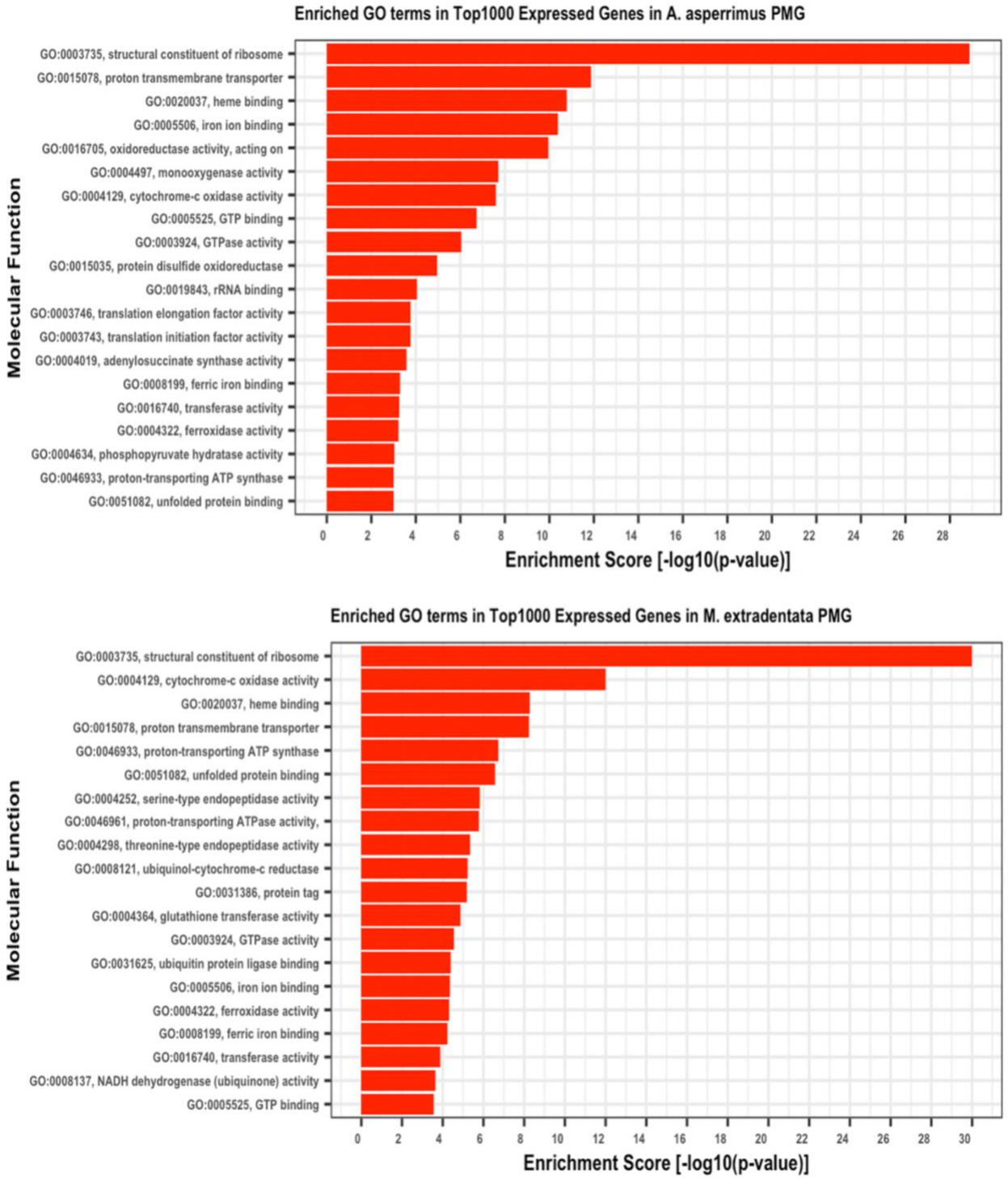


Figure 15. Top molecular function GO categories enriched for the 1000 most highly expressed genes in **posterior midgut** of *A. asperrimus* and *M. extradentata*. Enrichment score was derived from p values generated by weighted Fisher's Exact test, which was conducted based on species transcriptomes assembled from Chapter 1.

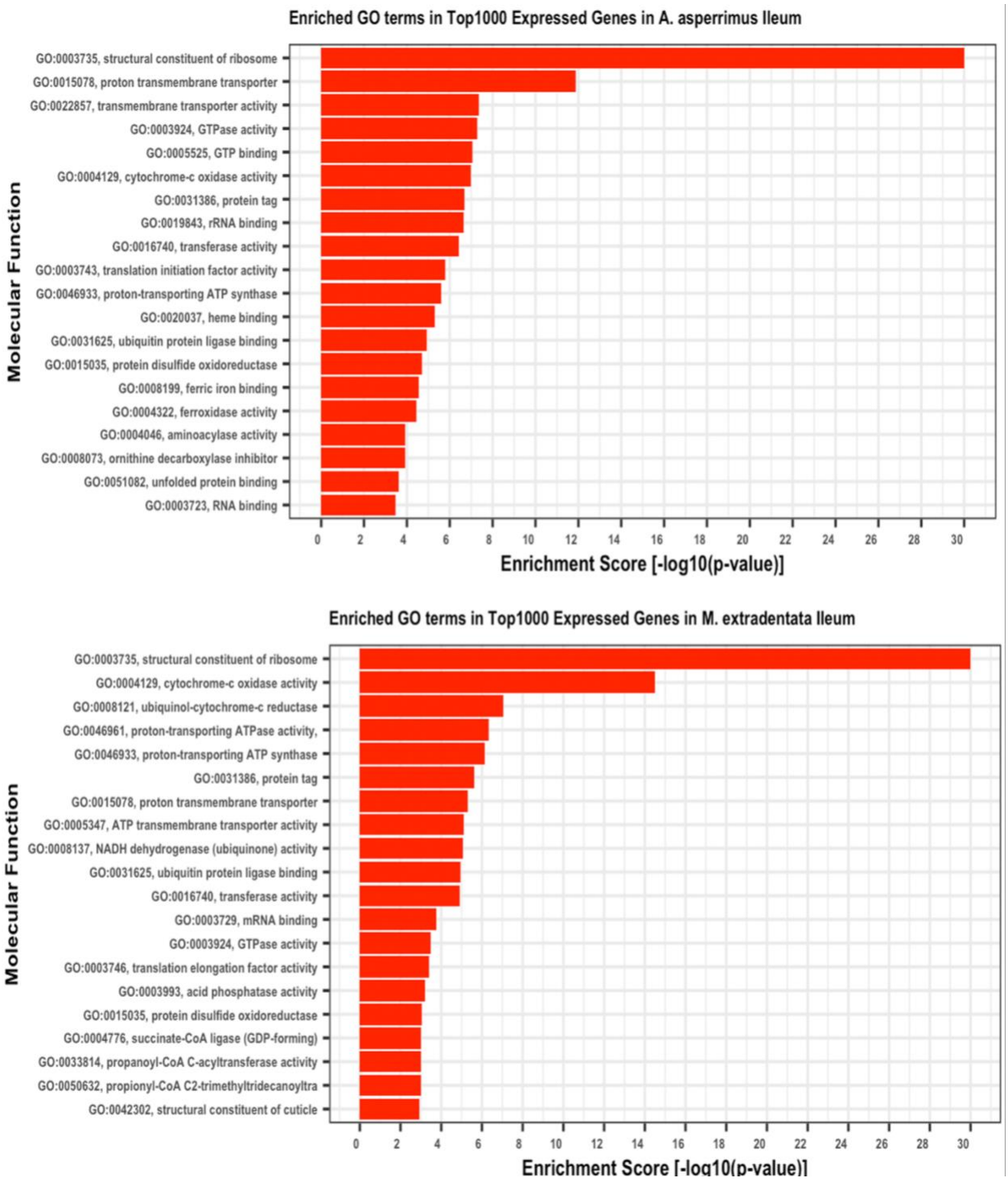


Figure 16. Top molecular function GO categories enriched for the 1000 most highly expressed genes in **ileum** of *A. asperrimus* and *M. extradentata*. Enrichment score was derived from p values generated by weighted Fisher's Exact test, which was conducted based on species transcriptomes assembled from Chapter 1.

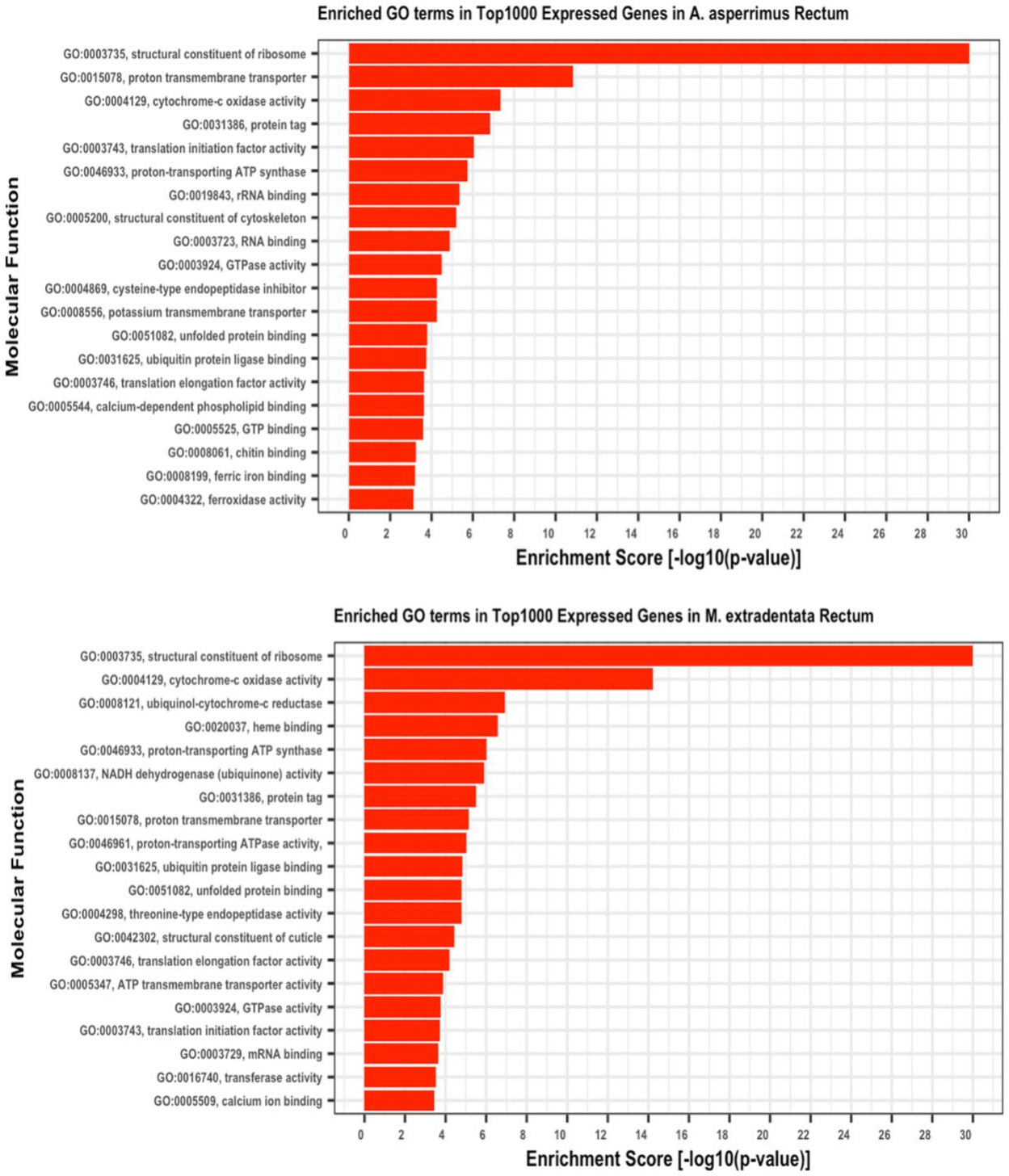


Figure 17. Top molecular function GO categories enriched for the 1000 most highly expressed genes in **rectum** of *A. asperrimus* and *M. extradentata*. Enrichment score was derived from p values generated by weighted Fisher's Exact test, which was conducted based on species transcriptomes assembled from Chapter 1.

Discussion

PCWDE Expressions

With high-quality transcriptomes assembled from high-depth reads and a wide selection of tissues, we were able to plot tissue-specific PCWDE expression profiles, including major tissues in the alimentary canal as well as the head and torso. De novo transcriptomes assembled by Trinity cannot differentiate homologous loci from various gene isoforms without mapping to a species genome (Wang, Gerstein et al. 2009). Therefore, overestimation of PCWDEs identified in *A. asperrimus* and *M. extradentata* was possible. Nevertheless, our results are easily interpreted given the strikingly high levels of differential expression of PCWDEs between the relevant compartments.

Previous studies claimed that cellulase activity falls to nearly nothing in phasmid PMGs, compared to AMGs that are highly digestive towards celluloses (Shelomi, Jasper et al. 2014). Our results suggested that in a phasmid's AMG, major PCWDE families like GH1 β-glucosidase, GH9 endoglucanase, and GH 28 polygalaturonase are highly expressed. However, in the PMGs of phasmids, cellulase families, such as β-glucosidases from GH1, GH2, and GH3, are also expressed, although at lower levels compared to the AMG and the other tissues in the digestive tract (Figures 5 and 6). As RNA-Seq data only confirms the expressions of these PCWDEs, further study is needed to confirm their enzymatic activities.

GH1 β-glucosidases (cellobiases) are endogenously expressed in all herbivorous insects. This cellulase family can break down cellobiose into two glucose residues. In phasmids, unlike GH9 endoglucanase, which is almost restrictedly expressed in AMG, GH1 cellobiases (and GH2 and GH3) have wide expression profiles through the alimentary canal and even in the head, presumably in the salivary gland. However, without activities from endoglucanases that break up the cellulose fibers, cellobiases have low efficiency in degrading cellulosic materials into monomers. This expression pattern was observed in a higher termite group where the only PCWDE enzymes expressed in their salivary glands were GH1 β-

glucosidases. Besides digestion, other biological functions played by β-glucosidases in termites have been reported, such as suppressing worker reproduction by the queen in *Cryptotermes secundus* (Korb, Weil et al. 2009) and producing egg pheromones for recognition in *Reticulitermes speratus* (Matsuura, Yashiro et al. 2009). It is promising that GH1 β-glucosidases have diversified functions, especially when there are abundant transcripts identified in multiple tissues with high expressions. However, such studies are still lacking.

