Fall 2013

The Termite Digestome: Understanding the Digestive Physiology Involved in Lignocellulosic Biomass Degradation

Zachary John Karl

Purdue University

Follow this and additional works at: https://docs.lib.purdue.edu/open_access_dissertations

Part of the Biochemistry Commons, Entomology Commons, and the Molecular Biology Commons

Recommended Citation
https://docs.lib.purdue.edu/open_access_dissertations/131

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact epubs@purdue.edu for additional information.
This is to certify that the thesis/dissertation prepared

By Zachary J. Karl

Entitled
The Termite Digestome: Understanding the Digestive Physiology Involved in Lignocellulosic Biomass Degradation

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Michael Scharf
Chair
Jonathan Neal
John Patterson
Nathan Mosier

To the best of my knowledge and as understood by the student in the Research Integrity and Copyright Disclaimer (Graduate School Form 20), this thesis/dissertation adheres to the provisions of Purdue University’s “Policy on Integrity in Research” and the use of copyrighted material.

Approved by Major Professor(s): Michael Scharf

Approved by: Steve Yaninek 11/24/2013
Head of the Graduate Program  Date
THE TERMITE DIGESTOME: UNDERSTANDING THE DIGESTIVE PHYSIOLOGY INVOLVED IN LIGNOCELLULOSIC BIOMASS DEGRADATION

A Dissertation
Submitted to the Faculty
of
Purdue University
by
Zachary J. Karl

In Partial Fulfillment of the
Requirements for the Degree
of
Doctor of Philosophy

December 2013
Purdue University
West Lafayette, Indiana
ACKNOWLEDGMENTS

Foremost, I would like to thank my major advisor Dr. Michael Scharf for his continuous support, motivation, patience, enthusiasm, and knowledge as I journeyed through my doctoral career. His guidance has helped me to mature both as a scientist and as an individual. I truly could not have asked for a better advisor and mentor for my Ph.D. study.

Besides my advisor, I would like to thank the rest of my dissertation committee: Dr. Jonathan Neal, Dr. John Patterson, and Dr. Nathan Mosier, for their encouragement, critiques, and insight that greatly contributed to the progression of my research.

I thank my fellow labmates and all of my friends for their continuous support, encouragement, and assistance throughout my time in graduate school.

Last but not the least, I would like to thank my family: my parents John and Elizabeth, for all of their unconditional support and love throughout my life. Words cannot express how truly grateful and lucky I am to have such wonderful parents. I would also like to thank my sisters Annie and Eva, for their all of their support and for dealing with my annoying tendencies as we grew up.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>vii</td>
</tr>
<tr>
<td>CHAPTER ONE: INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Bibliography</td>
<td>8</td>
</tr>
<tr>
<td>CHAPTER TWO: HOST/SYMBIONT COLLABORATION</td>
<td>12</td>
</tr>
<tr>
<td>Abstract</td>
<td>12</td>
</tr>
<tr>
<td>Introduction</td>
<td>13</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>18</td>
</tr>
<tr>
<td>The Termite</td>
<td>18</td>
</tr>
<tr>
<td>Chemicals and Reagents</td>
<td>19</td>
</tr>
<tr>
<td>Buffers and Solutions</td>
<td>20</td>
</tr>
<tr>
<td>Dissections</td>
<td>20</td>
</tr>
<tr>
<td>Glucose Detection Assay</td>
<td>21</td>
</tr>
<tr>
<td>Pentose Detection Assay</td>
<td>22</td>
</tr>
<tr>
<td>Reagent Specificity Test</td>
<td>23</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>23</td>
</tr>
<tr>
<td>Results</td>
<td>24</td>
</tr>
<tr>
<td>Reagent Specificity Test</td>
<td>24</td>
</tr>
<tr>
<td>Glucose/Pentose Detection Assays</td>
<td>24</td>
</tr>
<tr>
<td>Discussion</td>
<td>25</td>
</tr>
<tr>
<td>Termite Digestion</td>
<td>25</td>
</tr>
<tr>
<td>Biofuels Implications</td>
<td>30</td>
</tr>
<tr>
<td>Bibliography</td>
<td>40</td>
</tr>
<tr>
<td>CHAPTER THREE: GLUCOSE ABSORPTION</td>
<td>46</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>----------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Abstract</td>
<td>46</td>
</tr>
<tr>
<td>Introduction</td>
<td>46</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>56</td>
</tr>
<tr>
<td>The Termite</td>
<td>56</td>
</tr>
<tr>
<td>Chemicals and Reagents</td>
<td>57</td>
</tr>
<tr>
<td>Buffers and Solutions</td>
<td>57</td>
</tr>
<tr>
<td>Gene Selection</td>
<td>57</td>
</tr>
<tr>
<td>Dissections</td>
<td>59</td>
</tr>
<tr>
<td>Gene Expression</td>
<td>59</td>
</tr>
<tr>
<td>Glucose Feeding Assay</td>
<td>60</td>
</tr>
<tr>
<td>Glucose Detection Assay</td>
<td>60</td>
</tr>
<tr>
<td>Reduced Sugar Detection Assay</td>
<td>61</td>
</tr>
<tr>
<td>Optical Density Test</td>
<td>62</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>62</td>
</tr>
<tr>
<td>Results</td>
<td>63</td>
</tr>
<tr>
<td>Gene Selection</td>
<td>63</td>
</tr>
<tr>
<td>Gene Expression</td>
<td>64</td>
</tr>
<tr>
<td>Glucose Detection Assay</td>
<td>65</td>
</tr>
<tr>
<td>Reduced Sugar Detection Assay</td>
<td>65</td>
</tr>
<tr>
<td>Optical Density Test</td>
<td>65</td>
</tr>
<tr>
<td>Discussion</td>
<td>65</td>
</tr>
<tr>
<td>Bibliography</td>
<td>81</td>
</tr>
<tr>
<td>CHAPTER FOUR: FEEDSTOCK EFFECTS</td>
<td>89</td>
</tr>
<tr>
<td>Abstract</td>
<td>89</td>
</tr>
<tr>
<td>Introduction</td>
<td>90</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>96</td>
</tr>
<tr>
<td>Model Insect</td>
<td>96</td>
</tr>
<tr>
<td>Chemicals and Reagents</td>
<td>96</td>
</tr>
<tr>
<td>Buffers and Solutions</td>
<td>97</td>
</tr>
<tr>
<td>Feedstocks</td>
<td>97</td>
</tr>
</tbody>
</table>
# Chapter Five: Esterase Purification

## Abstract

## Introduction

## Materials and Methods

- Termites
- Chemicals and Reagents
- Buffers and Solutions
- Feedstock Native-PAGE
- Dissections
- Column Chromatography
- Fraction Esterase Activity Assay
- Fraction Native-PAGE
- Fraction SDS-PAGE
- Glucose Detection Assay
- Peptide Sequencing
- Data Analysis

## Results

- Feedstock Native-PAGE
- Fraction Esterase Activity Assay
- Fraction Native-PAGE
<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fraction SDS-PAGE</td>
<td>176</td>
</tr>
<tr>
<td>Glucose Detection Assay</td>
<td>177</td>
</tr>
<tr>
<td>Peptide Sequencing</td>
<td>177</td>
</tr>
<tr>
<td>Discussion</td>
<td>178</td>
</tr>
<tr>
<td>Bibliography</td>
<td>200</td>
</tr>
<tr>
<td>CHAPTER SIX: CONCLUSION</td>
<td>205</td>
</tr>
<tr>
<td>VITA</td>
<td>207</td>
</tr>
</tbody>
</table>
ABSTRACT


The purpose of this research was to advance the understanding of lower termite digestive physiology and discover potential biocatalysts that can aid in the degradation of lignocellulosic biomass. Various protein characterization and gene expression methods were used throughout this research in order to accomplish these objectives. The results of this dissertation indicate that: 1) termites and their symbionts act in a synergistic manner to degrade biomass in vitro, 2) the host fraction of the gut (i.e., foregut and midgut) is the likely site of glucose absorption, 3) the termite and its symbionts contribute specific enzymes to the digestive process, 4) diet impacts biochemical and molecular aspects of termite digestive physiology, 5) the genetic makeup of the termite digestome is highly conserved, and 6) potential esterase-based ligninases from this termite can enhance saccharification in the presence of recombinant cellulases. These results suggest that utilizing Reticulitermes flavipes as a source of feedstock-specific recombinant enzymes for utilization in commercial biorefinery processes can likely decrease biocatalyst input while increasing simple sugar output, resulting in a more
cost-effective approach to cellulosic ethanol production, making biofuels a more viable option in the near future.
CHAPTER ONE: AN INTRODUCTION TO TERMITE DIGESTION AND INDUSTRIAL BIOFUELS

Termites are eusocial insects known for their unique ability to feed upon, digest, and receive nourishment from materials composed of lignocellulose -- a recalcitrant biopolymer of lignin, hemicellulose, and cellulose found in the cell walls of all woody plant species (7, 10, 18, 20, 26, 27, 28, 29, 30, 33, 34, 36, 37, 40, 41, 42). The acquisition of nutrients from this structure involves biochemical processes that aid in the degradation of various bond types, the detoxification of hazardous by-products, the transport of metabolic nutrients such as monosaccharides and amino acids, and the conversion of these nutrients into energy (7, 10, 20, 26, 28, 30). Due to their efficiency in this degradation process, the biochemistries harbored within the termite digestive tract, both contributed by the termite host and its symbionts, have been an ongoing focus of investigation (4, 28, 30). This biochemical system also makes for an excellent model system from which researchers can gain insights and apply findings to large-scale in vitro production of commercial products from plant-based materials (28, 30). The following research set out to explore the specific biochemistries involved in the digestion of compositionally variable 2nd generation feedstocks by the lower termite Reticulitermes flavipes in order to 1) advance our understanding of termite digestive physiology and 2) discover potential biocatalysts that can enhance the cost-effective and energy-efficient
production of biofuels from lignocellulosic biomass. The information revealed through this research may also aid in the development of novel termite control technologies by revealing physiologically significant target sites, and produce novel catalytic candidates for the conversion of feedstocks into biomaterials.

*Reticulitermes flavipes*, more commonly referred to as the Eastern Subterranean termite, is a common termite species belonging to the family Rhinotermitidae (30). *R. flavipes* is the most abundant and widely-distributed species of termite in North America, inhabiting every state East of the Mississippi River and stretching North into lower parts of Canada (2). Due to their abundance, distribution, and habit of eating wooden structures, this species of termite is a very destructive and costly insect pest (20, 30). It is estimated that nearly 80% of the $3 billion spent annually on termite control is allocated for the control of this particular species of termite (2). Yet while their destructive behavior is considered a problem, understanding the degradation and digestion processes they utilize to accomplish this destruction is of major research interest within the biofuels and biomaterials industries, and may prove to be very beneficial (4, 10, 20, 21, 28, 30).

The first major goal of this research was to further the existing knowledge of termite digestive physiology by examining the collaborative relationship between the termite host and its symbionts in terms of food particle degradation and nutrient transport, and characterizing the specific enzymatic contributions to each fraction. Early investigations (1920 – 1940) into the digestive physiology of the termite gut suggested digestive collaboration to be occurring between the host termite and its symbionts (6, 12). Since then, however, the symbionts have become a main focus of research and have been receiving a bulk of the credit for degrading diet materials (4, 13, 17, 26). The current
dogmatic view of the termite digestive process begins with the sclerotized mandibles of worker termites (10, 28). These termites will “chew” or gnaw off particles of wood or plant material to be digested. Once the particle is inside the mouth parts/esophagus, it will encounter the first major region of the termite digestive tract, the foregut (7, 30, 37). The foregut is composed of the esophagus, the salivary glands, and the crop, and is the site of the first step in the degradation of the woody particles (lignocellulose), the depolymerization of lignin (7, 11, 26, 28, 30, 33, 37). In order to depolymerize this recalcitrant material and oxidize its toxic by-products, the salivary glands secrete such enzymes as laccases/phenol-oxidases, esterases, peroxidases, and cellulases (7, 10, 23, 37). Although cellulose digestion is also believed to occur in the foregut, this region of the digestive tract is currently under heavy investigation due to its potential lignin-degrading abilities (28).

The foregut invaginates into the next major region of the digestive tract called the midgut. The midgut is an elongated region of the digestive tract that contains the Malpighian tubules; long filamentous tubules that allow for the removal of waste products from the body and secretion into the digestive tract (5, 26, 29). While the woody particle travels through the midgut, hemicellulose and cellulose are acted upon by cellulolytic enzymes (29, 30, 33, 40, 42). Hemicellulose is degraded by enzymes such as β-mannosidase, β-xylosidase, β-arabinosidase, β-glucosidase, and endoglucanase (14, 24, 25, 32). The degradation of cellulose is carried out by two main families of glycosyl hydrolase enzymes: endoglucanase (GHF9) and β-glucosidase (GHF1) (29, 30, 36, 42). These two gut regions, the foregut and the midgut, are denoted as the “host fraction” throughout this research due to the understanding that the termite host is the main
contributor of functional proteins/enzymes within these regions (26, 27, 33). With that said, the third and final major region of the digestive tract, the hindgut, is referred to as the “symbiont fraction” for the same reasons (26, 27, 33). The hindgut, or paunch, begins at a junction with the midgut where the Malpighian tubules are located, and extends to the end of the rectum (5, 30).

The hindgut is the main region of distinction between higher and lower termites. The hindgut of lower termites, such as *R. flavipes*, contains both bacteria and protist populations, whereas higher termites lack protists and only house bacteria (4, 5, 13, 17, 18, 25, 33). A definitive number of protist and bacterial species found within the paunch of *R. flavipes* has yet to be determined, but estimates based on morphological data and 16s sequencing indicate that there may be as many as 12 unique protist species, and anywhere between 1,000 and 5,000 bacterial species (3, 17). Further hemi- and cellulolytic symbiont-derived degradation are reported to take place in this region (4, 20, 25, 26, 29, 30). This has been shown to be carried out by many of the same families of enzymes as previously mentioned as occurring in the midgut. However, exoglucanases (GHF7) are mainly associated with protists and other lower eukaryotes (4, 30, 33). The simple sugars, amino acids, and various other by-products released throughout this digestive process are mainly utilized by the termite and its symbionts as metabolic currency (20, 26, 28, 30, 33, 36, 41). The remaining feces, which contain fatty acids, lignin by-products, undigested sugars, bacteria, and some symbionts, are then excreted from the anus (4, 5, 10, 20, 30). Termites have been shown to be very efficient at this degradation process. Specifically, *R. flavipes* has the ability to degrade upwards of 99% of the lignocellulose structure and utilize roughly 80% of the energy stored in the
polysaccharides, making them the most efficient natural bioreactors studied to date (4).

Due to the efficiency of lignocellulose degradation exhibited by this termite species, and
the parallels between the chronological order of lignocellulose degradation within the
termite and industry, research is turning towards efficient models of biodegradation, such
as termites and their symbionts, for insights and answers (1, 9, 28, 30, 33, 35, 42).

The second major goal of this research was to discover novel catalysts from the
termite digestome (i.e., the pool of enzyme-encoding genes responsible for digestive
processes) that could aid in the efficient saccharification of woody biomass when
combined with already-established recombinant proteins for utilization in the production
of biofuels (26, 27, 28, 29, 30, 33, 42). Ethanol and other alcohol fuels can be derived
from natural lignocellulosic biomass, which accounts for over 50% of the world’s total
biomass (16, 21, 31). It has been well documented that the United States is the largest
consumer of Petroleum in the world with consumption continuing to increase every year
(8, 38). This in turn has led to a proportional increase in the amount of greenhouse gasses
released into the atmosphere, usually in the form of carbon dioxide (CO2) (8, 22, 38).
The use of biofuels is proposed to be a carbon neutral process (8, 38). This implies that
use of biofuels would not impact the accumulation of CO2 in the atmosphere like
currently used fossil fuels. The burning of biofuels is said to only release the CO2 that
the source plant had originally absorbed from the atmosphere (8, 9, 38). The entire
accumulation of harmful greenhouse gases caused by the production and use of biofuels
is estimated at less than 15% of the harmful gases released by petroleum production and
use (8, 40). The alternative use of cellulosic ethanol as a cleaner and more abundant fuel
source has been a main focus of biochemical and bioengineering research since the oil
crisis which occurred during the 1970s (15, 19, 38). From a societal viewpoint, the ultimate need that would be met through the increased use of biofuels is the direct impact it would have on economic growth. It has been estimated that an increase in biofuel production could create between 800 - 900k jobs within the next twenty years (8). The use of biofuels could ultimately lower fuel prices as well due to the resulting decrease in dependency on foreign oil (39). Both the increase in jobs and decrease in fuel prices have led economists to estimate a direct economic growth of $100 - $200 billion over the next twenty years (8).

For the biofuels industry, which is still searching for cost-effective biomass saccharification technologies, exploring the digestome of the termite can result in the discovery of proficient lignocellulases to be utilized in degradation of woody biomass. Thus, a central goal of this dissertation was to investigate and identify the collaboration of lignocellulosic biomass degradation catalysts from the digestive tract of the lower termite *R. flavipes*. This goal was accomplished through the utilization of innovative monosaccharide detection techniques, various enzyme activity assays, gene expression tests, and protein purification procedures. Eventually, recombinant lignocellulases derived from the host and symbiont fraction of this termite can serve as economically favorable and eco-friendly catalysts for lignocellulose degradation relative to current popular methods. In general, this research revealed that while there is saccharific synergy between host- and symbiont-derived lignocellulases, the termite host seems to be the main contributor of lignin and cellulose degrading enzymes, whereas the symbionts appear to be more responsible for hemicellulase input. This finding is supported by glucose transporter studies, which suggested that the lower termite itself may act
independently of the symbionts in order to liberate glucose from lignocellulosic materials. Also, termite-derived detoxification enzymes used in conjunction with cellulases appear to significantly increase the saccharification of lignocellulosic biomass, and this is an area that should thus receive additional focus in the future.


CHAPTER TWO: DEFINING HOST-SYMBIONT COLLABORATION IN THE DIGESTIVE TRACT OF THE LOWER TERMITE *RETICULITERMES FLAVIPES*

**ABSTRACT**

The eastern subterranean termite *Reticulitermes flavipes*, and its protozoan and bacteria symbionts, possess an array of lignocellulases which partake in the degradation and digestion of lignocellulosic biomass. However, the functional characterization of the saccharific relationship between these two symbiotic partners has yet to be established due to a narrow focus in genomic sequencing within this research field and a lack of targeted methods necessary to test this relationship. Through incubating pine wood lignocellulose with native gut tissue, and original monosaccharide detection techniques, a synergistic relationship was found between the termite host and its symbiotic protozoa in the saccharification of pine wood. This research aids in our understanding termite-symbiont digestion and can lead to novel termite-derived biomass conversion technologies.
INTRODUCTION

Greenhouse gas emissions from the combustion of petroleum fuels, the economic benefits of a sustainable biofuels society, such as Brazil, and the inevitable disappearance of fossil fuels has researchers around the world investigating possible alternative fuels and their means of production (8, 14, 19, 20, 44, 47, 57, 60, 67). Alcohol fuels, such as ethanol and butanol, are at the forefront of this investigation due to their low carbon emissions and carbon-natural means of production. The basic method of alcohol fuel production involves the depolymerization and degradation of the lignocellulose structure within plant cell walls in order to free monosaccharides for the purpose of fermentation into alcohol (13, 15, 31, 35). Within the biofuels industry, the degradation of this sturdy material follows one of two paths; a chemical path or an enzymatic path. The chemical path involves the use of acids, bases, salts, and extreme temperatures. Due to the energy inefficiency and abundance of toxic waste generated by the chemical path, the enzymatic path was developed. Enzymatic research investigates the degradation of lignocellulose through the use of organism-derived recombinant lignocellulases (28, 51, 57, 59). Many wood-feeding organisms are currently being investigated in order to gain an understanding of the biochemical process involved in the saccharification of plant-based material, and ultimately synthetically copy this process on a larger scale for use in the biofuels industry (3, 13, 27, 30, 44, 55, 63). Due to its well-established capacity as a wood-feeder, the model organism chosen for this investigation was the lower subterranean termite *Reticulitermes flavipes*. 
Termites are eusocial insects known for their remarkable ability to degrade and utilize the plant cell wall for various purposes, the most important of these purposes inevitably being nutrient acquisition (39, 42, 54, 61). The plant cell wall is structurally held together by lignocellulose; a very recalcitrant material consisting of cellulose (40%), hemicellulose (25%), and lignin (20%) (13, 42, 44). The degradation and digestion of this complex material is such an extraordinary feat, that it should come as no surprise that termites employ the help of other organisms, namely protozoa and bacteria, to aid in the saccharification process. The development of next-generation sequencing technologies within the past 10-15 years has given rise to metatranscriptomics as a way to identify potential relationships between a host and its symbiont(s). Recent studies, using both microarray analysis and pyrosequencing, have determined that both the termite and its symbionts contain lignocellulase-encoding genes within their respective digestomes (21, 36, 38, 44, 49, 53). The termite itself, referred to as the host fraction of the digestive tract (salivary glands, foregut, midgut) throughout the rest of this chapter (Figure 2.1), contains lignocellulase-encoding genes specific towards cellulose degradation and lignin depolymerization, such as endoglucanase, β-glucosidase, phenol-oxidase, laccase, and esterase (10, 27, 36, 43, 49, 58, 65, 66). The protists and bacteria, or symbiont fraction of the digestive tract (hindgut) (Figure 2.1), have been shown to contain lignocellulase-encoding genes responsible for the hydrolysis of hemicellulose and cellulose, such as exoglucanase, β-glucosidase, β-xylosidase, β-mannosidase, and β-arabinosidase (3, 19, 33, 36, 46, 49, 50). Both termite and symbiont gene discoveries and annotations have been supported with follow-up functional studies using substrate specific enzyme detection assays, gene silencing studies, or a combination of both methods (10, 27, 33,
43, 45, 50, 54, 58, 64, 66). However, aside from their theorized saccharific fraction-specific contributions based off of single-gene/enzyme selection studies, little is known about the digestive relationship between the termite and its symbionts, and whether or not they collaborate in order to achieve higher degradation efficiency (monosaccharide release). What is known is that termites possess the ability to degrade upwards of 99% of the lignocellulose structure and utilize roughly 80% of the energy stored in the polysaccharides, making them the most efficient natural bioreactors and exemplary models of biomass degradation (4, 19, 31, 44, 63). However, the termite itself is not solely responsible for this high degree of efficiency.

The presence of protozoa and bacteria within the gut of lower termites was first reported over a century ago (24), yet many aspects of this symbiosis, including digestion, immunity, and other intrinsic physiologies, have remained largely uncharacterized. Our understanding of the symbiont population is so limited that a definitive number of protist and bacterial species residing within the hindgut has not been determined (25, 34). Furthermore, a recent study suggests that the popularly accepted range of protist species diversity within the hindgut (10-12 species) has been greatly underestimated (21). However, steps have been taken to understand and better define the hindgut-specific role of symbionts through the understanding of the hindgut environment, which has been established as being an important component of lignocellulose degradation (34). The hindgut, or paunch, is one of the three major regions of the lower termite digestive tract. This region is referred to as the “symbiont fraction” due to this being the main location and functional protein release site of the microbial and protozoa communities (44, 49). However, studies have found a few clusters of microbes residing in the midgut using T-
RFLP fingerprints and clone libraries, but these findings have not been conclusive, suggesting that cross contamination of gut regions during dissections was inevitable (62) and that physical separation of host and symbiont for *in vitro* studies is impossible (49). Protozoa have also been found throughout the midgut using scanning electron microscopy, but these flagellates appear to be in an encysted non-functional state (25). The pH throughout the termite digestive tract is neutral, ranging from pH 6.5-7.2 (5). The tracheal system of the termite allows for the direct connection of individual cells of the gut epithelium to the atmosphere (32). This in turn generates a constant influx of oxygen into the system. The termite digestive tract, from the beginning of the foregut to the end of the midgut, is an aerobic environment, while the hindgut radially transitions from an outer aerobic environment to an inner anaerobic environment (12). However, the decreasing gradient of oxygen occurs over a large amount of space relative to the termite gut. This suggests a gradual decrease in oxygen which allows for microorganisms with a continuum oxygen requirements to reside throughout the anaerobic center to the more aerobic outer edge (5, 12). These conditions split the protozoa into the two main phyla found within the termite hindgut; parabasalids (anaerobic interior) and oxymonads (aerobic exterior) (6, 21, 25, 32, 34). Most of the oxygen consumption is due to the metabolic activity of the microbial community (12, 18). The influx and reduction of oxygen plays a direct role in the flow of carbon and electrons. Conversely, in the case of hydrogen, large amounts can be found deposited in the center of the hindgut lumen, being produced by the hydrogenosomes of the anaerobic protozoa, dissipating towards the epithelial edge as it is being consumed by the methanogens, acetogens, and sulfate reducing microbes (4, 5, 12). These microbes use hydrogen as a substrate for such
processes as methanogenesis and reductive acetogenesis. Hydrogen can be found throughout the hindgut, from the anterior to the posterior, with low hydrogen partial pressure in the posterior region of the gut creating, or as a result of, even more microhabitats (5, 12, 18, 25, 32, 34). Besides hydrogen, the anaerobic protozoa and bacteria also produce carbon dioxide during polysaccharide fermentation, which is also a necessary substrate in methanogenesis and acetogenesis (4, 6, 26, 30, 46). As portrayed, the symbionts appear to aid in the homeostasis of the hindgut environment, but still very little is known about the extent of their collaboration with termite-derived lignocellulases in biomass degradation. Only speculations based on gene sequencing data exist.

Past research has explored, and present research continues to explore, the termite and its symbiont populations independently as potential sources of lignocellulases for utilization in the degradation of plant-based feedstocks. However, a more in-depth understanding of the physiological and biochemical interactions between the termite and its symbiont populations may reveal novel proteins and genes that can be utilized to increase saccharific efficiency in the biofuels industry. This research focuses on taking the initial steps necessary for characterizing the relationship of the termite and its symbionts in the saccharification of 2nd generation feedstocks (processive waste) by investigating the collaboration between these two fractions in the release of monosaccharides from the naturally occurring substrate, pine wood lignocellulose (i.e., “sawdust”). The results of this study will be used to determine if both the termite and its symbiotic microbiota should be probed as sources of potential recombinant lignocellulases.
The main objective of this research was to quantify the degree to which the termite host and its symbionts collaborate in the saccharification of woody biomass. This relationship was investigated using sawdust incubation assays and colorimetric monosaccharide detection techniques. Data from this study indicate that enzymes derived from the termite host and its symbionts act upon lignocellulosic food sources in a synergistic manner in order to hydrolyze the plant cell wall into simple pentose and hexose sugars. Also, significantly more glucose was released from the pine wood substrate during the incubations than pentose monosaccharides. This is the first demonstration and quantification of host-symbiont synergy within a wood-feeding insect. These results suggest that both the termites and their symbionts participate in the saccharification of dietary plant matter, and that both termite- and symbiont-derived lignocellulases should be considered when searching for enzymes that will comprise maximally efficient recombinant protein cocktails. Also, glucose appears to be a more important nutritional component of the termite and symbiont diet than do pentose sugars.

