A STUDY OF ECOLOGICAL AND PHYSIOLOGICAL ROLES OF THE TERMITE ASSOCIATED *VERRUCOMICROBIUM* (TAV) STRAINS THROUGH A SYSTEM BIOLOGY APPROACH

by

JANTIYA ISANAPONG

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

MAY 2013

Acknowledgements

First of all, I would like to express my sense of gratitude to Dr. Jorge Rodrigues, my supervising professor, for giving me the opportunity to learn and work in his lab. He has given me scholarly advice, and support both professionally and personally throughout my graduate study at UT Arlington. Next, I would like to thank my committee, which includes Dr. James Grover, Dr. Woo-Suk Chang, Dr. Jeff Demuth and Dr. Andrew Hunt, for their guidance during my research.

I would also like to express my thanks to Dr. Maeli Melotto, Dr. Thomas Chrzanowski, Dr. Laura Mydlarz, and staff in the Genomics Core Facility for providing laboratory facilities and being a source of information. In addition, I wish to thank my past lab member, Dr. Babur Mirza, for his guidance, friendship and for being a role model of hardwork. Also, I'd like to thank Dr. Atcha Boonmee and Blaine Thomson for giving valuable advice on protein expression and purification.

I am thankful to all my lab members, including Aditya, Sealy, Austin, Malini, Wadud, and Dr. Maeli's lab members, Nisita, Shweta, and Debanjana for their assistance, encouragement and friendship. I am grateful to my research for introducing me to amazing Thai friends, Anchana, Lita, Tai, Noom, Fon, and Nit. Thank you all for making my time here wonderful.

I would like to thank the Royal Thai Government for providing me a scholarship to pursue my degrees in the United States, and also the Department of Biology and the Department of Earth and Environmental Sciences for endowing me with a graduate teaching fellowship. Also, I am immensely thankful to all the staff for their support and assistance.

I'd also like to express my deepest gratitude towards my parents, Zin and Orawan Isanapong, my brothers and sisters for encouraging and believing in me. Lastly, I would like to wholeheartedly thank my husband, Pisist. Thank you for your motivation during the times when I was discouraged. I couldn't have come this far without your support and encouragement.

April 17, 2013

Abstract

A STUDY OF ECOLOGICAL AND PHYSIOLOGICAL ROLES OF THE TERMITE ASSOCIATED *VERRUCOMICROBIUM* (TAV) STRAINS THROUGH A SYSTEM BIOLOGY APPROACH

Jantiya Isanapong, Ph.D

The University of Texas at Arlington, 2013

Supervising Professor: Jorge L.M Rodrigues

Termite hindguts are populated by a dense and diverse community of microbial symbionts working in concert to transform lignocellulosic plant material and derived residues into acetate, to recycle and fix nitrogen, and to remove oxygen. Although much has been learned about the breadth of microbial diversity in the hindgut, the ecophysiological roles of its members is less understood. In this study, we present new information about the ecophysiology of microorganism *Diplosphaera colotermitum* strain TAV2, an autochthonous member of the *Reticulitermes flavipes* gut community. An integrated high-throughput approach was used to determine the transcriptomic and proteomic profiles of cells grown under hypoxia (2% O_2) or atmospheric (20% O_2) concentrations of oxygen. Our results revealed that genes and proteins associated with energy production and utilization, carbohydrate transport and metabolism, nitrogen fixation, and replication and recombination were up-regulated under 2% O_2. The metabolic map developed for TAV2 indicates that this microorganism may be involved in biological nitrogen fixation, amino acid production, hemicellulose degradation, and consumption of O_2 in the termite hindgut. Variation of O_2 concentration explained 55.9% of the variance in proteomic profiles, suggesting an adaptive evolution of TAV2 to the hypoxic periphery of the hindgut. Our findings advance the current understanding of microaerophilic microorganisms in the termite gut and expand our understanding of the ecological roles for members of the phylum

Verrucomicrobia. Moreover, one of the enzymes in xylanolytic system involved in hydrolyzing acetyl group from hemicellulose xylan, acetyl xylan esterase (AXE), was found to be up-regulated from both transcriptomic and proteomic analyses, so the *axe* gene has been cloned and expressed into *E. coli* BL 21(DE3), and then the AXE enzyme was purified. From the enzyme activity assays, the enzyme shows activity toward acetylated xylan by releasing acetyl group from the xylan backbone. The optimum temperature and pH of the enzyme was at 50^{o}C and pH 7.0, respectively. The ability of the AXE in deacetylating xylan suggests the roles of AXE in open up the xylose backbone for other enzymes like xylanase, xylosidase, and cellulase to access and degrade hemicellulose and cellulose.

Table of Contents

Acknowledgements iii

Abstract…… iv

List of Illustrations x

List of Tables… xii

Preface…… xiii

Chapter 1 Introduction 1

1.1 Composition of Plant Cell Wall 2

1.2 Cellulase Enzymes 4

1.3 Roles of Termites in Biogeochemical Cycle 5

1.4 Physicochemical Characteristics of Termite's Hindgut 7

1.4.1 Oxygen Level 8

1.4.2 Hydrogen Partial Pressure 8

1.5 Cellulose Digestive System of Termites 9

1.5.1 Foregut and Midgut 9

1.5.2 Hindgut 10

1.6 Metabolic Activities in the Hindgut 12

1.6.1 Fermentation 12

1.6.2 Reductive Acetogenesis 13

1.6.3 Methanogenesis 14

1.6.4 Acetate Formation from Lactate 15

1.6.5 Nitrogen Metabolism 15

1.7 The Phylum *Verrucomicrobia* 16

1.8 Termite Associated *Verrucomicrobium* (TAV) 18

1.9 Why the Termite Associated *Verrucomicrobium* (TAV) Strains? 19

1.9.1 Phylogenetic and Evolutionary Gradients in these Five TAV Strains Allow us to Study Genomic and Functional Variations among Closely Related Species. ... 19