We found traces of α-glucosidases from GH31 and GH63 expressed in multiple tissues of phasmids. The primary function of α-glucosidases is in carbohydrate hydrolysis in hydrolyzing the α-linked glucose residue from the non-reducing terminus of a substrate. Polysaccharides in PCWs are mostly connected by β glycosyl linkages, with a few α linkages that present in the side chains in hemicelluloses and pectin. Therefore, little involvement in the PCW degradation of phasmids was suggested for these families. The global expressions of α-glucosidases in Figures 5 and 6 also supported other biological functions. In *Apis mellifera* L., three types of α-glucosidase have been identified and characterized with different substrate specificities and expression locations. Only type III α-glucosidase, which is expressed in hypopharyngeal glands, is involved in breaking sucrose into glucose and fructose with a high efficiency; the other two types showed no to low activity towards soluble sucrose in the ventriculus and hemolymph (Kubota, Tsuji et al. 2004). In *Rhodniu prolixus*, the kissing bug that consumes blood, α-glucosidase is involved in hemoglobin digestion and hemozoin formation (Mury, da Silva et al. 2009). A similar global expression pattern of GH31 α-glucosidase was reported in brown plant hoppers, *Nilaparvata lugens*, in which they claimed a whole-body function that is involved in detoxification towards phenolic glycosides from plant leaves (Miao, Jia et al. 2018). In phasmids, we assumed that α-glucosidase contributes to a similar biological process.

Overall, the *de novo* transcriptome analysis provided strong clues about putative gut-restricted or whole-body biological functions. Future experimental work will hopefully follow up on these suggestions. Additionally, with available genomic data, our work could enable evolutionary studies in GH gene families,

which would shed light on gene family expansion and the phylogenetic relationships within the lineage of Polyneoptera, or herbivorous insects.

In the biofuel industry, cellulases, hemicellulases, and pectinases are extremely critical and in high demand in both cellulosic fuel production and waste treatment (Carroll and Somerville 2009). As the high cost of industrial PCWDEs prioritizes the search for more active glycoside hydrolases from incompletely explored sources like termites (Carroll and Somerville 2009), Phasmatodea can be worth looking at, because phasmids strictly feed on leaves and have extensive PCWDEs.

Compartmentalization of Digestion in Phasmids

Our results based ontissue-specific expression of PCWDEs, GO annotations of DEGs, and enrichment analyses for DEGs suggest that phasmid AMGs are a digestively significant compartment of the midgut. This is consistent with claims made by previous morphological studies on phasmid alimentary structure, which presumed AMG as a major digestive sub-tissue because the pleated structure increases surface area and slows down food bolus (Shelomi, Watanabe et al. 2014). Major categories of PCWDE families include GH9 endoglucanases, GH1 cellobiases, and GH28 polygalacturonases. Synergistically, these three families can digest cellulosic polysaccharides down to single glucose residues (Watanabe and Tokuda 2010). Therefore, it is plausible that phasmids are capable of fully degrading major glycosyl compounds.

The 1,000 most highly expressed genes in each tissue belong to the midgut and hindgut provided further information on the compartmentalization of digestion in phasmids. Generally, in leaf-feeding insects, leaf material is chewed up by mouthparts lubricated by saliva. Salivary gland can secrete saliva with various digestive enzymes like α-glucosidases, amylases, cellulases, and laminarinases (Terra 1990, Watanabe and Tokuda 2010, Da Lage 2018). However, the digestive power these enzymes provide for the saliva is subsidiary. GH1 β-glucosidase were found expressed in phasmid head tissue in our results. As the saliva of

higher termites has enzymatic function from the same gene family (Scharf, Karl et al. 2011) , it is plausible that the saliva of phasmids plays a similar subsidiary role in the first step of food processing.

Hydrolase activities are clearly up-regulated and important in phasmid AMGs, according to our GO annotations and enrichment analysis on the most highly expressed genes. This was expected, given that the insect midgut has been known as the major tissue where the most intense digestive process happens (Waterhouse 1957). In phasmids, midgut tissue has two compartments based on their distinctive structures, as shown in Figures 5 and 6 (also Azevedo, Fialho et al. (2013), Shelomi and Kimsey (2014)). Until recently, however, the digestive compartmentalization of function between the phasmid anterior and posterior sections was poorly understudied.

The pleated structure of the AMG makes this section a highly possible place to perform molecule degradation (Shelomi, Jasper et al. 2014). We used our RNA-Seq data to further confirm this hypothesis and found that the GO terms of various hydrolase activities are the most enriched among the up-regulated genes in the AMG. On the contrary, the up-regulated genes we found in PMG were enriched in protein degradation. Annotations of these DEGs supported this suggestion, with protease activity being one of the top molecular function annotations in both species.

Based on the tissue-specific expression profiles of these DEGs (Figures 7 and 8), similar expression patterns of the DEGs were observed among phasmid PMGs and two hindgut tissues (ileum and rectum). Specifically, we found that a fraction of up-regulated genes in PMGs kept high expression levels in the ileum, and the fraction of these genes that were highly expressed in the rectum was smaller. Not only did this result suggest a unique digestive tissue in phasmid AMGs, but also suggested that the PMG is a sub-tissue that shares some biological functions with the hindgut. In other words, the PMG is not as functionally distinctive as the AMG to the rest of the alimentary canal.

Conclusion

In this chapter, we explored tissue-specific expression of PCWDEs and DEGs (between the AMG and PMG). We conducted GO annotations and enrichment analysis for these DEGs and explored the function of the 1,000 most highly expressed genes in each tissue. Major PCWDE families like GH9 endoglucanases, GH1 cellobiases, and GH28 polygalacturonases were highly expressed in the AMG, while fewer genes from these three families showed expression in the PMG. As for the hindgut tissues (ileum and rectum), GH9 and GH28 showed close-to-zero TPM values, while β-glucosidases from GH1, GH2, and GH3 showed light expression in the hindgut. We found the GH families like GH2, GH3, and GH30 that represent a variety of hemicellulases showed wider expression profiles in multiple tissues than the most abundant cellulase and pectinase families. Some glycoside hydrolase families, such as α-glucosidases from GH31 and GH63, and α-galactosidases from GH27, may not be involved in PCW degradation, however, their global expression profiles in all selected tissues suggest their putative whole-body biological functions.

The compartmentalization of phasmid digestion was studied based on the DEGs between the anterior and posterior midguts, as well as the 1,000 most highly expressed genes from each alimentary tissue. We observed significant differentiation between the anterior and posterior midguts in carbohydrate metabolism. The most up-regulated genes in the AMG are mainly various hydrolase activities, including polygalacturonases, β-glucosidases, β-galactosidases, and glucosylceramidases. On the other hand, the PMG is up-regulated in a variety of proteases such as metallopeptidases, aminopeptidases, and metallocarboxypeptidases. While the AMG and the PMG are highly differentiated in GO annotations for their most highly expressed genes, the ileum and rectum showed similar top GO terms distributions. More enzymes that are involved in energy generation and transmembrane transportation are highly expressed in these two hindgut tissues compared to the midgut.

In a nutshell, the compartmentalization of physiological expertise in phasmid alimentary canals was shown in our results. Anterior and posterior midguts are major localizations of the breaking down of carbohydrates and proteins respectively. Meanwhile, hindgut tissues are more focused on the transmembrane transportations of molecules and ATP/GTP synthesis.

CHAPTER 3 METAGENOMIC STUDY OF PHASMATODEA MIDGUT MICROBIOME

Overview

Metagenomics has been widely applied to study the composition and functional aspects of symbiotic microorganisms. various Metagenomics uses direct sequencing of genetic material extracted from the host and microbes (Riesenfeld, Schloss et al. 2004). In other words, metagenomic analysis can provide researchers with information about what microorganisms exist within a body and what are they doing. This technique makes it possible for scientists to study microorganisms that are not culturable, which consists of 99% of all known microbes (Handelsman 2004). Metagenomics can be used to study microbiomes in samples from various environments such as soil, water, and gastrointestinal tracts (Sleator, Shortall et al. 2008, Joynson, Pritchard et al. 2017).

In general, there are two approaches to metagenomic analysis. The first makes use of the targeted sequencing of genomes of all species present in an environment to identify taxonomic composition by matching 16S rDNA sequences (Handelsman 2004, Wooley, Godzik et al. 2010, Kim, Yoon et al. 2013). While this approach is currently the most widely used, it has limitations in that rRNA sometimes does not contain sufficient variation to differentiate between many classes of microbes. One also does not get any functional information about the genetic capacity of the microbes with this approach. The second approach, called the shotgun approach, is to make DNA libraries from microorganism containing samples (gut compartment contents in our case) and sequence everything (not just the rRNA). In this approach, whole gene sequences of all the present microbes are obtained (Liu, Gibbons et al. 2011, Kelley, Liu et al. 2012). Bioinformatics can further allow one to (1) use diagnostic genes present in restricted microbial lineages to make determinations of presence and absence, and (2) to use the sequences themselves (via alignment with genes of known function) and their expression levels as measures of the functional capabilities of the microbes.

Traditionally, people assumed symbiotic microorganism are responsible for the degradation of PCWs (Watanabe and Tokuda 2010) in herbivorous insects. Several metagenomic studies have been conducted on the hindguts and midguts of termites in support of this (Warnecke, Luginbuhl et al. 2007, Hongoh 2011, Liu, Zhang et al. 2013, Berlanga, Llorens et al. 2016, Hu, da Costa et al. 2019, Victorica, Soria et al. 2020). These studies provided strong evidence that bacterial symbionts are critical in termites' degradation of celluloses and hemicelluloses. Warnecke, Luginbuhl et al. (2007), for example, aligned shotgun reads from termite guts against known bacterial GH sequences, from which they identified more than 100 gene fragments analogous to glycoside hydrolases. Undoubtably, the contribution of symbiotic microbes to insect PCW digestion is vital, especially in termites, cockroaches, and Scarabaeoid beetles (Slaytor 1992, Lundgren and Lehman 2010, Ni and Tokuda 2013, Brune 2014, Li, Young et al. 2020). However, recent studies have reported endogenous PCWDEs in insects playing critical roles in digestion (Pauchet, Kirsch et al. 2014, Shelomi, Jasper et al. 2014, Chang, Yendrek et al. 2016, Antony, Johny et al. 2017, Gao, Liu et al. 2020). In phasmids, many papers and reports have suggested that endogenous PCWDEs, those produced by the insects, play dominant roles in PCW breakdown. Further, given that the structure of phasmid hindguts is unlikely to house populations of symbiotic microbes, phasmids may be especially dependent on PCWDEs (Shelomi, Watanabe et al. 2014, Wu, Crowhurst et al. 2016). However, phasmids are still quite understudied with respect to their gut microbiota.

Researchers have not ruled out the possibility that microbial symbionts might contribute to PCW digestion, especially in the midgut in Phasmids; even though a study claimed that no microbes were degrading PCWs in stick insects. This study used 454 pyro-sequencing (with shallow depth) and 16S classification (Shelomi, Lo et al. 2013). Here, we propose using the shotgun sequencing approach to conduct a metagenomic analysis to take a closer look at the microbiome of Phasmids with a higher coverage and read depth. Our study will illuminate the microbial composition of stick insect midguts. Furthermore, our data will allow us to determine if any PCWDEs originate from bacteria. If not, we could concur that

PCW breakdown is solely based on endogenous PCWDEs. Our hypothesis is that microbial PCWDEs do exist in phasmid midguts, although we are unsure whether they have PCW breakdown capacity.

Methods

Insect Rearing and Dissection

The same species of stick insects as in Chapters 1 and 2 were used in this chapter: *Aretaon asperrimus* and *Medauroidea extradentat*a. All insects were cultured at room temperature at the University of California, Davis. All individuals were fed with cleansed rose leaves collected from campus. The cleansing process for the feeding material included a first wash with soap water followed by two washes with deionized water due to the impact on folivorous insects' gut microbiomes from the soil environment of the host plant (Hannula, Zhu et al. 2019).

Midgut tissue for environmental DNA extraction was dissected in empty petri dishes using sterilized razor blades with careful observation under microscopy. After the removal of the head and the slicing of the ventral cuticle, the whole alimentary canal was pulled out. Anterior and posterior midguts were then dissected with their gut contents included for the following extraction. Each sample was pooled from the same tissues of three individuals and collected in 1.5 ml microtubes for subsequent DNA extraction.

DNA Isolation and Library Preparation

Pooled tissue was ground gently in buffer ATL provided by a DNeasy® Blood & Tissue Kit (QIAGEN, CA, USA). This was followed by DNA extraction guided by this kit's protocol. Isolated DNA went through quality control process on a NanoDrop ND-1000 spectrophotometer ([NanoDrop Technologies, Wilmington, DE, USA] for purity, Qubit 3 Fluorometer for concentration, and Agilent 2100 Bioanalyzer [Agilent Technologies, Palo Alto, CA, USA] for quality). The DNA samples were acoustically sheared to an average length of 200 bp using a Covaris E220 Focused Ultrasonicator (Covaris, MA, USA). This was followed library construction using AMPure XP beads and the NEBNext® Ultra DNA Library Prep Kit

(New England Biolabs, MA, USA), guided by the manufacturer's protocol. Library quality was accessed using an Agilent 2100 Bioanalyzer, and library concentration was measured by a Qubit 3 Fluorometer. The procedure described above was repeated three times for each tissue in each species.

Sequencing

Next, 150 base pair paired-end sequencing was performed on the HiSeq 4000 at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley. Raw reads were uploaded to the MG-RAST Web server [Project name: Real Metagenomics Project Aretaon and Medauroidea]. Three library replicates were prepared for all tissues except for the posterior midgut of *M. extradentata* due to a failure of library preparation.

Analyzing Platform

We conducted the metagenomic analysis on the MG-RAST server (Keegan, Glass et al. 2016) and the metagenomics rapid annotation using a subsystems technology server. This platform possesses multiple features that helped us address some of the most common issues in metagenomic study.