MATERIALS AND METHODS

The Termite

*Reticulitermes flavipes*, a lower termite species of the family Rhinotermitidae, was used throughout this study. Three separate *R. flavipes* colonies were used in order to determine the presence or absence of inter-colony disparity, along with four replicates from each colony in order to test intra-colony variation. The colonies B2, K5, and K6 were collected from the University of Florida campus (Alachua County, FL, USA) and a
Kanapaha residence (Alachua County, FL, USA). Each colony was identified as *R. flavipes* through the use of genomic DNA extraction, PCR, and 16-s rDNA gene sequencing (1, 48). Termites of the worker caste were used in this study due to their superior ability to digest lignocellulose over other castes present in termite colonies. Termites were considered workers based on previously described morphological distinctions such as the absence of wings, large mandibles, and distended abdomens (23). For no less than 6 months prior to this study, each colony was held in separate plastic containers on a diet of pine wood shims (Nelson Wood Shims; Cohasset, MN) and brown paper towels. Given that this termite species is subterranean, the conditions within the incubation chamber were set at a temperature of 22°C with 70% relative humidity and a 0:24 light:dark photoperiod.

**Chemicals and Reagents**

*Sugars:* glucose, mannose, arabinose, rhamnose, cellobiose, glucuronic acid, and trehalose were purchased from Sigma-Aldrich (St. Louis, MO); xylose, sucrose, and galacturonic acid were purchased from Fisher Scientific (Waltham, MA); all sugars are D-enantiomers except for arabinose which is L-enantiomer. *Glucose detection reagent:* Autokit Glucose (Wako Chemical; Richmond, VA). *Pentose detection reagent:* 0.6 g thiourea (Sigma-Aldrich; St. Louis, MO) and 0.3 g 4-Bromoaniline (Sigma-Aldrich; St. Louis, MO) in 15 mL of glacial acetic acid (Mallinckrodt Chemicals; St. Louis, MO).
Buffers and Solutions

*Homogenization buffer:* 1x phosphate buffered saline (PBS), pH 7.0 (Bio-Rad; Hercules, CA).  *Glucose assay incubation/reaction buffer:* 0.1M sodium acetate buffer, pH 7 (Mallinckrodt Pharmaceuticals; St. Louis, MO).  *Glucose assay stop solution:* 10 µL of 0.2M ethylenediaminetetraacetic acid (EDTA) (Sigma-Aldrich; St. Louis, MO) in 500 µL of 0.1 M sodium acetate assay buffer.  *Sugar serial dilutions:* All sugars used for reagent specificity tests and standard curves were dissolved in 0.1M sodium acetate buffer, pH 7 and purchased from Sigma-Aldrich (St. Louis, MO).

Digestive Tract Dissections and Tissue Preparations

Individual termites were randomly removed from each of the three colonies. The first 25 termites morphologically determined to be of the worker caste were set aside for digestive tract dissections. Whole guts were first removed according to an established procedure in which the hind section of the termite body was cut, resulting in an opening with which to grab the anus and pull out the entire intact digestive tract (64). The whole guts were then further dissected into host fractions (salivary gland (SG), foregut (FG), midgut (MG); Figure 2.1) and symbiont fractions (hindgut; Figure 2.1), rinsed of gut material, and placed in 1x PBS (pH 7.0) at a concentration of 10 µL per gut fraction. The gut fractions were then homogenized using 1.5 mL Pellet Pestles (Kimble-Kontes; Vineland, NJ) and centrifuged at 14,600 rcf with a set temperature of 4°C for 15 minutes. The supernatants were then separated into 50 µL aliquots and stored at -80°C until sawdust assays were prepared. All supernatants were stored for one day before use and
each frozen aliquot was only thawed once. Tests using gut homogenate and model substrates were performed to ensure protein freeze/thaw efficacy over time (Figure 2.6).

**Glucose Detection Assay**

Pinewood sawdust was placed in a vented 1.5 mL Eppendorf centrifuge tube with 500 µL sodium acetate buffer at 2% w/v. Sawdust assays contained four different sawdust treatments; (i) host fraction; pine wood sawdust incubated with 50 µL of host fraction homogenate (SG, FG, MG) and 50 µL of PBS homogenization buffer, (ii) symbiont fraction; pine wood sawdust incubated with 50 µL of symbiont fraction homogenate (HG) and 50 µL of phosphate saline buffer, (iii) whole gut; pine wood sawdust incubated with 50 µL of host homogenate and 50 µL of symbiont homogenate, and (iv) blank; pine wood incubated with 100 µL of phosphate buffer saline. The final volume within each treatment was 600 µL. Each treatment was divided into three technical replications per biological replication. The treatments were then incubated at 37 °C with an agitation rate of 220 rpm for 10 hours. Once the incubation time had expired, the treatments were centrifuged at 12,000 rcf and 23 °C for 5 minutes. To stop the reaction, the supernatant was removed and placed in a clean centrifuge tube with 0.2 M EDTA at a volume of 10 µL per 500 µL of 0.1 M sodium acetate assay buffer. This resulted in a final concentration of 4 mM EDTA per treatment. The supernatant (50 µL) was then pipetted in triplicate microplate wells. A standard curve of serially diluted glucose, from 5 mM of glucose downward to a buffer blank, was then generated in 0.1 M sodium acetate assay buffer + 4mM EDTA. Glucose detection reagent (200 µL) was then added to each well containing glucose dilutions or aliquots of treatment (52).
The microplate was then left to incubate at room temperature (24 °C) for 5 minutes before absorbance was read as an endpoint at a wavelength of 505 nm.

**Pentose Detection Assay**

This detection method followed a modified procedure previously established for the specific detection of pentose (i.e., 5-carbon) sugars (11, 37). Pentose detection reagent was first prepared by combining 0.6 g thiourea and 15 mL glacial acetic acid in a 15 mL Nalgene Falcon tube. This reagent was then mixed gently and centrifuged for 5 minutes at 1,000 xg. The resulting supernatant was transferred to a 50 mL Nalgene Falcon tube containing 0.3 g of 4-Bromoaniline and inverted gently 3-4 times. The thiourea pellet was saved for later re-use in reagent preparation. It was important to keep Falcon tube wrapped in foil throughout the experiment due to light sensitivity of the reagent.

Pinewood sawdust assays, including treatments, incubation, and microplate setup, were constructed following the procedures mentioned in the glucose assay above. However, in this assay standard curves were generated using a serial dilution of xylose in 0.1 M sodium acetate assay buffer + 4mM EDTA, from 5 - 0.078125 mM, including a buffer blank (0) as the 8th dilution. Pentose detection reagent was then added to wells containing xylose dilutions or treatment aliquots. The microplate was then covered with foil, placed within a baking dish, and moved to a drying oven in order to be heated at 70 °C for 10-15 mins. Once the heating time had ended, the plate was removed and left to cool in a fume hood for 70 minutes. Absorbance was then read as an endpoint at a
wavelength of 520 nm. All hazardous waste was stored in appropriate labeled containers and disposed of through official university waste handling facilities.

**Reagent Specificity**

Further tests were performed in order to ensure the specificity of the glucose and pentose detection reagents. Serial dilutions of glucose, mannose, xylose, arabinose, rhamnose, cellobiose, sucrose, galacturonic acid, glucuronic acid, trehalose, from 5 mM downward to 0.078125 mM, were tested with each of the monosaccharide detection reagents used in this study. Pure gut homogenate was also tested in order to ensure that sugars from the hemolymph and/or gut of the termite were not releasing sugars or substances that interfered with the colorimetric output of the detection reagent. Pure glucose solubilized in 0.1 M sodium acetate assay buffer + 4mM EDTA was used as the control in the glucose-specificity detection assay and pure xylose solubilized in 0.1 M sodium acetate assay buffer + 4mM EDTA was used as the control in the pentose-specificity detection assay.

**Data Analysis**

All data collected from the glucose and pentose detection assays were analyzed using the statistical program JMP version 8 (SAS Institute Inc.; Cary, NC). The data were first tested for normality using a Shapiro-Wilk test. All experimental categories (treatment, colony, replicate) were then tested against specific enzyme activity to determine significance ($\alpha = 0.05$) using a one-way ANOVA. Specific enzyme activity was calculated as micromoles of end product per minute of incubation per gut equivalent.
(µmol/min/gut) (29). If significance was found, an all-pairs Tukey’s HSD test was used
to determine specific means that were significantly different from each other. Expected
vs. observed monosaccharide release data was analyzed using a standard t-test.

RESULTS

Reagent Specificity

The colorimetric detection reagents used in this study demonstrated specificity
towards their purposed monosaccharides. The glucose-specific detection reagent
exhibited detectable color primarily within the glucose standard curve dilutions wells, but
also slightly detected mannose at very high concentrations. The pentose-specific
detection reagent displayed detectable color mainly within the wells containing xylose
and arabinose standard curve dilutions with minor, but negligible, detection also
occurring within the galactose standard curve dilution wells (Figure 2.2, 2.3). The
absorbance values of the experimental monosaccharides within each assay consistently
fell below the absorbance values of the lowest control monosaccharide serial dilution
(0.078125 mM).

Glucose and Pentose Detection Assays

The Shapiro-Wilk test determined the data to be normally distributed (p = .1931).
The colony and replicate variables did not demonstrate significant differences (p > 0.05)
within or between colonies, allowing for the data across all biological replicates to be
pooled (Tables 2.1 - 2.4). Glucose release was significantly greater than pentose release
within all three treatments of (a) host fraction, (b) symbiont fraction, and (c) whole gut (p < 0.0001) (Table 2.1; Figure 2.4). Across all three treatments, 80% more glucose was released than pentose monosaccharides (Figure 2.4). Quantities of glucose and pentose released also varied significantly between all three treatments (p < 0.0001) (Table 2.1, 2.2; Figure 2.4). The observed release of glucose (whole gut fraction) was significantly greater than what was expected (host fraction + symbiont fraction) (p < 0.0001) (Table 2.3, Figure 2.5), while the observed release of pentose sugars was also determined to be greater than expected (Table 2.4; Figure 2.5), but significance was not demonstrated in this case (p = 0.0881).

DISCUSSION

Termite Digestion

The portrayal of the symbiotic relationship between the lower termite and its hindgut symbionts has been an ongoing investigation since it was first characterized as a parasitic association (24). However, in the early 20th century the view of the termite-symbiont relationship changed to a mutualistic partnership when studies performed by Beckwith, Cleveland, and Hungate revealed that the protozoa possess the ability to degrade cellulose and possibly aid in the digestion of termite-phagocytized food particles (2, 9, 17). This depiction of termite-protozoan symbiosis still holds true today and has been further supported with sequencing technology and proteomics (4, 6, 19, 31, 36, 41, 44, 47, 63). However, this notion of mutualism has remained quantitatively unproven up until now. The present study was the first functional study that directly examined the
existence and degree to which termite and protozoan lignocellulases collaborate in the degradation of woody biomass.

The results of this study further support the long-held theory that both the termite and the hindgut symbionts possess the enzymatic machinery necessary to access and release monosaccharides, namely glucose, from woody biomass (2, 4, 6, 17, 19, 26, 27, 31, 39, 47, 50, 53, 55, 63). However, it is interesting and important to note that the symbiont fraction (HG) released significantly more glucose than the host fraction (SG/FG/MG) (Figure 2.4; Table 2.1). This finding agrees with lower termite metatranscriptome sequencing data which concluded that a majority (66%) of the cellulose-encoding genes found within both termite and symbiont cDNA libraries were of protist origin (49). The disparity of cellulose-encoding genes between these two fractions is possibly due to an immunological tradeoff within the host fraction, since this part of the digestive tract is the first line of defense against pathogens and toxins (30, 32, 36, 49). The host (termite) digestome has been shown to contain detoxification and antioxidant genes such as cytochrome p-450, esterase, glutathione-S-transferase, glutathione peroxidase, catalase, superoxide dismutase (etc.), while the protozoan digestome is predominately comprised of genes which encode enzymes involved in the acquisition of energy and nutrients (36, 49).

The separation of functions mentioned above can also be applied to the pentose release findings between these two fractions in which the symbiont fraction released significantly more 5-carbon monosaccharides than the host fraction (Figure 2.4; Table 2.2). Pentose hemicellulases have been well-studied in lower eukaryotes such as fungi, yeast, and protozoa, but appear to be absent from the lower termite digestome (3, 46, 49).
An oversight of this study that may have changed the appearance of host hemicellulolytic contributions (and should have been included due to the hemicellulose composition of the substrate pine wood), was the omission of a mannose detection test. Due to the specificity of the detection reagents towards glucose and the 5-carbon sugars xylose and arabinose, the 6-carbon hemicellulose monomer mannose was not quantitated. This may have been an important aspect of the degradation process to investigate since the hemicellulosic composition of gymnosperms, such as pinewood, is predominately comprised of mannose, and *R. flavipes* has been shown to contain α-mannosidase encoding-genes within its digestome, which ostensibly aids in glycosylation (3, 15, 49).

As noted earlier, the termite fraction and the symbiont fraction are independently capable of releasing simple sugars from the pine wood substrate. This result was predictable in the host fraction due to the results of sequencing data which revealed many candidate lignase- and cellulase-encoding genes (49). The stereotypically accepted pathway of lignocellulose degradation starts with the depolymerization of lignin in order to free hemicellulose and cellulose (7, 44, 56, 57). This is achieved with host lignases such as laccase/phenol-oxidase, esterase and peroxidase (10, 19, 20, 42, 58). Lower termite digestome sequencing studies have yet to expose protist lignase-coding genes (49), however, past research has demonstrated that some species of bacteria and protozoa do harbor lignases capable of depolymerizing lignin (22). During the bioinformatic characterization of the digestome studies, it is possible that genes could have been misannotated, did not demonstrate homology to any lignases currently stored in sequencing databases, or simply just are not present in the symbiont digestome. Another possibility is the movement of lignases from host fraction to symbiont fraction. The
degradation of a sturdy material, such as lignocellulose, requires enzymes to be in constant contact with the substrate. Due to this requirement, many enzymes, such as lignases, will adhere to particles of substrate and degrade the bonds as the particle travels through the digestive tract (14, 20). A third possibility as to how the symbiont fraction independently releases monosaccharides, but the least empirically supported, is the theory that lignin degradation may be an unnecessary step for the accession of hemicellulose and cellulose. Some research has suggested that lignin does not need to be degraded or modified in any way in order to release simple sugars (16). However, a majority of published research indicates that lignin is a major obstacle which needs to be depolymerized to some capacity in order to access the polysaccharides (4, 6, 7, 10, 13, 19, 20, 28, 31, 36, 38, 42, 44, 51, 56, 57).

Another outcome of this research showed that higher amounts of glucose were released in both the host and symbiont fractions than pentose, indicating that glucose may be a more important aspect of the termite and symbiont diet than pentose sugars (Figure 2.4; Table 2.1, 2.2). Monosaccharides such as glucose and constituents of hemicellulose, namely mannose, arabinose, xylose, rhamnose, and galactose, are essential components of the termite diet (4). They are utilized for their carbon, hydrogen, and oxygen which are converted into energy in the form of ATP through cellular respiration (3, 4, 6, 19, 39, 41). Due to the abundance of naturally occurring cellulose, organisms that readily convert sugar into usable energy have evolved digestomes which are more specialized towards cellulose degradation than hemicellulose degradation (26, 27, 50, 54, 55). The same can be said for the lower termite R. flavipes metatranscriptome which has been
shown to contain more cellulase-encoding genes \((n = 77)\) than hemicellulase-encoding genes \((n = 45)\) (49).

The most notable finding of this research was the first functional evidence of a synergistic collaboration between native host- and symbiont-derived enzymes within a wood-feeding insect in the saccharification of a natural substrate (Figure 2.5). Previous theories of termite-symbiont digestive collaboration have been formulated based on sequencing results and single-enzyme functional studies. The results of this study indicate that not only do host and symbiont enzymes collaborate in order release glucose, they do so in a synergistic manner (1.55-fold). Separately, these two fractions have the ability to release glucose from woody biomass, but when combining host and symbiont fractions the release of glucose is greater than simply just adding the outputs of the separate fractions held independently. However, due to the nature of this study \((in \ vitr)\) it is uncertain if the termite and/or the symbionts exploit this collaboration for nutrient acquisition. The termite digestive tract is a one-way path forward from the crop to the anus (4), which suggests that both host and symbiont enzymes act upon plant particles, but most likely not at the same time. Although, as mentioned previously pertaining to the possibilities of symbiont lignase capabilities, host-derived enzymes may attach to food particles and travel with these particles into the hindgut resulting in both host and symbiont enzymes acting together on the same plant matter. Follow-up studies devising a way to test glucose release along the digestive tract \(in \ vivo\) will give more insight of the true nature of the digestive collaboration occurring within the lower termite \(R. \ flavipes\). Also, the research on glucose transporter genes in Chapter 3 was undertaken to gain
possible insights into where glucose uptake takes place, which could reveal the relative importance of host and symbiont to host nutrition.

Lastly, it is of importance to note the lack significant differences found within and between the three termite colonies used throughout this study (Tables 2.1 – 2.4). The similarity of results shared among these colonies suggests that the genetic makeup of the termite and symbiont digestome is highly functionally conserved, which reinforces the notion that lignocelluase-encoding genes within the termite metatranscriptome are greatly evolved for efficiency at degrading lignocellulosic biomass. This could potentially have significant implications within the pest control industry where broad-based applications are of the highest importance (45).

**Biofuels Implications**

A major obstacle for the large-scale production and use of alcohol fuels, specifically ethanol, is the biomass conversion inefficiency of the current production methods (7, 8, 14, 28, 31, 35, 38, 44, 47, 51, 57, 59, 63, 67). The overall goal of production is to obtain the most amount of product at the least amount of cost. This cost is associated with the time, energy, money, and materials needed to convert 2nd generation feedstocks into alcohol fuel (31). When inefficiencies persist in synthetically created systems, it is often useful to look to nature for answers. However, an on-going problem that continues to occur within this field of research is the mind-set that there is one solution, or one model organism, from which all recombinant lignocelluloses utilized in biorefineries will be derived (19). Conversely, this research indicates that recombinant enzyme cocktails used for the depolymerization and hydrolysis of lignocellulose should
be carefully selected from various organisms based on their digestive specializations. An enzyme cocktail composed of enzymes which act to release simple sugars in a synergistic fashion, as demonstrated by *R. flavipes* and its symbionts, can aid in increasing the final product (fermentable sugars) while decreasing the cost. Furthermore, through this research, termites and their symbionts have been shown to be an excellent model system for the discovery of efficient biochemistries involved in both glucose and pentose liberation from woody plants. This is a significant finding due to the fact that while cellulose is the predominantly coveted polysaccharide, hemicellulose is currently the largest proportion of waste in the lignocellulolytic process owed to the lack of degradation technologies (15). To be able to utilize every product of the degraded lignocellulose structure would greatly increase the production of fermentable sugars, lignocellulose-derived bioplastics, and aromatic hydrocarbons for addition to fossil fuels.

In conclusion, the results of this research demonstrate a need for recombinant lignocellulase cocktails that are constituted of enzymes from multiple organisms, and designate the termite and its symbionts as excellent candidates to be further investigated and utilized in the large-scale saccharification of woody biomass.
Table 2.1. Global ANOVA output from an analysis of glucose monosaccharide release across sawdust incubation treatments.

<table>
<thead>
<tr>
<th>ANOVA</th>
<th>df</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>7</td>
<td>104.9</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Treatment</td>
<td>2</td>
<td>364.95</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Colony</td>
<td>2</td>
<td>1.81</td>
<td>0.1824</td>
</tr>
<tr>
<td>Rep</td>
<td>3</td>
<td>0.25</td>
<td>0.1824</td>
</tr>
<tr>
<td>Error</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>35</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.2. Global ANOVA output from an analysis of pentose monosaccharide release across sawdust incubation treatments.

<table>
<thead>
<tr>
<th>ANOVA</th>
<th>df</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>7</td>
<td>11.68</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Treatment</td>
<td>2</td>
<td>36.34</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Colony</td>
<td>2</td>
<td>0.67</td>
<td>0.5186</td>
</tr>
<tr>
<td>Rep</td>
<td>3</td>
<td>2.58</td>
<td>0.0737</td>
</tr>
<tr>
<td>Error</td>
<td>28</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>35</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2.3. Global ANOVA output from an analysis of glucose monosaccharide release from expected (host + symbiont fraction) vs. observed (whole gut).

<table>
<thead>
<tr>
<th>ANOVA</th>
<th>df</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>6</td>
<td>15.06</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Treatment</td>
<td>1</td>
<td>86.25</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Colony</td>
<td>2</td>
<td>1.71</td>
<td>0.2111</td>
</tr>
<tr>
<td>Rep</td>
<td>3</td>
<td>0.24</td>
<td>0.8706</td>
</tr>
<tr>
<td>Error</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 2.4. Global ANOVA output from an analysis of pentose monosaccharide release from expected (host + symbiont fraction) vs. observed (whole gut).

<table>
<thead>
<tr>
<th>ANOVA</th>
<th>df</th>
<th>F</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>6</td>
<td>1.81</td>
<td>0.1566</td>
</tr>
<tr>
<td>Treatment</td>
<td>1</td>
<td>3.27</td>
<td>0.0881</td>
</tr>
<tr>
<td>Colony</td>
<td>2</td>
<td>0.56</td>
<td>0.58</td>
</tr>
<tr>
<td>Rep</td>
<td>3</td>
<td>2.16</td>
<td>0.1308</td>
</tr>
<tr>
<td>Error</td>
<td>17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 2.1. A digitally-created drawing depicting the digestive tract of *R. flavipes* situated within a termite of the worker caste. The eastern subterranean termite digestive tract stretches the entire length of the body, starting at the salivary glands and ending at the anus. The digestive tract is divided into three major regions; the foregut, the midgut, and the hindgut. The foregut comprised of the oesophagus, the salivary glands, and the crop. The foregut invaginates into the midgut, which is an elongated region that contains the malpighian tubules at the junction with the hindgut. The final major region of the digestive tract is the hindgut, or paunch which is where the eukaryotic and prokaryotic symbionts are located. The digestive tract can be further divided into a host and a symbiont fraction; respectively named for the theoretized predominant contributor of functional proteins within each fraction based on published transcriptome and metatranscriptome results.
Figure 2.2. Test of glucose detection reagent specificity using serial dilutions of eight monosaccharides. Glucose was the only tested monosaccharide strongly acted upon by the glucose detection reagent, exhibiting detectable color above the buffer blank standard at an absorbance of 505 nm.
Figure 2.3. Test of pentose detection reagent specificity using serial dilutions of eight monosaccharides. Xylose and arabinose were the only tested monosaccharides acted upon by the pentose detection reagent, exhibiting detectable color above the buffer blank standard at an absorbance of 520 nm. However, the absorbance value of the highest concentration of galactose (5 mM) fell below the absorbance value of the lowest concentration of xylose (0.078125 mM).
Figure 2.4. Monosaccharide release from the three pine wood incubation treatments; host fraction (SG, FG, MG), symbiont fraction (HG), and whole gut (SG, FG, MG, HG). Significantly more glucose was released than pentose monosaccharides ($p < 0.0001$) from each incubated sawdust treatments; host fraction $= 3.43x$, symbiont fraction $= 3.66x$, and whole gut $= 2.81x$. Glucose release across treatments varied significantly ($p < 0.0001$), with whole gut releasing the most glucose, followed by symbiont fraction and then host fraction. The same pattern across treatments was demonstrated by pentose release as well.

* Denotes significance between glucose and pentose results within treatment fractions.
Figure 2.5. Expected (host fraction + symbiont fraction) versus observed (wholegut) monosaccharide release. The observed glucose release across all biological replicates was significantly greater than the expected glucose release ($p < 0.0001$), and demonstrated 1.55x synergy. Pentose release was greater in the observed release than the expected release, exhibiting 1.2x synergy, but the disparity was not found to be significant ($p = 0.0881$).

* Denotes significance between glucose and pentose results within treatment fractions.
Figure 2.5. **Protein homogenate freeze/thaw efficacy tests.** Exoglucanase activity and \(\beta\)-glucosidase activity were tested using the model substrates p-nitrophenyl-\(\beta\)-D-cellobioside and p-nitrophenyl-\(\beta\)-D-glucopyranoside, respectively. There were no differences found in the specific enzyme activity exhibited by termite gut homogenate frozen from 0 – 10 days using either substrate \((p = 0.8187, p = 0.7256)\).


CHAPTER THREE: INVESTIGATION OF GLUCOSE ABSORPTION CAPABILITIES WITH THE DIGESTIVE TRACT OF A LOWER TERMITE

ABSTRACT

This research investigated glucose localization and transport within the lignocellulose-degrading model organism *Reticulitermes flavipes*. Putative sugar transporter validation, gene expression assays, and colorimetric monosaccharide-detecting assays were used to investigate these aims. Results suggest that the majority of glucose and glucose transport occurs in the host fraction of the termite gut (foregut and midgut), while the hindgut symbiont fraction may rely more heavily on di- and oligosaccharide transport. This indicates the possibility that the termite host and the hindgut symbionts may be solely responsible for their own acquisition of carbohydrates for metabolic energy purposes.

INTRODUCTION

The digestion of plant matter within various organisms has recently become the focus of many research investigations due to the insights it can reveal about efficient lignocellulose degradation (25, 50, 51, 72). Conclusions drawn from these studies can be
applied to the biofuels and pest control industries (64). One of the model organisms used within these studies, mainly due to their superior ability to degrade and utilize a large percentage of the lignocellulose structure, is the lower termite *Reticulitermes flavipes* (9). Still, the exact digestive process occurring along the intestinal tract of this termite is largely unknown. Hypotheses have been formulated based on lignocellulase activity and gene expression demonstrated to occur within specific sections of the digestive tract (25, 51, 58). However, major caveats of these hypotheses are that degradation is not the only aspect of digestion that is occurring and enzymes involved in degradation can act downstream of their site of secretion (25, 38). Once monosaccharides and other important nutrients are released from the lignocellulose structure they need to be transported across membranes and into cells in order to be properly utilized (21). This chapter focuses on determining the location of glucose transporters in relation to current hypotheses pertaining to lignocellulose digestion within the lower termite gut.

Photosynthesis allows plants to seize the sun’s energy and use it to transform carbon dioxide and water into carbohydrates to be incorporated into the structural support system of the cell wall, lignocellulose (24). Termites possess the unique ability to release and digest the polysaccharides from this recalcitrant substance (25, 50, 51, 72). As the previous chapter established, lower termites accomplish this with the aid of symbiotic protozoa which play a role as significant contributors to this degradation process (5, 38, 51). As degradation strategies for woody biomass are becoming more and more defined, it is becoming apparent that there is a paucity of focus and progress in the understanding and characterization of transport systems for unbound nutrients within the insect digestive system (44). A majority of the published literature pertaining to sugar
transporters has focused on the transport system within mammalian systems (4, 13, 16, 32). However, within the past decade investigations into insect sugar transporter systems have been increasing and results suggest that many parallels can be drawn between insect and mammalian two systems (12, 14, 43).