1.9.2 The Genome Sequence of TAV 1 and TAV2 Reveals Important Genes for Nitrogen Metabolism, Carbon Metabolisms, and Ability to Control Oxygen Levels in the Hindgut. 21
Chapter 2 High-Quality Draft Genome Sequence of the *Opitutaceae* Bacterium Strain TAV1, a Symbiont of the Wood-Feeding Termite *Reticulitermes flavipes* 23
2.1 Introduction 23
2.2 Materials and Methods 23
2.3 Results 24
Chapter 3 Development of an Ecophysiological Model for *Diplosphaera colotermitum* TAV2, a Termite Hindgut *Verrucomicrobium* 26
3.1 Introduction 26
3.2 Materials and Methods 27
3.2.1 Culture Conditions 27
3.2.2 Microarray Design 28
3.2.3 Microarray Hybridization 28
3.2.4 Transcriptomic Data Analysis 29
3.2.5 Protein Extraction 29
3.2.6 Proteomic Data Analysis 30
3.2.7 Metabolic Reconstruction 31
3.2.8 Reverse-Transcription Quantitative PCR (RT-qPCR) 31
3.3 Results 31
3.3.1 Confirmation of the Microaerophilic Phenotype of TAV2 31
3.3.2 Transcriptome Analysis 32
3.3.3 Proteome Analysis 32
3.3.4 Transcriptomic and Proteomic Data Comparison 35
3.3.5 Metabolic Reconstruction 36
3.3.6 Validation of Expressed Genes through RT-qPCR 41

3.4 Discussion 41
3.4.1 A working model for Diplosphaera colotermitum strain TAV2 in the termite hindgut 46
Chapter 4 Cloning, Expression and Purification of Acetyl Xylan Esterase from the *Diplosphaera colotermitum* Strain TAV2 48
4.1 Introduction 48
4.2 Materials and Methods 50
4.2.1 Plasmid, Bacterial Strains, and Culture Conditions 50
4.2.2 Preparation of axe Gene Clones 50
4.2.3 Transformation of axe-pET-28a into E. coli 51
4.2.4 Confirmation of Cloned Insert and the Presence of Histidine Tag 51
4.2.5 Time-Course Analysis for Optimal Protein Expression 52
4.2.6 Culture Condition for Protein Expression 53
4.2.7 Histidine-Tagged Protein Purification 53
4.2.8 SDS-PAGE Analysis 54
4.2.9 Western Blot Analysis 54
4.2.10 Enzyme Assay 55
4.3 Results and Discussion 56
4.3.1 Nucleotide and Amino Acid Sequences of Acetyl Xylan Esterase 56
4.3.2 Confirmation of Cloned Insert 57
4.3.3 Determination of Optimal Protein Production 59
4.3.4 Purification of Acetyl Xylan Esterase 60
4.3.5 Characterization of Acetyl Xylan Esterase 62
4.4 Future Direction 65
Research Conclusions 66
Appendix A Family *Opitutaceae* 68

Appendix B List of Genes and Proteins that were Significantly Up- or Down-Regulated during Growth of *Diplosphaera colotermitum* Strain TAV2 Cells under 2% O_2 as Compared to Cells Grown under 20% O_2 82

Appendix C Scatter Plots Comparisons of Transcriptomic and Proteomic Whole Cell Profiles of *Diplosphaera colotermitum* under 2% and 20% O_2 Concentrations. 90

Appendix D Fold-Change Results for Selected Genes or Proteins with RT-qPCR, Microarray, and Log_2-Transformed Peptide Numbers for the *Diplosphaera colotermitum* Strain TAV2 92

Appendix E Auto-Induction Reagent Lists 94

References…. 97

Biographical information 111

List of Illustrations

Figure 1.1 Structure of cellulose 2

Figure 1.2 Structure of hemicellulose 3

Figure 1.3 Phenolic aromatic structure of lignin 3

Figure 1.4 Types of cellulase enzymes. 5

Figure 1.5 Type of termites and their digestive tract. 7

Figure 1.6 Oxygen and hydrogen gradients in the termite hindgut. 9

Figure 1.7 Overall metabolic activities in the termite hindgut. 12

Figure 1.8 Phylogenetic tree of five TAV strains 20

Figure 3.1 Heat map comparisons of calculated Z-scores for proteins differentially expressed by TAV2 cells grown under 2% or 20% O_2. 33

Figure 3.2 Relative proportions of unique proteins assigned to functional categories based on cluster of orthologous groups (COG). 34

Figure 3.3 Principal component analysis of the proteome results for TAV2 cells grown under 2% or 20% O_2. 35

Figure 3.4 Metabolic map of *Diplosphaera colotermitum* strain TAV2 constructed with selected differentially expressed genes and proteins. 37

Figure 3.5 Model representation of possible functional roles for the *Diplosphaera colotermitum* strain TAV2 in the hindgut of the termite *Reticulitermes flavipes*. 46

Figure 4.1 Confirmation of cloned insert, *axe* gene in pET-28a, using restriction enzyme digestion. 58