The analysis pipeline mainly included the removal of low-quality reads by SolexaQA (Cox, Peterson et al. 2010), gene calling, and protein identification. Because PCR amplification was included in the library construction protocol, duplicate read inferred sequencing error estimation (DRISEE) (Keegan, Trimble et al. 2012) was applied after the preprocessing to analyze the sets of ADRs (Gomez-Alvarez, Teal et al. 2009) and determine the degree of variation among prefix-identical sequences derived from the same template. MG-RAST gene calling is based on machine learning approaches and utilizes FragGeneScan (Trimble, Keegan et al. 2012), which was developed from a mature algorithm (Trimble, Keegan et al. 2012) to translate short and error-prone prokaryotic nucleic acid sequences for searches on protein similarity. The protein database used for annotation was derived from the M5nr (MD5-based non-redundant protein database) (Wilke, Harrison et al. 2012), a comprehensive database including several available data sources

including GO, the JGI (Joint Genome Institute), the KEGG (Kyoto Encyclopedia of Genes and Genomes), the NCBI (National Center for Biotechnology Information), SEED (The SEED Project), UniProt (UniProt Knowledgebase), VGI (Virginia Bioinformatics Institute), and eggnog (evolutionary genealogy of genes: Non-supervised Orthologous Groups) (Wilke, Harrison et al. 2012).

Results

Taxonomic Composition of Bacterial Community in Phasmid Midguts

We generated 20,992,128 reads that past QC and identified 313,472 proteins on average in each library, as shown in Table 4. Table 4 also listed library ids published on the MG-RAST server for future studies.

Table 4. Analytic Data of Metagenomic library of AMG and PMG in *A. asperrimus* (ARE) and *M. extradentata* (VIET)

Sample Name	Mgrast_Id	Seq Count	Post QC Mean Seq Length	Post QC Seq Count	Identified Protein Features
ARE_AMG_M_1	mgs816625	18,194,487	228 ± 37 bp	16,203,294	306,481
ARE_AMG_M_2	mgs816628	23,045,909	228 ± 37 bp	20,849,200	337,271
ARE_AMG_M_3	mgs816631	17,114,678	230 ± 37 bp	15,293,320	197,796
ARE_PMG_M_1	mgs816634	21,989,822	225 ± 38 bp	18,955,157	371,996
ARE_PMG_M_2	mgs816637	22,390,471	229 ± 37 bp	20,110,466	262,666
ARE_PMG_M_3	mgs816640	20,050,693	229 ± 37 bp	17,507,156	233,470
VIET_AMG_M_1	mgs816643	19,546,017	224 ± 38 bp	17,519,684	237,403
VIET_AMG_M_2	mgs816646	22,631,987	226 ± 37 bp	19,955,430	261,831
VIET_AMG_M_3	mgs816649	19,519,948	230 ± 37 bp	17,558,691	237,001
VIET_PMG_M_1	mgs816652	53,550,717	188 ± 47 bp	45,968,877	688,809
Mean		23,803,473		20,992,128	313,472

On the MG-RAST analyzing platform, we filtered all bacterial reads from the RefSeq database with an expectation value above e^{-5}, identities above 60%, and lengths above 20bps. The results in Tables 5 and 6 are the bacterial phyla present in each tissue replicate of *A. asperrimus* and *M. medauroidea*.

The most abundant bacterial phylum present in both phasmids' midguts in all replicates was Proteobacteria. The other dominant bacteria were Firmicutes, Actinobacteria, Bacteroidetes, and Cyanobacteria. Similar bacterial taxonomic compositions were shown for other less dominant species.

However, surprisingly, although Spirochaetes were at low levels in *A. asperrimus*, this phylum was the second-most abundant in *M. extradentata*'s midgut microbiome (Table 6).

Table 5. Bacterial Phyla present in *Aretaon asperrimus* Midgut Microbiome

	# of reads in each sample replicate											
phylum	**ARE_AMG_M_1**		**ARE_AMG_M_2**		**ARE_AMG_M_3**		**ARE_PMG_M_1**		**ARE_PMG_M_2**		**ARE_PMG_M_3**	
Acidobacteria	69	0.235%	55	0.876%	43	0.936%	70	0.122%	58	1.165%	68	0.635%
Actinobacteria	323	1.100%	463	7.370%	907	19.739%	491	0.859%	530	10.643%	5322	49.729%
Aquificae	3	0.010%	3	0.048%	1	0.022%	3	0.005%	1	0.020%	3	0.028%
Bacteroidetes	257	0.875%	370	5.890%	288	6.268%	317	0.554%	340	6.827%	270	2.523%
Candidatus Poribacteria	0	0.000%	0	0.000%	0	0.000%	1	0.002%	0	0.000%	0	0.000%
Chlamydiae	4	0.014%	5	0.080%	6	0.131%	3	0.005%	4	0.080%	2	0.019%
Chlorobi	48	0.163%	39	0.621%	35	0.762%	46	0.080%	47	0.944%	34	0.318%
Chloroflexi	29	0.099%	41	0.653%	25	0.544%	37	0.065%	33	0.663%	55	0.514%
Chrysiogenetes	1	0.003%	0	0.000%	0	0.000%	0	0.000%	0	0.000%	0	0.000%
Cyanobacteria	191	0.650%	201	3.200%	43	0.936%	255	0.446%	62	1.245%	78	0.729%
Deferribacteres	1	0.003%	0	0.000%	0	0.000%	2	0.003%	0	0.000%	1	0.009%
Deinococcus-Thermus	16	0.054%	12	0.191%	10	0.218%	16	0.028%	22	0.442%	20	0.187%
Dictyoglomi	3	0.010%	5	0.080%	2	0.044%	3	0.005%	2	0.040%	3	0.028%
Elusimicrobia	3	0.010%	0	0.000%	2	0.044%	0	0.000%	3	0.060%	0	0.000%
Fibrobacteres	0	0.000%	0	0.000%	0	0.000%	0	0.000%	0	0.000%	1	0.009%
Firmicutes	707	2.407%	791	12.592%	597	12.992%	1039	1.817%	757	15.201%	782	7.307%
Fusobacteria	61	0.208%	66	1.051%	33	0.718%	110	0.192%	56	1.124%	31	0.290%
Gemmatimonadetes	4	0.014%	0	0.000%	2	0.044%	3	0.005%	0	0.000%	3	0.028%
Lentisphaerae	0	0.000%	2	0.032%	0	0.000%	0	0.000%	0	0.000%	2	0.019%
Nitrospirae	2	0.007%	4	0.064%	1	0.022%	2	0.003%	0	0.000%	1	0.009%
Planctomycetes	22 (	0.075%	24	0.382%	20	0.435%	22	0.038%	18	0.361%	24	0.224%
Proteobacteria	27515	93.681%	4044	64.374%	2467	53.689%	54624	95.530%	2896	58.153%	3852	35.993%
Spirochaetes	47	0.160%	70	1.114%	40	0.871%	46	0.080%	57	1.145%	48	0.449%
Synergistetes	14	0.048%	13	0.207%	14	0.305%	16	0.028%	28	0.562%	21	0.196%
Tenericutes	9	0.031%	17	0.271%	11	0.239%	13	0.023%	9	0.181%	14	0.131%
Thermotogae	8	0.027%	10	0.159%	7	0.152%	15	0.026%	8	0.161%	4	0.037%
Verrucomicrobia	32	0.109%	44	0.700%	37	0.805%	41	0.072%	43	0.863%	55	0.514%
unclassified (derived from Bacteria)	2	0.007%	3	0.048%	4	0.087%	5	0.009%	6	0.120%	8	0.075%

Functional Annotation of Bacterial Community in Phasmid Midguts

We identified and annotated all bacterial sequences based on SEED subsystems (Overbeek, Olson et al. 2014) with the expectation value set at e^{-15}, minimal identity set at 95%, the minimal sequence length set at 80bps, and the minimal abundance at 15 reads. This is a stricter filter compared to the taxonomic alignment because it is more confident for us to infer the functional composition of bacterial genes.

The results in Figures 18 (AMG) and 19 (PMG) showed the compositions of the most abundant functional genes that originated from phasmid midgut (anterior and posterior) microbiomes. As shown in the figures, similar compositions of the most abundant functional annotations were present in the anterior

and posterior midgut, suggesting similar populations of microbiomes in the two sections of phasmid midguts. We found that carbohydrate metabolism plays a major part among these functional annotations for bacterial genes in both species. Protein and derivative metabolism were also pronounced. The functional annotation of "clustering-based subsystems" is one of the top three functional annotations that contains functions involved in ribosomes, proteosomes, and recombination-related clusters (Delmont, Prestat et al. 2012).

Among the reads that were categorized as being involved in carbohydrate metabolisms, we used SEED subsystems level 2 annotation to find the distribution of these bacterial genes in different types of

Table 6. Bacterial Phyla present in *Medauroidea extradentata* Midgut Microbiome

	# of reads in each sample replicate							
Phylum	**VIET_AMG_1**		**VIET_AMG_2**		**VIET_AMG_3**		**VIET_PMG_1**	
Acidobacteria	54	0.596%	72	0.704%	56	0.595%	148	0.692%
Actinobacteria	389	4.295%	476	4.656%	375	3.986%	1270	5.941%
Aquificae	3	0.033%	0	0.000%	2	0.021%	11	0.051%
Bacteroidetes	166	1.833%	176	1.722%	175	1.860%	556	2.601%
Candidatus Poribacteria	0	0.000%	0	0.000%	2	0.021%	0	0.000%
Chlamydiae	5	0.055%	5	0.049%	9	0.096%	12	0.056%
Chlorobi	40	0.442%	38	0.372%	38	0.404%	123	0.575%
Chloroflexi	22	0.243%	45	0.440%	35	0.372%	109	0.510%
Cyanobacteria	159	1.756%	104	1.017%	109	1.158%	617	2.886%
Deferribacteres	3	0.033%	1	0.010%	1	0.011%	3	0.014%
Deinococcus-Thermus	11	0.121%	9	0.088%	7	0.074%	21	0.098%
Dictyoglomi	2	0.022%	5	0.049%	2	0.021%	10	0.047%
Elusimicrobia	1	0.011%	0	0.000%	0	0.000%	1	0.005%
Fibrobacteres	0	0.000%	1	0.010%	0	0.000%	3	0.014%
Firmicutes	544	6.006%	484	4.734%	558	5.930%	1518	7.101%
Fusobacteria	6	0.066%	3	0.029%	4	0.043%	17	0.080%
Gemmatimonadetes	4	0.044%	1	0.010%	1	0.011%	3	0.014%
Lentisphaerae	3	0.033%	3	0.029%	3	0.032%	5	0.023%
Nitrospirae	2	0.022%	0	0.000%	1	0.011%	8	0.037%
Planctomycetes	18	0.199%	17	0.166%	10	0.106%	53	0.248%
Proteobacteria	6172	68.146%	7280	71.212%	6351	67.499%	14839	69.416%
Spirochaetes	1424	15.723%	1472	14.399%	1640	17.430%	1945	9.099%
Synergistetes	1	0.011%	3	0.029%	8	0.085%	8	0.037%
Tenericutes	8	0.088%	12	0.117%	5	0.053%	30	0.140%
Thermotogae	4	0.044%	2	0.020%	1	0.011%	9	0.042%
Verrucomicrobia	15	0.166%	10	0.098%	15	0.159%	48	0.225%
unclassified (derived from Bacteria)	1	0.011%	4	0.039%	1	0.011%	10	0.047%

sugar processing. The results are shown in Figure 20. The subsystems level 2 molecular functions that were

annotated for most bacterial reads were monosaccharides and Di-/oligosaccharide metabolism, as shown in Figure 20. The same dominant components of these two categories in carbohydrate metabolisms were present in bacterial genes in phasmid PMGs, as shown in Figure 21. This result indicated that bacteria in phasmid midguts plays a putative role in processing degraded polysaccharides, or that the microbes are commensals or parasites that feed of these nutrients.

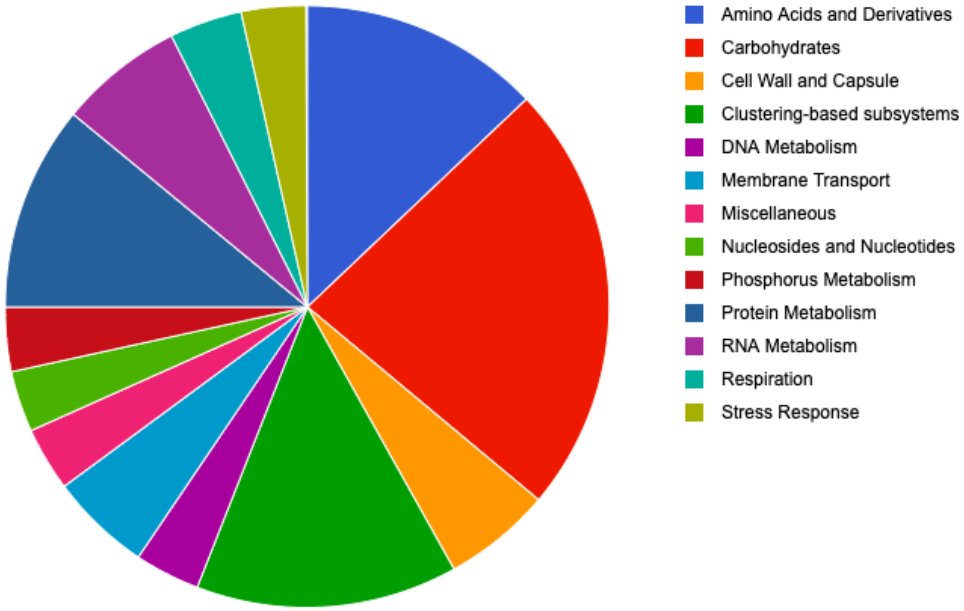


Figure 18. Composition of Most Abundant Functional Annotations of Genes from Phasmid's AMG microbiome. Reads were filtered sequences taxonomically belong to Bacteria domain in The Reference Sequence (RefSeq) collection. SEED database alignment filters include E-value being 15, minimal read length being 80bps, minimal identity being 95% and minmal abundance being 15 reads.