Mammalian and insect cells, like the cells of most carbon-consuming organisms, require a constant influx of monosaccharides, predominantly glucose, into their cells in order to produce the energy necessary to maintain functional homeostasis and metabolism (11, 13, 47). Glucose transporters are found in every cell of most eukaryotes (6). A majority of cells within both mammalian and insect systems are not able to produce their own free glucose through the glucose-generating process of gluconeogenesis due to a lack of the specific enzyme necessary to accomplish this task, glucose-6-phosphatase (5, 28). The basic model of sugar transport within these two systems describes the transfer of sugar molecules through cell membranes, both into and out of the cell, by means of passive and active transport (13, 28, 32, 39, 43, 66, 69, 71). Hexose diffusion (i.e., glucose diffusion) across membranes is known to primarily occur through facilitated diffusion, or passive transport (10, 12, 13, 19, 28, 32). Passive transport requires protein transporters and moves down concentration gradients into or out of the cell, while active transport moves molecules against concentration gradients using cellular energy either in the form of ATP or as a result of differences in electrochemical potentials (32, 39). Interestingly, the passive component of glucose transport was initially rejected when it was first hypothesized. The “solvent drag”, or “paracellular flow” theory, proposed that the passive aspect of glucose absorption was merely an artifact of nutrients being carried or “dragged” through intracellular spaces due
to high rates of water absorption caused by the disparity of glucose concentrations (≥100 mM) on one side of the membrane (41). However, this theory was not upheld due to additional evidence demonstrating that active transporters saturate at 30-50 mM glucose concentrations. These studies revealed that sodium-dependent (active) cotransporters became saturated in the presence of 30-50 mM of glucose, yet absorption of glucose continued in a linear fashion at concentrations of ≥50 mM (20, 33). These findings provided evidence that other mechanisms of glucose transport were at work. Other research in the mammalian system demonstrated that glucose is linearly absorbed at concentrations as high as ~300 mM (28). These investigations revealed that two possible transport components were at work; one that lacked the transport of water molecules but was more constant, and one that contrasported water molecules but was inconsistent at times (13, 32, 69). The first component was shown to be the main transporter of glucose at low concentrations (1-50 mM) while the second component was as much as 5x more active in the presence of high glucose concentrations (≥50 mM) (28, 32, 68). These two components were later determined to be active and passive transport, respectively. The accepted theory in present-day literature is that active transport is continuously occurring (39). When glucose concentrations become too high and saturate this process, facilitated transporters (GLUT2 and GLUT2-like) are recruited to the membrane (12, 13, 32). However, the solvent drag theory has not been completely dismissed. Paracellular flow has been a proven mode of transport for amino acids and may transport hexose molecules at very low rates in conjunction with passive transport in the presence of very high glucose concentrations (≥200 mM), but the amounts transported by this process are not of any nutritional significance (41).
The mammalian system has been the most thoroughly studied transport system to date. A large reason for this is due to the theory that sugar transporter dysregulation may trigger some sugar-based diseases such as monogenic diabetes (3). Within the mammalian system, the popular model of glucose absorption, as previously stated, includes the utilization of both active and passive modes of sugar transport. The mammalian hexose transporters belong to one of two super families; the sodium-solute symporter family (SSSF), or the major facilitative superfamily (MFS) (26, 28). The MFS (Pfam CL0015, IPR016196) was first thought to contain only contain members that functioned as sugar transporters, however, this family has expanded to now include drug efflux systems, citric acid cycle metabolites, hydrogen symport permeases, and phosphate exchangers (40). Evidence suggests this family is extensive and has many members still undiscovered. A majority of the glucose transport within the mammalian system occurs by means of the facilitated D-glucose transporter proteins (4, 13, 28, 32). The gene family SLCA2 encodes the major family of protein sugar transporters known as the GLUT family (43, 61). Protein members of the GLUT family are uniporters that possess the motif-identifying 12 transmembrane domain (TM) segments (helices) connected by loops with 8 segments acting amphipathically and 4 segments demonstrating hydrophobicity (4, 36, 37, 42, 53). These properties help sugar substrates navigate the lipid bilayer. The N-terminal and C-terminal regions of these proteins face the cytosolic side of the membrane (37, 61). The GLUT family includes fourteen members, GLUT1-12, 14, and HMIT, but only seven have been well-documented (4, 12, 23, 28, 36, 40). They are further split into classes with GLUT1-4 composing class 1, GLUT 5, 7, 9, and 11 composing class 2, and GLUT6, 8, 1, 12, HMIT constituting class
3 (40). GLUT1-4, 8, and 10 have been established as glucose transporters whereas GLUT5 has been characterized as a fructose transporter (4, 32). GLUT1 transporter has also have been shown to play a role in nutrient (glucose)-sensing (20, 23, 35, 36, 37, 42, 57). However, substrates cannot be absolutely determined from protein sequences for these transporters because structure-function associations in the proteins are not characterized, but they can be fit to the transmembrane topological model to ensure they belong in the MFS clan (36, 40). Moreover, GLUT transporters do not show absolute specificity towards a single substrate, with galactose, glucosamine, and 2-deoxyglucose also determined as substrates (7, 28, 32). These proteins are highly related but come from distinct genes, are pH-independent, have different subcellular localizations, and are very tissue specific (32, 40, 48, 68). The mechanistic action of the GLUT proteins employs a simple conformational change. The protein reveals a binding site towards either the outside or inside of the cell. The binding of the sugar substrate to the revealed site causes conformational changes in the transport protein allowing the sugar molecule to move to the opposite side of where attachment occurs (4, 23, 32, 33, 57). The active glucose transporters (SGLT1-3) are Na+-dependent symport cotransporters that use the energy associated with ion pumps to move sugar molecules across membranes (39, 64). SGLT transporters are found in the membranes of the small intestines and kidney, and an isoform was found in the central nervous system (16). The SGLT transporters have been shown to also carry molecules of water across the membrane with the sugar substrate. In the mammalian system, net fluid absorption was found to be proportional to the rate of solute transport (glucose and NaCl) (71). Fluid could be absorbed even against osmotic pressures in the presence of glucose. However, some suggest that the water is transported
through osmosis due to the higher sugar concentrations and may not be transported across with the sugars (69). The SGLT transporters move monosaccharides into the epithelium and across the basolateral membrane while the facilitative sugar transporter (GLUT2) assists in the transport of glucose into the epithelium, through the basolateral membrane and into erythrocytes (32, 33, 66, 67, 69). The fructose transporter (GLUT5) is also active in the BLM but at rates much less that of GLUT2 (27, 47). Although these transporters were first found and characterized within the mammalian system, they have since been discovered in other organisms as well. GLUT5-like transporters have been found in birds, SGLT1-like transporters have been found in snails and crustaceans, and GLUT2-like transporters have been found in tapeworms and various insects (4, 7, 11, 12, 14, 16, 17, 19, 43, 48, 55, 60).

It has been well established that glucose absorption is typically completed through passive transport within insects (11, 12, 14, 17, 19, 21, 43, 44). However, sugar absorption in insects was first thought to be a purely diffusional process, taking advantage of low hemolymph concentrations of glucose due to rapid conversion into trehalose (8, 47). Insects contain very little glucose and fructose in their hemolymph due to their conversion into trehalose (5, 8, 29, 34). The sugar gradients caused by this lack of monosaccharides in the hemolymph leads to the simple facilitative diffusion of glucose out of the gut (34). However, in flying insects, like the honeybee, the concentrations of usable energy (glucose and fructose) are much higher and the storage form of energy (trehalose) is much lower due to their need for constant energy fuel during flight (8, 17, 29). Trehalose is the preferred form of stored energy within the insect system (5). It is a non-reducing disaccharide of two α-glucose molecules bound together through an α-
glucosidic bond (8). The synthesis of this sugar is triggered by neuropeptides and
generally occurs in the fat body of the insect where trehalose-6-phosphate synthase is
active (47). Trehalose can act as an antioxidant protecting against extreme temperatures,
varying oxygen gradients, and desiccation (34). The chironomid, Polypedilum
vanderplanki, uses trehalose as a desiccation protectant during periods of anhydrobiosis
(8). This chironomid is the only organism known to possess facilitated trehalose
transporters (29). Trehalose transporters have also been found in archae, bacteria, and
yeast, but these are only the active transporter type which needs cellular energy to
transfer the disaccharide (49, 67). Glucose is a product of hydrolyzed trehalose and is
present in much lower concentrations than trehalose in the hemolymph, but it is the more
regulated sugar of the two (8, 43, 47). Despite the initial theory that insects lacked
facilitated sugar transporters, recent research of many different insect systems has
revealed a passive process to exist and closely mirror the mammalian sugar transport
system (5, 6, 7, 12, 14).

Glucose is utilized within the insect system for energy purposes, growth and
development, and reproduction (12, 14, 43, 60). The only functional analysis of an insect
transporter has been performed in the brown rice planthopper, Nilaparvata lugens (43).
The investigators sequenced and characterized the function of a facilitated GLUT-like
protein which was determined to be a fructose transporter (GLUT5-like protein). The pea
aphid, Acyrthosiphon pisum, has other methods for regulating sugar physiology when
transport alone is not sufficient. The plant phloem on which the pea aphid feeds produces
sap with very high concentrations of sucrose (~1M) (5). Homopteran insects feed on this
sap, but face the hazards of desiccation due to the osmotic pressure caused by extreme
sugar concentrations within the gut (44, 47). Transglycosidase has been proposed as a way to osmoregulate the gut of these insects (27). Converting single glucose monosaccharides into oligosaccharides decreases the osmotic pressure per unit of sugar (34). Sucrose is hydrolyzed into the monosaccharides glucose and fructose by $\alpha$-glucosidase (44). Fructose and a small amount of glucose are transported across the membrane into the hemolymph, while the remaining glucose in the gut is converted into an oligosaccharide by transglucosidase. Another insect which has received attention due to its carbohydrate diet is the aphid parasitoid, *Aphidius ervi* (11, 12, 21). The larvae and adults of this parasitoid absorb sugars and amino acids through the epidermis of their midguts with the aid of sugar transporters (11, 21). Protein sequence analysis has revealed that *A. ervi* contains a GLUT2-like sugar transporter in the midgut apical and basolateral cell membranes (12). An active transporter sequence (SGLT1-like) and fructose transporter sequence (GLUT5-like) were also found in this insect. This transport system was shown to allow for the absorption of glucose and fructose, but not galactose (11). These characteristics closely reflect those of the mammalian transport system. Other insect transport systems have also shown homology to the mammalian transport system. Genomic sequencing and bioinformatics efforts within *Drosophila melanogaster* have revealed protein sequence motifs of sugar transporters highly similar to the mammalian facilitated glucose transporters (GLUTs) (6, 19, 33, 66). The red imported fire ant has also been the subject of extensive nutritional transport studies since understanding the process of energy utilization could reveal novel control technologies. A full length GLUT-like gene was sequenced from red imported fire ant cDNA (14). Sugar transporter localization studies have also revealed information regarding the
digestive process along the gut of insects. Within the cockroach and honey bee, the lack of sugar transporters located in the crop suggests that glucose absorption occurs mainly in the midgut (17, 60). Further, research demonstrated that little to no absorption occurs in the crop, and food transport from crop to midgut is regulated by midgut glucose concentration.

Tapeworms, which are mammalian parasites, have been shown to absorb glucose through facilitated diffusion with 2 mammalian-like (GLUT-like) glucose transporters (48). The absorption of galactose into the tissue of snails has been characterized as a sodium-dependent (SGLT-like) active transport process (7). Three distinct sugar transporters were found in *Leishmania mexicana* that are members of the major facilitator superfamily, and the yeast, *Saccharomyces cerevisiae*, has also been shown to transport glucose with the aid of facilitated GLUT-like proteins (10, 49, 68). However, there is still a significant lack of discovery, characterization, and understanding of both eukaryotic and prokaryotic sugar transport systems. One such system that merits an in-depth investigation is the carbohydrate transport system of the lower termite *R. flavipes*.

The objective of this study was to investigate sites of epithelial glucose absorption within the lower termite digestive tract. This research was conducted in order to gain insights into locations along the digestive tract where degradation may occur, and also to further characterize the type of lignocellulolytic relationship that exists between the termite and its symbionts. Gene expression analysis, monosaccharide detection assays, and optical density detection assays were implemented to investigate glucose transport localization thereby inferring potential sites of glucose absorption. Results suggest that the main site of glucose absorption within the entire digestive tract is in the host fraction
(foregut and midgut). This in turn suggests that the host termite itself may meet its own caloric energy needs through its own endogenous digestive processes. Termite- and symbiont-derived enzymes may demonstrate synergy when co-expressed, but the fraction-specific degradation processes may occur independently of each other in vivo. Also, these results imply that di- and oligosaccharides reaching the hindgut may be completely allocated for microbial symbionts and rapidly converted into the metabolic intermediate acetate.

MATERIALS AND METHODS

Termites

*Reticulitermes flavipes* workers were used throughout this study. The identification of species and caste was verified using procedures mentioned in Chapter 2 (56). One *R. flavipes* laboratory colony, Whistler 1 (W1), collected from Purdue University campus (Tippecanoe County, IN, USA), was used throughout this study. This colony was held in a plastic container on a diet of pine wood (Nelson Wood Shims; Cohasset, MN) and moist paper towels. This container was held in an incubation chamber with a constant temperature of 24°C, ~70% R.H., and a 0:24hr light:dark ratio. Three separate biological replicates of 100 termite workers were removed from the lab colony and used in this research.
Chemicals and Reagents

Glucose detection reagent: Autokit Glucose (Wako Chemical; Richmond, VA). Nucleic acid dye: SYBR green (Bio-Rad; Hercules, CA). Glucose dye: allura red (Sigma-Aldrich; St. Louis, MO).

Buffers and Solutions

Homogenization buffer: 1x phosphate buffered saline, pH 7.0 (Bio-Rad; Hercules, CA).

Sugar serial dilution: D-glucose powder (Sigma-Aldrich; St. Louis, MO) dissolved in homogenization buffer. Carboxymethyl cellulose solution: 2% (w/v) carboxymethyl cellulose (CMC) (Sigma-Aldrich; St. Louis, MO) dissolved in Nanopure water. DNSA stop solution: 1% 3,5-dinitrosalicylic acid (DNSA) (Eastman Chemical Company; Kingsport, TN), 0.2% crystalline phenol (Sigma-Aldrich; St. Louis, MO), 0.05% sodium sulfide (Sigma-Aldrich; St. Louis, MO), 30% sodium potassium tartrate (Fischer Scientific; Waltham, MA) and 0.4 M sodium hydroxide (Mallinckrodt Chemicals; St. Louis, MO).

Selection of Glucose Transporter Gene Contigs

All contigs annotated as sugar or carbohydrate transporters, from a previously sequenced R. flavipes gut cDNA library, underwent screening using various database searches to ensure the annotations as sugar transporters were correct, and to determine if the candidates contained the hexose-specific transporter signatures. To ensure that the nucleotide sequences were annotated correctly, searches for homologs were first performed using the basic local alignment search tool to search for nucleotide matches
The cutoff for determined homology was an E-value of $<1e^{-10}$. Candidate sugar transporters that passed the first screening underwent three further screenings to ensure they contained the proper glucose transporter motifs. The contigs were first translated by the online software ExPASy (Swiss Institute of Bioinformatics; Lausanne, Switzerland), and the best open reading frames were chosen. The online database InterPro (EMBL; Hinxton, Cambridge) (70) was then used to determine if the candidate(s) contained the Major Facilitator Superfamily (MFS) signature (IPR016196) (40). Secondly, InterPro was used to determine if the candidate(s) contained the General Substrate Transporter (GST) signature (IPR005828) (43). The final informatic analysis used InterPro to determine if the candidate(s) contained the hexose transporter signature (IPR003663) (43, 61). BLASTp (protein BLAST) was then used to confirm these findings, validating the InterPro results using both BLASTp signatures and homologous database matches. To ensure the candidate(s) were membrane proteins, they were checked for hydrophobic regions using a Kyte-Doolittle hydrophilicity plot in the program Protean within the bioinformatics software package DNASTAR (Madison, Wisconsin). The candidate(s) that passed all of these screening tests were advanced to gut section gene expression assays (see below). Quantitative real-time PCR primers were designed for the candidate sugar transporter(s) using the software Primer3 (Whitehead Institute for Biomedical Research) with specific parameters such as a melting temperature of $60 \pm 1$ °C, a product range size of 150-250 base pairs, and a primer size of 18-23 base pairs (51). Before use, a melt gradient was performed in order to determine the ideal melting and annealing temperature of each primer set.
**Digestive Tract Dissections**

Intact digestive tracts were removed from termites using previously established procedures in Chapter 2. The guts were further dissected into the sections (i) foregut + salivary gland, (ii) midgut, (iii) hindgut rinsed of all microbial symbionts, and (iv) microbial symbionts. Each section was placed immediately in section-specific 1.5 mL centrifuge tubes (Eppendorf; Hamburg, Germany) with PBS homogenization buffer at a volume of 5 µL per section. The gut sections were homogenized using 1.5 mL Pellet Pestles (Kimble-Kontes; Vineland, NJ) and centrifuged at 14,600 rcf and 4°C for 15 minutes. The supernatants were collected and stored at -80°C until needed. This dissection method was used for both gene expression and glucose feeding assays.

**Sugar Transporter Gene Expression**

The digestive tracts of 50 workers were dissected and sectioned, and protein homogenate was collected using the methods stated earlier. RNA was isolated from the foregut + salivary gland, midgut, hindgut, and symbiont homogenates using the SV Total RNA Isolation Kit (Promega; Madison, WI). Equal concentrations of RNA across all sections (empirically determined using a NanoDrop spectrophotometer), having been normalized to the lowest RNA concentration, were then reverse transcribed into cDNA using an iScript cDNA synthesis kit (Bio-Rad; Hercules, CA). Quantitative real-time PCR gene expression assays were carried out using 20 µL reactions. SYBR green was used as the detection reagent and NADH-dh (dehydrogenase; Genbank Accession No. BQ788185) was used as the reference gene throughout this study. The resulting Ct values were first normalized to the reference gene (NADH-dh) in order to obtain ΔCt.
values. These values were then further normalized to the lowest expressed gut section value to obtain ΔΔCt values. The ΔΔCt values were then transformed to $2^{-\Delta\Delta Ct}$ in order to determine relative fold expression differences among gut regions.

**Glucose Feeding Assay**

Biological replicates of 25 worker termites were randomly removed from the laboratory colony W1. Each sample of 25 termites was placed in an individual 35 x 10 mm suspension dish (Thermo Scientific; Middletown, VA) with either 50 mg of D-glucose powder (Sigma-Aldrich; St. Louis, Missouri) or 50 mg of pine wood (Nelson Wood Shims; Cohasset, MN) as a control. Within each dish, the food source was placed at one edge of the dish and 100 µL of water was placed at the opposite edge. The Petri dishes were then set in an incubation chamber with environmental settings matching that of the colony incubation chamber. The termites were left to feed on the glucose or pine wood for five days before being removed for gut region dissections as described above.

**Glucose Detection Assay**

Gut region homogenates from both glucose- and wood-fed termites were pipetted in triplicates at a volume of 50 µL per microplate well. A standard curve of serially diluted glucose was generated with a maximum concentration of 5 mM downward to 0.078125 mM, with a buffer blank as the final dilution. The standard curve was produced in 0.1 M sodium acetate assay buffer + 4mM EDTA. A glucose detection reagent (Wako Chemical; Richmond, VA) was then added to each well containing glucose dilutions or aliquots of treatment at a volume of 200 µL of glucose detection reagent per well (50).
The microplate was then left to shake for 5 minutes before absorbance was read as an endpoint at a wavelength of 505 nm.

**Reduced Sugar Detection Assay**

A solution of 2% (w/v) carboxymethyl cellulose (Sigma-Aldrich; St. Louis, MO) was made in Nanopure water. A standard curve was then generated (100 µL per well) by serial diluting D-glucose (Sigma-Aldrich; St. Louis, MO) in Nanopure water from a maximum concentration of 5 mM downwards to 0.078125 mM, including a buffer blank as the 8th dilution. Section-specific protein homogenate from glucose-fed and wood-fed termites was added to empty wells (5 µL per well). The 2% CMC solution (95 µL) was then added to any well containing protein homogenate, which resulted in a total volume of 100 µL per protein homogenate well. The plate was allowed to incubate at room temperature for 1 hour. After 1 hour, 100 µL of 1% DNSA stop solution (1% 3,5-dinitrosalicylic acid, 0.2% crystalline phenol, 0.05% sodium sulfide, 30% sodium potassium tartrate and 0.4 M sodium hydroxide) was added to wells containing both standards and protein, and was then immediately placed in a boiling water bath for 10 minutes. Following the heating phase, the plate was placed on ice for 15 minutes in order to cool. The plate was then read as an endpoint at a wavelength of 540 nm. The standard curve of glucose was used to deduce the amount of reduced sugar in each well containing gut homogenate.
Optical Density Detection Test

A total of 50 mg of D-glucose was placed in 5 mL liquid red dye (allura red) in a 15 mL Nalgene Falcon tube for 24-hr to allow adsorption to occur (65). The dye liquid was then poured off and the remaining dyed glucose granules were placed on a small foil-covered tray. The tray was placed in a drying chamber and allowed to dry at 50 °C for 24 hours. The dyed glucose was then used in a glucose feeding assay following the same procedures previously described. Liquid red dye was also given to termites as a control. The assay was allowed to run for five days before the termites were dissected. The dissections and homogenate preparations also followed the aforementioned techniques. The homogenates were directly pipetted into microplate wells in 50 µL volumes per well and read as an end point at 505 nm.

Data Analysis

GENE EXPRESSION

The resulting Ct values were first normalized to the reference gene (NADH-dh) in order to obtain ΔCt values. These values were then analyzed using a one-way ANOVA with the statistical program JMP version 9 (SAS Institute Inc.; Cary, NC) to determine statistical differences in gene expression across termite digestive tract fraction (foregut, midgut, hindgut, symbiont cells), biological replicate, and technical replicate. The ΔCt values were then further normalized to the lowest expressed gut section value to obtain ΔΔCt values. The ΔCt values were then transformed to $2^{-\Delta\Delta C_T}$ in order to determine relative expression.
GLUCOSE AND REDUCED SUGAR DETECTION ASSAYS

Glucose and oligosaccharide concentrations were calculated as micromoles of end product per gut equivalent (µmol/gut). The data collected were then analyzed using JMP9. The data were first tested for normality using a Shapiro-Wilk test. The experimental categories (digestive tract fraction, biological replicate, technical replicate) were then tested against sugar concentration to determine significance ($\alpha = 0.05$) using a one-way ANOVA. If significance was found for the whole-model ANOVA ($p<0.05$), an all-pairs Tukey’s HSD test was used for separation of means.

OPTICAL DENSITY DETECTION TEST

The OD values (end point absorbance) detected at 505nm underwent the same analysis as the previous tests. JMP9 was used to perform all statistical analyses. A Shapiro-Wilk test was used to test the data for normality. The previously established experimental categories were then tested against the OD value output to determine significance using a one-way ANOVA. An all-pairs Tukey’s HSD test was then used to determine mean separation significance.

RESULTS

Contig Selection

Four contigs annotated as sugar transporters were found in the termite host library digestome database; ST228, ST256, ST388, and ST501. The theoretical sugar
transporters ST501, ST228, and ST388 were determined to be members of the major facilitator superfamily (MFS) (IPR016196) (Figure 3.1, 3.2, 3.3). The hypothetical sugar transporter ST256 did not show congruency with the motifs necessary to fit within the MFS domain, and was therefore excluded from any future validation steps. The sugar transporters ST501 and ST228 contain the General substrate transporter (GST) signature (IPR005828) (Figure 3.1, 3.2). The sugar transporter ST388 did not contain this signature and was therefore removed from the potential glucose transporter candidates. The sugar transporter ST501 contained the hexose transporter signature (IPR003663) (Figure 3.1). ST228 did not contain this signature and thus it was excluded from any future experiments. The BLASTp searches used to validate the InterPro signatures found in ST501 demonstrated congruency with the InterPro findings (Figure 3.5). BLASTp also revealed that ST501 demonstrated high degrees of homology to sugar transporters found in other insects with carbohydrate-rich diets such mosquitoes, butterflies, and plant hoppers (Table 3.2). The final informatics validation, a Kyte-Doolittle hydrophilicity plot, revealed many hydrophobic regions of the ST501 protein, strongly suggesting that it is a membrane-bound protein.

**Sugar Transporter Gene Expression**

The host fraction (foregut and midgut) of the *R. flavipes* termite digestive tract demonstrated 6.83-fold higher ST501 expression (P<0.0001) than did the symbiont fraction (hindgut tissue and symbiont cells). The foregut showed 1.13-fold higher expression than the midgut fraction (p<0.0007). No difference was found between biological replicates (p=0.7593).
**Glucose Detection Assay**

The foregut and midgut homogenates contained higher *in vivo* concentrations of glucose (p<0.0001) than the hindgut and symbiont fractions, in which glucose levels were below limits of detection (Figure 3.2). The foregut and midgut glucose concentrations did not differ (p=0.7873). Glucose concentrations were negligible (≤ buffer blank) in the pine wood-fed replicates. No difference was found between biological replicates (p=0.8758).

**Reduced Sugar Detection Assay**

Foregut and midgut homogenates contained higher (p<0.0001) concentrations of reduced cellulose residues than the hindgut and symbiont homogenates (Figure 3.3). Reduced sugar concentrations were negligible (≤ buffer blank) in the pine wood-fed termite replicates. No difference was found between biological replicates (p=0.8921).

**Optical Density Detection Test After Feeding on Dyed Glucose**

Spectrophotometer absorbance readings (OD) were highest in the foregut and midgut fractions (host fraction) (Figure 3.4). Color detection was negligible (≤ buffer blank) in the termites fed only dye. No difference was found between biological replicates (p=0.6435).

**DISCUSSION**

This study considered four candidate sugar transporter genes from a termite. Of these four genes, the ST501 gene appears to encode a member of the Major Facilitator Superfamily of proteins (Figure 3.1). This is not surprising since ST501 also contains the
hexose transporter signature, and hexose transporters are predominantly classified as passive transport proteins (28). The expression of this transporter occurs mainly in the host fraction of the digestive tract (foregut and midgut) (Figure 3.7). Together, this evidence suggests that ST501 is a termite-derived sugar transporter. The glucose detection assay, reduced sugar detection assay, and optical density detection assay results all agree with gene expression findings in that the release and absorption of glucose seems to take place in the foregut and midgut of *R. flavipes* (Figure 3.8, 3.9, 3.10). The midgut expression results agree with traditional histological conclusions from other insect studies, which have found that the midgut is the primary site of sugar absorption (5, 17, 43, 60).