Figure 4.2 SDS-PAGE [8-12% (w/v)] of time course for optimal protein induction (A), and comparison of induced proteins from the *E. coli* BL21 (DE3) containing empty plasmid and *axe*-inserted plasmid (B). 59

Figure 4.3 Gravitational flow purification of AXE protein using Ni-NTA resin (A), and purified AXE eluted from cation exchange column (B). 61

Figure 4.4 Western blot analysis of AXE. 61

Figure 4.5 Optimum temperature of AXE toward acetylated xylan 62

Figure 4.7 Optimum pH of AXE toward acetylated xylan 64

Figure A - 1 Maximum likelihood genealogy reconstruction based on the 16S rRNA gene from isolated members of the family *Opitutaceae*. 70

Figure C- 1 Log_2 transformed normalized scatter plots comparisons of (A) transcriptomic and (B) proteomic whole cell profiles of *Diplosphaera colotermitum* under 2% and 20% O_2 concentrations. 91

List of Tables

Table A - 1 Genomic properties of members of the family *Opitutaceae* 72

Table A - 2 Correlation coefficients of tetranucleotide usage patterns (upper triangle) and % average nucleotide identity (lower triangle) calculated between *Opitutaceae* genomes. 73

Table A - 3 Comparison of selected phenotypic characteristics of the family *Opitutaceae*. 77

Table B - 1 List of genes and proteins that were significantly up- or down-regulated under 2% and 20% O_2 83

Table D- 1 Fold-change results for selected genes or proteins with RT-qPCR, microarray, and log_2-transformed peptide numbers for the *Diplosphaera colotermitum* strain TAV2. Locus IDs were selected based on fold induction or number of peptides detected. . 93

Preface

This dissertation is proposed for a Doctor of Philosophy degree at the University of Texas at Arlington in spring 2013. The following chapters in this dissertation have been presented in the following journals and book.

Chapter 2

Isanapong J, Goodwin L, Bruce D, Chen A, Detter C, Han J, Han CS, Held B, Huntemann M, Ivanova N, Land ML et al. 2012. High-quality draft genome sequence of the Opitutaceae Bacterium Strain TAV1, a symbiont of the wood-feeding termite *Reticulitermes flavipes.* Journal of Bacteriology 194(10):2744-2745. Permission to reuse the content in this dissertation was already obtained.

Chapter 3

Jantiya Isanapong, W. Sealy Hambright, Austin G. Willis, Atcha Boonmee, Stephen J. Callister, Kristin E. Burnum, Ljiljana Paša-Tolić, Carrie D. Nicora, John T. Wertz, Thomas M. Schmidt, Jorge L.M. Rodrigues. 2013. Development of an ecophysiological model for *Diplosphaera colotermitum* TAV2, a termite hindgut *Verrucomicrobium.* The ISME Journal (*in press*).

Appendix A

Rodrigues J, Isanapong J. 2013. Family *Opitutaceae*. In: Rosenberg E, Stackebrant E, Thompson F, DeLong E, Lory S, editors. The Prokaryotes. 4th ed. New York, NY (*in press*): Springer.

Chapter 1

Introduction

Petroleum based fossil fuel is a finite source of energy. Burnings of fossil fuels create greenhouse gases like carbon dioxide (CO_2), methane (CH_4) in the atmosphere, which can cause global warming and climate change phenomena. The Intergovernmental Panel on Climate Change (IPCC, 2007) has revealed that major global greenhouse gases are CO_2 from fossil fuel use (57%), and deforestation and decay of biomass (17%), CH_4 (14%), N_2O (8%). Main sources of emission are from energy supply (26%), industry (19%), and the rests come from forestry, agricultural and transportation activities (Metz *et al.*, 2007). Globally, combustion of fossil fuel has significantly increased the atmospheric CO_2 concentration since 1900. The emissions have been raised by over 16 times (from 2,000 to 32,000 tetragrams) from 1900 to 2008 (Boden, 2010). One of the strategies to solve these problems is to produce a renewable energy such as ethanol. Ethanol is considered as a clean energy comparing to gasoline because it contains about 35% of oxygen. This oxygenated fuel can decrease emission of CO_2, particulate matters, and oxide of nitrogen (NO_x) from the combustion process of fossil fuel (Badger, 2002).

The most abundant source for renewable fuel production is plant biomass. Annually, plants yield 136 x 10^{15} grams of dry plant materials (Ljungdahl and Eriksson, 1985). Food crops, such as corn, soybean, and sugar cane, which are rich in starch and/or sugar contents, are currently major sources for ethanol production. However, extensive use of these crops could increase food prices (Farrell *et al.*, 2006); moreover, corn kernel contains 60% of starch, which is considered as inefficient starting material for ethanol production (Somerville, 2007). On the other hand, non-food crops and low value plants, such as wood chips, leaf litters and grass, contain up to 90% of degradable portion which can be converted to sugar, a starting material for fermentation. In term of ethanol production process, an environmental unfriendly acid hydrolysis

(using sulfuric acid or phosphoric acid) is a widely used method for converting starchy and cellulosic materials into sugars. For this reason, studies of effective utilization of low value cellulosic materials using environmental friendly microbial enzymatic hydrolysis are of increasing interest.