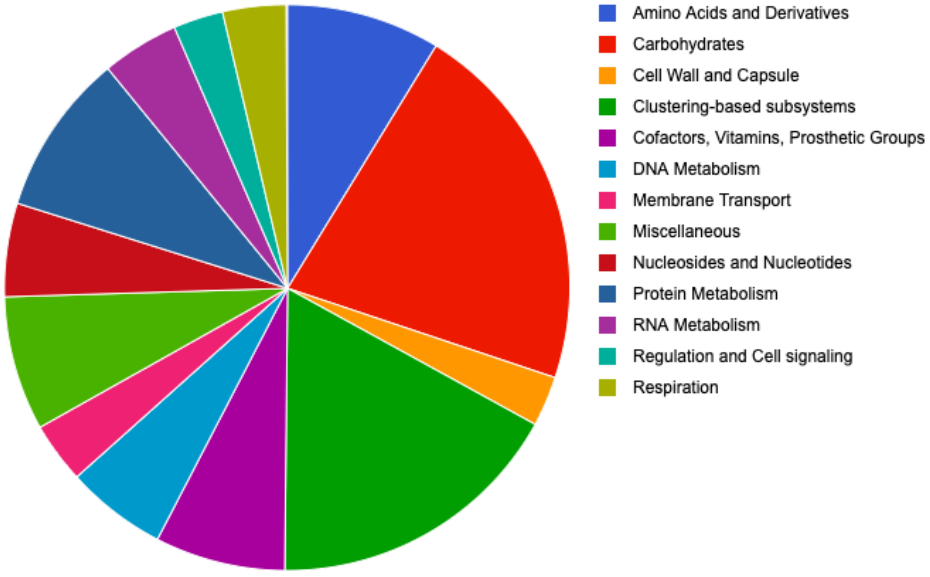


Figure 19. Composition of Most Abundant Functional Annotations of Genes from Phasmid's PMG microbiome. Reads were filtered sequences taxonomically belong to Bacteria domain in The Reference Sequence (RefSeq) collection. SEED database alignment filters include E-value being 15, minimal read length being 80bps, minimal identity being 95% and minmal abundance being 15 reads.

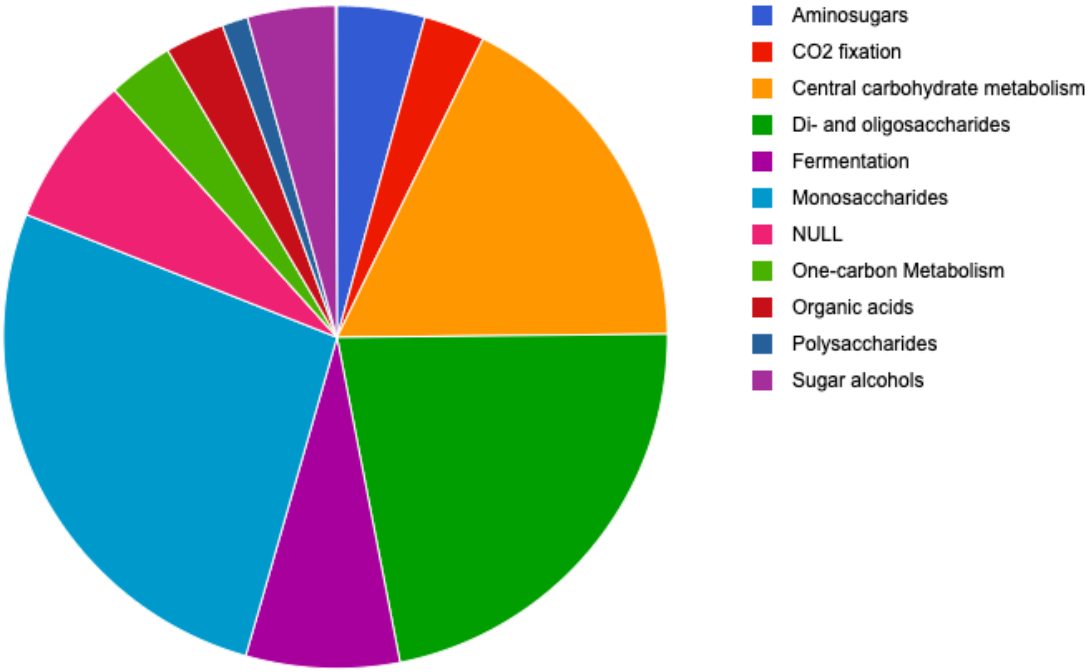


Figure 20. Composition of SEED Subsystems Level-2 Functional Annotations of Genes that are involved in Carbohydrates metabolism from Phasmid's AMG microbiome.

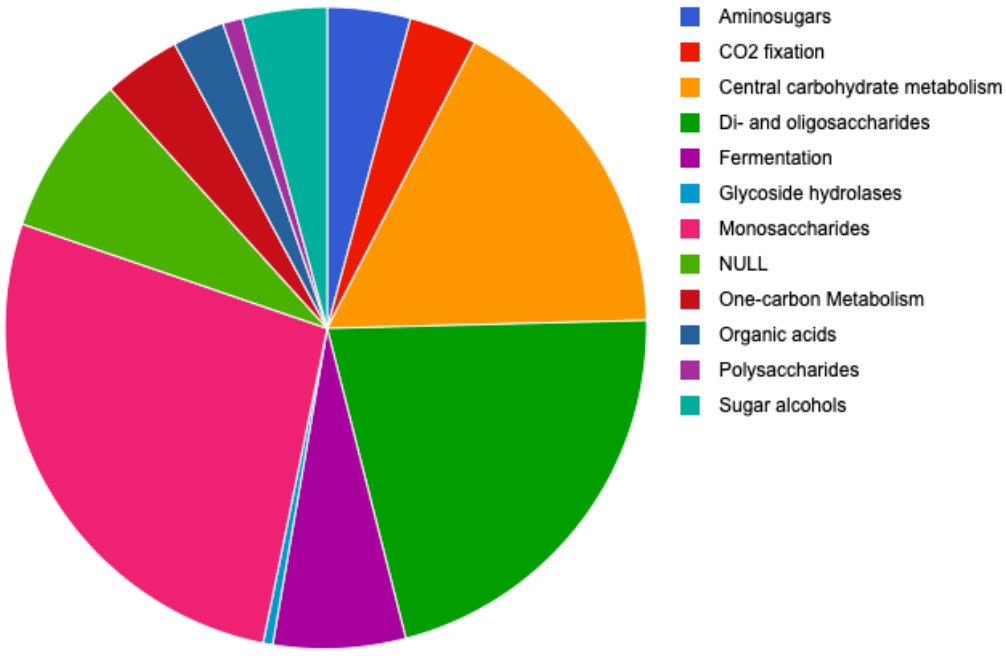


Figure 21. Composition of SEED Subsystems Level-2 Functional Annotations of Genes that are involved in Carbohydrates metabolism from Phasmid's PMG microbiome.

Putative PCWDEs of Bacterial Origin

To identify putative PCWDE genes from phasmid midgut bacteria, we applied the following filters on all available databases that MG-RAST provided and downloaded these sequences. For SEED subsystems level 2, we filtered sequences that were annotated to "glycoside hydrolases." For COG, we filtered sequences annotated as "cellulase and related proteins" (functions) and "carbohydrate transport and metabolism" (level 2). For KO and NOG, we chose sequences annotated as "carbohydrate metabolism."

We downloaded all known bacterial PCWDE sequences from NCBI and made a local blast database. Tblastx and InterPro scan were conducted on Blast2GO (Conesa, Gotz et al. 2005) against our local known bacterial PCWDE database and the public EMBL-EBI InterPro server. We successfully identified a variety of putative cellulases, pectinases, and hemicellulases that originated from bacteria in phasmid midguts, as shown in Table 7.

Beta-glucosidases were the dominant cellulases from bacteria in both phasmids' anterior midguts. endo-1,4-glucanases only showed in *A. asperrimus*' midguts and were aligned to genes from Firmicutes and unknown bacteria collected from environmental samples. Source bacterial phylum of beta-glucosidases include Proteobacteria, Actinobacteria, Firmicutes, Bacteroidetes, and Spirochaetes. In the microbiomes from two sections of midgut in *A. asperrimus*, 73 beta-glucosidase sequences were found in the PMG, whereas 35 were found in the AMG. A more abundant source phylum of beta-glucosidases was Proteobacteria, with 33 sequences in the PMG, compared to nine sequences from this phylum in the AMG. Interestingly, in microbiomes of *M. extradentata*'s midgut, only seven sequences of beta-glucosidases in total were identified over two phyla: Actinobacteria and Firmicutes. Together with the result that there were no endo-1,4-glucanases identified, a lack of bacterial cellulases in *M. extradentata* midguts was inferred.

A few bacterial hemicellulases were identified from *A. asperrimus*' midgut microbiome, and all of them belonged to Firmicutes. One beta-1,3-glucanase and one beta-xylosidase were present in the AMG, whereas nine beta-xylosidases were present in the PMG.

Abundant bacterial pectinases were identified in both species' midgut microbiomes: 135 sequences from *A. asperrimus* and 524 sequences from *M. extradentata* (Table 7). In *A. asperrimus*, all polygalacturonases identified were from the PMG microbiome, whereas in *M. extradentata*, microbiomes of both sections showed sequences of this enzyme. Comparing the PMG against the AMG of both species, we noticed more bacterial polygalacturonases present in the posterior section of midgut – 513 sequences in the PMG versus 146 sequences in the AMG. These pectinases came from similar bacterial phyla in both species, which included Proteobacteria, Acidobacteria, and Firmicutes. Two additional sequences were from Spirochaetes in the PMG of *M. extradentata*. Among these bacteria that possessed a great number of pectinases, Proteobacteria was the most significant source that comprised more than 90% of all sequences.

Table 7. Number of Bacterial Sequences Identified as PCWDEs in Phasmatodea Midgut Microbiome

		Cellulase				Pectinase		Hemicellulase			
		beta-glucosidase		beta-1,4-endoglucanase		polygalacturonase		beta-1,3 glucanase		beta-xylosidase	
Species	**Tissue of Microbiome**	**# of sequencees**	**Source phylum**	**# of sequencees**	**Source phylum**	**# of sequences**	**Source Phylum**	**# of sequences**	**Source Phylum**	**# of sequences**	**Source Phylum**
***Aretaon asperrimus* (Are)**	Anterior Midgut	12	Actinobacteria	2	unknown			1	Firmicutes	1	Firmicutes
		11	Firmicutes								
		9	Proteobacteria								
		3	Bacteroidetes								
	Posterior Midgut	16	Actinobacteria	1	Firmicutes	130	Proteobacteria			9	Firmicutes
		5	Bacteroidetes	9	unknown	4	Acidobacteria				
		18	Firmicutes			1	Firmicutes				
		33	Proteobacteria								
		1	Spirochaetes								
***Medauroidea extradentata* (Viet)**	Anterior Midgut	2	Actinobacteria			2	Firmicutes				
		2	Firmicutes			12	Acidobacteria				
						132	Proteobacteria				
	Posterior Midgut	3	Actinobacteria			5	Firmicutes				
						17	Acidobacteria				
						354	Proteobacteria				
						2	Spirochaetes				

Discussion

Taxonomic and Functional Composition of Phasmid Midgut Microbiome

The composition of the midgut microbiota of insects is influenced by several factors, such as insect development, structural and physiological conditions in different alimentary tissues regions, and diet references and conditions (Wang, Wang et al. 2020). Generally, the hindgut of insects has been the focal tissue to study insect gut microbiota due to its projections in which microbes live, and due to being the last part of the alimentary canal to contain degraded bolus of food. In Phasmatodea, however, not only are there few studies on gut microbiota, but several characteristics of its alimentary canal have also suggested that it is unlikely to house symbionts. Moreover, the most likely tissue to house gut symbionts to help phasmid digestion is the midgut (Shelomi, Lo et al. 2013).

In this chapter, we successfully assessed the taxonomic compositions of the midgut microbiome in two species of stick insects using high throughput deep sequencing. Our results showed that in both species, Proteobacteria was significantly abundant in two sections of the midgut microbiome. It was reported before that Proteobacteria is the most dominant phylum in insects' gut microbial communities (Yun, Roh et al. 2014). The other major bacterial phyla (Actinobacteria, Bacteroidetes, Firmicutes, and Cyanobacteria) were also known to be abundant in the gastrointestinal environments of other phytophagous insects (Victorica, Soria et al. 2020). Our data showed a significant difference between the numbers of Spirochaetes that were identified in the two species' gut microbiomes (Tables 5 and 6). The phylum of Spirochaete, especially one of its genera, Treponema, contributes critically to plant biomass digestion in termite hindguts, as reported in Warnecke, Luginbuhl et al. (2007); moreover, most genes that encode CAZymes in termite hindguts were taxonomically belonging to Sprichaetes and Fibrobacteres (He, Ivanova et al. 2013). The almost complete lack of Fibrobacteria in phasmid gut microbiota indicated a different PCW-degrading scheme verses termite. Given that we identified a great number of PCWDE transcripts strongly expressed in phasmid midguts, it is possible that endogenous PCWDEs compensated for the lack of Fibrobacteria and Spirochaetes in phasmid midguts. Due to same feeding material and the living environment we provided

for both species, the much lower number of Spirochaetes identified in *A. asperrimus*, compared to *M. exradentata*, showed us putative different PCW-degrading designs in the midguts of these two species.

The presence of a microbial taxon in an insect's gut does not necessarily mean that it contributes to physiological functions of the host insect. However, most phytophagous insects house midgut or hindgut microbes with PCW-degrading capabilities (Hatefi, Makhdoumi et al. 2017, Peterson and Scharf 2018, Pothula, Shirley et al. 2019, Amiri, Bandani et al. 2020). Our study, the result of a global functional annotation of all bacterial sequences in phasmid gut microbiomes, showed us strong capabilities of phasmid's gut bacteria in metabolizing carbohydrates (Figures 18-21).

Putative Bacterial PCWDEs

From all bacterial sequences, we identified cellulases and pectinases. Specifically, bacterial PCWDEs present in phasmid's midgut microbiomes include beta-glucosidases, endoglucanases, polygalacturonases, beta-xylosidases, and beta-1,3-glucanases. Among these PCWDEs, Polygalacturonase was the most abundant one identified in both species' midgut microbiomes. Polygalacturonase is an important pectinase that helps phytophagous insects to degrade plant leaves. In *Bombyx mori*, pectin degradation is dependent on the polygalacturonases provided by gut symbiotic bacteria (Anand, Vennison et al. 2010). In Chapter 2, we observed that enriched genes encoded for polygalacturonases were highly expressed in the anterior section of midgut. Together with the abundant bacterial pectinase sequences from phasmid posterior midguts, it suggested a possible synergism between endogenous and bacterial enzymes in pectin degradation where endogenous pectinases are major players in AMG, and bacterial pectinases provide more digestive power towards pectin in PMG.

The major sources of bacterial cellulases are Actinobacteria, Proteobacteria, and Firmicutes. Proteobacteria and Firmicutes are reportedly dominant bacteria phyla in insect's gastrointestinal environments (Engel and Moran 2013), and beta-glucosidase is one of the most common glycoside

hydrolases possessed by bacteria (Slaytor 1992, Liu, Li et al. 2019). The suggestion that bacterial beta-glucosidases are helping the host to fully degrade celluloses to monomers is unlikely. However, given that beta-glucosidases provide much digestive capacity in termites, cockroaches, and beetles (Hongoh 2011, Liu, Zhang et al. 2013, Ni and Tokuda 2013, Yun, Roh et al. 2014, Liao, Guo et al. 2020), this suggestion is not impossible. Further studies on phasmid gut microbiota require biochemical analysis on these bacterial PCWDEs.