Surprisingly, the gene expression results also point towards the foregut as a site of glucose absorption, which has not been demonstrated in previous sugar transporter research in other insect systems (Figure 3.7) (17). However, the sugar transporters in this section of the digestive tract could be performing a well-recognized function in addition to sugar transport, which is the glycosylation of hazardous waste or xenobiotics. This step is generally performed to make xenobiotics and other harmful molecules more soluble and therefore more easily removed from the system (6, 22). Due to the termites’ apparent arsenal of ligninases, and potentially their ability to degrade lignin in the foregut, the removal of the hazardous phenolic byproducts of this process would be necessary (58). Thus, conjugative detoxification functions potentially explain the presence of glucose accumulation (and glucose transporter expression) in the foregut (Figure 3.8). Two other explanations exist to explain the occurrence of glucose in the foregut; 1) the lower termite possesses ligninases and cellulases in the foregut that are
able to degrade and release glucose monomers from the lignocellulose structure (51, 58),
2) termites perform proctodeal trophallaxis as a way to replenish symbiont communities
after molting, but may also acquire previously degraded and defecated simple sugars in the
process (39).

A study performed to investigate glucose presence in the foregut (crop) of the
honey bee *Apis mellifera* found that glucose monomers were present in the crop of this
social insect as well (17). However, another finding from this study determined that
glucose transmembrane transport did not occur in this region. In order to investigate this
possibility, radiolabeled glucose feeding and time-course dissections could be utilized to
determine how much glucose is leaving the foregut before entering the midgut.

The lack of sugar transporter gene expression, sugar transport activity and
detectable glucose in the hindgut tissue was expected since this region is generally, from
a histological view point, considered to contain gut microbes and their metabolic by-
products, and obstruct them from reaching the termite hemolymph (Figure 3.7, 3.8) (9).
This region generally lacks any absorptive capabilities in other insect systems as well (38,
43, 47, 60). However, the detection of reduced cellulose sugar in the hindgut (Figure
3.9) was unanticipated based on glucose staining results (Figure 3.8). This is most likely
due to sugar residues not being washed away during tissue rinses. Alternatively, the lack
of sugar transporter expression in the symbiont cell fraction was not surprising as all
contigs selected for this study were from a termite host cDNA library (58); therefore,
ST501 should have been primarily expressed in the host fraction. No contigs from a
previously sequenced symbiont library (58) demonstrated homology to any known
eukaryotic sugar transporters; however, transporter sequences were present in a newer
454 pyrosequencing database containing mixed host and symbiont sequences (54) that was probed in silico for annotated sugar transporters. The genome of R. flavipes has not been sequenced. The current databases contain the termite’s “digestome”, or rather the set of genes (or partial genes) that demonstrated quantitatively significant differences between termites fed on diverse substrates; allowing for the determination of genes more likely to be responsive, and thus responsible, for a certain aspect of the digestive process (46, 54).

The lack of symbiont transporters found in our various sequence databases suggests that they demonstrate similar responsiveness across a diet of varying compositions. Although, the lack of symbiont sugar transporters containing the MFS domain is not uncommon. A study that investigated the complete genome of the parasitic protozoan Entamoeba histolytica, a protist known to transport glucose across membranes, did not find any homologs to known eukaryotic glucose (hexose) transporters in the protist’s genome (2). However, sugar transporters of a different family, the glycoside–pentoside–hexuronide:cation symporter family (GPH), primarily responsible for the transport of di- and oligosaccharides, were present in the genome instead (2). Transporters belonging to the GHP family are characterized as active transporters due to their dependence on cation support, such as Na⁺, for transporting sugars across membranes; although some researchers still argue that they should belong to the MFS due to their high degree of sequence homology to members of the MFS domain (61). The lack of a true glucose transporter but the possible presence of a di- oligosaccharide transporter is a potential explanation as to why glucose was not detected in the symbiont cell fluid while reduced sugars were present. This result suggests that the termite host
(foregut and midgut) possesses the capability to release cellulose from the lignocellulose structure and further degrade it into the monosaccharide constituent glucose for the termite’s own energy needs. The remaining undigested oligosaccharides are possibly moved into the hindgut, taken up by the symbionts through the utilization of di- and oligosaccharide transporters and/or endocytosis, and directly converted into the metabolic end product acetate, resulting in consistently low glucose concentrations in the hindgut. However, BLAST searches with possible GHP sequences did not reveal any significant matches to any termite/symbiont annotated genes in the NCBI database. But again, our lab’s contributed sequences to this database are from the gut digestome, and not the complete genome, so these genes just may have not be sequenced yet. Thus, the results of this chapter not only answer digestive process-related questions, but also have further implications in the biofuels and pest control industries in regard to termite digestome mining.

The previous chapter, Chapter 2, demonstrated that both the termite host and the protist/bacterial symbionts play a role in the digestion of plant matter within the termite digestive tract. However, questions still remained pertaining to (a) the specific role of each fraction and (b) the degree of dependence by each fraction on the other. This chapter addresses both questions. As mentioned above, it appears that the termite host itself has the ability to release and transport glucose without the aid of the symbiont fraction. It is still unclear whether or not the symbiont fraction needs any aspect of the upstream host processing to occur in order to gain access to the di- and oligosaccharides arriving in the hindgut, but evidence suggests that modifications to the lignocellulose structure are made prior to the food particles reaching the hindgut (9, 69). This finding
suggests that from a digestive aspect, the termite itself may act independently of the symbionts in order to obtain energy (in the form of glucose) necessary to maintain its functional homeostasis, while the symbionts may rely on the host termite to some degree. This implies that the termite itself is the key to the “pretreatment”, or lignin disassociation step, necessary to gain access to the polysaccharide component of lignocellulose.

The pretreatment step is considered one of the primary bottlenecks for the development of biorefinery technology due to the expense of this step compared to the expense of the entire process (~70%) (30). This points towards the host termite as the most likely source for potential pretreatment enzymes (50). Also, the lack of glucose and glucose transport in the symbiont cells, and conversely the higher levels of both in the host fraction, provides additional evidence that the host termite should be a viable source of source carbohydrate hydrolyzing enzymes when searching for and selecting enzymes for the saccharification of biomass feedstocks. Even though past sequencing studies have found glycosyl hydrolase family 7 (GHF7) present in the symbiont digestome (58), it is still unclear if these cellulases confer true exoglucanase activity towards substrates in the digestive process. In terms of pest control, it appears that any chemical or molecular control methods (i.e., cellulase inhibitors or dsRNAs) should be geared towards termite-specific physiology and the host digestome. If the termite host fraction truly acts independently of symbionts in the obtaining of nutrients, targeting the symbionts may prove to be a futile endeavor.

This chapter lays a foundation for future research investigating specific termite host and symbiont digestive physiology and degradation contributions. In order to truly
understand the specific contributions of the host and symbiont fractions to the digestive process, further investigations comparing enzymatic activity and gene expression across a range of substrates will need to be performed. Gene silencing studies employing RNA interference (e.g., 51) may also be particularly informative in future characterization studies.
Table 3.1. Homologous matches for the ST501 sugar transporter in the NCBI data base using BLASTp (non-redundant and position-specific iterated) as the search platform.

<table>
<thead>
<tr>
<th>Description</th>
<th>Query cover</th>
<th>E value</th>
<th>Ident</th>
<th>Accession</th>
</tr>
</thead>
<tbody>
<tr>
<td>sugar transporter [Aedes aegypti]</td>
<td>98%</td>
<td>7.00E-58</td>
<td>46%</td>
<td>XP_001652873.1</td>
</tr>
<tr>
<td>sugar transporter [Culex quinquefasciatus]</td>
<td>99%</td>
<td>3.00E-57</td>
<td>46%</td>
<td>XP_001843966.1</td>
</tr>
<tr>
<td>putative sugar transporter [Danaus plexippus]</td>
<td>98%</td>
<td>1.00E-55</td>
<td>46%</td>
<td>EHJ75059.1</td>
</tr>
<tr>
<td>sugar transporter 12 [Nilaparvata lugens]</td>
<td>93%</td>
<td>2.00E-50</td>
<td>45%</td>
<td>BAI83426.1</td>
</tr>
<tr>
<td>sugar transporter protein 3 [Bombyx mori]</td>
<td>99%</td>
<td>4.00E-49</td>
<td>41%</td>
<td>NP_001182631.1</td>
</tr>
<tr>
<td>glucose transporter 8 [Solenopsis invicta]</td>
<td>100%</td>
<td>9.00E-40</td>
<td>38%</td>
<td>AAX92638.1</td>
</tr>
</tbody>
</table>
Table 3.2. Sugar transporter ST501 and reference gene primer sequences which were derived using the software Primer3.

<table>
<thead>
<tr>
<th>Contig</th>
<th>Primer Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST501 F</td>
<td>ACGACCGCTCTCTGTCACCTT</td>
</tr>
<tr>
<td>ST501 R</td>
<td>GTCACACACTGCGCAGAAC</td>
</tr>
<tr>
<td>NADH-dh F</td>
<td>GCTGGGGGTITATTCTCTCTA</td>
</tr>
<tr>
<td>NADH-dh R</td>
<td>GGCATACCACAAAGAGCAAAA</td>
</tr>
</tbody>
</table>

Figure 3.1. InterPro output analysis of the selected ST501 best open reading frame.

The putative sugar transporter ST501 shows sequence signatures consistent with the Major Facilitator Superfamily, General substrate transporter, and hexose transporter proteins.
Figure 3.2. InterPro output analysis of the selected ST228 best open reading frame.

The putative sugar transporter ST228 shows sequence signatures consistent with the Major Facilitator Superfamily and General substrate transporter protein.

Figure 3.3. InterPro output analysis of the selected ST388 best open reading frame.

The putative sugar transporter ST388 shows sequence signatures consistent with the Major Facilitator Superfamily.
**Figure 3.4.** InterPro output analysis of the selected ST256 best open reading frame.

The putative sugar transporter ST256 shows no sequence signatures consistent with traditional sugar transporters.

**Figure 3.5.** BLASTp output analysis of the selected ST501 best open reading frame.

The putative sugar transporter ST501 shows BLASTp sequence signatures consistent with the Major Facilitator Superfamily and General substrate transporter protein.
Figure 3.6. Kyte-Doolittle hydrophilicity plot of ST501 using the bioinformatics program Protean. The putative sugar transporter ST501 shows many hydrophobic regions (i.e., amino acids having negative hydrophilicity values), which characterize it as a likely membrane-bound protein, a characteristic common among sugar transporters. Numbers across the top indicate amino acid positions.
Figure 3.7. Normalized gene expression of the putative glucose transporter ST501 across foregut, midgut and hindgut tissue, and symbiont cells. The host fraction (foregut and midgut) demonstrated higher expression of the putative glucose transporter ST501 than did the symbiont fraction (hindgut and symbiont cells) (p<0.0001). The foregut [FG] showed the highest expression, followed by the midgut [MG], symbiont cells [S], and lastly the hindgut [HG] tissue.
Figure 3.8. Free glucose detected in the four digestive tract fractions after the 5-day feeding assays were complete. More glucose was detected in the foregut [FG] and midgut [MG] protein homogenates than in the hindgut [HG] and symbiont [S] cell homogenates (p<0.0001). The amount of glucose detected between the foregut and midgut did not differ (p=0.7873).
Figure 3.9. Detected cellulose oligo- and monosaccharides from the same protein homogenates used in the glucose detection assay. The foregut [FG] and midgut [MG] homogenates contained more reduced cellulose oligo- and monosaccharides than the hindgut [HG] tissue and symbiont [S] cells (p<0.0001). The foregut and midgut values did not differ (p=0.2367).
Figure 3.10. Spectrophotometric detection of dyed glucose in gut fraction homogenates of termites fed dyed glucose. The midgut [MG] contained the highest amount of detectable red dye, followed by the foregut [FG], symbiont [S] cells, and lastly the hindgut [HG] tissue. All four digestive tract fractions differed significantly (p<0.0001).
BIBLIOGRAPHY


CHAPTER FOUR: ENZYME ACTIVITY AND GENE EXPRESSION IN RESPONSE TO VARIATIONS IN FEEDSTOCK COMPOSITION

ABSTRACT

The goals of this research were to investigate feedstock compositional effects on *Reticulitermes flavipes* digestive biochemistries, characterize host and symbiont fraction-specific enzyme contributions, and confirm these findings with molecular techniques. Feeding bioassays, enzyme assays utilizing both model and natural substrates, and complimentary gene expression assays were used to investigate these aims. Findings suggest that feedstocks greatly augment termite digestive biochemistries and underlying gene expression. The termite host appears to be the main contributor of cellulases and ligninases while the symbionts supply the majority of the hemicellulases. These results further advocate the lower termite *R. flavipes* as an excellent model for feedstock-specific protein research and source of lignocellulases for the derivation of feedstock-specific recombinant protein cocktails.
INTRODUCTION

Lignocellulosic biomass has been heavily researched as a sustainable source of simple sugars for fermentation to biofuels and biomaterials (1, 2, 3, 9, 13, 18, 19, 27, 29, 34, 37, 43, 47, 54). Within the past few decades, several technologies have been established that permit this conversion process to happen on an industrial scale (27, 28, 34, 56, 60). The largest objective now is to further enhance existing technologies, and discover new technologies, that allow for this conversion process to become more cost-competitive in the present petroleum-dominated markets. The use of various feedstocks for biofuel production, more specifically 2nd generation feedstocks, has become the main platform for alcohol fuel production within the last decade (3, 29, 47). Due to varying temperate zones within the United States, a single crop solution, comparable to that of sugar cane in Brazil, will not suffice as a sole fermentable sugar source for large-scale production of alcohol fuels. Instead, researchers within the U.S. are focusing on technologies that are tailored to individual feedstock groups based on their lignocellulose composition (1, 5, 13, 27, 34, 35, 37). Many 2nd generation feedstocks are being investigated as potential sources of fermentable simple sugars for conversion into ethanol, other alcohol fuels, and biomaterials (47, 54). These feedstocks are being assessed based on two key factors; 1) the abundance of cellulose and hemicellulose within the plant’s waste residues and 2) the energy and cost input needed to release these polysaccharides from the lignocellulose structure. Due to the variation of the lignocellulose structure from feedstock to feedstock, technologies catered to specific structural components of each feedstock makes the saccharification process more
efficient (41). Among the technologies, two overall approaches have been established in order to uncover efficient degradation methods. The first approach investigates degradation catalysts external to the feedstock; such as enzymes, strong acids and bases, and pyrolysis techniques (2, 19, 25, 41, 43, 48, 55, 63). The second approach utilizes the recent advancements in sequencing technology in order to investigate internal degradation catalysts by looking into the plant feedstock’s own genome; such as genetically modifying the feedstock to alter its own gene expression, or incorporating a lignocellulase-encoding gene(s) from other organisms into the feedstock’s genome (8, 19, 35). The broad goal of this chapter was to characterize the specific enzymatic contributions of the termite host (salivary glands, foregut, midgut) and the symbiont fraction (hindgut) to the degradation process of various biomass feedstocks. This chapter also sought to characterize a feedstock-specific external degradation platform by investigating feedstock composition effects on termite and symbiont-derived lignocellulase gene expression and activity. As a potential source of saccharific catalysts, termites are excellent model systems due to their efficient lignocellulose degradation capabilities. A detailed investigation of this feedstock-specific degradation platform can increase the efficiency of the degradation process, which in turn could lead to lower costs of processing and allow large-scale alcohol fuel production to be a more feasible endeavor.

The biofuels industry utilizes three main types of 2nd generation feedstocks: 1) agricultural wastes such as corn stover, soy bean residues, and sugar cane bagasse, 2) forestry wastes which include sawdust or wood fractions not suitable for the market, 3) paper wastes such as “black liquor” (mainly lignin and hemicellulose) and discarded
paper pulp (47). The agricultural and forestry wastes are the most abundant of the three
types and can be further broken down into hardwoods, softwoods, and grasses. The
composition of the main components of lignocellulose (lignin, hemicellulose, and
cellulose) varies among these three groups (5, 6, 8, 13, 41, 43). Cellulose, which is a β-
1,4-linked polymer of D-glucose, comprises the largest portion of the structure (30-50%),
followed by hemicellulose, a β-1,4-linked polymer of various 5-carbon and 6-carbon
sugars (20-40%), lignin, a biopolymer of various phenylpropanoids (15-25%), and lastly
resins, fatty acids, phytosterols, ash, and other compounds (5-10%) (32, 33). Not only do
these three main components vary, but hemicellulose and lignin composition vary as well
(8, 12, 18). Hemicellulose is typically composed of various ratios of the 5-carbon and 6-
carbon sugars xylose, arabinose, rhamnose, galactose, mannose, and fucose (18, 32).
Lignin is composed of varying ratios of p-hydroxyphenyl, guaiacyl, and syringyl
phenylpropanoids (8, 11, 16). Due to the broad aim of this study, which was to
investigate new feedstock-specific degradation technologies, agricultural and forestry 2nd
generation feedstocks of varying composition were chosen. These feedstocks included
pine wood shims and beech wood xylan as the forestry waste residues, corn stover and
soybean hulls as agricultural waste residues, and Whatman filter paper (> 98% pure
cellulose) as the control.

The forestry waste residues of pine wood and beech wood are primarily composed
of 40% cellulose, 30% hemicellulose, and 25% lignin (32, 33). The differences in the
lignocellulose structure between these two feedstocks lies within the hemicellulose and
lignin fractions. Pinewood hemicellulose, which is characteristic of most softwoods, is
predominantly comprised of galactoglucomannan with only a small amount of xylan;
while beech wood hemicellulose, which is typical of hardwood species, is primarily composed of xylan with some glucomannan (12, 15, 16, 18, 32, 33, 47, 56). In terms of lignin composition, pine wood is primarily composed of guaiacyl with a small amount of \( p \)-hydroxyphenyl residues, while beech wood is a combination of guaiacyl and syringyl with trace amounts of \( p \)-hydroxyphenyl residues (8, 32, 33). The agricultural waste residues of corn stover and soybean hulls differs in their primary composition with corn stover consisting of 40% cellulose, 33% hemicellulose, and 15% lignin; while soybean hulls consists of 20% cellulose, 50% hemicellulose, and 20% lignin (3, 5, 6, 30, 32, 33, 45, 50). Hemicellulose and lignin compositions also vary between these two feedstocks. Corn stover hemicellulose is predominantly composed of xylan with a small amount of arabinan; while soybean hull hemicellulose is more of an even mix of xylan, arabinan, and mannans (3, 6, 32, 33). Concerning lignin composition, corn stover mainly consists of guaiacyl while soybean hull is a mix of syringyl and guaiacyl residues with a trivial amount of \( p \)-hydroxyphenyl residues (3, 5, 8, 32, 33). The control of this study, Whatman filter paper, is 98% cellulose. Thus, the compositional differences of these feedstocks vary substantially, making them appropriate experimental feedstocks for comparison in this study.

In order to investigate the effect of feedstock composition on termite digestive biochemistries, and determine relative termite and symbiont digestive contributions, eight enzymes were selected as digestive catalytic representatives. Each enzyme studied is a member of one of the three main groups of lignocellulases; 1) cellulases, 2) hemicellulases, and 3) ligninases. Cellulases act on hydrogen bonds at various points of the cellulose polysaccharide chain, cleaving the chain into shorter oligo-, di-, and
monosaccharides (2, 7, 19, 23, 24, 41, 43, 48, 52, 56, 63, 65, 66). Hemicellulases act in much of the same manner, except their substrates are the 5- and 6-carbon hemicellulose chains and they remove successive monosaccharides from the non-reducing end of the chain (12, 18, 25, 38, 41, 43, 50). Ligninases are mainly responsible for breaking the ester bonds within the lignin biopolymer, acting on the ester bonds linking lignin to hemicellulose, and/or radicalizing phenols in order to prevent the inhibition of hemicellulases and cellulases (4, 10, 11, 15, 16, 28, 36, 41, 46, 50, 56, 60). The cellulase glycosyl hydrolase families (GHF) chosen for this study were endoglucanase (GHF9), exoglucanase (GHF7), and β-glucosidase (GHF1). The hemicellulose glycosyl hydrolase families chosen for this study were β-mannosidase (GHF26), β-xylosidase (GHF43), and β-arabinosidase (GHF62). Lastly, the ligninases chosen for this study were phenol-oxidase and esterase. Though the enzymes within each group act on similar substrates, they tend to demonstrate different modes of action.

The cellulases all act upon hydrogen bonds linking monomers of D-glucose, but they act at various regions of the cellulose chain; 1) endoglucanases cut internal beta-1,4-glucosidic bond into shorter, more manageable oligosaccharides, 2) exoglucanases cut disaccharides (cellobiose) from the non-reducing end of the cellulose chain, and 3) β-glucosidases hydrolyze cellobiose and other short oligosaccharides (3-4 D-glucose monomers) to glucose (2, 7, 19, 23, 24, 41, 43, 48, 52, 56, 63, 65, 66). The hemicellulases chosen all have the same mode of action in that they remove successive monosaccharides from the non-reducing end of the polymer chain (12, 18, 25, 38, 41, 43, 50). However, the hemicellulose targets (xylan, mannan, arabinan) of the hemicellulases are all different. Although the ligninases both act on lignin, they have different modes of
action. Esterases hydrolyze lignin and monophenol-carbohydrate bonds by using water to split esters into an alcohol and an acid (11, 15, 16, 28, 36, 41, 46, 50, 56, 60). Phenol oxidases catalyze the \( o \)-hydroxylation of monophenols to \( o \)-diphenols in order to disrupt free phenols from inhibiting cellulases and hemicellulases (4, 10, 11, 15, 16, 28, 41, 46, 56, 60).

The objective of this research was to explore the enzymatic responses within the \( R. \ flavipes \) gut when fed upon lignocellulosic feedstocks of varying lignin and carbohydrate compositions, and also to characterize the specific enzymatic contributions made by the termite host and protozoan symbionts in the degradation and depolymerization of woody biomass. In order to investigate enzymatic activity and gene expression variation, feedstock bioassays, enzyme assays using model and natural substrates, and gene expression assays were conducted. Due to the nature of some lignocellulases to attach to plant particles as they act upon them, enzyme activity may be detected in regions of the gut where the enzyme was not produced, which made it necessary to corroborate enzyme activity profiles with gene expression analyses. The results from this experiment strongly suggest that enzymatic production within the termite varies from feedstock to feedstock in order to overcome compositional differences and efficiently release simple sugars. Contrary to dogmatic theories on termite digestion, it appears that host contributions of lignocellulases are greater than those of symbiotic contributions. This study also provided direction as to which types of genomic resources, termite or symbiont, to probe when selecting candidate genes for future purification and recombinant protein work.
MATERIALS AND METHODS

Model Insect

*Reticulitermes flavipes*, a lower termite species of the family Rhinotermitidae, was used throughout this study. Three separate *R. flavipes* colonies were used as biological replicates. These colonies were collected from the University of Florida campus (Alachua County, FL, USA), a Kanapaha residence (Alachua County, FL, USA) and Purdue University campus (Tippecanoe County, IN, USA). Each colony was identified as *R. flavipes* through the use of genomic DNA extraction, PCR, and mitochondrial 16-s rDNA gene sequencing (49). Termites of the worker caste were used exclusively due to their superior ability to digest lignocellulose over other castes that make up a termite colony. Termites were considered workers if they lacked wings, large mandibles, and distended abdomens (21). For no less than 6 months prior to this study, each colony was held in separate plastic containers on a diet of pine wood (Nelson Wood Shims; Cohasset, MN) and paper, and at a temperature of 24°C, 70% relative humidity under 0:24 light:dark conditions.

Chemicals and Reagents

*Enzyme-specific model substrates*: p-nitrophenyl-β-D-cellobioside, p-nitrophenyl-β-D-glucopyranoside, p-nitrophenyl-β-D-xylopyranoside, p-nitrophenyl-β-D-mannopyranoside, p-nitrophenyl-β-L-arabinopyranoside, 1,2,3-Trihydroxybenzene, p-nitrophenyl acetate; all model substrates were purchased from Sigma-Aldrich (St. Louis,
MO). *Nucleic acid dye:* SYBR green (Bio-Rad; Hercules, CA). *Protein detection reagent:* Bradford protein assay (Bio-Rad; Hercules, CA).

**Buffers and Solutions**

*Homogenization buffer:* 1x phosphate buffered saline (PBS), pH 7.0 (Bio-Rad; Hercules, CA). *Carboxymethyl cellulose solution:* 2% (w/v) carboxymethyl cellulose (CMC) (Sigma-Aldrich; St. Louis, MO) dissolved in Nanopure water. *DNSA stop solution:* 1% 3,5-dinitrosalicylic acid (DNSA), 0.2% crystalline phenol, 0.05% sodium sulfide, 30% sodium potassium tartrate and 0.4 M sodium hydroxide. *Enzyme activity detection solutions:* 4mM model substrate in homogenization buffer.

**Feedstocks**

Five different feedstocks of various lignin and sugar compositions were used during this study. Filter paper (FP) (110mm 98% pure cellulose) (Whatman; Maidstone, UK) was used as the control feedstock while beech wood xylan (X) (>90% xylose residues) (Sigma-Aldrich; St. Louis, Missouri), pine wood shims (PW) (Nelson Wood Shims; Cohasset, MN), corn stover (CS) (Specialty 3557), and soybean hull residues (SB) (Williams 82) were used as the experimental feedstocks. The corn stover was donated by Dr. Nathan Mosier of the Purdue University Agricultural and Biological Engineering department and the soybean hull residues were donated by Dr. Karen Hudson of the Purdue University USDA-agronomy department. Corn stover and soybean hull residues were kept at -20°C until bioassays were conducted.
**Bioassay Setup**

All termites and feedstocks were weighed before assay setup in order to assess termite fitness and feedstock consumption. Feedstocks were held for 24 hours in a 45°C drying chamber prior to being weighed in order to assure all water was removed. For each biological replicate (colony) three technical replicates of each feedstock were tested. A known amount of each feedstock was placed in the center of its own respective 100mm x 20mm Petri dish (Thermo Scientific; Middletown, VA). The feedstocks were then saturated with 150µL of deionized water and allowed to sit for 4 hours in order to absorb the moisture. 50 termite workers and one termite soldier were then added to each Petri dish. The Petri dishes were placed in a 240mm x 90mm sealable plastic container lined with moist paper towels to prevent desiccation and further placed in a 30°C incubation chamber under 0:24 light:dark conditions. Deionized water (100µL) was added to the feedstocks on days 2, 4, 6, 8, and 10 of each 12 day assay to prevent termite desiccation. In total, there were 3 biological replicates of each of the five feedstocks per colony, 50 worker termites per biological replicate, and 9 biological replicates, giving an overall number of 750 termites used per biological replicate and 2,250 worker termites used throughout the entire study.

**Dissections**

Upon completion of the twelfth feeding day, all surviving termites were removed from feedstock Petri dishes, placed in separate new Petri dishes, and weighed. Feedstocks were placed in a 45°C drying oven for another 24 hours and weighed the next day. Thirty whole guts from surviving worker termites of each biological replicate were
removed (39, 66). The whole guts were then split into host fractions (salivary gland, foregut, midgut) and symbiont fractions (hindgut) and placed in a sodium phosphate saline buffer solution (PBS) at 10µL per gut fraction equivalent. The gut tissues were then homogenized and centrifuged at 14,600 rcf, 4°C for 15 minutes. The remaining supernatants were aliquoted into 20µL samples and stored at -80°C until used in enzyme activity assays. All supernatants were stored for 1 week before enzyme assays began and each frozen aliquot was only thawed once. The remaining whole guts of surviving worker termites were placed in centrifuge tubes and stored at -80°C for gene expression assessment as detailed below.