1.1 Composition of Plant Cell Wall

Higher plants can fix carbon dioxide from atmosphere into their cells via photosynthesis. Annually, they can fix 120 $*10^{12}$ grams of carbon and use it to build their structures such as cell wall (Ciais *et al.*, 1997). Plant cell wall is a resilient and hardly degradable part. It consists of three major components, cellulose, hemicellulose, and lignin (Thompson, 1983). The complex form of cellulose, hemicelluloses, and lignin is called together as lignocellulose (Faulon *et al.*, 1994).

Cellulose (28-50%) is linear carbohydrate homo-polymers of several anhydrous glucose units $(C_6H_{10}O_5)_n$ which are usually formed as a long chain (n= 2,000-26,000). Each glucose monomer is polymerized by β-1,4-glucosidic bond. Beta linkage is a strong bond and characterized as ether oxygen in C_1 position is pointed in the same position as C_6, and thus forms an un-branched, rigid crystalline structure which is resistant to cellulolysis (Ophardt, 2003).

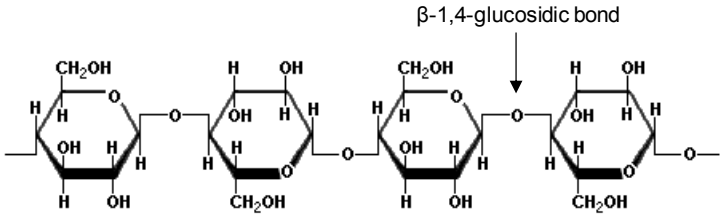

Figure 1.1 Structure of cellulose

Picture adapted from http://www.scientificpsychic.com/fitness/carbohydrates2.html

Hemicellulose (20-30%) is well known as xylan. It composes of hetero-polymers of carbon sugars, such as xylose (majority), mannose, arabinose, galactose, rhamnose, and uronic

acids. These matrix components also have β-1,4-linked backbone with amorphous, short branches of polysaccharides (up to 3,000 units) distributing between cellulose microfibrils.

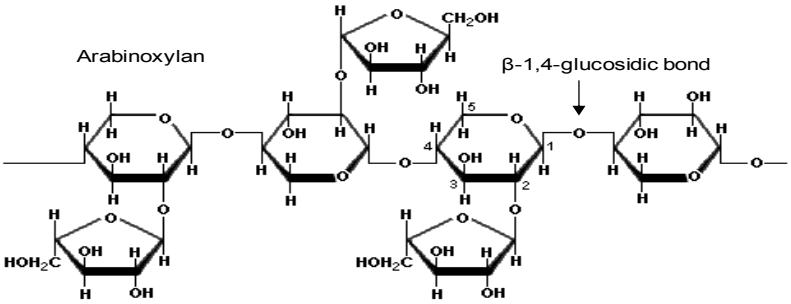

Figure 1.2 Structure of hemicellulose

Picture adapted from http://www.scientificpsychic.com/fitness/carbohydrates2.html

Lignin (18-30%) is phenolic aromatic polymer linked with other components resulting in a strong meshwork. Lignin is formed in lignification process when plants require strong cell walls during their growth (Watanabe and Tokuda, 2010).

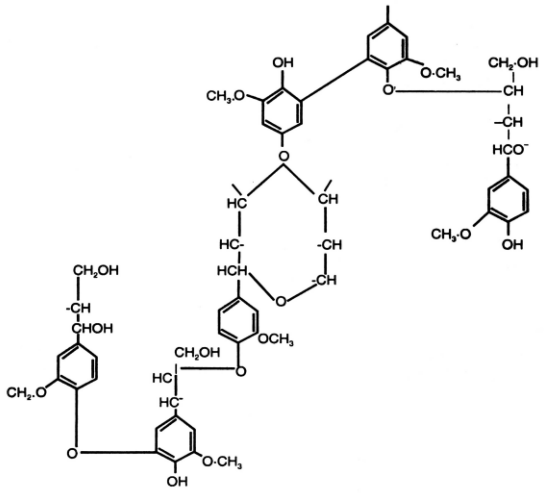

Figure 1.3 Phenolic aromatic structure of lignin

Picture adapted from http://www.eplantscience.com/index_files/biotechnology

1.2 Cellulase Enzymes

Most animals and vertebrates have enzymes to digest alpha-linkage containing compound like starches, but they lack enzymes to breakdown beta-linkage of lignocellulose. On the other hand, some animals and insects like termites, cows, horses, sheep, are able to breakdown rigid structures of beta-linkage lignocellulose-containing compounds using cellulase enzymes. These cellulases are mostly produced from symbiotic microbes (protists, bacteria, and fungi) living in host's digestive tract. Cellulases hydrolyze β-1,4-glycosidic bonds in lignocellulose and release molecules of sugar monomers. In general, there are three categorized cellulase enzymes based on substrate specificity and enzymatic action as shown in Figure 1.4 (Watanabe and Tokuda, 2010). 1) endoglucanases (endo-β-1,4-glucanases or 1,4-β-D-glucan-4-glucanohydrolases) randomly hydrolyze β-1,4 bonds from inside of cellulose chain and release short termini chain like cellotriose, cellobiose, and small amount of glucose. 2) exoglucanases (1,4-β-D-glucan cellobiohydrolases or 1,4-β-D-glucan glucohydrolases) cleave the short termini from outside of both reducing and non-reducing ends of cellulose chain. The enzyme cellobiohydrolases release cellobiose (two glucose molecules linkage) and glucohydrolases release glucose. 3) β-glucosidases(1,4-β-D-glucosidases or cellobiases) hydrolyze cellobiose or cello-oligomers to glucose molecules from reducing ends. The synergetic activities of various types of cellulases with a wide range of substrate specificities contribute to complete cellulose degradation process. Another cellulolytic system found in anaerobic bacteria like *Clostridium* sp. is a cellulosome. The cellulosome consists of multiple catalytic modules and carbohydrate-binding module (CBMs) on a cellulose-integrating protein (CIP). The cellulosome can be found anchored on S-layer of bacterial cell walls or found independently within the cells.