Conclusion

In this chapter, we assessed the taxonomic and functional composition of the midgut microbiota of two Phasmatodea species, *Aretaon asperrimus* and *Medauroidea extradentata*. Dominant bacteria phyla including Proteobacteria, Actinobacteria, and Firmicutes were identified with a vast majority of genes encoding enzymes involved in carbohydrate metabolism. We further identified putative PCWDEs that originated from phasmid's midgut bacteria, including beta-glucosidase that attacks cellulosic oligomers, and polygalacturonase that attacks pectin polymers. Our data suggested a major bacterial source of polygalacturonases present in both phasmids' posterior sections of midgut. This conclusion is a further interpretation of phasmid PCW digestive physiology. The lack of Fibrobacteres and Spirochaetes in *A. asperrimus* suggested a lesser PCW-degrading power than the microbiota from termites. In general, the relationship between midgut bacterial and endogenous PCWDEs is still unconfirmed; however, here this study suggests potential contributions from midgut bacteria to digestion in Phasmids.

SUMMARY

In this dissertation, we applied next generation sequencing technologies to generate transcriptomic and metagenomic data from various alimentary tissues in *A. asperrimus and M. extradentata.* After this, we assembled transcriptomes for both species, and both had good quality of assembly. We successfully identified major endogenous PCWDEs, including cellulases from GH3, GH6, and mainly GH1 and GH9; hemicellulases from GH2, GH30 and GH31; and pectinases from GH28. All genes were confirmed with the presence of the catalytic residues. Compared to other herbivorous insects and Polyneopterans, stick insects have uncommonly high numbers of PCWDE genes. Ancient gene duplication and family expansion of cellulases and pectinases were also reported in previous studies (Shelomi, Danchin et al. 2016, Shelomi, Heckel et al. 2016). Their digestive physiology and mechanisms based on abundant endogenous PCWDEs are still under-studied, however. This is particularly true for tissue- and sub-tissue specific enzymatic expressions of these families. While we demonstrate a full complement of endogenous and microbial PCWDEs in stick insect's guts, a lack of biochemical studies of these PCWDEs limits our ability to draw strong conclusions. Studies of enzymatic activity on different substrates could strengthen our claims and further our understandings of the PCW digestion in stick insects.

Based on the RNA-Seq analysis, we plotted tissue-specific expression profiles of identified PCWDEs and DEGs (AMG versus PMG). GO annotations and enrichment analyses were conducted for the 1,000 most highly expressed genes in each tissue. Major PCWDE families like GH9 endoglucanases, GH1 cellobiases, and GH28 polygalacturonases were highly expressed in the AMG, while fewer genes from these three families showed significant expression in the PMG. As for the hindgut tissues (ileum and rectum), GH9 and GH28 showed close-to-zero TPM values, while β-glucosidases from GH1, GH2, and GH3 showed light expressions in the hindgut. We found GH families like GH2, GH3, and GH30, that represent a variety of hemicellulases, showed wider expression profiles in multiple tissues than the most abundant cellulase and pectinase families. Some glycoside hydrolase families, such as α-glucosidases from

GH31 and GH63, and α-galactosidases from GH27, may not be involved in PCW degradation. Their global expression profiles in all selected tissues suggest their putative whole-body biological functions.

The compartmentalization of phasmid digestion was explored by identifying DEGs between the anterior and posterior midguts, as well as by annotating the 1,000 most highly expressed genes from each alimentary tissue. We observed significant differentiation between anterior and posterior midguts in carbohydrate metabolism. The most up-regulated genes in AMG were mainly various hydrolases including polygalacturonases, β-glucosidases, β-galactosidases, and glucosylceramidases. On the other hand, in the PMG up-regulated of a variety of proteases such as metallopeptidases, aminopeptidases, and metallocarboxypeptidases was found. While the AMG and PMG were highly differentiated in GO annotations for their most highly expressed genes, the ileum and rectum showed similar top GO term distributions. More enzymes that are involved in energy generation and transmembrane transportation were highly expressed in these two hindgut tissues, compared to the midgut. In a nutshell, our data suggested the compartmentalization of digestive expertise in phasmid alimentary canals. Specifically, the anterior and posterior midguts are major localizations for breaking down carbohydrates and proteins respectively. Meanwhile, the hindgut tissues are more focused on the transmembrane transportations of molecules and ATP/GTP synthesis.

Finally, we applied high throughput sequencing to investigate the midgut microbiota of these two stick insect species. We assessed taxonomic and functional compositions of the midgut microbiota. Dominant bacteria phyla including Proteobacteria, Actinobacteria, and Firmicutes were identified with a vast majority of gene encoding enzymes involved in carbohydrate metabolism. We further identified putative PCWDEs that originated from phasmid midgut bacteria, including beta-glucosidase, which attacks cellulosic oligomers, and polygalacturonase, which attacks pectin polymers. Our data suggested a major bacterial source of polygalacturonases present in both phasmids' posterior sections of midgut, which inferred a potential pectin degradation synergism between endogenous and bacterial pectinases in different

sections of phasmid midguts. The lack of Fibrobacteres and Spirochaetes in *A. asperrimus* suggested a lesser PCW-degrading power than that of the microbiota in termites. In general, the relationship between midgut bacteria and endogenous PCWDEs is still unconfirmed. However, here we provide deep sequencing data to question the role that midgut bacteria play in PCW degradation and suggest a potential contribution of midgut bacteria to the digestion of their hosts.

For future research, with more and more sequences of endogenous PCWDEs being reported in phasmids, a direction to study the evolutions of different enzyme families in this lineage is highly attractive. Moreover, further demonstrations of the expression profiles and the enzymatic activities of bacterial PCWDEs in phasmids could be helpful, given that we revealed a variety of gut bacteria that possess these genes. A potential synergism of PCW degradation in stick insects is likely, considering a collaborative mechanism of PCW digestion is performed by endogenous and bacterial enzymes in termites, which is a close order to phasmids. Therefore, future studies that focus on the synergism of PCW digestion in phasmids can be significant.

REFERENCES

Aanen, D. K., P. Eggleton, C. Rouland-Lefevre, T. Guldberg-Froslev, S. Rosendahl and J. J. Boomsma (2002). "The evolution of fungus-growing termites and their mutualistic fungal symbionts." Proceedings of the National Academy of Sciences of the United States of America **99**(23): 14887-14892.

Adams, M. D., J. M. Kelley, J. D. Gocayne, M. Dubnick, M. H. Polymeropoulos, H. Xiao, C. R. Merril, A. Wu, B. Olde, R. F. Moreno, A. R. Kerlavage, W. R. Mccombie and J. C. Venter (1991). "Complementary-DNA Sequencing - Expressed Sequence Tags and Human Genome Project." Science **252**(5013): 1651-1656.

Alam, M. S., S. Nakashima, Y. Deyashiki, Y. Banno, A. Hara and Y. Nozawa (1996). "Molecular cloning of a gene encoding acid alpha-glucosidase from Tetrahymena pyriformis." Journal of Eukaryotic Microbiology **43**(4): 295-303.

Albersheim, P., A. Darvill, K. Roberts, R. Sederoff and A. Staehelin (2011). "Plant Cell Walls: From Chemistry to Biology." Plant Cell Walls: From Chemistry to Biology: 1-430.

Alexa, A. and J. Rahnenfuhrer (2010). "topGO: enrichment analysis for gene ontology." R package version **2**(0): 2010.

Amiri, A., A. R. Bandani and M. Kafil (2020). "Gut compartments and ovary bacterial symbionts of the Sunn pest." Journal of Asia-Pacific Entomology **23**(3): 723-730.

Anand, A. A. P., S. J. Vennison, S. G. Sankar, D. I. G. Prabhu, P. T. Vasan, T. Raghuraman, C. J. Geoffrey and S. E. Vendan (2010). "Isolation and characterization of bacteria from the gut of Bombyx mori that degrade cellulose, xylan, pectin and starch and their impact on digestion." Journal of Insect Science **10**.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data, Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom.

Antony, B., J. Johny, S. A. Aldosari and M. M. Abdelazim (2017). "Identification and expression profiling of novel plant cell wall degrading enzymes from a destructive pest of palm trees, Rhynchophorus ferrugineus." Insect Molecular Biology **26**(4): 469-484.

Araujo, F. A., D. Barh, A. Silva, L. Guimaraes and R. T. J. Ramos (2018). "GO FEAT: a rapid web-based functional annotation tool for genomic and transcriptomic data." Scientific Reports **8**.

Armenteros, J. J. A., K. D. Tsirigos, C. K. Sonderby, T. N. Petersen, O. Winther, S. Brunak, G. von Heijne and H. Nielsen (2019). "SignalP 5.0 improves signal peptide predictions using deep neural networks." Nature Biotechnology **37**(4): 420-+.

Aylward, F. O., K. E. Burnum, J. J. Scott, G. Suen, S. G. Tringe, S. M. Adams, K. W. Barry, C. D. Nicora, P. D. Piehowski, S. O. Purvine, G. J. Starrett, L. A. Goodwin, R. D. Smith, M. S. Lipton and C. R. Currie (2012). "Metagenomic and metaproteomic insights into bacterial communities in leaf-cutter ant fungus gardens." Isme Journal **6**(9): 1688-1701.

Azadi, P., M. A. ONeill, C. Bergmann, A. G. Darvill and P. Albersheim (1995). "The backbone of the pectic polysaccharide rhamnogalacturonan I is cleaved by an endohydrolase and an endolyase." Glycobiology **5**(8): 783-789.

Azevedo, D. O., M. D. Q. Fialho, N. C. Vargas, E. F. Vilela, J. C. Zanuncio and J. E. Serrao (2013). "Morphology of the Digestive Tract of Cladomorphus Phyllinus (Phasmatodea: Phasmidae)." Florida Entomologist **96**(4): 1417-1423.

Barcoto, M. O., C. Carlos-Shanley, H. Fan, M. Ferro, N. S. Nagamoto, M. Bacci, C. R. Currie and A. Rodrigues (2020). "Fungus-growing insects host a distinctive microbiota apparently adapted to the fungiculture environment." Scientific Reports **10**(1).

Barrett, T., C. G. Suresh, S. P. Tolley, E. J. Dodson and M. A. Hughes (1995). "The crystal structure of a cyanogenic beta-glucosidase from white clover, a family 1 glycosyl hydrolase." Structure **3**(9): 951-960.

Bedford, G. O. (1978). "Biology and Ecology of Phasmatidea." Annual Review of Entomology **23**: 125-149.

Berkov, A., J. Feinstein, J. Small and M. Nkamany (2007). "Yeasts isolated from neotropical wood-boring beetles in SE Peru." Biotropica **39**(4): 530-538.

Berlanga, M., C. Llorens, J. Comas and R. Guerrero (2016). "Gut Bacterial Community of the Xylophagous Cockroaches Cryptocercus punctulatus and Parasphaeria boleiriana." Plos One **11**(4).

Bernays, E. A. (1998). "Evolution of feeding behavior in insect herbivores - Success seen as different ways to eat without being eaten." Bioscience **48**(1): 35-44.

Boyd, D. W., A. C. Cohen and D. R. Alverson (2002). "Digestive enzymes and stylet morphology of Deraeocoris nebulosus (Hemiptera : Miridae), a predacious plant bug." Annals of the Entomological Society of America **95**(3): 395-401.

Bradler, S. and T. R. Buckley (2018). "Biodiversity of Phasmatodea." Insect Biodiversity: Science and Society, Vol Ii: 281-+.

Bradler, S., N. Cliquennois and T. R. Buckley (2015). "Single origin of the Mascarene stick insects: ancient radiation on sunken islands?" Bmc Evolutionary Biology **15**.

Bradler, S., J. A. Robertson and M. F. Whiting (2014). "A molecular phylogeny of Phasmatodea with emphasis on Necrosciinae, the most species-rich subfamily of stick insects." Systematic Entomology **39**(2): 205-222.

Brand, P., W. Lin and B. R. Johnson (2018). "The Draft Genome of the Invasive Walking Stick, Medauroidea extradendata, Reveals Extensive Lineage-Specific Gene Family Expansions of Cell Wall Degrading Enzymes in Phasmatodea." G3-Genes Genomes Genetics **8**(5): 1403-1408.

Bray, N. L., H. Pimentel, P. Melsted and L. Pachter (2016). "Near-optimal probabilistic RNA-seq quantification." Nature Biotechnology **34**(5): 525-527.

Breznak, J. A. and A. Brune (1994). "Role of Microorganisms in the Digestion of Lignocellulose by Termites." Annual Review of Entomology **39**: 453-487.

Brune, A. (2014). "Symbiotic digestion of lignocellulose in termite guts." Nature Reviews Microbiology **12**(3): 168-180.

Buckley, T. R., D. Attanayake and S. Bradler (2009). "Extreme convergence in stick insect evolution: phylogenetic placement of the Lord Howe Island tree lobster." Proceedings of the Royal Society B-Biological Sciences **276**(1659): 1055-1062.

Buckley, T. R., D. Attanayake, J. A. A. Nylander and S. Bradler (2010). "The phylogenetic placement and biogeographical origins of the New Zealand stick insects (Phasmatodea)." Systematic Entomology **35**(2): 207-225.

Busch, A., E. G. J. Danchin and Y. Pauchet (2019). "Functional diversification of horizontally acquired glycoside hydrolase family 45 (GH45) proteins in Phytophaga beetles." Bmc Evolutionary Biology **19**.

Buscher, T. H., T. R. Buckley, C. Grohmann, S. N. Gorb and S. Bradler (2018). "The Evolution of Tarsal Adhesive Microstructures in Stick and Leaf Insects (Phasmatodea)." Frontiers in Ecology and Evolution **6**.

Calderon-Cortes, N., M. Quesada, H. Watanabe, H. Cano-Camacho and K. Oyama (2012). "Endogenous Plant Cell Wall Digestion: A Key Mechanism in Insect Evolution." Annual Review of Ecology, Evolution, and Systematics, Vol 43 **43**: 45-71.

Camacho, C., G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer and T. L. Madden (2009). "BLAST plus : architecture and applications." Bmc Bioinformatics **10**.

Cameron, A. E. (1912). Structure of the Alimentary Canal of the Stick-Insect, Bacillus rossii Fabr.; with a Note on the Parthenogenesis of this Species. Proceedings of the Zoological Society of London, Wiley Online Library.