**Enzyme Activity Assays**

Colorimetric enzyme assays were used to investigate enzymatic activity in host and symbiont fractions (10, 64, 66). Both model (synthetic) and natural substrates were used for enzyme assays that measure cellulase, hemicellulase, phenol oxidase, and esterase activities (see Table 4.1 for substrates and abbreviations). These assays were performed using aliquoted gut homogenate from the feedstock assay dissections. Protein concentrations for each gut homogenate were determined using Bradford protein assays (Bio-Rad, Hercules, CA) with bovine serum albumin as the standard.

*Endoglucanase activity.*

Endoglucanase activity was determined using common CMC assay procedures (64). A solution of 2% (w/v) carboxymethylcellulose (CMC) was made in Nanopure water. A 5mM concentration of pure glucose in Nanopure water was also prepared and
used as a standard. The first three columns of the 96-well microplate consisted of 100µL serial dilutions of the 5mM glucose mixture. Host or symbiont homogenate (5µL) from each technical replicate of feedstock was added to the remaining wells in triplicate. The 2% CMC solution (95µL) was then added to any well containing either host or symbiont fraction, which resulted in a total volume of 100µL per well. The plate was allowed to sit at room temperature for 1 hour. After an hour, 100µL of 1% DNSA stop solution (1% 3,5-dinitrosalicylic acid, 0.2% crystalline phenol, 0.05% sodium sulfide, 30% sodium potassium tartrate and 0.4M sodium hydroxide) was added to wells containing both standards and samples, and was then placed in boiling water for 10 minutes. Following the heating phase, the plate was placed on ice for 15 minutes in order to cool. The plate was then read immediately as an endpoint assay at a wavelength of 540nm. The glucose standard curve was used to deduce the amount of reduced sugar in each well.

Exoglucanase, β-glucosidase, β-mannosidase, β-arabinosidase, β-xylosidase, and esterase activity.

These activity assays all followed the same protocol with the only difference being the substrate involved (66). Four millimolar concentrations of substrate were prepared in sodium acetate buffer (0.1M, pH 7.0). Host or symbiont fraction homogenate (5µL) from each technical replicate of feedstock was added to open wells. The samples were pipetted in triplicates to increase sampling accuracy. Substrate (95µL) was then added to any well containing host or symbiont homogenate resulting in a total volume of 100µL (3.8mM substrate) per well. Three wells were dedicated as buffer blanks and contained only 5µL PBS and 95µL substrate. These were kinetic assays, being read at a
wavelength of 420nm every 30 seconds for 45 minutes at a controlled temperature of 30°C. Mean velocities (mOD / sec) were recorded and transformed into specific enzyme activity units (µmol/min/mg) using results from the Bradford protein assay to correct for protein variations between wells and the \( p \)-nitrophenyl molar extinction coefficient of 0.6605mM\(^{-1}\)cm\(^{-1}\) (67).

**Phenol-oxidase activity.**

This assay employed procedures modified from previous research (10). A 4mM concentration of the substrate pyrogallol was prepared in Nanopure water. Host or symbiont homogenate (5µL) was pipetted onto a 96-well microplate. Substrate solution (57.5µL) was then added to each well along with 187.5µL of the buffer to give a total volume of 250µL per well (0.9mM substrate). The kinetic assay was read at a wavelength of 420nm every 30 seconds for 45 minutes at a controlled temperature of 30°C. Results from the Bradford protein assay along with the pyrogallol molar extinction coefficient of 24.70mM\(^{-1}\)cm\(^{-1}\) were used to determine specific enzyme activity.

**Lignocellulase-encoding Gene Selection**

A *Reticulitermes flavipes* 454 pyrosequencing database, containing both termite and symbiont contigs, was bioinformatically probed for candidate lignocellulase-encoding genes (46). The contigs were chosen based on the enzyme groups that were previously investigated using enzyme assays above. Three groups of enzyme-encoding genes were chosen from the 454 database, with specific families of enzymes representing each group: 1) Cellulases (Endoglucanase, Exoglucanase, \( \beta \)-glucosidase), 2)
Hemicellulases (β-mannosidase, β-arabinosidase, β-xylosidase), and 3) Ligninases (laccase, esterase). Each gene was chosen to represent one of the nine enzyme families based on the read count in the 454 database; the contig with the highest read count in each family was selected for this study (Table 4.2). Along with lignocellulase genes, a known stretch (M13) of the pGEM plasmid vector (Promega; Madison, WI) was used as an introduced reference gene (Table 4.2). Quantitative real-time PCR primers were designed for the candidate lignocellulases using the software Primer3 (Whitehead Institute for Biomedical Research). Parameters were set for primer design including a melting temperature of 60 ± 1 °C, a product range size of 150-250 base pairs, and a primer size of 18-23 base pairs (44).

**Lignocellulase Gene Expression Assays**

The frozen whole guts that were saved from enzyme assay dissections were thawed at room temperature (24°C) and further dissected into host fraction (salivary gland, foregut, midgut) and symbiont fraction (hindgut). Homogenates from the dissected gut fractions were collected using the methods stated earlier. RNA was isolated from the host fraction and symbiont fraction homogenates using the SV Total RNA Isolation Kit (Promega; Madison, WI). Equal concentrations of RNA across both sections (empirically determined using a NanoDrop spectrophotometer), having been normalized to the lowest RNA concentration, were then reverse transcribed into cDNA using the iScript cDNA synthesis kit (Bio-Rad; Hercules, CA). Quantitative real-time PCR gene expression assays were carried out using 20 µL reactions. SYBR green (Bio-
Rad; Hercules, CA) was used as the detection reagent and pGEM as a “spike-in” reference gene.

Data Analysis

**ENZYME ACTIVITY ASSAYS**

Specific enzyme activity of whole guts was calculated as micromoles of end product per minute of enzyme reaction per milligram of protein (µmol/min/mg) (26). All data collected from the enzyme activity assays were analyzed using the statistical program JMP version 9 (SAS Institute Inc.; Cary, NC). The data were first tested for normality using a Shapiro-Wilk test. All experimental categories (feedstock, colony, replicate) were then tested against specific enzyme activity to determine significance (α = 0.05) using a one-way ANOVA followed by an all-pairs Tukey’s HSD test for separation of means. This analysis was carried out for both whole gut results and host vs. symbiont results across feedstocks. Host vs. symbiont enzyme activity data within feedstocks was analyzed using a standard pairwise t-test.

**LIGNOCELLULASE GENE EXPRESSION ASSAYS**

These methods followed previously described methods (44). The resulting Ct values from whole gut Q-PCR were first normalized to the reference gene (pGEM) in order to obtain ΔCt values. These values were then analyzed using the statistical program JMP version 9 (SAS Institute Inc.; Cary, NC) to determine statistical differences in gene expression. A global analysis using a one-way ANOVA was performed of lignocellulase
gene expression across feedstocks, biological replicates, and colony. The ΔCt values were then further normalized to the control feedstock filter paper in order to obtain ΔΔCt values. The ΔΔCt values were then transformed to $2^{-\Delta\Delta Ct}$ in order to determine normalized fold expression. The resulting Ct values from the host vs. symbiont Q-PCR were first normalized to the reference gene (pGEM) in order to obtain ΔCt values. A global ANOVA analysis for these values was then performed using the same procedures and software that were used for whole gut analysis, except fraction (host vs. symbiont) was also taken into account. The ΔCt values were then further normalized to the lowest overall expressed gut fraction (symbiont) value to obtain ΔΔCt values. The ΔΔCt values were then transformed to $2^{-\Delta\Delta Ct}$ in order to determine normalized fold expression (22).

RESULTS

ENDOGLUCANASE

Endoglucanase activity within the whole gut did not vary significantly across feedstocks (Figure 4.1). Endoglucanase activity primarily occurs in host fraction (salivary glands, foregut, midgut) of the digestive tract (Figure 4.2). Whole gut endoglucanase gene expression was highest in termites that fed on filter paper, followed by pine wood, corn stover, beech wood, and lastly soybean hulls (Figure 4.3). Endoglucanase gene expression was mainly detected in the host fraction (Figure 4.4). Colony and biological replicates did not demonstrate a significant effect on enzyme activity or gene expression (Table 4.3-4.6).
EXOGLUCANASE

Exoglucanase whole gut activity was higher in gut homogenate of termites that fed on pine wood and corn stover than the rest of the feedstocks, and it was lowest for beech wood xylan (Figure 4.5). Exoglucanase activity mainly occurs in the symbiont fraction (hindgut) of the digestive tract (Figure 4.6). Whole gut exoglucanase gene expression was highest in termites fed filter paper, followed by corn stover, pine wood, and then soy bean (Figure 4.7). Expression of exoglucanase was detected primarily in the symbiont fraction (Figure 4.8). There were no significant differences found between the colonies or the biological replicates of enzyme and gene expression assays (Table 4.3-4.6).

β-GLUCOSIDASE

β-glucosidase whole gut activity was highest in pine wood, followed by filter paper, then corn stover, beech wood xylan and lastly soybean hulls (Figure 4.9). β-glucosidase activity was shown to mainly occur in the host fraction, with some activity also taking place in the symbiont fraction (Figure 4.10). Termites that fed on pine wood demonstrated the highest β-glucosidase gene expression, followed by soybean hulls, corn stover, and lastly beech wood xylan (Figure 4.11). β-glucosidase gene expression was higher in the host fraction than in the symbiont fraction (Figure 4.12). No significant colony effect or biological replicate effect was found in either the enzyme or gene expression assays (Table 4.3-4.6).
**β-XYLOSIDASE**

β-xylosidase whole gut activity was highest in termites that fed on pine wood, followed by corn stover, filter paper and beech wood xylan, and lastly soybean hulls (Figure 4.13). β-xylosidase activity occurred evenly across the host and symbiont fractions of the digestive tract (Figure 4.14). β-xylosidase gene expression was highest in termites that were fed soybean hulls, followed by filter paper, corn stover, pine wood, and lastly beech wood xylan (Figure 4.15). Both the host and symbiont fractions demonstrated equal levels of β-xylosidase gene expression (Figure 4.16). There were no differences found between the colonies or the biological replicates of the enzyme and gene expression assays (Table 4.3-4.6).

**β-MANNOSIDASE**

β-mannosidase whole gut enzyme activity was highest in termites that fed on pine wood and filter paper, followed by corn stover, soybean hulls, and lastly beech wood xylan (Figure 4.17). β-mannosidase activity occurred primarily in the host fraction (Figure 4.18). β-mannosidase gene expression was highest in termites that fed on soybean hulls, followed by termites that fed on filter paper, corn stover, pine wood, and lastly beech wood xylan (Figure 4.19). β-mannosidase gene expression was much higher in the symbiont fraction, and almost non-existent in the host fraction (Figure 4.20). No significant colony or biological replicate effect was found in either the enzyme or gene expression assays (Table 4.3-4.6).
β-arabinosidase whole gut activity was highest in termites that fed on pine wood, followed by filter paper, soybean hulls, corn stover and finally beech wood xylan (Figure 4.21). β-arabinosidase activity occurred mostly in the host fraction, with activity also occurring in the symbiont fraction (Figure 4.22). β-arabinosidase gene expression was highest in termites that fed on soybean hulls, followed by corn stover, and pine wood; beech wood xylan did not show any detectable gene expression (Figure 4.23). β-arabinosidase gene expression was found to be primarily in the symbiont fraction (Figure 4.24). There were no differences found between the colonies or the biological replicates of the enzyme and gene expression assays (Table 4.3-4.6).

ESTERASE

Whole gut esterase activity was highest with beech wood xylan feeding, followed by pine wood and filter paper, corn stover, and lastly soybean hulls (Figure 4.25). Esterase activity occurred mainly in the host fraction (Figure 4.26). Esterase gene expression was highest in termites that fed on soybean hulls, followed by pine wood, corn stover, and then beech wood xylan (Figure 4.27), while esterase gene expression primarily occurred in the host fraction (Figure 4.28). There was no significant colony or biological replicate effect found in both the enzyme and gene expression assays (Table 4.3-4.6).
**PHENOLOXIDASE**

Phenoloxidase whole gut activity was highest in termites fed on beech wood xylan, followed by pine wood and filter paper, then corn stover and lastly soybean hulls (Figure 4.29). Phenoloxidase enzyme activity is mostly in the host fraction, but the symbiont fraction also demonstrates a substantial amount of activity as well, particularly with beech wood xylan feeding (Figure 4.30). Phenoloxidase gene expression was highest in termites that fed on pine wood, followed by corn stover, beech wood, soybean hulls, and lastly filter paper (4.31). Phenoloxidase gene expression was primarily found in the host fraction of the digestive tract (Figure 4.32). There were no significant differences found between the colonies or the biological replicates of the enzyme and gene expression assays (Table 4.3-4.6).

**DISCUSSION**

The termite digestome is a literal genetic pool filled with lignocellulase-encoding genes capable of producing highly efficient biomass degradation catalysts (31, 39, 40, 41, 43, 48, 51, 57, 58, 65). As demonstrated in the previous chapters, both the termite host and the protist and bacteria symbionts contribute to the degradation of the lignocellulose structure in order to free simple sugars for energy needs and other metabolic/intermediary pathways (38, 39, 40, 31). These simple sugars and by-products can be utilized in the biotech industry for the production of alcohol fuels and other useful biomaterials (9, 37). This research focused on eight lignocellulases and five feedstocks which were previously noted and described above. The main purpose of this chapter was to further investigate
termite digestive physiology, and more specifically, to characterize diet effects at the biochemical and molecular levels. Another goal of this chapter was to define specific enzymatic contributions of industry-relevant lignocellulases within the host and symbiont fractions of the termite gut.

The three cellulases studied in this chapter, endoglucanase, exoglucanase, and \( \beta \)-glucosidase, all revealed both surprising and expected results (Table 4.3-4.6; Figure 4.1-4.12). Endoglucanase activity of the termite whole gut protein homogenates towards carboxymethylcellulose did not show any significant variation across termites that were fed on compositionally different feedstocks (Table 4.3; Figure 4.1). This result suggests that endoglucanases may be produced consistently and continuously regardless of substrate cues. However, gene expression validation of a particular endoglucanase from the previously sequenced 454 pyrosequencing database did demonstrate significant variation across feedstocks (Table 4.5; Figure 4.3). The disparity between the lack of endoglucanase activity variation and the presence of gene expression variation could be due to the fact that only one gene within this family was used in this study. The full picture of endoglucanase gene expression may not have been captured. The whole gut gene expression results indicated that expression of this gene was highest in termites that fed on filter paper, pine wood and corn stover, and lowest in termites that fed on soybean hulls. This result is complimentary to the compositional characteristics of each feedstock. In this study, filter paper (98%), pine wood (40%) and corn stover (40%) had the highest percentage of cellulose, while soybean hulls have the lowest percentage (20%) (6, 32). Since this enzyme primarily acts on amorphous cellulose (33), it was expected to reveal feedstocks with the largest amount of cellulose to have the largest
impact on endoglucanase-encoding genes (Figure 4.3). The host fraction, in agreement with what is known for other wood feeding insects (24), insects with sugar-based diets, and our own 454 pyrosequencing database (46), was the main source of both endoglucanase activity and gene expression (Figure 4.2, 4.4). The finding indicates that the termite host is mainly responsible, if not fully responsible, for the contribution of endoglucanases. The opposite can be said about exoglucanase activity and gene expression. These enzymes are typically contributed by lower eukaryotes in nature and they were primarily found in our symbiont contig library (51, 52), so it was expected that the majority of the exoglucanase activity and gene expression would occur in the symbiont fraction (Figure 4.6, 4.8). However, greater than 40 protist exoglucanase genes from GHF7 are expressed in the *R. flavipes* gut, making this a complicated enzyme family (46, 51). Similarly to endoglucanase gene expression results, exoglucanase gene expression was highest in termites that fed on filter paper, pine wood, and corn stover due to their cellulose content, some of which has a crystalline structure (Figure 4.7). Exoglucanases are the only known cellulases that termites possess that are capable of degrading crystalline cellulose. Unlike endoglucanase, the exoglucanase activity assays demonstrated significant differences across feedstocks which corresponded to differences found in gene expression profiles (Figure 4.5).

The β-glucosidase results were interesting in that there was a key difference between whole gut activity results and whole gut gene expression results. While termites that were fed on pine wood demonstrated the highest levels in both tests, soybean hulls showed the lowest β-glucosidase activity results yet the second highest gene expression results (Figure 4.9, 4.11). The discord between these two results could potentially be due
to cellulase inhibitors being present in the soybean hulls (30, 61, 62, 67). This could potentially induce expression of the β-glucosidase gene, but ultimately inhibit function of β-glucosidase enzymes. The most probable inhibiting compounds would be phenols that are released when lignin is depolymerized (30, 61, 62). β-glucosidase activity and gene expression were highest in the symbiont fraction, which agrees with our previous digestome sequencing data in which we found β-glucosidase genes in the host termite contig library (Figure 4.10, 4.12) (42, 51). Another interesting and noteworthy finding from these cellulase studies, was that filter paper, which is 98% pure cellulose, did not induce the highest activity results in any of the three cellulase families. This could potentially be a mechanism utilized by the termite in order to overcome digestive barriers. It has been demonstrated that the components responsible for the recalcitrant nature of lignocellulose, namely lignin and certain hemicellulose polymers, are responsible for increased cellulase gene expression to compensate for said barriers (16, 19). This could be the reason why more active cellulases are being produced for the lignin- and hemicellulose-rich pine wood, corn stover, and soybean hulls, and less are being produced for the relatively barrier-free filter paper. Another explanation for this result could be feedback inhibition of the glucose released (39). High levels of glucose have been shown to repress translation of cellulases and hemicellulases (67) and directly inhibit β-glucosidases (39). Due to the high percentage of glucose potentially released from this substrate, feedback inhibition is very plausible.

The cellulases in this study also revealed very fraction-specific differences. While endoglucanase and β-glucosidase appear to be contributed primarily by the termite host, exoglucanases are mainly symbiont-derived enzymes; for example, over 40 GHF7
exoglucanases from protist symbionts were previously identified in *R. flavipes* gut sequence datasets (46, 51). Also, endoglucanase, which was unchanged across feedstocks, looks to be a prime candidate lignocellulase for the addition to any feedstock-specific enzyme cocktail, while exoglucanase and β-glucosidase should be targeted more to feedstocks containing high percentages of cellulose. Lastly, but not surprisingly, feedstocks with high cellulose content are primary targets for all three of these cellulases and thus they should be included in cocktails for such feedstocks. However, as in the case of filter paper, glucose and lignin should be continuously removed during processing in order to avoid any inhibitory effects as they are released from the lignocellulose structure.

The hemicellulases β-xylosidase, β-mannosidase, and β-arabinosidase all demonstrated activity and gene expression differences across feedstocks. β-xylosidase activity was highest in termites that fed on pine wood and corn stover, and lowest in termites that fed on soybean hulls (Figure 4.13). This result is both expected and surprising at the same time. Expectedly, the hemicellulose fraction of corn stover is predominantly xylan, which likely explains why β-xylosidase activity was high in termites that fed on corn stover (3, 33). However, the high activity in pine wood is perplexing due to the limited quantity of xylan in its hemicellulose fraction; however studies have demonstrated that cellulose can induce xylosidase gene expression (18, 25, 50). The combination of some xylan and a large proportion of cellulose could be responsible for the high β-xylosidase activity found in termites that fed on pine wood. Another confounding result was the difference found between β-xylosidase activity and β-xylosidase gene expression found in termites that were fed soybean hulls. The enzyme
activity was the lowest in soybean hulls out of the feedstocks tested while the gene expression was the highest (Figure 4.13, 4.15). Like the $\beta$-glucosidase result, hemicellulases have also been shown to be inhibited by phenols produced from the depolymerization of lignin (20, 30, 61, 62). Another interesting result was the congruency between unexpected results in the fraction-specific activity and gene expression assays. Hemicellulases are notoriously associated with lower eukaryotes such as protists and fungi, as well as prokaryotes (23, 25, 38, 45, 51, 53, 57). Thus it is interesting that both the activity assays and gene expression assays $\beta$-xylosidase activity and gene expression were equal in both the host and symbiont fractions (Table 4.4, 4.6; Figure 4.14, 4.16). These results suggests that the termite host contains xylosidase-encoding genes within its digestome, which were not revealed in our microarray data set, and are not typically found in insect systems (51). This result could also be due to the PCR primers and substrate not being specific enough, but neither possibility seems likely. Investigating the termite digestome further for $\beta$-xylosidase encoding genes will likely reveal genes that were previously overlooked or even misannotated.

$\beta$-mannosidase and $\beta$-arabinosidase tests both demonstrated fraction-specific differences between activity assays and gene expression results (Figure 4.17, 4.19, 4.21, 4.23). Both enzymes revealed higher activity in the host fraction (Figure 4.18, 4.22), while gene expression was determined to primarily occur in the symbiont fraction (Figure 4.20, 4.24). As mentioned previously about hemicellulases, the latter gene expression result agrees with past research finding that these enzymes are typically associated with lower eukaryotes and prokaryotes (51, 52). The $\beta$-mannosidase enzyme activity results, which showed a majority of the activity in the host fraction (Figure 4.18), could
potentially be due to termite derived enzymes that are acting on the p-nitrophenyl
substrate. According to past sequencing efforts, both symbionts and termites encode
mannosidase genes (51). A simple explanation for the discourse between the activity and
gene expression results is that the termite-derived mannosidases may be produced
upstream of the hindgut in order to perform the task of increasing the defenses of the
termite gut against harmful bacterial (23, 24). The contig chosen for the gene expression
study may have represented a symbiont-specific gene. This scenario is the most probable
cause of the differences found between the β-mannosidase activity and gene expression
assays.

The β-arabinosidase differences are more perplexing since R. flavipes host-
derived arabinosidases were not found through previous sequencing (51). This could
again be due to misannotation or simply from not capturing arabinosidases in the termite
fraction with the sequencing platform. Other termite-derived enzymes may also be acting
on the substrate which could be giving false positive β-arabinosidase activity results.
Eventually sequencing the entire R. flavipes genome will shed light on these fraction-
specific differences. The whole gut activity results for β-mannosidase were not
surprising in that pinewood and filter paper demonstrated the highest activity, while
beech wood xylan was the lowest in all cases (Figure 4.17). Pine wood hemicellulose is
mannan-rich (61, 62), which explains the high activity of β-mannosidase activity in the
homogenate of termites that fed on pine wood. Interestingly, soybean hull feeding
produced the highest β-mannosidase gene expression (Figure 4.19). Soybean hull
hemicellulose is a biopolymer which contains 5- and 6-carbon polysaccharides including
mannan (6, 45). The β-arabinosidase whole gut activity and gene expression profiles
were very similar to that of the β-mannosidase results. Pine wood and filter paper-fed termites demonstrated the highest β-arabinosidase activity, while cDNA from termites that fed on soybean hulls were shown to have the highest β-arabinosidase gene expression (Figure 4.21, 4.23). Since soybean hull hemicellulose is a mixture of hemicellulose polymers, which include arabinan, the gene expression results are expected. However, there is disagreement with the activity assay results which revealed pine wood and filter paper as the feedstocks with the highest β-arabinosidase activity (Figure 4.21). An explanation for this reverts back to the point that only one arabinosidase gene was tested in the gene expression assays. This gene could very well respond favorably to compounds in soybean hulls, while other untested β-arabinosidase genes may actually be responding to the high cellulose content in pine wood and filter paper.

The hemicellulases studied demonstrated significant variation across feedstocks in both the activity assays and the gene expression assays. This result is not surprising since hemicellulose is the most variable component among tested feedstocks. The discourse found between the activity assay and gene expression results in all three hemicellulases tested, suggests a need to remove certain enzymatic inhibitory substances, namely phenols, as the saccharification process is occurring. Also, the differences found in the fraction-specific assays between activity and gene expression suggest a further need to sequence the entire genomes of model organisms in order to capture the full landscape of relevant lignocellulases.

The ligninases studied in this chapter, esterase and phenoloxidase, demonstrated very similar activity and gene expression results. As expected, both enzymes
demonstrated host fraction-specificity in terms of activity and gene expression (Figure 4.26, 4.28, 4.30, 4.32). In order to access the polysaccharides, lignin must first be depolymerized and radicalized (1, 15, 16, 59, 60). Due to this process, the insect host, in this case the termite, generally contributes ligninases that act upstream of cellulases and hemicellulases contributed by both the host and their symbionts (10, 16, 31, 41, 43, 65). Esterase had highest activity in beech wood xylan and highest gene expression in soybean hulls, followed closely by pine wood (Figure 4.25, 4.27). These three feedstocks have high percentages of lignin and lignin associated to hemicellulose (6, 32, 33, 45), which would require esterases to break internal and external ester linkages between phenols, and amid lignin and hemicellulose (11, 16, 28, 36, 41, 50, 60).

Similarly, phenoloxidase results revealed highest activity and gene expression in lignin-rich pine wood and beech wood xylan (Figure 4.29, 4.31). Phenoloxidase does not play a role in degradation nor depolymerization like the other enzymes in this study, but instead this enzyme radicalizes phenolic byproducts to prevent them from inhibiting downstream cellulases and hemicellulases (4, 10, 61, 62). Due to this characteristic, the presence of both cellulose and hemicellulose have been confirmed as enhancers of phenoloxidase gene expression and activity (4, 50). The ligninases under investigation in this chapter revealed very comparable results. They both vary significantly across feedstocks, with lignin and polysaccharide content appearing to be direct contributors to their gene expression, and expectedly, they seem to be host-contributed enzymes.

In conclusion, this dissertation chapter revealed that dietary feedstocks do have an impact on termite digestive physiology, both at the biochemical and the molecular level. In agreement with earlier bioinformatic predictions (51), the termite host appears to be
the main contributor of ligninases and cellulases, while symbionts appear to be the main contributor of hemicellulases. These results can drive future research into termite-derived lignocellulases, whether it be for production of biofuels or biomaterials materials or for control purposes, as well as indicate which gut fraction (host or symbiont) to explore for such enzymes. The feedstock impacts found throughout this study point towards a need to explore the potential energy-efficient and cost-saving development of feedstock-specific lignocellulase cocktails. The termite digestive tract, based on the results from this study, remains an excellent candidate for the selection of such lignocellulases. This research also demonstrated a need to routinely investigate lignocellulases at both the biochemical (activity) and molecular (gene expression) levels in order to truly grasp the overall processes of digestion. Due to the inhibitory or induction effects of certain substances in vivo, gene expression and enzyme activity are not always congruent. Finally, in an era when many protein-based speculations are founded on results obtained through large-scale nucleotide sequencing studies, this research points towards a need for more functional-based validation of high-throughput / next-generation sequencing results.
Table 4.1. Model and natural substrates used for enzyme activity assays.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Substrate</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endoglucanase</td>
<td>Carboxymethylcellulose</td>
<td>CMC</td>
</tr>
<tr>
<td>Exoglucanase</td>
<td>p-nitrophenyl-β-D-xylosidase</td>
<td>pNPC</td>
</tr>
<tr>
<td>β-glucosidase</td>
<td>p-nitrophenyl-β-D-glucopyranoside</td>
<td>pNPG</td>
</tr>
<tr>
<td>β-xylosidase</td>
<td>p-nitrophenyl-β-D-xylopyranoside</td>
<td>pNPX</td>
</tr>
<tr>
<td>β-mannosidase</td>
<td>p-nitrophenyl-β-D-mannopyranoside</td>
<td>pNPM</td>
</tr>
<tr>
<td>β-arabinosidase</td>
<td>p-nitrophenyl-β-L-arabinopyranoside</td>
<td>pNPA</td>
</tr>
<tr>
<td>Phenoloxidase</td>
<td>1,2,3-Trihydroxybenzene</td>
<td>Pyrogallol</td>
</tr>
<tr>
<td>Esterase</td>
<td>p-nitrophenyl acetate</td>
<td>pNPAc</td>
</tr>
</tbody>
</table>

Table 4.2. Lignocellulase and reference gene primer list.