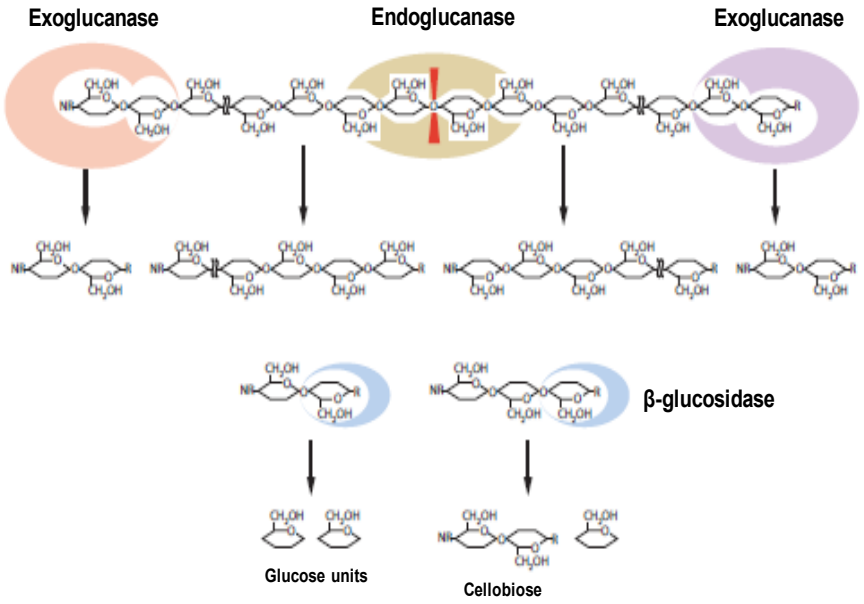

Figure 1.4 Types of cellulase enzymes.

R, reducing end; NR, nonreducing end. Pictured adapted from (Watanabe and Tokuda, 2010).

Cellulases have been grouped into glycoside hydrolase (GH) family based on amino acid sequence similarity and substrate specificity. The GHs are enzymes hydrolyzing the glycosidic bonds of carbohydrates and non-carbohydrates (Cantarel *et al.*, 2009). In term of structure, many of these enzymes have a CBM linked to a catalytic module. It is believed that CBMs enhances activity of catalytic domain by binding firmly on the cellulose chain during hydrolysis, then moving to the next available positions with the aid of processive enzyme. The combination of these modules and the processive enzyme are released from microorganism and their role is to hydrolyze the crystalline region of cellulose, while the one released from host termites commonly hydrolyze amorphous region.

1.3 Roles of Termites in Biogeochemical Cycle

Termites are a small to medium sized, colorless insect. They have a strong mouth path for chopping recalcitrant plant materials into small pieces. They are one of the insects developing social behavior in a caste system: king and queen, soldiers, and workers, which each individual

group has its own duty. Among the members, only the worker termites can actually digest wood and provide the partially digested food to other members in a colony via regurgitated or fecal pellets (http://www.amentsoc.org/insects/fact-files/orders/isoptera.html). It was estimated that two-third of land surface worldwide was occupied by one of the termite species. They are more abundant in tropical and subtropical environments, and less abundant in temperate regions. Termites play important roles in global carbon cycle as a decomposer of ecological system. They are well known in digesting dead plants, decaying woods, grass, and leaf litter within terrestrial ecosystems. Annually, all termites worldwide can consume significant amount of lignocelluloses up to 3-7x10^5 grams (Collins and Wood, 1984; Khalil *et al.*, 1990) . For this reason, they are considered as one of the most important lignocellulose-degrading insect.

Termites are classified in the order of Isoptera. Seven families and 2600 species have been recognized (Engel and Krishna, 2004). Based on termite's molecular phylogenetic and physiological studies, feeding habits, and lifestyle, they can be categorized into two major groups, lower and higher termites (Ohkuma, 2008). The lower termites compose of six families, Hodotermitidae, Kalotermitidae, Mastotermitidae, Rhinotermitidae, Serritermitidae, and Termopsidae (Hongoh, 2011). All families are wood-feeding termites, except the members of Hodotermitidae, which mainly feed on dead grasses. In general, the lower termites have a simple digestive tract consisting of foregut, midgut, and hindgut as shown in Figure 1.5. The hindgut has a single dilated paunch which is the site where microbial symbionts like protists, archaea, and bacteria reside. Many lower termites nest the mound and feed within the same food source indicating their restricted and limited food source. This is an indication that the lower termites have highly efficient digestive system in which they can almost completely degrade cellulose and hemicellulose as well as their outstanding ability in recycling nutrients (Ohkuma, 2008). The higher termites contain only one family of Termitidae but members of this family cover 70% of all termite species (Engel and Krishna, 2004). The higher termites can be grouped into four subfamilies, Apicotermitinae, Macrotermitinae, Nasutitermitinae, and Termitinae (Hongoh, 2011). Contrary to the lower termites, the higher termites feed on various food sources, such as wood,

litter, grass, soil, and lichen. Most of them forage for food in different places outside their nest. The higher termites have several gut paunches. The main difference between the lower and higher termites is that the higher termites lack of flaggelated protists in their paunch.