Carroll, A. and C. Somerville (2009). "Cellulosic Biofuels." Annual Review of Plant Biology **60**: 165-182.

Cazemier, A. E., H. J. M. OpdenCamp, J. H. P. Hackstein and G. D. Vogels (1997). "Fibre digestion in arthropods." Comparative Biochemistry and Physiology a-Molecular & Integrative Physiology **118**(1): 101-109.

Chang, H. X., C. R. Yendrek, G. Caetano-Anolles and G. L. Hartman (2016). "Genomic characterization of plant cell wall degrading enzymes and in silico analysis of xylanses and polygalacturonases of Fusarium virguliforme." Bmc Microbiology **16**.

Chapman, R. F. (1998). The Insects: Structure and Function, Cambridge university press.

Cheeseman, M. T. and G. Pritchard (1984). "Proventricular Trituration in Adult Carabid Beetles (Coleoptera, Carabidae)." Journal of Insect Physiology **30**(3): 203-209.

Cherqui, A. and W. F. Tjallingii (2000). "Salivary proteins of aphids, a pilot study on identification, separation and immunolocalisation." Journal of Insect Physiology **46**(8): 1177-1186.

Clark, J. (1976). "The dissection of a stick insect." Journal of Biological Education **10**(1): 31-34.

Conesa, A., S. Gotz, J. M. Garcia-Gomez, J. Terol, M. Talon and M. Robles (2005). "Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research." Bioinformatics **21**(18): 3674-3676.

Cook, D. M., E. D. Henriksen, R. Upchurch and J. B. D. Peterson (2007). "Isolation of polymer-degrading bacteria and characterization of the hindgut bacterial community from the detritus-feeding larvae of Tipula abdominalis (Diptera : Tipulidae)." Applied and Environmental Microbiology **73**(17): 5683-5686.

Cosgrove, D. J. (2005). "Growth of the plant cell wall." Nature Reviews Molecular Cell Biology **6**(11): 850-861.

Cox, M. P., D. A. Peterson and P. J. Biggs (2010). "SolexaQA: At-a-glance quality assessment of Illumina second-generation sequencing data." Bmc Bioinformatics **11**.

Cragg, S. M., G. T. Beckham, N. C. Bruce, T. D. H. Bugg, D. L. Distel, P. Dupree, A. G. Etxabe, B. S. Goodell, J. Jellison, J. E. McGeehan, S. J. McQueen-Mason, K. Schnorr, P. H. Walton, J. E. M. Watts and M. Zimmer (2015). "Lignocellulose degradation mechanisms across the Tree of Life." Current Opinion in Chemical Biology **29**: 108-119.

Da Lage, J. L. (2018). "The Amylases of Insects." Int J Insect Sci **10**: 1179543318804783.

Davison, A. and M. Blaxter (2005). "Ancient origin of glycosyl hydrolase family 9 cellulase genes." Molecular Biology and Evolution **22**(5): 1273-1284.

Delmont, T. O., E. Prestat, K. P. Keegan, M. Faubladier, P. Robe, I. M. Clark, E. Pelletier, P. R. Hirsch, F. Meyer, J. A. Gilbert, D. Le Paslier, P. Simonet and T. M. Vogel (2012). "Structure, fluctuation and magnitude of a natural grassland soil metagenome." Isme Journal **6**(9): 1677-1687.

Donovan, S. E., P. Eggleton and D. E. Bignell (2001). "Gut content analysis and a new feeding group classification of termites." Ecological Entomology **26**(4): 356-366.

Douglas, A. E. (2009). "The microbial dimension in insect nutritional ecology." Functional Ecology **23**(1): 38-47.

Eddy, S. R. (2011). "Accelerated Profile HMM Searches." Plos Computational Biology **7**(10).

Edgar, R. C. (2004). "MUSCLE: multiple sequence alignment with high accuracy and high throughput." Nucleic Acids Research **32**(5): 1792-1797.

Egert, M., U. Stingl, L. D. Bruun, B. Pommerenke, A. Brune and M. W. Friedrich (2005). "Structure and topology of microbial communities in the major gut compartments of Melolontha melolontha larvae (Coleoptera : Scarabaeidae)." Applied and Environmental Microbiology **71**(8): 4556-4566.

Engel, M. S., B. Wang and A. S. Alqarni (2016). "A thorny, 'anareolate' stick-insect (Phasmatidae s.l.) in Upper Cretaceous amber from Myanmar, with remarks on diversification times among Phasmatodea." Cretaceous Research **63**: 45-53.

Engel, P. and N. A. Moran (2013). "The gut microbiota of insects - diversity in structure and function." Fems Microbiology Reviews **37**(5): 699-735.

Fischer, R., R. Ostafe and R. M. Twyman (2013). "Cellulases from Insects." Yellow Biotechnology Ii: Insect Biotechnology in Plant Protection and Industry **136**: 51-64.

Friedemann, K., B. Wipfler, S. Bradler and R. G. Beutel (2012). "On the head morphology of Phyllium and the phylogenetic relationships of Phasmatodea (Insecta)." Acta Zoologica **93**(2): 184-199.

Fujita, A., M. Hojo, T. Aoyagi, Y. Hayashi, G. Arakawa, G. Tokuda and H. Watanabe (2010). "Details of the digestive system in the midgut of Coptotermes formosanus Shiraki." Journal of Wood Science **56**(3): 222-226.

Furnham, N., G. L. Holliday, T. A. P. de Beer, J. O. B. Jacobsen, W. R. Pearson and J. M. Thornton (2014). "The Catalytic Site Atlas 2.0: cataloging catalytic sites and residues identified in enzymes." Nucleic Acids Research **42**(D1): D485-D489.

Gabrisko, M. (2013). "Evolutionary History of Eukaryotic alpha-Glucosidases from the alpha-Amylase Family." Journal of Molecular Evolution **76**(3): 129-145.

Gao, P., Z. K. Liu and J. B. Wen (2020). "Expression Profiling of Plant Cell Wall-Degrading Enzyme Genes inEucryptorrhynchus scrobiculatusMidgut." Frontiers in Physiology **11**.

Geib, S. M., T. R. Filley, P. G. Hatcher, K. Hoover, J. E. Carlson, M. D. Jimenez-Gasco, A. Nakagawa-Izumi, R. L. Sleighter and M. Tien (2008). "Lignin degradation in wood-feeding insects." Proceedings of the National Academy of Sciences of the United States of America **105**(35): 12932-12937.

Genta, F. A., I. Bragatto, W. R. Terra and C. Ferreira (2009). "Purification, characterization and sequencing of the major beta-1,3-glucanase from the midgut of Tenebrio molitor larvae." Insect Biochemistry and Molecular Biology **39**(12): 861-874.

Genta, F. A., W. R. Terra and C. Ferreira (2003). "Action pattern, specificity, lytic activities, and physiological role of five digestive beta-glucanases isolated from Periplaneta americana." Insect Biochemistry and Molecular Biology **33**(11): 1085-1097.

Gomez-Alvarez, V., T. K. Teal and T. M. Schmidt (2009). "Systematic artifacts in metagenomes from complex microbial communities." Isme Journal **3**(11): 1314-1317.

Gottardo, M., D. Mercati and R. Dallai (2012). "The spermatogenesis and sperm structure of Timema poppensis (Insecta: Phasmatodea)." Zoomorphology **131**(3): 209-223.

Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. D. Zeng, Z. H. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman and A. Regev (2011). "Full-length transcriptome assembly from RNA-Seq data without a reference genome." Nature Biotechnology **29**(7): 644-U130.

Gunning, G. E. (1987). "Behavioral Observations of the Walking Stick, Anisomorpha-Buprestoides (Phasmatodea, Phasmatidae)." Florida Entomologist **70**(3): 406-408.

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

Handelsman, J. (2004). "Metagenomics: Application of genomics to uncultured microorganisms." Microbiology and Molecular Biology Reviews **68**(4): 669-+.

Hannula, S. E., F. Zhu, R. Heinen and T. M. Bezemer (2019). "Foliar-feeding insects acquire microbiomes from the soil rather than the host plant." Nature Communications **10**.

Hatefi, A., A. Makhdoumi, A. Asoodeh and O. Mirshamsi (2017). "Characterization of a bi-functional cellulase produced by a gut bacterial resident of Rosaceae branch borer beetle, Osphranteria coerulescens (Coleoptera: Cerambycidae)." International Journal of Biological Macromolecules **103**: 158-164.

He, S. M., N. Ivanova, E. Kirton, M. Allgaier, C. Bergin, R. H. Scheffrahn, N. C. Kyrpides, F. Warnecke, S. G. Tringe and P. Hugenholtz (2013). "Comparative Metagenomic and Metatranscriptomic Analysis of Hindgut Paunch Microbiota in Wood- and Dung-Feeding Higher Termites." Plos One **8**(4).

Herscovics, A. (1999). "Importance of glycosidases in mammalian glycoprotein biosynthesis." Biochimica Et Biophysica Acta-General Subjects **1473**(1): 96-107.

Hochuli, D. F. (1996). "The ecology of plant/insect interactions: Implications of digestive strategy for feeding by phytophagous insects." Oikos **75**(1): 133-141.

Hongoh, Y. (2011). "Toward the functional analysis of uncultivable, symbiotic microorganisms in the termite gut." Cellular and Molecular Life Sciences **68**(8): 1311-1325.

Hrdlickova, R., M. Toloue and B. Tian (2017). "RNA-Seq methods for transcriptome analysis." Wiley Interdisciplinary Reviews-Rna **8**(1).

Hu, H. F., R. R. da Costa, B. Pilgaard, M. Schiott, L. Lange and M. Poulsen (2019). "Fungiculture in Termites Is Associated with a Mycolytic Gut Bacterial Community." Msphere **4**(3).

Huang, S. W., H. Y. Zhang, S. Marshall and T. A. Jackson (2010). "The scarab gut: A potential bioreactor for bio-fuel production." Insect Science **17**(3): 175-183.

Joynson, R., L. Pritchard, E. Osemwekha and N. Ferry (2017). "Metagenomic Analysis of the Gut Microbiome of the Common Black Slug Arion ater in Search of Novel Lignocellulose Degrading Enzymes." Frontiers in Microbiology **8**.

Keegan, K. P., E. M. Glass and F. Meyer (2016). "MG-RAST, a Metagenomics Service for Analysis of Microbial Community Structure and Function." Microbial Environmental Genomics (Meg) **1399**: 207-233.

Keegan, K. P., W. L. Trimble, J. Wilkening, A. Wilke, T. Harrison, M. D'Souza and F. Meyer (2012). "A Platform-Independent Method for Detecting Errors in Metagenomic Sequencing Data: DRISEE." Plos Computational Biology **8**(6).

Kelley, D. R., B. Liu, A. L. Delcher, M. Pop and S. L. Salzberg (2012). "Gene prediction with Glimmer for metagenomic sequences augmented by classification and clustering." Nucleic Acids Research **40**(1).

Kim, M., H. Yoon, Y. H. You, Y. E. Kim, J. R. Woo, Y. Seo, G. M. Lee, Y. J. Kim, W. S. Kong and J. G. Kim (2013). "Metagenomic Analysis of Fungal Communities Inhabiting the Fairy Ring Zone of Tricholoma matsutake." Journal of Microbiology and Biotechnology **23**(10): 1347-1356.

Kirsch, R., L. Gramzow, G. Theissen, B. D. Siegfried, R. H. Ffrench-Constant, D. G. Heckel and Y. Pauchet (2014). "Horizontal gene transfer and functional diversification of plant cell wall degrading polygalacturonases: Key events in the evolution of herbivory in beetles." Insect Biochemistry and Molecular Biology **52**: 33-50.

Kirsch, R., N. Wielsch, H. Vogel, A. Svatos, D. G. Heckel and Y. Pauchet (2012). "Combining proteomics and transcriptome sequencing to identify active plant-cell-wall-degrading enzymes in a leaf beetle." BMC Genomics **13**: 587.

Kjer, K. M. (2004). "Aligned 18S and insect phylogeny." Systematic Biology **53**(3): 506-514.

Klass, K. D., O. Zompro, N. P. Kristensen and J. Adis (2002). "Mantophasmatodea: A new insect order with extant members in the afrotropics." Science **296**(5572): 1456-1459.

Korb, J., T. Weil, K. Hoffmann, K. R. Foster and M. Rehli (2009). "A Gene Necessary for Reproductive Suppression in Termites." Science **324**(5928): 758-758.

Kristensen, N. P. (1975). "The phylogeny of hexapod "orders". A critical review of recent accounts." Journal of Zoological Systematics and Evolutionary Research **13**(1): 1-44.

Kriventseva, E. V., D. Kuznetsov, F. Tegenfeldt, M. Manni, R. Dias, F. A. Simao and E. M. Zdobnov (2019). "OrthoDB v10: sampling the diversity of animal, plant, fungal, protist, bacterial and viral genomes for evolutionary and functional annotations of orthologs." Nucleic Acids Research **47**(D1): D807-D811.

Kubota, M., M. Tsuji, M. Nishimoto, J. Wongchawalit, M. Okuyama, H. Mori, H. Matsui, R. Surarit, J. Svasti, A. Kimura and S. Chiba (2004). "Localization of alpha-glucosidases I, II, and III in organs of European honeybees, Apis mellifera L., and the origin of alpha-glucosidase in honey." Biosci Biotechnol Biochem **68**(11): 2346-2352.

Letsch, H. and S. Simon (2013). "Insect phylogenomics: new insights on the relationships of lower neopteran orders (Polyneoptera)." Systematic Entomology **38**(4): 783-793.

Li, H., S. E. Young, M. Poulsen and C. R. Currie (2020). "Symbiont-Mediated Digestion of Plant Biomass in Fungus-Farming Insects." Annu Rev Entomol.

Li, L. L., S. R. McCorkle, S. Monchy, S. Taghavi and D. van der Lelie (2009). "Bioprospecting metagenomes: glycosyl hydrolases for converting biomass." Biotechnology for Biofuels **2**.

Liao, Q. L., Y. J. Guo, J. S. Zhou, Y. Wan, R. Carballar-Lejarazu, L. J. Sheng, F. P. Zhang, S. Q. Wu and S. Q. Zou (2020). "Characterization of Bacterial Communities Associated withRhynchophorus

ferrugineusOlivier (Coleoptera: Curculionidae) and its HostPhoenix sylvestris." Current Microbiology **77**(11): 3321-3329.