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Contig</th>
<th>Primer Sequence</th>
<th>Accession Number</th>
<th>Recombinant Enzymes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Endoglucanase</td>
<td>ENDO F</td>
<td>CATACGGCGGATAGTAGCA</td>
<td>AY572862</td>
<td>Cell-1</td>
</tr>
<tr>
<td></td>
<td>ENDO R</td>
<td>TCAACAGGCAAGGCAACATAC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exoglucanase</td>
<td>EXO F</td>
<td>GTTTCCAGGCGTACACAGT</td>
<td>FL645416</td>
<td>GHF7-3</td>
</tr>
<tr>
<td></td>
<td>EXO R</td>
<td>GTCTTGGCCCGGACCAGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>β-glucosidase</td>
<td>GLUC F</td>
<td>GGAGATGCTGAAGGCAATT</td>
<td>HM152540</td>
<td>β-glu</td>
</tr>
<tr>
<td></td>
<td>GLUC R</td>
<td>TGTGCTGCACTCTTGGAATC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>β-xylosidase</td>
<td>XYLO F</td>
<td>CGAGAGATTTGGGTGATT</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>XYLO R</td>
<td>CCAGAACCACTGTGACTCTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>β-xylosidase</td>
<td>MANN F</td>
<td>CATACCATGGTCCACATCGAACA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MANN R</td>
<td>TGTGCTGCCGCTGATGATT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>β-arabinosidase</td>
<td>ARAB F</td>
<td>CCAAATGTGACGGACATCGA</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ARAB R</td>
<td>TGTTCTTTCACACGTCC</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phenoloxidase</td>
<td>LAC F</td>
<td>CTGGTGCCGACAGGATGATTCT</td>
<td></td>
<td>LacA</td>
</tr>
<tr>
<td></td>
<td>LAC R</td>
<td>ATCTCCGGCTCCTCTGCA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Esterase</td>
<td>ESTER F</td>
<td>TGTCCGGACGGGCTCTGTC</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ESTER R</td>
<td>TAAACGGACCTCTGCTGATT</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pGEM</td>
<td>M13F</td>
<td>GCAAACACGCGCGACGCTGCTG</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M13R</td>
<td>AACAGCTATGACCGCTGC</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4.3. Whole gut global ANOVA of enzymatic activity across feedstocks.

<table>
<thead>
<tr>
<th>Endoglucanase</th>
<th>Exoglucanase</th>
<th>β-glucosidase</th>
<th>β-mannosidase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replicate</td>
<td>8</td>
<td>3.11</td>
<td>0.4289</td>
</tr>
<tr>
<td>Colony</td>
<td>2</td>
<td>0.38</td>
<td>0.6827</td>
</tr>
<tr>
<td>Feedstock</td>
<td>4</td>
<td>2.43</td>
<td>0.0674</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>β-arabinosidase</th>
<th>β-xylosidase</th>
<th>Phenoloxidase</th>
<th>Esterase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replicate</td>
<td>8</td>
<td>2.14</td>
<td>0.1297</td>
</tr>
<tr>
<td>Colony</td>
<td>2</td>
<td>0.12</td>
<td>0.8914</td>
</tr>
<tr>
<td>Feedstock</td>
<td>4</td>
<td>14</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>
Table 4.4. Gut Fraction global ANOVA of enzymatic activity across feedstocks.

<table>
<thead>
<tr>
<th></th>
<th>Endoglucanase</th>
<th>Exoglucanase</th>
<th>β-glucosidase</th>
<th>β-mannosidase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replicate</td>
<td>8</td>
<td>0.82</td>
<td>0.5932</td>
<td>Replicate</td>
</tr>
<tr>
<td>Colony</td>
<td>2</td>
<td>0.25</td>
<td>0.7679</td>
<td>Colony</td>
</tr>
<tr>
<td>Feedstock</td>
<td>4</td>
<td>78.98</td>
<td>&lt;0.0001</td>
<td>Feedstock</td>
</tr>
<tr>
<td>Fraction</td>
<td>1</td>
<td>545.03</td>
<td>&lt;0.0001</td>
<td>Fraction</td>
</tr>
<tr>
<td>β-arabinosidase</td>
<td>Source</td>
<td>d.f.</td>
<td>F</td>
<td>P value</td>
</tr>
<tr>
<td>Replicate</td>
<td>8</td>
<td>1.55</td>
<td>0.2181</td>
<td>Replicate</td>
</tr>
<tr>
<td>Colony</td>
<td>2</td>
<td>0.09</td>
<td>0.9139</td>
<td>Colony</td>
</tr>
<tr>
<td>Feedstock</td>
<td>4</td>
<td>18.03</td>
<td>&lt;0.0001</td>
<td>Feedstock</td>
</tr>
<tr>
<td>Fraction</td>
<td>1</td>
<td>29.27</td>
<td>&lt;0.0001</td>
<td>Fraction</td>
</tr>
</tbody>
</table>
Table 4.5. Whole gut global ANOVA of gene expression across feedstocks.

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>F</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replicate</td>
<td>8</td>
<td>0.07</td>
<td>0.9298</td>
</tr>
<tr>
<td>Colony</td>
<td>2</td>
<td>0.54</td>
<td>0.5839</td>
</tr>
<tr>
<td>Feedstock</td>
<td>4</td>
<td>59.13</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>F</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replicate</td>
<td>8</td>
<td>0.27</td>
<td>0.7153</td>
</tr>
<tr>
<td>Colony</td>
<td>2</td>
<td>0.02</td>
<td>0.9764</td>
</tr>
<tr>
<td>Feedstock</td>
<td>4</td>
<td>156.5</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>F</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replicate</td>
<td>8</td>
<td>0.15</td>
<td>0.8613</td>
</tr>
<tr>
<td>Colony</td>
<td>2</td>
<td>0.19</td>
<td>0.8232</td>
</tr>
<tr>
<td>Feedstock</td>
<td>4</td>
<td>46.27</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>F</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replicate</td>
<td>8</td>
<td>0.02</td>
<td>0.9808</td>
</tr>
<tr>
<td>Colony</td>
<td>2</td>
<td>0.2</td>
<td>0.8174</td>
</tr>
<tr>
<td>Feedstock</td>
<td>4</td>
<td>127.91</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>F</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replicate</td>
<td>8</td>
<td>0.05</td>
<td>0.9587</td>
</tr>
<tr>
<td>Colony</td>
<td>2</td>
<td>0.01</td>
<td>0.9858</td>
</tr>
<tr>
<td>Feedstock</td>
<td>4</td>
<td>193.8</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>F</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replicate</td>
<td>8</td>
<td>0.18</td>
<td>0.8389</td>
</tr>
<tr>
<td>Colony</td>
<td>2</td>
<td>0.04</td>
<td>0.9613</td>
</tr>
<tr>
<td>Feedstock</td>
<td>4</td>
<td>130.6</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>F</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replicate</td>
<td>8</td>
<td>0.324</td>
<td>0.7253</td>
</tr>
<tr>
<td>Colony</td>
<td>2</td>
<td>1.81</td>
<td>0.1774</td>
</tr>
<tr>
<td>Feedstock</td>
<td>4</td>
<td>10.74</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>d.f.</th>
<th>F</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Replicate</td>
<td>8</td>
<td>0.01</td>
<td>0.9812</td>
</tr>
<tr>
<td>Colony</td>
<td>2</td>
<td>0.11</td>
<td>0.8938</td>
</tr>
<tr>
<td>Feedstock</td>
<td>4</td>
<td>52.84</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>
Table 4.6. Gut Fraction global ANOVA of gene expression across feedstocks.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Replicate</td>
<td>8</td>
<td>0.03</td>
<td>0.9672</td>
<td>Replicate</td>
<td>8</td>
<td>0.25</td>
<td>0.7037</td>
<td>Replicate</td>
<td>8</td>
<td>0.06</td>
<td>0.9438</td>
<td>Replicate</td>
<td>8</td>
<td>0.08</td>
<td>0.9232</td>
</tr>
<tr>
<td>Colony</td>
<td>2</td>
<td>0.04</td>
<td>0.962</td>
<td>Colony</td>
<td>2</td>
<td>0.02</td>
<td>0.9847</td>
<td>Colony</td>
<td>2</td>
<td>0.04</td>
<td>0.9642</td>
<td>Colony</td>
<td>2</td>
<td>0.17</td>
<td>0.8485</td>
</tr>
<tr>
<td>Feedstock</td>
<td>4</td>
<td>112.23</td>
<td>&lt;0.0001</td>
<td>Feedstock</td>
<td>4</td>
<td>695.73</td>
<td>&lt;0.0001</td>
<td>Feedstock</td>
<td>4</td>
<td>43.66</td>
<td>&lt;0.0001</td>
<td>Feedstock</td>
<td>4</td>
<td>180.52</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Fraction</td>
<td>1</td>
<td>258.91</td>
<td>&lt;0.0001</td>
<td>Fraction</td>
<td>1</td>
<td>41.36</td>
<td>&lt;0.0001</td>
<td>Fraction</td>
<td>1</td>
<td>40.36</td>
<td>&lt;0.0001</td>
<td>Fraction</td>
<td>1</td>
<td>122.11</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>β-arabinosidase</td>
<td></td>
<td></td>
<td></td>
<td>β-xylosidase</td>
<td></td>
<td></td>
<td></td>
<td>Phenoloxidase</td>
<td></td>
<td></td>
<td></td>
<td>Esterase</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Replicate</td>
<td>8</td>
<td>0.06</td>
<td>0.9483</td>
<td>Replicate</td>
<td>8</td>
<td>0.09</td>
<td>0.9074</td>
<td>Replicate</td>
<td>8</td>
<td>0.19</td>
<td>0.8275</td>
<td>Replicate</td>
<td>8</td>
<td>0.12</td>
<td>0.8845</td>
</tr>
<tr>
<td>Colony</td>
<td>2</td>
<td>0.12</td>
<td>0.8824</td>
<td>Colony</td>
<td>2</td>
<td>0.03</td>
<td>0.9743</td>
<td>Colony</td>
<td>2</td>
<td>0.23</td>
<td>0.7916</td>
<td>Colony</td>
<td>2</td>
<td>0.06</td>
<td>0.9386</td>
</tr>
<tr>
<td>Feedstock</td>
<td>4</td>
<td>192.43</td>
<td>&lt;0.0001</td>
<td>Feedstock</td>
<td>4</td>
<td>666.96</td>
<td>&lt;0.0001</td>
<td>Feedstock</td>
<td>4</td>
<td>24.84</td>
<td>&lt;0.0001</td>
<td>Feedstock</td>
<td>4</td>
<td>70.57</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Fraction</td>
<td>1</td>
<td>163.43</td>
<td>&lt;0.0001</td>
<td>Fraction</td>
<td>1</td>
<td>3.44</td>
<td>0.0669</td>
<td>Fraction</td>
<td>1</td>
<td>67.97</td>
<td>&lt;0.0001</td>
<td>Fraction</td>
<td>1</td>
<td>50.03</td>
<td>&lt;0.0001</td>
</tr>
</tbody>
</table>
Figure 4.1. Whole gut endoglucanase activity profile across feedstocks. No endoglucanase activity differences were found among homogenates of termites that were fed various feedstock diets (p = 0.0674).
Figure 4.2. Gut fraction endoglucanase activity profile across feedstocks. The host fraction homogenate demonstrated significantly more endoglucanase activity than symbiont fraction homogenate (p < 0.0001).

Symbiont mean separation = A, B, C. Host mean separation = W, X.

*** denotes a p – value of < 0.0001 in pairwise comparisons of host and symbiont specific activity.
Figure 4.3. Whole gut endoglucanase gene expression profile across feedstocks.

Significant differences in endoglucanase gene expression profiles found with pinewood-fed termites demonstrated the highest gene expression relative to filter paper (p < 0.0001).
Figure 4.4. Gut fraction endoglucanase gene expression profile across feedstocks.

In agreement with the enzyme activity assays, the gene expression profiles reveal a majority of the endoglucanase gene expression occurs in the host fraction (p < 0.0001).

Symbiont mean separation = A, B, C, D. Host mean separation = V, W, X, Y, Z.

*** denotes a p – value of < 0.0001 in pairwise comparisons of host and symbiont gene expression.
Figure 4.5. Whole gut exoglucanase activity profile across feedstocks. Exoglucanase activity was detected to be greatest in the homogenate of termites that fed on pine wood and corn stover. Exoglucanase activity varied across feedstocks ($p < 0.0001$).
Figure 4.6. Gut fraction exoglucanase activity profile across feedstocks. The symbiont fraction of the termite digestive tract contained more exoglucanase activity than the host fraction (p < 0.0001).

Symbiont mean separation = A, B, C. Host mean separation = W, X, Y.

*** denotes a p – value of < 0.0001 in pairwise comparisons of host and symbiont specific activity. NS denotes a p – value of > 0.05.
Figure 4.7. Whole gut exoglucanase gene expression profile across feedstocks.

Exoglucanase gene expression varied across gut homogenate collected from termites that upon different feedstocks feedstocks (p < 0.0001), and was greatest in termites that fed on corn stover and lowest in termites that fed on beech wood xylan.
Figure 4.8. Gut fraction exoglucanase gene expression profile across feedstocks. In congruency with the gut fraction-specific exoglucanase activity results, gene expression profiles using exoglucanase-specific primers revealed exoglucanase expression to be greatest in the symbiont fraction (p < 0.0001).

Symbiont mean separation = A, B, C, D, E. Host mean separation = V, W, X, Y.

*** and ** denotes p – values of < 0.0001 and < 0.001 respectively in pairwise comparisons of host and symbiont gene expression.
Figure 4.9. Whole gut β-glucosidase activity profile across feedstocks. β-glucosidase activity varied across homogenates of termites that were fed different feedstocks (p < 0.0001), was greatest in termites that fed on pine wood and filter paper, and lowest in termites that fed on beech wood xylan and soybean hulls.
Figure 4.10. Gut fraction β-glucosidase activity profile across feedstocks. The host fraction of termite digestive tracts contained more β-glucosidase activity than symbiont fractions (p < 0.0001).

Symbiont mean separation = A, B, C. Host mean separation = W, X, Y.

*** and * denotes p – values of < 0.0001 and < 0.01, respectively in pairwise comparisons of host and symbiont specific activity.
Figure 4.11. Whole gut β-glucosidase gene expression profile across feedstocks. β-glucosidase gene expression varied across termites that fed upon different feedstocks (p < 0.0001), and was highest in termites that fed on pine wood and soybean hulls, and lowest in termites that fed on beech wood xylan.
Figure 4.12. Gut fraction $\beta$-glucosidase gene expression profile across feedstocks.

Gene expression of $\beta$-glucosidase primarily occurred in the host fraction of the termite digestive tract (p < 0.0001).

Symbiont mean separation = A, B, C, D. Host mean separation = V, W, X, Y.

*** and * denotes p – values of < 0.0001 and < 0.01, respectively in pairwise comparisons of host and symbiont gene expression.
Figure 4.13. Whole gut β-xylosidase activity profile across feedstocks. Specific β-xylosidase activity varied across termites that fed on different feedstocks (p < 0.0001), and was highest in termites that fed upon pine wood and relatively even across the other feedstocks.
Figure 4.14. Gut fraction $\beta$-xylosidase activity profile across feedstocks. Both the host and symbiont fractions demonstrated equivalent levels of specific $\beta$-xylosidase activity ($p = 0.2249$).

Symbiont mean separation = A, B. Host mean separation = W, X, Y.

***, **, and * denotes $p$ – values of $< 0.0001$, $< 0.001$, $p < 0.01$, respectively in pairwise comparisons of host and symbiont specific activity. NS denotes a $p$ – value of $p > 0.05$. 
Figure 4.15. Whole gut β-xylosidase gene expression profile across feedstocks. β-xylosidase gene expression varied across termites that fed on different feedstocks (p < 0.0001). Termites that fed on soybean hulls were found to contain the highest gene expression of β-xylosidase, while termites that fed on beech wood xylan contained the lowest levels of gene expression.
Figure 4.16. Gut fraction β-xylosidase gene expression profile across feedstocks.

The host and symbiont fractions of the termite digestive tract showed equal levels of β-xylosidase gene expression (p = 0.0669), which agreed with the activity assay findings. Symbiont mean separation = A, B, C, D. Host mean separation = V, W, X, Y.

*** and ** denotes p – values of < 0.0001 and < 0.001, respectively in pairwise comparisons of host and symbiont gene expression. NS denotes a p – value of p > 0.05.
Figure 4.17. Whole gut $\beta$-mannosidase activity profile across feedstocks. $\beta$-mannosidase specific activity varied across homogenates collected from termites that fed upon different feedstocks ($p < 0.0001$), was greatest in termites that fed on filter paper and pine wood, and lowest in termites that fed on beech wood xylan.
Figure 4.18. Gut fraction β-mannosidase activity profile across feedstocks. Specific β-mannosidase activity was predominantly found in the host fraction of the termite gut (p < 0.0001).

Symbiont mean separation = A, B, C. Host mean separation = W, X, Y.

*** denotes a p-value of < 0.0001 in pairwise comparisons of host and symbiont specific activity.
Figure 4.19. Whole gut β-mannosidase gene expression profile across feedstocks. β-mannosidase gene expression varied across termites that fed upon different feedstocks ($p < 0.0001$), and was highest in termites that fed on soybean hulls and lowest in termites that fed on beech wood xylan.
Figure 4.20. Gut fraction β-mannosidase gene expression profile across feedstocks.

β-mannosidase gene expression was primarily found in the symbiont fraction of the termite digestive tract (p < 0.0001).

Symbiont mean separation = A, B. Host mean separation = W, X, Y.

*** denotes a p-value of < 0.0001 in pairwise comparisons of host and symbiont gene expression.
**Figure 4.21.** Whole gut β-arabinosidase activity profile across feedstocks. β-arabinosidase activity varied across termites that fed on feedstock (p < 0.0001), and was found to be highest in termites that fed upon pine wood and lowest in termites that fed on beech wood xylan.
Figure 4.22. Gut fraction $\beta$-arabinosidase activity profile across feedstocks. $\beta$-arabinosidase activity was highest in the host fraction of the termite gut ($p < 0.0001$). Symbiont mean separation = A, B. Host mean separation = W, X.

***, **, and * denotes $p$ – values of $< 0.0001$, $< 0.001$, $p < 0.01$, respectively in pairwise comparisons of host and symbiont enzyme activity. NS denotes a $p$ – value of $p > 0.05$. 
Figure 4.23. Whole gut β-arabinosidase gene expression profile across feedstocks.

β-arabinosidase gene expression varied across termites that fed upon different feedstocks (p < 0.0001), and was highest in termites that fed on soybean hulls and lowest in termites that fed on beech wood xylan.
Figure 4.24. Gut fraction β-arabinosidase gene expression profile across feedstocks.

Gene expression of β-arabinosidase was predominantly found in the symbiont fraction of the termite digestive tract (p < 0.0001).

Symbiont mean separation = A, B, C, D. Host mean separation = V, W.

*** and * denotes p – values of < 0.0001 and < 0.01, respectively in pairwise comparisons of host and symbiont gene expression.
**Figure 4.25. Whole gut Esterase activity profile across feedstocks.** Esterase activity varied across termites that fed upon different feedstocks ($p < 0.0001$), was highest in termite that fed on beech wood xylan and lowest in termites that fed on soybean hulls.
Figure 4.26. Gut fraction esterase activity profile across feedstocks. Esterase gene expression was primarily found in the host fraction of the termite digestive tract (p < 0.0001).

Symbiont mean separation = A, B, C. Host mean separation = W, X, Y, Z.

*** denotes a p – value of < 0.0001 in pairwise comparisons of host and symbiont enzyme activity.
Figure 4.27. **Whole gut esterase gene expression profile across feedstocks.** Esterase gene expression varied across termites that fed on different lignocellulosic feedstocks ($p < 0.0001$), was greatest in termites that fed on soybean hulls and pine wood, and lowest in termites that fed on beech wood xylan.
**Figure 4.28. Gut fraction esterase gene expression profile across feedstocks.** The esterase gene expression profile revealed greater esterase gene expression in the host fraction of the termite gut than in the symbiont fraction (p < 0.0001).

Symbiont mean separation = A, B, C, D, E. Host mean separation = V, W, X, Y.

*** denotes a p – value of < 0.0001 in pairwise comparisons of host and symbiont gene expression.
Figure 4.29. Whole gut phenoloxidase activity profile across feedstocks.

Phenoloxidase activity varied across termites that fed upon different feedstocks ($p < 0.0001$), was highest in termites that fed on beech wood xylan and lowest in termites that fed upon corn stover and soybean hulls.
Figure 4.30. Gut fraction phenoloxidase activity profile across feedstocks. The host fraction of the termite gut had greater phenoloxidase activity than the symbiont fraction (p < 0.0001).

Symbiont mean separation = A, B. Host mean separation = W, X, Y.

*** and ** denotes p – values of < 0.0001 and < 0.001, respectively in pairwise comparisons of host and symbiont enzyme activity. NS denotes a p – value of p > 0.05.
Figure 4.31. Whole gut phenoloxidase gene expression profile across feedstocks.

Phenoloxidase gene expression varied significantly across termites that fed on different feedstocks (p < 0.0001), and was greatest in termites that fed upon pine wood and corn stover.
Figure 4.32. Gut fraction phenoloxidase gene expression profile across feedstocks.

Phenoloxidase gene expression was predominantly found in the host fraction of the digestive tract (p < 0.0001).

Symbiont mean separation = A, B, C, D, E. Host mean separation = V, W, X, Y.

*** denotes a p – value of < 0.0001 in pairwise comparisons of host and symbiont gene expression.
BIBLIOGRAPHY


CHAPTER FIVE: PURIFICATION AND CHARACTERIZATION OF A POTENTIAL ACCESSORY LIGNINASE FROM THE DIGESTIVE TRACT OF R. FLAVIPES

ABSTRACT

The purpose of this research was to purify and characterize a potential ligninase from digestive tract homogenate of the lower termite Reticulitermes flavipes. Colorimetric activity assays, ion-exchange column chromatography, tandem mass spectrometry, and subsequent bioinformatic analysis were used in an effort to meet this objective. Findings indicate that lignin-associated detoxification proteins with esterase-like mechanisms, or proteins that are able to act upon and degrade α-naphthyl-ester and pNP-ester substrates, were purified from the gut homogenate of R. flavipes. Additionally, these proteins significantly increase the biomass saccharification output of recombinant cellulases (endoglucanase and β-glucosidase). However, results also indicate that further bioinformatic analysis using deeper sequencing databases and recombinant protein validation will need to be used in order to determine the exact protein(s) of interest purified in this chapter. These results strongly suggest that an enriched protein in the purified fractions would make an excellent addition to any recombinant protein cocktail utilized for the degradation of a lignin-rich substrate.
INTRODUCTION

The production of biofuels and biomaterials from woody biomass, specifically 2nd generation wastes, has been an ongoing focus of research for the past few decades (3, 4, 11, 16, 19, 24, 29, 31, 32, 37). A dominant goal of research in this area has been to make these production methods, and therefore the end-products, more cost-competitive in the current petroleum-dominated markets (15, 40, 42). In order to achieve this goal, research is focusing on making production methods more efficient in terms of putting less time, money, and resources into production while obtaining more end-product. Lignin, which is mainly responsible for the recalcitrant nature of lignocellulose, is currently the main obstacle preventing biofuels and biomaterials from competing with petroleum-based products (4, 11, 12, 15, 19, 24, 29, 31, 32, 37, 40, 42). A large fraction of the time, money, and resources necessary to produce nonpetroleum-based products is presently allocated for biomass pretreatment; the step in which lignin is depolymerized and disassociated from the lignocellulose structure in order to access and hydrolyze the hemicellulose and cellulose polysaccharides (11, 19, 29, 31, 37, 40). Increasing the efficiency of the pretreatment step will greatly improve the overall efficiency of the production process, which should make biofuels and biomaterials a more economically viable option in the near future.

As mentioned in previous chapters, biofuels can be produced from lignocellulosic forestry and agricultural wastes. The cell wall of every woody material contains the rigid structural support system lignocellulose, which is a biopolymer of lignin, hemicellulose,
and cellulose in varying ratios depending on the residue (16, 29, 43). The lignin structure primarily consists of the three monolignol monomers p-coumaryl alcohol, sinapyl alcohol, and coniferyl alcohol (36). The monolignols are converted into the phenylpropanoids p-hydroxyphenyl, syringyl, and guaiacyl, respectively, after having gone through the phenylpropanoid biosynthesis pathway, and are then polymerized to form the lignin structure. As demonstrated in the previous chapter, the ratios and combinations of these phenylpropanoids varies from feedstock to feedstock (13, 15, 36). Characterizing and catering these variations to pretreatment technologies may lead to increased efficiency in the lignocellulose saccharification process.

The ultimate goal of lignocellulose degradation is capturing and further processing the simple sugars which compose the polysaccharides of the structure. In order to accomplish this task, the polysaccharides must be made accessible to hydrolytic enzymes (16, 40). This accessibility is only possible with the removal of lignin from the plant tissue (3, 29, 37). Lignin is an amorphous substance used by plants for structural support, a barrier against oxidative stress, hydrophobicity for water transport, and microbial resistance (13, 36, 37, 43). The amorphous characteristic of lignin, or rather the lack of a distinct principal structure, is what makes lignin tough to break down. However, being able to degrade lignin is very important not only for access to the polysaccharides, but also for utilization of the lignin itself. Lignin is the largest renewable source of carbon on Earth next to hemicellulose and cellulose (42). Recovered lignin byproducts from the degradation of lignocellulose are currently used in many processes such as in the production of ceramics and oils, and it is also used in the
derivation of other products such as metal cleaners, rubber reinforcements, insecticides, HIV inhibitors, and fuel stabilizers (12, 15, 37).