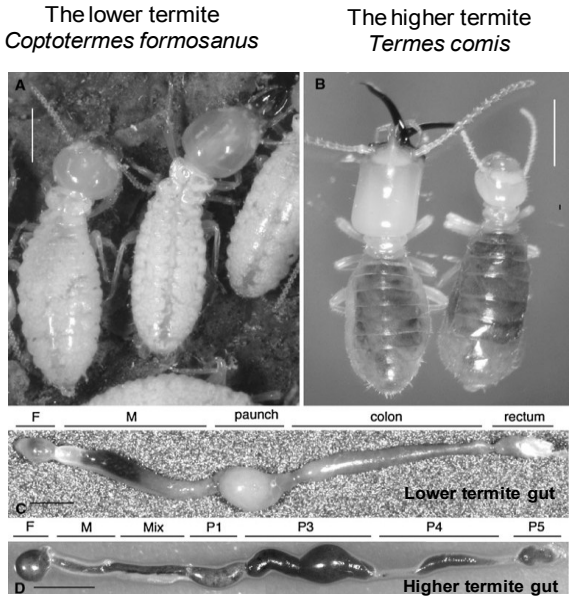

Figure 1.5 Type of termites and their digestive tract.

A, the lower termites *Coptotermes formosanus*; B, the higher termites *Termes comis*; C, digestive tract of *C. formosanus*; D, digestive tract of *T. comis*. F, foregut; M, midgut; P, paunch. Picture adapted from (Hongoh, 2011)

1.4 Physicochemical Characteristics of Termite's Hindgut

Since this research focuses on bacteria isolated from lower termite, the following information will target on major physicochemical characteristics of the lower termite only. The hindgut of lower termite has a single dilated paunch. This dilation increases the residence time of cellulose in the hindgut, and thus facilitates its degradation by increasing the contact time with microbes.

1.4.1 Oxygen Level

Studies of oxygen level in the hindgut of *Reticulitermes flavipes* using an oxygen microelectrode have revealed that the hindgut is not completely anoxic but 60% of the hindgut contains oxygen at peripheral region (approximately 150-200 μm below the epithelial surface), and 40% at the gut core is completely anoxic (Brune *et al.*, 1995). Creating this environmental condition is not easy since the hindgut is small and its surface area to volume ratio is high, so O_2 concentration in the gut tends to reach equilibrium with surrounding environment. For this reason, efficient O_2 barriers are extremely important. Microbes living in the cuticle or a cup-like indentation of the gut peripheral (Breznak and Pankratz, 1977) are believed to take up oxygen penetrating through epithelial tissue for their respiration. By doing so, they create a microoxic periphery around the anoxic core as shown in Figure 1.6 (Brune, 1998). In the paunch of *R. flavipes*, it was estimated that O_2 used for microbial respiratory activity in the hindgut tissue accounted for 21% of the termite's respiration (Brune *et al.*, 1995). On the contrary to hindgut, the center of midgut and posterior region of the hindgut are completely aerobic by containing considerable amount of O_2 up to 50 and 30% air saturation, respectively (Brune *et al.*, 1995). The hindgut microenvironments are spatially separated as well as the microbes distributed in each region. This is evidenced by the presence of various types of microbes, such as anaerobic bacteria and protozoa, strictly aerobic and facultative anaerobic bacteria in the gut.

1.4.2 Hydrogen Partial Pressure

Several studies have measured the partial pressure of H_2 using microelectrodes in agarose-embedded paunch of lower termites, they found that the H_2 partial pressure is very high up to 30–72 kPa in *Zootermopsis nevadensis*, 15–30 kPa in *Reticulitermes santonensis*, and 2–5 kPa in *R. flavipes* (Ebert and Brune, 1997; Pester and Brune, 2007). The hydrogen gradient is opposite to O_2 gradient as shown in Figure 1.6. In *R. flavipes,* hydrogen partial pressure is found to be higher at the center (about 4.5 kPa) and decreases to 1 kPa at the gut wall.

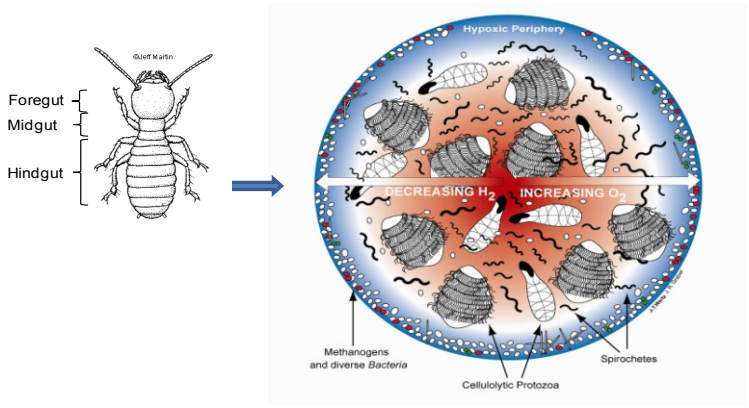

Figure 1.6 Oxygen and hydrogen gradients in the termite hindgut.