Licht, H. H. D. and P. H. W. Biedermann (2012). "Patterns of functional enzyme activity in fungus farming ambrosia beetles." Frontiers in Zoology **9**.

Liou, Y. F., P. Charoenkwan, Y. S. Srinivasulu, T. Vasylenko, S. C. Lai, H. C. Lee, Y. H. Chen, H. L. Huang and S. Y. Ho (2014). "SCMHBP: prediction and analysis of heme binding proteins using propensity scores of dipeptides." Bmc Bioinformatics **15**.

Liu, B., T. Gibbons, M. Ghodsi, T. Treangen and M. Pop (2011). "Accurate and fast estimation of taxonomic profiles from metagenomic shotgun sequences." Bmc Genomics **12**.

Liu, J., K. Q. Song, H. J. Teng, B. Zhang, W. Z. Li, H. J. Xue and X. K. Yang (2015). "Endogenous cellulolytic enzyme systems in the longhorn beetle Mesosa myops (Insecta: Coleoptera) studied by transcriptomic analysis." Acta Biochimica Et Biophysica Sinica **47**(9): 741-748.

Liu, N., H. J. Li, M. G. Chevrette, L. Zhang, L. Cao, H. K. Zhou, X. G. Zhou, Z. H. Zhou, P. B. Pope, C. R. Currie, Y. P. Huang and Q. Wang (2019). "Functional metagenomics reveals abundant polysaccharide-degrading gene clusters and cellobiose utilization pathways within gut microbiota of a wood-feeding higher termite." Isme Journal **13**(1): 104-117.

Liu, N., L. Zhang, H. K. Zhou, M. L. Zhang, X. Yan, Q. Wang, Y. H. Long, L. Xie, S. Y. Wang, Y. P. Huang and Z. H. Zhou (2013). "Metagenomic Insights into Metabolic Capacities of the Gut Microbiota in a Fungus-Cultivating Termite (Odontotermes yunnanensis)." Plos One **8**(7).

Lo, N., H. Watanabe and M. Sugimura (2003). "Evidence for the presence of a cellulase gene in the last common ancestor of bilaterian animals." Proceedings of the Royal Society B-Biological Sciences **270**: S69-S72.

Loman, N. J., R. V. Misra, T. J. Dallman, C. Constantinidou, S. E. Gharbia, J. Wain and M. J. Pallen (2012). "Performance comparison of benchtop high-throughput sequencing platforms." Nature Biotechnology **30**(5): 434-+.

Lundgren, J. G. and M. Lehman (2010). "Bacterial Gut Symbionts Contribute to Seed Digestion in an Omnivorous Beetle." Plos One **5**(5).

Malgas, S., S. J. van Dyk and B. I. Pletschke (2015). "beta-Mannanase (Man26A) and alpha-galactosidase (Aga27A) synergism - A key factor for the hydrolysis of galactomannan substrates." Enzyme and Microbial Technology **70**: 1-8.

Martin, M. M. (1983). "Cellulose Digestion in Insects." Comparative Biochemistry and Physiology a-Physiology **75**(3): 313-324.

Martin, M. M. (1991). "The Evolution of Cellulose Digestion in Insects." Philosophical Transactions of the Royal Society B-Biological Sciences **333**(1267): 281-288.

Matos, I. T. S. R., E. N. Assuncao, E. do Carmo, V. M. Soares and S. Astolfi (2017). "Isolation and Characterization of Yeasts Able to Assimilate Sugarcane Bagasse Hemicellulosic Hydrolysate and Produce Xylitol Associated with Veturius transversus (Passalidae, Coleoptera, and Insecta)." International Journal of Microbiology **2017**.

Matsuura, K., T. Yashiro, K. Shimizu, S. Tatsumi and T. Tamura (2009). "Cuckoo Fungus Mimics Termite Eggs by Producing the Cellulose-Digesting Enzyme beta-Glucosidase." Current Biology **19**(1): 30-36.

McDonough, M. A., R. Kadirvelraj, P. Harris, J. C. N. Poulsen and S. Larsen (2004). "Rhamnogalacturonan lyase reveals a unique three-domain modular structure for polysaccharide lyase family 4." Febs Letters **565**(1-3): 188-194.

Miao, Y. T., H. K. Jia, Z. Li, Y. D. Liu and M. L. Hou (2018). "Transcriptomic and Expression Analysis of the Salivary Glands in Brown Planthoppers, Nilaparvata lugens (Hemiptera: Delphacidae)." Journal of Economic Entomology **111**(6): 2884-2893.

Mikaelyan, A., T. Kohler, N. Lampert, J. Rohland, H. Boga, K. Meuser and A. Brune (2015). "Classifying the bacterial gut microbiota of termites and cockroaches: A curated phylogenetic reference database (DictDb)." Systematic and Applied Microbiology **38**(7): 472-482.

Misof, B., S. L. Liu, K. Meusemann, R. S. Peters, A. Donath, C. Mayer, P. B. Frandsen, J. Ware, T. Flouri, R. G. Beutel, O. Niehuis, M. Petersen, F. Izquierdo-Carrasco, T. Wappler, J. Rust, A. J. Aberer, U. Aspock, H. Aspock, D. Bartel, A. Blanke, S. Berger, A. Bohm, T. R. Buckley, B. Calcott, J. Q. Chen, F. Friedrich, M. Fukui, M. Fujita, C. Greve, P. Grobe, S. C. Gu, Y. Huang, L. S. Jermiin, A. Y. Kawahara, L. Krogmann, M. Kubiak, R. Lanfear, H. Letsch, Y. Y. Li, Z. Y. Li, J. G. Li, H. R. Lu, R. Machida, Y. Mashimo, P. Kapli, D. D. McKenna, G. L. Meng, Y. Nakagaki, J. L. Navarrete-Heredia, M. Ott, Y. X. Ou, G. Pass, L. Podsiadlowski, H. Pohl, B. M. von Reumont, K. Schutte, K. Sekiya, S. Shimizu, A. Slipinski, A. Stamatakis, W. H. Song, X. Su, N. U. Szucsich, M. H. Tan, X. M. Tan, M. Tang, J. B. Tang, G. Timelthaler, S. Tomizuka, M. Trautwein, X. L. Tong, T. Uchifune, M. G. Walzl, B. M. Wiegmann, J. Wilbrandt, B. Wipfler, T. K. F. Wong, Q. Wu, G. X. Wu, Y. L. Xie, S. Z. Yang, Q. Yang, D. K. Yeates, K. Yoshizawa, Q. Zhang, R. Zhang, W. W. Zhang, Y. H. Zhang, J. Zhao, C. R. Zhou, L. L. Zhou, T. Ziesmann, S. J. Zou, Y. R. Li, X. Xu, Y. Zhang, H. M. Yang, J. Wang, J. Wang, K. M. Kjer and X. Zhou (2014). "Phylogenomics resolves the timing and pattern of insect evolution." Science **346**(6210): 763-767.

Moller, I. E., H. H. D. F. Licht, J. Harholt, W. G. T. Willats and J. J. Boomsma (2011). "The Dynamics of Plant Cell-Wall Polysaccharide Decomposition in Leaf-Cutting Ant Fungus Gardens." Plos One **6**(3).

Monteiro, E. C., F. K. Tamaki, W. R. Terra and A. F. Ribeiro (2014). "The digestive system of the "stick bug" Cladomorphus phyllinus (Phasmida, Phasmatidae): A morphological, physiological and biochemical analysis." Arthropod Structure & Development **43**(2): 123-134.

Morandin, C., U. Pulliainen, N. Bos and E. Schultner (2018). "De novo transcriptome assembly and its annotation for the black ant Formica fusca at the larval stage." Scientific Data **5**.

Mueller, U. G. and N. Gerardo (2002). "Fungus-farming insects: Multiple origins and diverse evolutionary histories." Proceedings of the National Academy of Sciences of the United States of America **99**(24): 15247-15249.

Mury, F. B., J. R. da Silva, L. S. Ferreira, B. D. Ferreira, G. A. de Souza, J. A. de Souza-Neto, P. E. M. Ribolla, C. P. Silva, V. V. do Nascimento, O. L. T. Machado, M. A. Berbert-Molina and M. Dansa-Petretski (2009). "Alpha-Glucosidase Promotes Hemozoin Formation in a Blood-Sucking Bug: An Evolutionary History." Plos One **4**(9).

Nakashima, K., H. Watanabe, H. Saitoh, G. Tokuda and J. I. Azuma (2002). "Dual cellulose-digesting system of the wood-feeding termite, Coptotermes formosanus Shiraki." Insect Biochemistry and Molecular Biology **32**(7): 777-784.

Nation, J. L. (1983). "Specialization in the Alimentary Canal of Some Mole Crickets (Orthoptera, Gryllotalpidae)." International Journal of Insect Morphology & Embryology **12**(4): 201-210.

Naumoff, D. G. (2004). "Phylogenetic analysis of alpha-galactosidases of the GH27 family." Molecular Biology **38**(3): 388-399.

Ni, J. F. and G. Tokuda (2013). "Lignocellulose-degrading enzymes from termites and their symbiotic microbiota." Biotechnology Advances **31**(6): 838-850.

Nygaard, S., H. Hu, C. Li, M. Schiott, Z. Chen, Z. Yang, Q. Xie, C. Ma, Y. Deng, R. B. Dikow, C. Rabeling, D. R. Nash, W. T. Wcislo, S. G. Brady, T. R. Schultz, G. Zhang and J. J. Boomsma (2016). "Reciprocal genomic evolution in the ant-fungus agricultural symbiosis." Nat Commun **7**: 12233.

Oh, H. W., S. Y. Heo, D. Y. Kim, D. S. Park, K. S. Bae and H. Y. Park (2008). "Biochemical characterization and sequence analysis of a xylanase produced by an exo-symbiotic bacterium of Gryllotalpa orientalis, Cellulosimicrobium sp HY-12." Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology **93**(4): 437-442.

Okonechnikov, K., O. Golosova, M. Fursov and U. Team (2012). "Unipro UGENE: a unified bioinformatics toolkit." Bioinformatics **28**(8): 1166-1167.

Oppert, C., W. E. Klingeman, J. D. Willis, B. Oppert and J. L. Jurat-Fuentes (2010). "Prospecting for cellulolytic activity in insect digestive fluids." Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology **155**(2): 145-154.

Ordaz-Ortiz, J. J. and L. Saulnier (2005). "Structural variability of arabinoxylans from wheat flour. Comparison of water-extractable and xylanase-extractable arabinoxylans." Journal of Cereal Science **42**(1): 119-125.

Overbeek, R., R. Olson, G. D. Pusch, G. J. Olsen, J. J. Davis, T. Disz, R. A. Edwards, S. Gerdes, B. Parrello, M. Shukla, V. Vonstein, A. R. Wattam, F. F. Xia and R. Stevens (2014). "The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST)." Nucleic Acids Research **42**(D1): D206-D214.

Pauchet, Y., D. Freitak, H. M. Heidel-Fischer, D. G. Heckel and H. Vogel (2009). "Glucanase Activity in a Glucan-Binding Protein Family from Lepidoptera." Journal of Biological Chemistry **284**(4): 2214-2224.

Pauchet, Y., R. Kirsch, S. Giraud, H. Vogel and D. G. Heckel (2014). "Identification and characterization of plant cell wall degrading enzymes from three glycoside hydrolase families in the cerambycid beetle Apriona japonica." Insect Biochemistry and Molecular Biology **49**: 1-13.

Pauchet, Y., P. Wilkinson, R. Chauhan and R. H. Ffrench-Constant (2010). "Diversity of Beetle Genes Encoding Novel Plant Cell Wall Degrading Enzymes." Plos One **5**(12).

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

Peterson, B. F. and M. E. Scharf (2018). "Metatranscriptomic Techniques for Identifying Cellulases in Termites and their Symbionts." Cellulases: Methods and Protocols **1796**: 85-101.

Pimentel, H., N. L. Bray, S. Puente, P. Melsted and L. Pachter (2017). "Differential analysis of RNA-seq incorporating quantification uncertainty." Nature Methods **14**(7): 687-+.

Pothula, R., D. Shirley, O. P. Perera, W. E. Klingeman, C. Oppert, H. M. Y. Abdelgaffar, B. R. Johnson and J. L. Jurat-Fuentes (2019). "The digestive system in Zygentoma as an insect model for high cellulase activity." Plos One **14**(2).

Poulsen, M., H. F. Hu, C. Li, Z. S. Chen, L. H. Xu, S. Otani, S. Nygaard, T. Nobre, S. Klaubauf, P. M. Schindler, F. Hauser, H. L. Pan, Z. K. Yang, A. S. M. Sonnenberg, Z. W. de Beer, Y. Zhang, M. J. Wingfield, C. J. P. Grimmelikhuijzen, R. P. de Vries, J. Korb, D. K. Aanen, J. Wang, J. J. Boomsma and G. J. Zhang (2014). "Complementary symbiont contributions to plant decomposition in a fungus-farming termite." Proceedings of the National Academy of Sciences of the United States of America **111**(40): 14500-14505.

Rahle, W. (1970). Untersuchungen an Kopf und Prothorax von Embia ramburi Rimsky-Korsakow 1906 (Embioptera, Embiidae).

Riesenfeld, C. S., P. D. Schloss and J. Handelsman (2004). "Metagenomics: Genomic Analysis of Microbial Communities." Annual Review of Genetics **38**(1): 525-552.

Robertson, J. A., S. Bradler and M. F. Whiting (2018). "Evolution of Oviposition Techniques in Stick and Leaf Insects (Phasmatodea)." Frontiers in Ecology and Evolution **6**.

Sakon, J., D. Irwin, D. B. Wilson and P. A. Karplus (1997). "Structure and mechanism of endo/exocellulase E4 from Thermomonospora fusca." Nature Structural Biology **4**(10): 810-818.

Scharf, M. E., Z. J. Karl, A. Sethi and D. G. Boucias (2011). "Multiple Levels of Synergistic Collaboration in Termite Lignocellulose Digestion." Plos One **6**(7).

Scrivener, A. M., H. Watanabe and H. Noda (1997). "Diet and carbohydrate digestion in the yellow-spotted longicorn beetle Psacothea hilaris." Journal of Insect Physiology **43**(11): 1039-1052.

Seppey, M., M. Manni and E. M. Zdobnov (2019). "BUSCO: Assessing Genome Assembly and Annotation Completeness." Gene Prediction: Methods and Protocols **1962**: 227-245.