The core methods currently used to accomplish this task are enzyme degradation, extreme heat treatments (pyrolysis), the use of strong acids and bases, and plant genetic modifications (3, 4, 11, 12, 13, 15, 16, 19, 24, 29, 31, 37, 40, 42). However, there are severe downsides to some of these methods. Extreme heat treatment is still energy inefficient due to the fact that the process currently uses more units of energy for production than units of energy gained as an end product (15, 19, 40). Temperatures of ≥ 180 °C are needed in order to solubilize lignin (15, 42). The solubilized lignin can lead to the formation of phenolic compounds like furfural which have inhibitory effects on other cellulolytic enzymes and are toxic to other hydrolytic organisms used in bioreactors such as yeast and bacteria (14). Extreme heat treatment can also lead to the transformation of amorphous cellulose into the more recalcitrant crystalline cellulose, making the structure even more rigid and inaccessible to cellulases (37). Similar to the problems faced by extreme heat, strong acid and base treatments also form large amounts of inhibitory and toxic compounds (12). This process also produces many volatile compounds through which carbon is lost, taking away from the overall sum of substrate to be converted into usable products. The waste associated with this pretreatment method is also of concern. Using strong acids and bases produces significantly more hazardous waste materials than any other form of pretreatment, and properly disposing of the waste is expensive (11, 12, 40).

A technology that may prove to be fruitful in the future is lignin degradation through plant genetic modifications (1, 4, 13, 16, 37). However, as it currently stands,
gene regulating technologies discovered within the past decade, both in terms of upregulating and downregulating lignin biosynthesis-encoding genes, generally result in the stunting of growth, vascular system damage, and other harmful plant phenotypes (11, 16, 42). Although there are also a few downsides to the utilization of enzymes for pretreatment, such as the degree of crystallinity of the cellulose, feedback inhibition, and lack of substrate surface area, enzymatic pretreatment is still proving to be the most viable option (3, 11, 12, 14, 15, 16, 19, 24, 28, 29, 31, 40, 42). While other pretreatment methods are still trying to develop a model system, the perfect enzymatic model systems are likely all around us in nature. It is only a matter of time before in vivo degradation processes become perfected as in vitro biomass saccharification methods.

Many organisms, both eukaryotic and prokaryotic, have evolved efficient arsenals of enzymes capable of degrading and removing lignin from their respective food sources in order to access the carbon-rich polysaccharides as an energy source (2, 7, 9, 12, 14, 27, 28, 29, 31, 33, 39, 43). These enzymes typically fall into two different categories based on their mode of action: 1) acting directly upon the lignin structure, either breaking internal bonds or external bonds with hemicellulose, or 2) oxidizing phenols that have been released during degradation in order to protect the termite and to stop cellulase-inhibitory effects from occurring (9, 12, 14, 29). Some of the more well-studied ligninases include esterase, etherase, peroxidase, and phenoloxidase (3, 19, 29, 39). Based on past sequencing results, coupled with the enzymatic and molecular findings in Chapter 4, termites possess ligninases that appear to be feedstock specific (31). This chapter focuses on esterase activity due to structural characteristics of lignin (ester
linkages), as well as previous esterase activity and gene expression results which showed profile variations across feedstocks (see Chapter 4).

The main objective of this chapter was to purify and characterize a potential esterase-based ligninase from the termite digestive system. Proteins with esterase activity were targeted due previous sequencing findings, protein activity results, and gene expression conclusions; all of which show esterase enzymes playing a critical role in the degradation of lignocellulose, specifically lignin. Protein purification was achieved using column chromatography and SDS-PAGE separation techniques, while characterization was determined using colorimetric assays, Native PAGE gel analysis, and tandem mass spectrometry. The data indicate that at least one of the purified proteins possesses the ability to act upon ester linkages and potentially makes lignocellulose polysaccharides more accessible to recombinant cellulases and hemicellulases. These results suggest that at least one of the purified proteins is an excellent candidate for recombinant protein production for utilization in biomass-degrading enzyme cocktails that also include recombinant cellulases already produced by our lab.

MATERIALS AND METHODS

Termites

*Reticulitermes flavipes* termites of the worker caste were used in this study. The identification of species and caste was verified using methods described in Chapter 2 (34). A laboratory colony, Whistler 1 (W1), collected from Purdue University’s campus (Tippecanoe County, IN, USA), was used throughout this study. This colony was held in
a plastic container on a diet of pine wood shims (Nelson Wood Shims; Cohasset, MN) and moist paper towels. This container was held in an incubation chamber at a constant temperature of 24°C, ~70% R.H., and a 0:24 hr light:dark ratio.

**Chemicals and Reagents**

*Native-PAGE loading buffer*: Native-PAGE sample buffer (Bio-Rad; St. Louis, MO).

*Anion-exchange resin*: DEAE-Sepharose Fast Flow (Sigma-Aldrich; St. Louis, MO).

*Protein ladder*: Precision Plus Protein Kaleidoscope ladder (Bio-Rad; Hercules, CA).

*Glucose detection reagent*: Autokit Glucose (Wako Chemical; Richmond, VA).

**Buffers and Solutions**

*Native-PAGE 7% Resolving gel*: 5.15 mL Nanopure water, 2.5 mL 1.5M Tris-HCl (Sigma-Aldrich; St. Louis, MO), pH 8.8, 2.31 mL 30% Acrylamide (Bio-Rad; Hercules, CA), 0.05 mL 10% ammonium persulfate (APS) (Bio-Rad; Hercules, CA) in Nanopure water, 0.005 mL TEMED (Sigma-Aldrich; St. Louis, MO).

*Native-PAGE 4% stacking gel*: 3.22 mL Nanopure water, 1.2 mL 0.5M Tris-HCl, pH 6.8, 0.672 mL 30% Acrylamide (Bio-Rad; Hercules, CA), 0.05 mL 10% ammonium persulfate (APS) (Bio-Rad; Hercules, CA) in Nanopure water, 0.005 mL TEMED (Sigma-Aldrich; St. Louis, MO).

*Native-PAGE running buffer*: 4.04g Tris (Fisher Scientific; Waltham, MA), 14.4g Glycine (Sigma-Aldrich; St. Louis, MO), 1L Nanopure water.

*Esterase substrate solution*: 60 mM α-naphthyl propionate (Sigma-Aldrich; St. Louis, MO) in acetone (Mallinckrodt Chemicals; St. Louis, MO).

*Stop solution*: Fast Blue BB (Sigma-Aldrich; St. Louis, MO).

*Incubation buffer*: 0.1M sodium phosphate (Fisher Scientific; Waltham,
MA), pH 7. *Native-PAGE storage buffer:* 10% glacial acetic acid (Mallinckrodt Chemicals; St. Louis, MO) in Nanopure water. *Homogenization buffer:* 50 mM Tris-HCl, pH 7.4, 20% glycerol (Applied Biosystems; Foster City, CA). *Elution buffer:* 20 mM Tris-HCl, pH 7.4. *Protein detection reagent:* Bradford protein assay (Bio-Rad; Hercules, CA). *Esterase activity detection solution:* 4mM p-nitrophenyl acetate (p-NPA) (Sigma-Aldrich; St. Louis, MO) in 0.1 M sodium acetate, pH 7.0. *SDS-PAGE 10% resolving gel:* 4.06 mL Nanopure water, 2.5 mL Tris-HCl, pH 8.8., 3.3 mL 30% acrylamide, 0.1 mL 10% SDS (Fisher Scientific; Waltham, MA), 0.05 mL 10% ammonium persulfate, 0.005 mL TEMED. *SDS-PAGE 4% stacking gel:* 3 mL Nanopure water, 1.2 mL Tris-HCl, pH 6.8, 0.84 mL 30% acrylamide, 0.05 mL 10% SDS, 0.05 mL 10% ammonium persulfate, 0.005 mL TEMED. *SDS-PAGE loading buffer:* 950 µL Laemmli sample buffer (Bio-Rad; Hercules, CA), 50 µL beta-mercaptoethanol (βME) (Bio-Rad; Hercules, CA). *Gel fixing buffer:* 50% methanol (Omnisolv; McLean, VA), 10% glacial acetic acid (Mallinckrodt Chemicals; St. Louis, MO), 40% Nanopure water. *SDS-PAGE gel staining reagent:* 0.25% Coomassie Blue R-250 (Bio-Rad; Hercules, CA), 45% methanol (Omnisolv; McLean, VA), 45% Nanopure water, 10% glacial acetic acid. *SDS-PAGE de-staining reagent:* 5% methanol (Omnisolv; McLean, VA), 7.5% glacial acetic acid (Mallinckrodt Chemicals; St. Louis, MO), 87.5% Nanopure water. *SDS-PAGE storage buffer:* 7% glacial acetic acid in Nanopure water. *Glucose assay stop solution:* 10 µL 0.2M ethylenediaminetetraacetic acid (EDTA) (Sigma-Aldrich; St. Louis, MO) in 500 µL of 0.1 M sodium acetate assay buffer. *Glucose assay incubation/reaction buffer:* 0.1M sodium acetate buffer, pH 7 (Mallinckrodt Pharmaceuticals; St. Louis, MO).
Feedstock Esterase Native-PAGE Gel

These methods followed a previously established protocol with slight modifications (39). Aliquots of stored termite gut homogenate from the feeding bioassays in Chapter 4 were thawed and run on an esterase Native PAGE gel. Aliquots of host and symbiont fractions from each of the five feedstocks previously tested in Chapter 4 (pine wood, filter paper, beech wood xylan, corn stover, and soybean hulls) containing 5 µg of total protein were diluted 1:1 with Native PAGE sample buffer (Bio-Rad; Hercules, CA) and loaded onto native PAGE gels (7% resolving gel and 4% stacking gel). Electrophoresis was conducted in 1x Tris-Glycine running buffer for 1.5 hr at 120V and 40 °C. The gel was then placed in a glass incubating container and gently rocked for 20 min in 100 mL of sodium phosphate buffer (0.1 M, pH 7) containing 1 mL of 60 mM α-naphthyl propionate (Sigma-Aldrich; St. Louis, MO) in acetone (1 mM final concentration). In order to detect the hydrolysis product α-naphthol (bands), 20 mg of filtered Fast Blue BB (Sigma-Aldrich; St. Louis, MO) reagent made in 1 ml water (2.0% final conc.) was added to the buffer. The gel was allowed to incubate in the staining reagent for 45 minutes before fixing and storing. The gel was fixed and destained with 10% acetic acid in Nanopure water for 5 minutes and then photographed. The gel was then stored in the 10% acetic acid solution at 4°C.
**Dissections**

Individual termites (200) morphologically determined to be of the worker caste (20) were randomly removed from colony Whistler 1 and placed in a 100mm x 20mm Petri dish (Thermo Scientific; Middletown, VA) with beech wood xylan and allowed to feed for 10 days. Upon the 10th day, 150 termites were set aside for digestive tract dissections. Whole guts were first removed according to an established procedure in which the hind section of the termite body was cut resulting in an opening with which to take hold of the anus and pull out the entire intact digestive tract. The whole guts were then placed in a 1.5 mL centrifuge tube (Eppendorf; Hamburg, Germany) in 1 mL of esterase buffer (50 mM Tris-HCl pH 7.4 & 20% glycerol). The guts were homogenized using 1.5 mL Pellet Pestles (Kimble-Kontes; Vineland, NJ) and centrifuged at 14,600 rcf with a set temperature of 4°C for 15 minutes. The supernatant was removed, placed in a clean 1.5 mL centrifuge tube, and stored at -80°C until column chromatography was performed. The supernatant was stored for one day before use and only thawed once.

**Chromatography**

These methods followed an ion-exchange chromatography drip column system was established in order to separate and concentrate various beetle esterases (5). A 5 mL plastic syringe with the plunger removed was the column used in this chromatography setup. A metal titration stand with clamps were used to hold the column upright. Glass wool was rinsed with 100% ethanol and placed at the bottom of the syringe in order to hold the resin in place. DEAE-Sepharose Fast Flow (Sigma-Aldrich; St. Louis, MO) was used as the anion-exchange resin. The 1 mL of resin was loaded on top of the glass wool
and packed into a uniform mass using the elution buffer (20 mM Tris-HCl pH 7.4). Once the resin was packed, 1 mL of the previously dissected termite whole gut homogenate was loaded onto the column, resting on top of the resin. Six pass through fractions were then collected at a volume of 1 mL per fraction using a suction-controlled flow valve to control the drip rate. After the sixth pass-through fraction, the drip rate was stopped and a gradient former was connected to the drip column. The gradient former contained two columns with the first column holding 11 mL of elution buffer and the second column holding 11 mL of the elution buffer with 0.6 M of NaCl. The gradient former and the chromatography column were then set at a constant drip rate and allowed to run. In all, twenty three esterase fractions were collected, across a salt concentration range of 0 M to 0.6 M NaCl. Protein concentrations for each fraction were immediately determined using Bradford protein assays (Bio-Rad, Hercules, CA) with bovine serum albumin as the standard.

**Esterase Colorimetric Activity Assay**

A four millimolar concentration of p-nitrophenyl acetate was prepared in sodium acetate (0.1 M, pH 7). Samples from each collected fraction (10 µL) were added to open wells in a 96-well microplate. The p-nitrophenyl acetate substrate (90 µL) was then added to any well containing a fraction sample resulting in a total volume of 100 µL (3.8 mM substrate) per well. Three wells were dedicated as buffer blanks and contained only 10 µL of PBS and 90 µL of substrate. This was a kinetic assay, being read at a wavelength of 420 nm every 30 seconds for 45 minutes at a controlled temperature of 30°C. Mean velocities (mOD / sec) were recorded and transformed into specific enzyme
activity units (µmol/min/mg) using the p-nitrophenyl molar extinction coefficient of 0.6605mM⁻¹cm⁻¹ and protein concentration results from the Bradford protein assay to correct for protein variations between the fractions (41).

**Fraction Esterase Native-PAGE Gel**

Aliquots of the collected fractions 1 - 12, and crude gut homogenate containing 5 µg of total protein, were diluted 1:1 with Native PAGE sample buffer (Bio-Rad; Hercules, CA) and loaded onto native PAGE gels (7% resolving gels and 4% stacking gels). Electrophoresis, incubation, staining, photographing, and storing all followed the same protocol as previously described for the feedstock esterase Native-PAGE gel.

**Fraction SDS-PAGE Protein Gel**

Aliquots of the collected fractions and crude gut homogenate previously run on the Native-PAGE esterase gel were also run on SDS-PAGE gels. The proteins contained in the fractions were first precipitated using a previously established method which utilized rinses with methanol and chloroform in order to obtain 30 µg of protein per aliquot (38). The precipitated protein was then re-suspended in 10 µL of SDS loading buffer (950 µL Laemmli sample buffer: 50 µL βME) and loaded on an SDS gel (10% resolving gel and 4% stacking gel) with 30 µg of crude homogenate and 10 µL of Precision Plus Protein Kaleidoscope ladder (Bio-Rad; Hercules, CA). Electrophoresis was conducted in Tris-glycine SDS running buffer (1x) and allowed to run at 200V for 1 hr. The gel was then placed in a glass incubating container and fixed overnight in a fixing buffer (50% methanol, 10% acetic acid, 40% Nanopure water). After the 24-hr
period, the fixing buffer was poured off and 100 mL of a gel staining reagent (0.25% Coomassie Blue R-250, 45% methanol, 45% Nanopure water, 10% acetic acid) was added to gel container and allowed to stain for 4 hrs. After the staining process, the staining reagent was poured off and 100 mL of a de-staining reagent (5% methanol, 7.5% acetic acid, 87.5% Nanopure water) was added to the container and allowed to de-stain the gel for 24 hrs. The gel was then photographed and stored in 7% glacial acetic acid at 4°C until gel band excision.

**Glucose Detection Assays**

In order to verify which fractions to excise bands from, glucose detection assays using both collected fractions and already developed recombinant cellulases were performed following previously described methods (28). Pinewood sawdust was placed in a vented 1.5 mL Eppendorf centrifuge tube with 500 µL sodium acetate buffer (0.1 M, pH 7) at 2% w/v. Sawdust assays contained different sawdust treatments; (i) pine wood sawdust incubated with 8 µg of recombinant Cell-1 (endoglucanase) in 150 µL of sodium acetate, (ii) pine wood sawdust incubated with 8 µg of recombinant β-glu (β-glucosidase) in 150 µL of sodium acetate, (iii) pine wood sawdust incubated with a combination of 8 µg of recombinant cell-1 and 8 µg of recombinant β-glu in 150 µL of sodium acetate, (iv) pine wood sawdust incubated with 8 µg of fraction, 8 µg of recombinant cell-1 and 8 µg of recombinant β-glu in 150 µL of sodium acetate, and (v) blank; pine wood incubated with 150 µL of sodium acetate. The final volume within each treatment was 650 µL. Each treatment was divided into three technical replications. The treatments were then incubated at 37 °C with an agitation rate of 220 rpm for 10 hours. Once the incubation
time had expired, the treatments were centrifuged at 12,000 rcf and 23 °C for 5 minutes. To stop the reaction, the supernatant was removed and placed in a clean centrifuge tube with 0.2 M EDTA at a volume of 10 µL per 500 µL of 0.1 M sodium acetate assay buffer. This resulted in a final concentration of 4 mM EDTA per treatment. The supernatant (50 µL) was then pipetted in triplicate microplate wells. A standard curve of serially-diluted glucose, from 5 mM of glucose downward to a buffer blank, was then generated in 0.1 M sodium acetate assay buffer + 4 mM EDTA. A glucose detection reagent (Wako Chemical; Richmond, VA) (200 µL) was then added to each well containing glucose dilutions or aliquots of treatment (52). The microplate was then left to incubate at room temperature (24 °C) for 5 minutes before absorbance was read as an endpoint at a wavelength of 505 nm. The same process was then repeated with beech wood xylan at 6% w/v as the substrate instead of pine wood sawdust.

**Peptide Sequencing**

Gel bands of interest based on esterase activity assay results and glucose detection assay results (Figure 5.3, 5.5, 5.6) were excised using a sterile razor and a sterile cutting apparatus and sent to the Bindley Bioscience Center located on the Purdue University campus (Tippecanoe County, IN, USA) for amino acid sequencing. The bands were first digested with trypsin, fixed with Carbamidomethyl, and then run on a TripleTOF 5600 mass spectrometer using liquid chromatography-tandem mass spectrometry. The amino acid results from the spectrometry run were first blasted using BLASTp against the entire NCBI protein database to determine amino acid homology. The sequences were also back-translated and blasted against a local *R. flavipes* 454 pyrosequencing database to
determine nucleotide homology using the bioinformatics software Mascot (Matrix Science; Boston, MA) (6). The parameters for these searches were set at BLASTp, nr protein database, expect=20000, no filter, and PAM30.

Data Analysis

GLUCOSE DETECTION ASSAYS

All data collected from the glucose detection assays were analyzed using the statistical program JMP version 9 (SAS Institute Inc.; Cary, NC). The data were first tested for normality using a Shapiro-Wilk test. All experimental categories (replicate & fraction) were then tested against specific enzyme activity to determine significance (α = 0.05) using a one-way ANOVA. Specific enzyme activity was calculated as micromoles of end product per minute of incubation per gut equivalent (µmol/min/mg) (22). If significance was found, an all-pairs Tukey’s HSD test was used to determine specific means that varied significantly between replicates and fractions.

RESULTS

Feedstock Esterase Native-PAGE Gel

A band was detected in lanes of corn stover- and soybean hull-fed termite homogenate that was absent in the other feedstocks, while pine wood, beech wood xylan, and soybean hulls also had a banding pattern that is different from filter paper and corn stover feeding (Figure 5.1).
Fraction Esterase Activity and Protein Concentrations

Of the 29 ion exchange fractions collected, 7 fractions contained detectable protein concentrations and esterase activity (Figure 5.2). The pass-through phase, which serves as a step to elute unwanted proteins, was successful in that the protein concentration was detectable in fraction 1, peaked in fraction 2, dropped off in fraction 3, and was undetectable in the remaining pass-through fractions. The protein concentrations and esterase activity were again detectable in salt-gradient fractions 8-11, with a majority of the protein eluting into fractions 8 and 9.

Fraction Esterase Native-PAGE Gel

Esterase activity was most pronounced in the gel lanes containing crude gut homogenate and fractions 2, 8, 9, and 10 (Figure 5.3). Aside from the crude gut preparation, the highest amount of activity, measured in detectable α-naphthyl propionate hydrolysis, occurred in fractions 8 and 9. These results are consistent with the esterase activity assay results in that fractions 2, 8, 9, and 10 all contained activity, with fractions 8 and 9 having the highest activity.

Fraction SDS-PAGE Protein Gel

Strong protein bands were detectable in the crude homogenate and fractions 2, 8, 9, and 10 (Figure 5.4). Again, these findings are consistent with the enzyme activity assay and Native PAGE results. The pass-through fraction 2 contains the most protein, followed by the crude homogenate, fraction 9, 10, and lastly fraction 8. There are two
bands in the middle of fractions 8, 9, and 10 that look to be consistent among all three of these fractions.

**Glucose Detection Assays**

The pine wood sawdust assay revealed that glucose levels were detectable in the treatments $\beta$-glu + cell-1, fraction 8 + $\beta$-glu + cell-1, and fraction 9 + $\beta$-glu + cell-1 (Figure 5.5). Fraction 9 released significantly more glucose when combined with $\beta$-glu and cell-1 than any other fraction ($p < 0.0001$), with a 6-fold greater release of glucose than the next highest treatment ($\beta$-glu + cell-1). Of the fractions with detectable glucose release, the treatment $\beta$-glu + cell-1 by themselves released the lowest amount of glucose.

The beech wood xylan assay revealed detectable levels of glucose in all treatments tested (Figure 5.6). Again, the fraction 9 treatment released significantly more glucose when combined with $\beta$-glu and cell-1 than all other treatments ($p < 0.0001$). Fractions 8 and 10 released the second highest amounts of glucose.

**Excised Band Sequencing**

In total, four bands were excised from the SDS-PAGE gel; E1, E2, B1, and B2 (Figure 5.4). Bands B1 and B2 contained a much greater assortment of proteins than E1 and E2. Bands B1 and B2 demonstrated the highest degree of homology to 14-3-3 proteins, which are conserved regulatory proteins (Table 5.1, 5.2). Bands E1 and E2 demonstrated homology to such proteins as cytochrome p450 cyp6, n-acetylmannosamine-6-phosphate 2-epimerase, aldo-keto reductase, alcohol dehydrogenase, and glyceraldehyde-3-phosphate dehydrogenase (Table 5.3, 5.4).
Interestingly, the same contig found to encode an aldo-keto reductase in band E2 was also found in bands B1 and B1, but at much lower degrees of homology. Also, a contig encoding an alcohol dehydrogenase and a contig encoding a glyceraldehyde-3-phosphate dehydrogenase in band E2 was also found in band B1, but again, at a much lower degree of homology compared to E2. The top matches in band E2 demonstrated the highest degrees of homology (ion score and e-value) compared to the top matches of the other three bands, and encode proteins with ligninase capabilities (Table 5.4-5.7; Figure 5.7 - 5.9).

DISCUSSION

It is well-established that termites produce protein catalysts involved in the efficient degradation and removal of the recalcitrant substance lignin from the lignocellulose polymer (29, 31). This is accomplished through the degradation of lignin by hydrolysis and radicalization of ester linkages, and/or the oxidation and glycosylation of free phenolic compounds (32, 43). Based on data collected from multiple sequencing platforms, activity assays, gene expression analysis, and esterase native PAGE, the lower termite *R. flavipes* contains esterases that respond to feedstock compositional differences (32, 35, 39; Chapter 4; Figure 5.1). Past research indicates that these enzymes are capable of breaking ester linkages (39). The main goal of this chapter was to purify (fractionate) potential esterases from the termite gut and characterize their activity in order to find lignin-associated candidates for recombinant protein production.
In the previous chapter (Chapter 4), two candidate ligninases (esterase and phenoloxidase) were shown to have greater activity and gene expression in termites that fed on lignin-rich diets (beech wood and pine wood), and very low gene expression profiles in termites that fed filter paper diets that are lignin-free (> 98% pure cellulose). The enzyme activity and gene expression profiles varied across feedstocks and were found primarily in the host fraction (salivary glands, foregut, and midgut); indicating that feedstock composition can induce or inhibit a ligninase response, and the termite host is the main contributor (source) of these proteins. Due to the proposed ligninase mechanisms of each enzyme, esterase was chosen as the protein of focus due to its potential roles in either the direct degradation of the lignin structure or hydrolysis of phenylpropanoids esterified onto the hemicellulose polymer, and the plethora of established purification methods and data existing in esterase research (5, 7, 8, 27, 39).

In order to further investigate the previous esterase results, a Native-PAGE esterase gel was run using aliquots from the same homogenates used previously in Chapter 4. The banding patterns and intensities from this gel corroborate the esterase activity and gene expression findings from Chapter 4; specifically the host fraction contains greater amounts of esterase activity and termites that fed upon compositionally different feedstocks demonstrated different esterase biochemistries as indicated by the absence or presence of bands across feedstocks (Figure 5.1). In order to take a deeper look into the esterase activity and identify potentially differentially expressed esterase isoforms, ion-exchange column chromatography was performed.

The column chromatography system used DEAE-Sepharose as the resin, which has been widely used as an anion-exchange reactive group to separate esterases from
heterogeneous protein mixtures (5). Although the gradient of NaCl ranged from 0 – 0.6 M, all detectable proteins eluted within the 0 – 0.4 M range, which is the typical elution salt range of esterases (5, 7, 27). The next step was to determine if unwanted proteins eluted in the pass-through phase and to investigate which fractions contained esterase proteins for further investigation. The protein concentration appeared to correlate with the amount of esterase activity detected, except for fraction 2 in which the protein concentration was proportionally much greater than the esterase activity indicating that the proteins eluted in this pass-through fraction were predominantly non-esterase proteins (Figure 5.2). Fractions 8 – 11 contained protein concentration profiles very similar to that of the specific esterase activity profiles. All fractions containing protein (1 – 3, 8 – 11) were further investigated, with special attention paid to the esterase-containing fractions 8 – 10, using both activity assays and separation techniques. Fractions 7 and 12 were also tested in case smaller, less reactive esterases escaped the detection of the Bradford and esterase activity assays.

In order to confirm the activity assay findings, a Native-PAGE esterase gel was run using α-naphthyl propionate as the substrate. The results of this gel confirmed the previous findings that fractions 2, 8, 9, and 10 all contained esterase or esterase-like proteins, while fractions 1, 3, and 11 also contained esterase proteins, but at a much lower concentration (Figure 5.3). Using the same fractions, multiple SDS-PAGE gels were run to separate proteins for sequencing. Again, fractions 2, 8, 9, and 10 repeatedly displayed distinct bands (Figure 5.4). Four intense bands were seen in fractions 8-10 that do not appear to be present in the pass-through fraction. Two of the bands appeared to be present in all three fractions (8 – 10), while two were most noticeable in fraction 10. The
two bands consistent in fractions 8 – 10 (E1 and E2) appear to have a molecular mass of ~45 – 55 kDa. The other two bands, more specific to fraction 10 (B1 and B2), have molecular masses of ~25 – 35 kDa. While the four bands mentioned were of interest due to their intensity and absence from the pass-through fraction column, the sizes of the bands didn’t provide any further validation as to whether or not they were esterase proteins, i.e., esterases have been recorded as varying in size from 1 – 160 kDa (2).

However, bands B1 and B2 were in the typical size range of feruloyl esterases (~30 kDa) (2, 27). To better determine which bands to excise, further activity tests were performed.