Picture adapted from http://www.desertmuseum.org/books/nhsd_termites.html

1.5 Cellulose Digestive System of Termites

While many insects cannot degrade recalcitrant lignocellulose-containing woods, termites do. Termites can breakdown 74-99% of cellulose and 65-87% of hemicellulose (Breznak and Brune, 1994) to oligosaccharides and short chain fatty acids, and use these compounds as important energy sources. Wood-feeding lower termites contain a simple but highly effective decomposing system. They have developed a proficient symbiotic lignocellulose degradation involving both host termite and their gut microbiota. Termites have evolved two independent lignocellulose-digesting sites, 1) the foregut and midgut, and 2) the hindgut. Cellulose degradation in the foregut and midgut is mainly accomplished by endogenous cellulases secreted from termite itself. In the hindgut, cellulose degradation is largely contributed to cellulase enzymes produced from microbial symbionts such as flagellated protists and bacteria living in the hindgut (Brune, 2006; Fujita *et al.*, 2010).

1.5.1 Foregut and Midgut

After woods are ingested, termites masticate wood particles in the foregut using mechanical grinding structures like mandibles and cuticular of the gizzard (Fujita *et al.*, 2010).

Mastication (chewing) can help increase surface area, thus cellulose fibers will be more exposed and easily accessed by endogenous cellulases secreted from salivary glands. A study in the lower termite *Coptotermes formosanus* has shown that average size of wood particles in the foregut is about 20 μm which is larger than those in the midgut (10 μm). This result implies the role of gizzard in grinding the ingested wood particles into smaller pieces (Fujita *et al.*, 2010). Termites can partially digest woods using their own enzymes called endogenous cellulases which are endo-β-1,4-glucanase (endoglucanase) and β-glucosidase secreted from the salivary gland and/or midgut (Watanabe and Tokuda, 2010). In general, functional digestion of endoglucanase is to digest amorphous region of lignocelluloses, and the glucose released from digestion will be absorbed by termite. The enzyme endoglucanase is accounted for 6% of total soluble proteins in the midgut of *C. formosanus* (Fujita *et al.*, 2010). Based on glycoside hydrolase classification, the endogenous endoglucanases are members of glycoside hydrolase family 9 (GHF9) (Hongoh, 2011), which is usually found in many insects like termites, cockroaches, beetles, crickets, and etc. (Watanabe and Tokuda, 2010).

1.5.2 Hindgut

The termite hindgut is a site where cellulose degradation and absorption of nutrients occur. Complete cellulose degradation is accomplished because termites and microbes present in its hindgut have developed a symbiotic relationship, in which termites grind woods into small pieces, and then gut microbial symbionts degrade cellulosic materials and transform them into acetate (Ohkuma, 2003). The symbiotic relationship is an obligatory one and termites fed with antibiotics will have their overall cellulose degradation activities significantly decreased (Tokuda and Watanabe, 2007). Gut microbes like protists, bacteria, and archaea are found to colonize and are structurally distributed in the hindgut. Gut protists are previously well known for their important roles in degrading cellulose by endocytosis followed by catalyzing through their cellulase enzymes. Protist cells of *R. speratus* have predominant genes encoding GHF 7, 5, 45, β-glucosidase and several hemicellulase (Todaka *et al.*, 2007). Microbial cellulase enzymes

secreted in the hindgut play important roles in digesting crystalline portion of the cellulose, while endogenous cellulase secreted by termite partially digests amorphous regions of cellulose fibers (Hongoh, 2011).

Not only the gut protists play significant roles in cellulolytic digestion system in the hindgut, but also gut prokaryotes being responsible for the vitality of termites (Schultz and Breznak, 1978). The abundance of gut prokaryotes can level up the persistence of gut protists and the termite nutrition (Breznak *et al.*, 1973) because several prokaryotes can fix nitrogen for the termite host. Moreover, their localization at the gut epithelium might help with the biochemical interaction (Breznak and Pankratz, 1977), for example, creating a microoxic condition in the hindgut facilitates acetate formation instead of lactate.

Prokaryotes living in the hindgut can be broadly categorized into three groups: prokaryotes having cellular association with gut protist, prokaryotes living in the gut wall, and free living prokaryotes. Prokaryotes living intracellular or inside the protists are called endosymbionts which are bacteria in phyla *Bacteroidetes*, *Firmicutes*, *Proteobacteria*, and etc. (Hongoh, 2011). Ectosymbionts are prokaryotes commonly living on protists' cell surfaces and usually have motility symbiosis with protists; for example, bacterial members in the phyla *Bacteroidetes*, *Spirochaetes*, and *Synergistetes* (Hongoh, 2011). Prokaryotes attaching to the gut wall account for 3-20% of total gut prokaryotes, and the free-living prokaryotes account for relatively small portion (Nakajima *et al.*, 2005). In terms of microbial diversity, it is believed that there are about 250 microbial species in 1 µL of hindgut fluid (Nadin, 2007). The hindgut of *R. flavipes* harbors at least six species of flagellates, 20–30 different bacterial morphotypes, and 5 x 10^8 viable bacteria per ml of hindgut fluid. The majority (63.3%) of the bacteria is anaerobic heterotrophic bacteria named *Streptococci* (now *Lactococci*) which are *S. lactis* and *S.cremoris*. Another 27.9% is *Bacteroides* and *Enterobacteriaceae* and minority of the isolates are *Lactobacillus* and *Fusobacterium* (Schultz and Breznak, 1978). However, 90% of prokaryotes in the hindgut is uncultivable, so many studies of microbial diversity has employed molecular-based methods to investigate microbial diversity. Analysis of 16S rRNA genes of lower termite *Cryptotermes*

domesticus revealed that the *Methanobrevibacter*, *Leuconostoc*, *Bacteroides* and *Treponema* were major groups discovered (Ohkuma and Kudo, 1998).