Shelomi, M., E. G. J. Danchin, D. Heckel, B. Wipfler, S. Bradler, X. Zhou and Y. Pauchet (2016). "Horizontal Gene Transfer of Pectinases from Bacteria Preceded the Diversification of Stick and Leaf Insects." Scientific Reports **6**.

Shelomi, M., D. G. Heckel and Y. Pauchet (2016). "Ancestral gene duplication enabled the evolution of multifunctional cellulases in stick insects (Phasmatodea)." Insect Biochemistry and Molecular Biology **71**: 1-11.

Shelomi, M., W. C. Jasper, J. Atallah, L. S. Kimsey and B. R. Johnson (2014). "Differential expression of endogenous plant cell wall degrading enzyme genes in the stick insect (Phasmatodea) midgut." Bmc Genomics **15**.

Shelomi, M. and L. S. Kimsey (2014). "Vital staining of the stick insect digestive system identifies appendices of the midgut as novel system of excretion." J Morphol **275**(6): 623-633.

Shelomi, M., W. S. Lo, L. S. Kimsey and C. H. Kuo (2013). "Analysis of the gut microbiota of walking sticks (Phasmatodea)." BMC Res Notes **6**: 368.

Shelomi, M., H. Watanabe and G. Arakawa (2014). "Endogenous cellulase enzymes in the stick insect (Phasmatodea) gut." Journal of Insect Physiology **60**: 25-30.

Shelomi, M., B. Wipfler, X. Zhou and Y. Pauchet (2020). "Multifunctional cellulase enzymes are ancestral in Polyneoptera." Insect Molecular Biology **29**(1): 124-135.

Sikora, J., J. Urinovska, F. Majer, H. Poupetova, J. Hlavata, M. Kostrouchova, J. Ledvinova and M. Hrebicek (2010). "Bioinformatic and biochemical studies point to AAGR-1 as the ortholog of human acid alpha-glucosidase in Caenorhabditis elegans." Molecular and Cellular Biochemistry **341**(1-2): 51-63.

Simao, F. A., R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva and E. M. Zdobnov (2015). "BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs." Bioinformatics **31**(19): 3210-3212.

Simon, S., H. Letsch, S. Bank, T. R. Buckley, A. Donath, S. L. Liu, R. Machida, K. Meusemann, B. Misof, L. Podsiadlowski, X. Zhou, B. Wipfler and S. Bradler (2019). "Old World and New World Phasmatodea: Phylogenomics Resolve the Evolutionary History of Stick and Leaf Insects." Frontiers in Ecology and Evolution **7**.

Singh, S., M. Gupta, S. Pandher, G. Kaur, N. Goel and P. Rathore (2019). "Using de novo transcriptome assembly and analysis to study RNAi in Phenacoccus solenopsis Tinsley (Hemiptera: Pseudococcidae)." Scientific Reports **9**.

Slaytor, M. (1992). "Cellulose Digestion in Termites and Cockroaches - What Role Do Symbionts Play." Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology **103**(4): 775-784.

Sleator, R. D., C. Shortall and C. Hill (2008). "Metagenomics." Letters in Applied Microbiology **47**(5): 361-366.

Soria-Carrasco, V., Z. Gompert, A. A. Comeault, T. E. Farkas, T. L. Parchman, J. S. Johnston, C. A. Buerkle, J. L. Feder, J. Bast, T. Schwander, S. P. Egan, B. J. Crespi and P. Nosil (2014). "Stick Insect Genomes Reveal Natural Selection's Role in Parallel Speciation." Science **344**(6185): 738-742.

Soto-Robles, L. V., V. Torres-Banda, F. N. Rivera-Orduna, E. Curiel-Quesada, M. E. Hidalgo-Lara and G. Zuniga (2019). "An Overview of Genes From Cyberlindnera americana, a Symbiont Yeast Isolated From the Gut of the Bark Beetle Dendroctonus rhizophagus (Curculionidae: Scolytinae), Involved in the Detoxification Process Using Genome and Transcriptome Data." Frontiers in Microbiology **10**.

Suh, S. O., C. J. Marshall, J. V. McHugh and M. Blackwell (2003). "Wood ingestion by passalid beetles in the presence of xylose-fermenting gut yeasts." Molecular Ecology **12**(11): 3137-3145.

Suh, S. O., N. H. Nguyen and M. Blackwell (2008). "Yeasts isolated from plant-associated beetles and other insects: seven novel Candida species near Candida albicans." Fems Yeast Research **8**(1): 88-102.

Szinwelski, N., M. S. Rodrigues, M. R. Pereira, J. E. Serrão and C. F. Sperber (2009). "Proventriculus of three nemobiinae crickets (Orthoptera: Grylloidea: Trigonidiidae)." Journal of Orthoptera Research **18**(1): 59-63.

Tartar, A., M. M. Wheeler, X. G. Zhou, M. R. Coy, D. G. Boucias and M. E. Scharf (2009). "Parallel metatranscriptome analyses of host and symbiont gene expression in the gut of the termite Reticulitermes flavipes." Biotechnology for Biofuels **2**.

Tassone, E. E., S. M. Geib, B. Hall, J. A. Fabrick, C. S. Brent and J. J. Hull (2016). "De novo construction of an expanded transcriptome assembly for the western tarnished plant bug, Lygus hesperus." Gigascience **5**.

Taylor, J. S. and J. Raes (2004). "Duplication and divergence: The evolution of new genes and old ideas." Annual Review of Genetics **38**: 615-643.

Terra, W. R. (1990). "Evolution of Digestive Systems of Insects." Annual Review of Entomology **35**: 181-200.

Terra, W. R. and C. Ferreira (1994). "Insect Digestive Enzymes - Properties, Compartmentalization and Function." Comparative Biochemistry and Physiology B-Biochemistry & Molecular Biology **109**(1): 1-62.

Terry, M. D. and M. F. Whiting (2005). "Mantophasmatodea and phylogeny of the lower neopterous insects." Cladistics **21**(3): 240-257.

Tibbot, B. K. and R. W. Skadsen (1996). "Molecular cloning and characterization of a gibberellin-inducible, putative alpha-glucosidase gene from barley." Plant Molecular Biology **30**(2): 229-241.

Tilgner, E. H., T. G. Kiselyova and J. V. McHugh (1999). "A morphological study of Timema cristinae vickery with implications for the phylogenetics of Phasmida." Deutsche Entomologische Zeitschrift **46**(2): 149-162.

Todaka, N., C. M. Lopez, T. Inoue, K. Saita, J. Maruyama, M. Arioka, K. Kitamoto, T. Kudo and S. Moriya (2010). "Heterologous Expression and Characterization of an Endoglucanase from a Symbiotic Protist of the Lower Termite, Reticulitermes speratus." Applied Biochemistry and Biotechnology **160**(4): 1168-1178.

Tokuda, G. (2019). "Plant cell wall degradation in insects: Recent progress on endogenous enzymes revealed by multi-omics technologies." Advances in Insect Physiology, Vol 57 **57**: 97-136.

Trimble, W. L., K. P. Keegan, M. D'Souza, A. Wilke, J. Wilkening, J. Gilbert and F. Meyer (2012). "Short-read reading-frame predictors are not created equal: sequence error causes loss of signal." BMC Bioinformatics **13**: 183.

Urbina, H., J. Schuster and M. Blackwell (2013). "The gut of Guatemalan passalid beetles: a habitat colonized by cellobiose- and xylose-fermenting yeasts." Fungal Ecology **6**(5): 339-355.

Vallotto, D., J. Bresseel, J. Constant and M. Gottardo (2016). "Morphology of the terminalia of the stick insect Dajaca napolovi from Vietnam (Insecta: Phasmatodea)." Entomological Science **19**(4): 376-382.

van Santen, Y., J. A. Benen, K. H. Schroter, K. H. Kalk, S. Armand, J. Visser and B. W. Dijkstra (1999). "1.68-A crystal structure of endopolygalacturonase II from Aspergillus niger and identification of active site residues by site-directed mutagenesis." J Biol Chem **274**(43): 30474-30480.

Varnai, A., L. Huikko, J. Pere, M. Siika-aho and L. Viikari (2011). "Synergistic action of xylanase and mannanase improves the total hydrolysis of softwood." Bioresource Technology **102**(19): 9096-9104.

Vatanparast, M., V. Hosseininaveh, M. Ghadamyari and S. M. Sajjadian (2012). "Pectinase and cellulase activity in the digestive system of the elm leaf beetle, Xanthogaleruca luteola Muller (Coleoptera: Chrysomelidae)." Journal of Asia-Pacific Entomology **15**(4): 555-561.

Velculescu, V. E., L. Zhang, B. Vogelstein and K. W. Kinzler (1995). "Serial Analysis of Gene-Expression." Science **270**(5235): 484-487.

Victorica, M. R., M. A. Soria, R. A. Batista-Garcia, J. A. Ceja-Navarro, S. Vikram, M. Ortiz, O. Ontanon, S. Ghio, L. Martinez-Avila, O. J. Q. Garcia, C. Etcheverry, E. Campos, D. Cowan, J. Arneodo and P. M. Talia (2020). "Neotropical termite microbiomes as sources of novel plant cell wall degrading enzymes." Scientific Reports **10**(1).

Wang, S. C., L. Y. Wang, X. Fan, C. Yu, L. Feng and L. Yi (2020). "An Insight into Diversity and Functionalities of Gut Microbiota in Insects." Current Microbiology **77**(9): 1976-1986.

Wang, Z., M. Gerstein and M. Snyder (2009). "RNA-Seq: a revolutionary tool for transcriptomics." Nature Reviews Genetics **10**(1): 57-63.

Warnecke, F., P. Luginbuhl, N. Ivanova, M. Ghassemian, T. H. Richardson, J. T. Stege, M. Cayouette, A. C. McHardy, G. Djordjevic, N. Aboushadi, R. Sorek, S. G. Tringe, M. Podar, H. G. Martin, V. Kunin, D. Dalevi, J. Madejska, E. Kirton, D. Platt, E. Szeto, A. Salamov, K. Barry, N. Mikhailova, N. C. Kyrpides, E. G. Matson, E. A. Ottesen, X. N. Zhang, M. Hernandez, C. Murillo, L. G. Acosta, I. Rigoutsos, G. Tamayo, B. D. Green, C. Chang, E. M. Rubin, E. J. Mathur, D. E. Robertson, P. Hugenholtz and J. R. Leadbetter

(2007). "Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite." Nature **450**(7169): 560-U517.

Watanabe, H., M. Nakamura, G. Tokuda, I. Yamaoka, A. M. Scrivener and H. Noda (1997). "Site of secretion and properties of endogenous endo-beta-1,4-glucanase components from Reticulitermes speratus (Kolbe), a Japanese subterranean termite." Insect Biochemistry and Molecular Biology **27**(4): 305-313.

Watanabe, H., H. Noda, G. Tokuda and N. Lo (1998). "A cellulase gene of termite origin." Nature **394**(6691): 330-331.

Watanabe, H. and G. Tokuda (2010). "Cellulolytic Systems in Insects." Annual Review of Entomology **55**: 609-632.

Waterhouse, A. M., J. B. Procter, D. M. A. Martin, M. Clamp and G. J. Barton (2009). "Jalview Version 2-a multiple sequence alignment editor and analysis workbench." Bioinformatics **25**(9): 1189-1191.

Waterhouse, D. F. (1957). "Digestion in Insects." Annual Review of Entomology **2**: 1-18.

Waterhouse, R. M., M. Seppey, F. A. Simao, M. Manni, P. Ioannidis, G. Klioutchnikov, E. V. Kriventseva and E. M. Zdobnov (2018). "BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics." Molecular Biology and Evolution **35**(3): 543-548.

Whiting, M. F., S. Bradler and T. Maxwell (2003). "Loss and recovery of wings in stick insects." Nature **421**(6920): 264-267.

Wiens, J. J., R. T. Lapoint and N. K. Whiteman (2015). "Herbivory increases diversification across insect clades." Nature Communications **6**.

Wilke, A., T. Harrison, J. Wilkening, D. Field, E. M. Glass, N. Kyrpides, K. Mavrommatis and F. Meyer (2012). "The M5nr: a novel non-redundant database containing protein sequences and annotations from multiple sources and associated tools." Bmc Bioinformatics **13**.

Wipfler, B., H. Letsch, P. B. Frandsen, P. Kapli, C. Mayer, D. Bartel, T. R. Buckley, A. Donath, J. S. Edgerly-Rooks, M. Fujita, S. L. Liu, R. Machida, Y. Mashimo, B. Misof, O. Niehuis, R. S. Peters, M. Petersen, L. Podsiadlowski, K. Schutte, S. Shimizu, T. Uchifune, J. Wilbrandt, E. Yan, X. Zhou and S. Simon (2019). "Evolutionary history of Polyneoptera and its implications for our understanding of early winged insects." Proceedings of the National Academy of Sciences of the United States of America **116**(8): 3024-3029.

Wooley, J. C., A. Godzik and I. Friedberg (2010). "A Primer on Metagenomics." Plos Computational Biology **6**(2).

Wu, C., R. N. Crowhurst, A. B. Dennis, V. G. Twort, S. L. Liu, R. D. Newcomb, H. A. Ross and T. R. Buckley (2016). "De Novo Transcriptome Analysis of the Common New Zealand Stick Insect Clitarchus hookeri (Phasmatodea) Reveals Genes Involved in Olfaction, Digestion and Sexual Reproduction." Plos One **11**(6).

Yuki, M., S. Moriya, T. Inoue and T. Kudo (2008). "Transcriptome analysis of the digestive organs of Hodotermopsis sjostedti, a lower termite that hosts mutualistic microorganisms in its hindgut." Zoological Science **25**(4): 401-406.

Yun, J. H., S. W. Roh, T. W. Whon, M. J. Jung, M. S. Kim, D. S. Park, C. Yoon, Y. D. Nam, Y. J. Kim, J. H. Choi, J. Y. Kim, N. R. Shin, S. H. Kim, W. J. Lee and J. W. Bae (2014). "Insect Gut Bacterial Diversity Determined by Environmental Habitat, Diet, Developmental Stage, and Phylogeny of Host." Applied and Environmental Microbiology **80**(17): 5254-5264.

Zhou, W. L., D. C. Irwin, J. Escovar-Kousen and D. B. Wilson (2004). "Kinetic studies of Thermobifida fusca Cel9A active site mutant enzymes." Biochemistry **43**(30): 9655-9663.

Zompro, O. (2004). "A key to the stick-insect genera of the 'Anareolatae' of the new world, with descriptions of several new taxa (Insects : Phasmatodea)." Studies on Neotropical Fauna and Environment **39**(2): 133-144.