Glucose release detection assays were performed using fractions 2, 7, 8, 9, 10, 11, and 12 in order to determine which fractions contained proteins directly involved in the saccharification of lignocellulose. These tests were performed on each fraction alone and in combination with the recombinant cellulases Cell-1 (a GHF9 endoglucanase) and β-glu (a GHF1 β-glucosidase), which, when combined, previously showed high levels of glucose release from both pine wood lignocellulose and beech wood xylan (28, 32). Two types of incubations were performed; one using pine wood sawdust as the substrate and the other using beech wood xylan. All fractions were incubated separately with both recombinant proteins. In both tests, when combined with β-glu + Cell-1, fraction 9 released significantly more glucose than any other fraction, making it the likely source of an efficient lignocellulase candidate (Figure 5.5, 5.6). The two bands previously of interest from this fraction (E1 and E2) were thus marked as bands to be excised for sequencing. Fraction 8 also released glucose in both tests, however, the concentrated bands of interest in fraction 8 were the same two bands that appeared in fraction 9, thus this fraction was excluded from further probing. Fraction 10 did not release detectable
amounts of glucose when incubated with pine wood sawdust, but released glucose amounts equal to that of fraction 8 when incubated with beech wood xylan. Due to the beech xylan results, the two bands of interest in fraction 10 (B1 and B2) were also marked as bands to be excised for sequencing.

The bioinformatics approach used to determine protein homology of the sequenced bands (E1, E2, B1 and B2) involved searching a back-translated nucleotide database of contigs (~9,500) from a previous 454 pyrosequencing study on the *R. flavipes* database gut (32). This 454 database was enriched for lignin-associated genes. The 454 database was primarily composed of genes responsive to the presence of lignin and hemicellulose (32).

The sequencing results did not reveal any homologous matches to annotated esterases in the 454 database. Bands B1 and B2 contained homologous matches to the regulatory family of 14-3-3 proteins (Table 5.1, 5.2). These proteins are found in all eukaryotic cells and are known to bind and modify a varied range of cellular proteins important in many biological processes such as bacterial development, apoptosis, and cell growth (10). This family of proteins gets its name (14-3-3) from the fraction in which it elutes from bovine brain homogenate in a DEAE-cellulose chromatography column (the 14th fraction) and the position it’s found on electrophoresis gels (3.3). Due to the resin used in this study (DEAE-Sepharose), it is not surprising that this protein family was captured and concentrated.

The sequencing of bands E1 and E2 revealed homologous protein matches that have lignin associations (Table 5.3, 5.4). However, some of these lignin associations are not known to be involved in degradation at present time. In band B1, a cytochrome p450
of the family cyp6 was found. These proteins are involved in the degradation of xenobiotics and other toxic compounds in organismal systems (14), and were previously found to be significantly responsive to lignin feeding in the *R. flavipes* gut (26, 32). These enzymes could play an important role in protecting other enzymes, such as cellulases and hemicellulases, and other organisms within the gut environment, such as bacteria and yeast, from the inhibitory and toxic effects of phenols and other toxic metabolites released from lignin (26, 31, 32). Other matches found for peptide sequences contained in this band, such as n-acetylmannosamine-6-phosphate 2-epimerase which plays a role in aminosugar metabolism, but this protein was not previously characterized as playing roles in lignin degradation. The sequencing results of band E2 revealed the most likely candidates involved in the glucose release from the pine and beech wood substrates (Table 5.4). One aldo-keto reductase (AKR) and two alcohol dehydrogenases (ADH) revealed ion-matches of > 50 (Table 5.5 – 5.7; Figure 5.7 – 5.9), which demonstrated the highest degree of homology of any matches found in all four bands. This result is not surprising. Aside from cytochrome p450 (28 contigs), the most lignin-associated contigs found in the 454 pyrosequencing study belonged to the AKR (17 contigs) and ADH families (19 contigs) (32).

AKRs are quickly coming to the forefront as essential catalysts in the efficient degradation of lignocellulose (9, 18). The binding site of AKR proteins is positioned in a hydrophobic fold at in the C-terminal of the β-sheet (18). Due to this hydrophobicity, AKRs prefer apolar and aromatic substrates, such as those contained in lignin monomers. The AKR revealed in this study (contig 00057) was the 18th highest expressed contig found in the 454 database (32). This AKR transcript was recently used to produce a
recombinant AKR protein. This protein, when combined with the Cell-1 and $\beta$-glu

cellulases of *R. flavipes*, significantly enhanced glucose release by this cellulase cocktail
by over 4-fold (32). AKRs have also been found in lignin-degrading yeast and have been
shown to act on lignin and the freed phenols (9, 18).

Alcohol dehydrogenase is also interesting in regard to the proposed function of
cinnamyl alcohol dehydrogenase (CAD) found in plants, which plays an important role in
the biosynthesis of lignin by converting cinnamyl aldehydes into cinnamyl alcohols (1).
However, it can also act as a detoxifier by converting alcohols back into aldehydes and
ketones (13, 21). Since cinnamyl alcohols are the basic building blocks of lignin, this
could be the role that ADHs are playing within the termite digestive system. Although
the mentioned enzymes may play a role in lignin by-product management, they have not
been directly linked to lignin degradation (breaking internal and external lignin bonds).
This could be because these proteins simply do not perform this task, or possibly because
it just has not yet been observed due to limitations in analytical chemistry or detection.

The last possibility, and one that should be explored further, is the possibility that
the bands that were sequenced contained bacterial-derived proteins that would not be
represented in the 454 database. Carboxyl esterases have been found in many species of
bacteria (2, 8, 12, 43). These esterases have molecular masses ranging from 25 – 55 kDa,
which the four bands of interest fall into. Many of the esterases found in bacteria
demonstrate substrate specificities toward pNP-esters. Feruloyl esterase has been
characterized in certain species of *Clostridium* bacteria, which is a genus of bacteria that
lower termites are known to harbor in their hindguts (17). The 454 database used for the
homologous protein searches may be useful when probing for termite and protist protein
matches, but the sequencing may not have gone deep enough to capture bacterial lignocellulase-encoding transcripts. However, deep sequencing was recently performed on *R. flavipes* using the Illumina platform (Rajarapu & Scharf unpublished), which should reveal a greater amount of bacterial sequences relevant to the digestion of lignocellulosic materials. Once annotated, this database should be queried using the amino acid sequences in the four bands in this study.

Cytochrome p450, aldo-keto reductase, and alcohol dehydrogenase are enzymes that have the potential to indirectly aid in the degradation of lignocellulose. Through degradation of lignin by-products in the gut microenvironment, these enzymes may enable cellulolytic enzymes to perform at maximum capacity. Further bioinformatic characterization of the proteins from the four bands and recombinant protein activity assays should be performed in order to determine the protein(s) responsible for the large increase in glucose release when incubated with recombinant cellulases. Efficient disassociation and removal of lignin continues to be an obstacle to making biofuels production a feasible endeavor. However, an increasing number of enzymatic cocktail components, such as the candidate aldo-keto reductase identified in this chapter, are being rapidly discovered and arranged in ways to make them fit together into a cohesive solution.
Table 5.1. Top homologous BLAST hits for tandem mass spectrometry sequencing results of band B1 against a local *R. flavipes* 454-pyrosequencing database

<table>
<thead>
<tr>
<th>Contig</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>contig03251</td>
<td>14-3-3</td>
</tr>
<tr>
<td>contig03895</td>
<td>14-3-3</td>
</tr>
<tr>
<td>contig05115</td>
<td>14-3-3</td>
</tr>
<tr>
<td>contig03011</td>
<td>14-3-3</td>
</tr>
<tr>
<td>contig03767</td>
<td>14-3-3</td>
</tr>
<tr>
<td>contig03714</td>
<td>14-3-3</td>
</tr>
<tr>
<td>contig02353</td>
<td>14-3-3</td>
</tr>
<tr>
<td>contig03968</td>
<td>14-3-3</td>
</tr>
<tr>
<td>contig02776</td>
<td>monooxygenase</td>
</tr>
</tbody>
</table>

Table 5.2. Top homologous BLAST hits for tandem mass spectrometry sequencing results of band B2 against a local *R. flavipes* 454-pyrosequencing database

<table>
<thead>
<tr>
<th>Contig</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>contig01447</td>
<td>takeout-like protein</td>
</tr>
<tr>
<td>contig02543</td>
<td>peptidase</td>
</tr>
<tr>
<td>contig09386</td>
<td>N/A</td>
</tr>
<tr>
<td>contig02156</td>
<td>hydrolase</td>
</tr>
<tr>
<td>contig01051</td>
<td>hydrolase</td>
</tr>
<tr>
<td>contig07923</td>
<td>proteasome</td>
</tr>
<tr>
<td>contig02293</td>
<td>heat shock protein 90</td>
</tr>
<tr>
<td>contig02353</td>
<td>14-3-3</td>
</tr>
<tr>
<td>contig03011</td>
<td>14-3-3</td>
</tr>
</tbody>
</table>
**Table 5.3.** Top homologous BLAST hits for tandem mass spectrometry sequencing results of band E1 against a local *R. flavipes* 454-pyrosequencing database.

<table>
<thead>
<tr>
<th>Contig</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>contig00190</td>
<td>serine or cysteine proteinase inhibitor</td>
</tr>
<tr>
<td>contig02046</td>
<td>N/A</td>
</tr>
<tr>
<td>contig03575</td>
<td>cytochrome p450 cyp6</td>
</tr>
<tr>
<td>contig04894</td>
<td>n-acetylmannosamine-6-phosphate 2-epimerase</td>
</tr>
<tr>
<td>contig02030</td>
<td>vacuolar protein sorting-associated protein</td>
</tr>
<tr>
<td>contig02952</td>
<td>N/A</td>
</tr>
<tr>
<td>contig04472</td>
<td>N/A</td>
</tr>
<tr>
<td>contig02936</td>
<td>pyruvate ferredoxin flavodoxin oxidoreductase</td>
</tr>
<tr>
<td>contig02363</td>
<td>N/A</td>
</tr>
</tbody>
</table>

**Table 5.4.** Top homologous BLAST hits for tandem mass spectrometry sequencing results of band E2 against a local *R. flavipes* 454-pyrosequencing database.

<table>
<thead>
<tr>
<th>Contig</th>
<th>Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>contig00057</td>
<td>aldo-keto reductase</td>
</tr>
<tr>
<td>contig03241</td>
<td>aldo-keto reductase</td>
</tr>
<tr>
<td>contig00012</td>
<td>alcohol dehydrogenase</td>
</tr>
<tr>
<td>contig00356</td>
<td>glyceraldehyde-3-phosphate dehydrogenase</td>
</tr>
<tr>
<td>contig00538</td>
<td>alcohol dehydrogenase</td>
</tr>
<tr>
<td>contig08876</td>
<td>N/A</td>
</tr>
</tbody>
</table>
Table 5.5. Aldo-keto reductase (contig 00057) scoring matrix when aligned with tandem mass spectrometry amino acid sequencing results in band E2.

<table>
<thead>
<tr>
<th></th>
<th>b</th>
<th>b**</th>
<th>b*</th>
<th>b*++</th>
<th>b0</th>
<th>b0**</th>
<th>Seq.</th>
<th>y</th>
<th>y*</th>
<th>y**</th>
<th>y***</th>
<th>y0</th>
<th>y0**</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>58.0393</td>
<td>44.5233</td>
<td>70.0287</td>
<td>35.5180</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>201.1234</td>
<td>101.0653</td>
<td>153.1128</td>
<td>92.0600</td>
<td>I</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>255.1448</td>
<td>120.5761</td>
<td>120.7508</td>
<td>57.6337</td>
<td>G</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>257.1322</td>
<td>179.1103</td>
<td>170.1050</td>
<td>107.6292</td>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>444.2523</td>
<td>222.6263</td>
<td>242.2347</td>
<td>123.6210</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>558.2882</td>
<td>279.6477</td>
<td>541.2617</td>
<td>271.1345</td>
<td>N</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>705.3590</td>
<td>353.1819</td>
<td>688.3691</td>
<td>344.0687</td>
<td>F</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>792.3886</td>
<td>596.6980</td>
<td>775.3621</td>
<td>358.1847</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>879.4207</td>
<td>440.2140</td>
<td>862.3947</td>
<td>431.7007</td>
<td>S</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1007.4792</td>
<td>504.2433</td>
<td>990.4572</td>
<td>495.7030</td>
<td>Q</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1135.5378</td>
<td>568.2726</td>
<td>1118.5113</td>
<td>559.7593</td>
<td>L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>1248.6219</td>
<td>624.8146</td>
<td>1231.5953</td>
<td>616.3013</td>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>1377.6645</td>
<td>689.3359</td>
<td>1360.6379</td>
<td>680.8226</td>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>R</td>
<td>175.1190</td>
<td>88.0631</td>
<td>158.0924</td>
<td>79.5498</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MSVRLTNSVRRFAATVMAFKLEKTPTVKNNGIEFPIFGLGTWSKPGEVTQAVKDAIDVGYRHIDCAHVYGNEPEVGAAIKAUKGEKVKKREDLFITSKILWNFTHRPDLVAPA IQTLDLGYLDDYHLHWPMAYKEGDDL.FPEKDKTLYSDVYVDTWKELEKLVDQGLTKSIGVSNSQSLERVLANARIKPVTNQVECHPYLNQKKLI EFSKAKGVTITAYSPLGPSRDPRWAPPDPPQLLEDPKVAAKYEKTPAQILLRYQVQQGNITIPSVTKSRIVENAQIFDFELSAEDVATIDSFDCNGRNVCHLDWIKDHKDYPFNIEF
Table 5.6. Alcohol dehydrogenase 1 (contig 00012) scoring matrix when aligned with tandem mass spectrometry amino acid sequencing results in band E2

<table>
<thead>
<tr>
<th></th>
<th>b</th>
<th>b**</th>
<th>b*</th>
<th>b**</th>
<th>h0</th>
<th>h0**</th>
<th>Seq.</th>
<th>y</th>
<th>y**</th>
<th>y*</th>
<th>y**</th>
<th>y0</th>
<th>y0**</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>130.0499</td>
<td>65.5286</td>
<td>112.0393</td>
<td>56.5233</td>
<td>E</td>
<td>7</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>231.0975</td>
<td>116.0324</td>
<td>213.0870</td>
<td>107.0471</td>
<td>T</td>
<td>1255.6630</td>
<td>618.3358</td>
<td>1218.3630</td>
<td>609.0219</td>
<td>1217.6250</td>
<td>609.3290</td>
<td>11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>330.1660</td>
<td>165.5866</td>
<td>312.1554</td>
<td>156.5813</td>
<td>V</td>
<td>1134.6154</td>
<td>567.8113</td>
<td>1117.5888</td>
<td>559.2980</td>
<td>1116.6048</td>
<td>558.8060</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>431.2136</td>
<td>216.1105</td>
<td>413.2031</td>
<td>207.1052</td>
<td>10</td>
<td>1085.5469</td>
<td>518.2771</td>
<td>1018.5204</td>
<td>509.7630</td>
<td>1017.5364</td>
<td>509.2718</td>
<td>9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>559.2722</td>
<td>280.1397</td>
<td>542.2457</td>
<td>271.6265</td>
<td>10</td>
<td>931.4993</td>
<td>467.7533</td>
<td>917.4727</td>
<td>459.2400</td>
<td>916.4887</td>
<td>458.7480</td>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>610.2997</td>
<td>306.0209</td>
<td>599.2801</td>
<td>300.1414</td>
<td>G</td>
<td>895.6140</td>
<td>403.1249</td>
<td>895.4144</td>
<td>395.2101</td>
<td>888.4301</td>
<td>594.1838</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>763.3621</td>
<td>382.1847</td>
<td>746.3355</td>
<td>373.6714</td>
<td>10</td>
<td>749.4192</td>
<td>375.2132</td>
<td>732.3927</td>
<td>366.7000</td>
<td>731.4087</td>
<td>366.2080</td>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>892.4947</td>
<td>444.7060</td>
<td>875.3781</td>
<td>438.1927</td>
<td>F</td>
<td>602.3508</td>
<td>301.6790</td>
<td>585.3243</td>
<td>293.1658</td>
<td>584.3402</td>
<td>292.6738</td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1006.4476</td>
<td>503.7374</td>
<td>989.4211</td>
<td>495.2142</td>
<td>N</td>
<td>673.3082</td>
<td>327.1577</td>
<td>456.2817</td>
<td>228.6445</td>
<td>672.3082</td>
<td>327.1577</td>
<td>456.2817</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1119.5317</td>
<td>560.2695</td>
<td>1102.5051</td>
<td>551.7562</td>
<td>10</td>
<td>559.2652</td>
<td>180.1363</td>
<td>342.2387</td>
<td>171.6230</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1218.6001</td>
<td>609.8037</td>
<td>1201.5735</td>
<td>601.2904</td>
<td>10</td>
<td>246.1812</td>
<td>123.5942</td>
<td>229.1547</td>
<td>115.0810</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>147.1128</td>
<td>74.0600</td>
<td>130.0863</td>
<td>65.5468</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

MVKTKKFILAKHVFGEPKDDDFTLVEEELPPLKEGIELCEAVWFSVDPYMRPS
VNSPTGVTMIGSQVARVIESRHPDVNYGIVHVGFWQRTIFSDQPSSLWLS
KPFIIIPIFSGLPLSLALGVLGMPGNTAYFGLEICPKPGDGVVVSAGAGAVGS
VGQLAKVKGCKVIGFAGSDKVKWLTDMLKFDVAFNYKTDITAALKKAAPEG
VDCYFDNVGGTLSSAVIAQMKEGRGRISVCGISYSEDYKNIPLAPILQPSFWK
QLTMEMGFIIVRRWSDKWEEGIDYNLQLIKEGKLIHRETVTQGFENIVKAFIMGMLRG
ENVGKAVVKA
Table 5.7. Alcohol dehydrogenase 2 (contig 00538) scoring matrix when aligned with tandem mass spectrometry amino acid sequencing results in band E2

<table>
<thead>
<tr>
<th>#</th>
<th>b</th>
<th>b++</th>
<th>b*</th>
<th>b*++</th>
<th>Seq.</th>
<th>y</th>
<th>y++</th>
<th>y*</th>
<th>y*++</th>
<th>#</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>129.1022</td>
<td>65.0548</td>
<td>112.0757</td>
<td>56.5415</td>
<td>K</td>
<td>1457.9202</td>
<td>729.4638</td>
<td>1440.8937</td>
<td>720.9505</td>
<td>14</td>
</tr>
<tr>
<td>2</td>
<td>242.1863</td>
<td>121.5968</td>
<td>225.1598</td>
<td>113.0835</td>
<td>L</td>
<td>1344.8362</td>
<td>672.9217</td>
<td>1327.8096</td>
<td>664.4085</td>
<td>13</td>
</tr>
<tr>
<td>3</td>
<td>339.2391</td>
<td>170.1232</td>
<td>322.2125</td>
<td>161.6099</td>
<td>P</td>
<td>1457.9202</td>
<td>729.4638</td>
<td>1440.8937</td>
<td>720.9505</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>467.2976</td>
<td>234.1525</td>
<td>450.2711</td>
<td>225.6392</td>
<td>Q</td>
<td>1247.7834</td>
<td>624.3953</td>
<td>1230.7569</td>
<td>615.8821</td>
<td>11</td>
</tr>
<tr>
<td>5</td>
<td>538.3348</td>
<td>269.6710</td>
<td>521.3080</td>
<td>261.1577</td>
<td>A</td>
<td>1119.7248</td>
<td>560.3661</td>
<td>1102.6583</td>
<td>551.8528</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>635.3875</td>
<td>318.1974</td>
<td>618.3610</td>
<td>309.6841</td>
<td>P</td>
<td>1048.6877</td>
<td>524.8475</td>
<td>1031.6612</td>
<td>516.3342</td>
<td>9</td>
</tr>
<tr>
<td>7</td>
<td>734.4559</td>
<td>367.7316</td>
<td>717.4294</td>
<td>359.2183</td>
<td>V</td>
<td>951.6350</td>
<td>476.3211</td>
<td>934.0804</td>
<td>467.8078</td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td>847.5400</td>
<td>424.2736</td>
<td>830.5135</td>
<td>415.7604</td>
<td>L</td>
<td>852.5665</td>
<td>426.7869</td>
<td>835.5400</td>
<td>418.2736</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>975.5986</td>
<td>488.3029</td>
<td>958.5720</td>
<td>479.7897</td>
<td>Q</td>
<td>739.4825</td>
<td>370.2449</td>
<td>722.4559</td>
<td>361.7136</td>
<td>6</td>
</tr>
<tr>
<td>10</td>
<td>1072.6513</td>
<td>536.8293</td>
<td>1055.6248</td>
<td>528.3160</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1185.7354</td>
<td>593.3713</td>
<td>1168.7089</td>
<td>584.8581</td>
<td>L</td>
<td>514.3711</td>
<td>257.6892</td>
<td>497.3466</td>
<td>249.1759</td>
<td>4</td>
</tr>
<tr>
<td>12</td>
<td>1298.8195</td>
<td>649.9134</td>
<td>1281.7929</td>
<td>641.4001</td>
<td>L</td>
<td>401.2871</td>
<td>201.1472</td>
<td>384.2605</td>
<td>192.6339</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>1411.9035</td>
<td>706.4554</td>
<td>1394.8770</td>
<td>697.9421</td>
<td>L</td>
<td>288.2030</td>
<td>144.0601</td>
<td>271.1765</td>
<td>136.0919</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>R</td>
<td>175.1190</td>
<td>88.0631</td>
<td>158.0924</td>
<td>79.5498</td>
<td>1</td>
</tr>
</tbody>
</table>

MVKTKKFVLAKHFAGEPKESDLKLVEEDLPALKDREILCEAVWLSVDPYMRPYSSTLDTGVTMMGTQVAKVVESRHHEYEYVGIHVGVHFGWQJRTRTVVNADDPGAWIWLKKPYIIPDFDGQPLSLALGVLGMPGNTAYFGFLEICDPKPGEVVVSAGAGAVGSHVGQIARLKGCKVIGFTGSDDKVKLVDVLKFDAAFYKTTTDITEALKQAAPEGVDCYFDNVGTLSSAVISQMKDRGRISVCGCISIYNAFDKLKLPQAPVLQPLLLRKQLKMEGFIYSRWHDRWDEGIKQNLEWIEKEGLVYRETETEGFNMPKAFFMGMLRGENVGKAVVKV
Figure 5.1. Native-PAGE esterase gel comparing host and symbiont esterases across feedstock diets. Esterase activity is greater in the host fraction than in the symbiont fraction, with band differences occurring among the host fraction feedstock treatments. PW = Pine Wood, FP = Filter Paper, X = Beech Wood Xylan, CS = Corn Stover, SB = Soybean Hull.
Figure 5.2. Chromatograph of the Bradford protein results of each purified protein fraction collected, overlaid with the results of esterase colorimetric activity assays for each fraction. The protein concentration results correlate with the esterase activity results. Protein concentration is much greater in fraction 2 than esterase activity, suggesting that proteins besides esterases were eluted into that fraction.
Figure 5.3. Native-PAGE esterase gel containing the fractions collected during the column chromatography run. Degradation products of α-naphthyl propionate (i.e., α-naphthol) were detected in the crude homogenate and fractions 2, 8, 9, and 10. The most intense coloration of the fractions occurs in fraction 9.
Figure 5.4. SDS-PAGE gel containing the fractions collected during the column chromatography run. Separated proteins were detectable in the crude homogenate and fractions 2, 8, 9, and 10. The gel bands E1 and E2 were excised from fraction 9 while bands B1 and B2 were excised from fraction 10.
Figure 5.5. **Glucose detection results from pine wood sawdust assays.** Fraction 9, in combination with $\beta$-glu and Cell-1, released more glucose than any other treatment ($p < 0.0001$). Fraction 8 was the only other fraction to release detectable amounts of glucose from pine wood sawdust. All incubations depicted on the left half of the figure included column chromatography fractions alone and those on the right half included fractions plus the $\beta$-glu + Cell-1 mixture.
Figure 5.6. Glucose detection results from beech wood xylan assay. Fraction 9 released more glucose than any other fraction (p < 0.0001). All treatments contained detectable amounts of glucose, with fractions 8, 9, and 10 being higher than any other treatment (p < 0.0001). All incubations depicted on the left half of the figure included column chromatography fractions alone and those on the right half included fractions plus the β-glu + Cell-1 mixture.
Figure 5.7. MS/MS Fragmentation of potential aldo-keto reductase protein.

Fragment ion matches (21/144) used for scoring, which are labeled in the graph. The fragmented ions exhibited an ion score of 108 with an e-value of $10^{-10}$. 
Figure 5.8. MS/MS Fragmentation of potential alcohol dehydrogenase protein 1.

Fragment ion matches (17/116) used for scoring, which are labeled in the graph. The fragmented ions exhibited an ion score of 52 with an e-value of $4.8\times10^{-05}$. 
Figure 5.9. MS/MS Fragmentation of potential alcohol dehydrogenase protein 2.

Fragment ion matches (28/104) used for scoring, which are labeled in the graph. The fragmented ions exhibited an ion score of 60 with an e-value of $1 \times 10^{-6}$. 
BIBLIOGRAPHY


CHAPTER SIX: CONCLUSION

The research involved in the assembly of this dissertation accomplished the goals that were previously set forth, namely (1) advancing our understanding of termite digestive physiology and (2) discovering potential biocatalysts that can add to the cost-effective and energy-efficient production of biofuels from lignocellulosic biomass. Innovative glucose detection assays revealed that the host and symbiont fractions of the termite digestive tract collaborate in the digestion of woody biomass, validated previous “digestome” sequencing results, and shed light on the hierarchical importance of certain carbohydrates in the termite diet (Chapter 2). Simple glucose-localization assays, in conjunction with glucose transporter gene expression analysis, revealed that lignocellulosic degradation and transportation of degraded by-products primarily occurs in the host fraction, indicating that the termite acts independently of its symbiotic microfauna in order to acquire its own energy in the form of glucose (Chapter 3). Specific enzyme activity assays and gene expression analyses demonstrated that feedstock composition significantly impacts biochemical and molecular aspects of digestive physiology, and the contributions of various lignocellulases are fraction specific (Chapter 4). Lastly, the purification and characterization of a potential ligninase from the termite digestive tract was achieved through protein separation techniques, activity assays, peptide sequencing, and bioinformatic analysis (Chapter 5). This latter approach
was used in order to purify an accessory protein(s) that enhanced glucose liberation from lignocellulosic feedstocks when incubated with recombinant cellulases. The overall conclusion of this dissertation is that utilizing the lower termite, *Reticulitermes flavipes*, as a source of feedstock-specific recombinant enzymes for use in commercial biorefinery processes can likely decrease biocatalyst input while increasing simple sugar output, resulting in a more cost-effective approach to cellulosic ethanol production, and perhaps making biofuels a more viable option in the near future.
VITA
Zachary John Karl was born in Lakewood, Ohio, the son of John and Elizabeth Karl. After completing his schoolwork at St. Edward High School in Lakewood in 2004, Zachary went on to Wittenberg University where he studied biology and received his Bachelor of Science in May 2008. In July 2009, he entered The Graduate School at The University of Florida, and in January 2011, he joined his major professor Dr. Michael Scharf at The Graduate School at Purdue University.