1.6 Metabolic Activities in the Hindgut

The anoxic environment of the termite hindgut is extremely important for microbial communities to carry out their functions, such as cellulose degradation, nitrogen fixation, acetogenesis, and methanogenesis. These metabolic activities could not occur in the presence of oxygen.

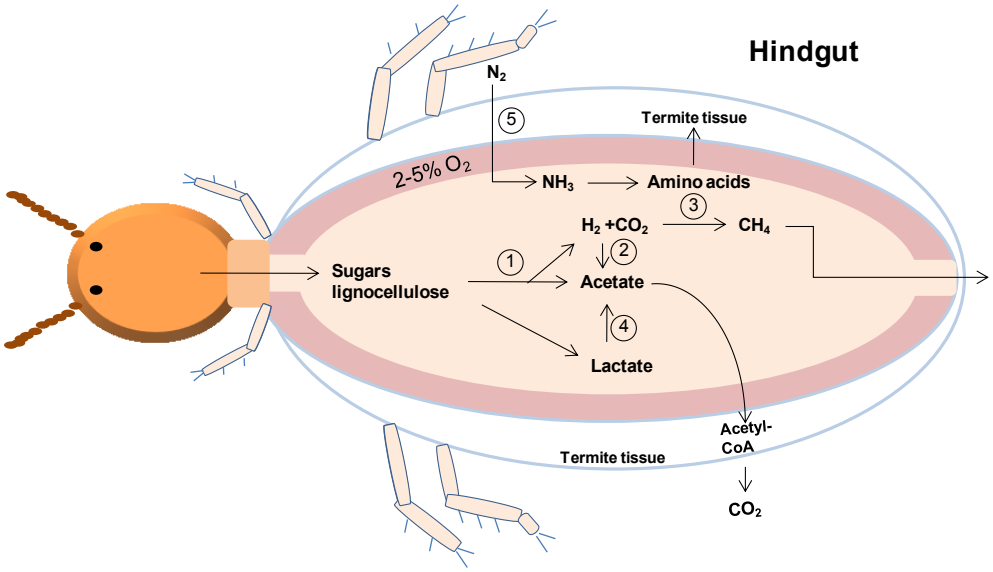

Figure 1.7 Overall metabolic activities in the termite hindgut.

Not drawn to scale

1.6.1 Fermentation

Partially digested cellulose and hemicellulose are moved from the midgut to hindgut. These cellulosic materials are then phagocytosed into protist's food vacuoles and further catalyzed by cellulase enzymes secreted from the gut protists and anaerobic bacteria living at the anaerobic center region. The released oligosaccharides are fermented to acetate, H_2, and CO_2 as

shown in Equation 1-1. Acetate is further absorbed through termite tissue and termite uses it as a main energy and carbon source. The fermentation end products, H_2, and CO_2, are then used by other microbes.

$$C_6H_{12}O_6 + 2H_2O \rightarrow 2CH_3COOH + 4H_2 + 2CO_2 \qquad \text{Equation 1-1}$$

The termite hindgut contains a high concentration of hydrogen. There is evidence that both gut protists and bacteria are responsible for H_2 production. Inoue et. al, (2007) found genes encoding for Fe-hydrogenase in the gut of the lower termite *C. formosanus* and they were identified as belonging to the protist *Pseudotrichonympha grassii* (Inoue *et al.*, 2007). Warnecke and co-authors (2007) studied the metagenome of the higher termite *Nasutitermes ephratae*, and they found genes encoding for Fe-hydrogenases identified as bacteria within the genus *Treponema* (Warnecke *et al.*, 2007). Their result suggests that gut protists are significant H_2-producers in the lower termites (Hongoh, 2011), while treponemes are main H_2 producers in wood-feeding higher termites.

1.6.2 Reductive Acetogenesis

$$4H_2 + 2CO_2 \rightarrow CH_3COOH + 2H_2O \qquad \text{Equation 1-2}$$

Acetogenesis and methanogenesis are the metabolic processes acting as H_2 sinks in the hindgut. Because accumulated fermentation byproducts especially H_2 can suppress fermentation, the produced H_2 has to be removed rapidly to prolong fermentation. Therefore, H_2 is considered as a key intermediate in the gut community. Various anaerobic bacteria, acetogens, and ectosymbionts (spirochetes) especially *Treponema primitia*, are in charge of reductive acetogenesis by utilizing H_2 and CO_2 to produce acetate as shown in Equation 1-2 (Graber and Breznak, 2004; Leadbetter *et al.*, 1999). This metabolic process can generate one-third of total

acetate production in the gut (Pester and Brune, 2007). Bacteria like treponemes utilize a key enzyme of reductive acetogenesis, formyl tetrahydrofolate synthetase (FTHFS), in the Wood-Ljungdahl pathway in order to synthesize acetate (Pester and Brune, 2006; Salmassi and Leadbetter, 2003). In general, H_2-dependent acetogens surpass H_2-dependent methanogens in termite gut because acetogens can utilize both H_2 and organic compounds as their substrates, while methanogens cannot. In addition, due to their difference in specific localization, acetogens usually inhabit in the central region by attaching onto surface of protists or inhabit in the gut fluid where H_2 concentration is highest (Ebert and Brune, 1997). Thus, they have more potential to utilize H_2 than methanogens living in the gut peripheral where H_2 concentration is low (Pester and Brune, 2007).

1.6.3 Methanogenesis

$$4H_2 + CO_2 \rightarrow CH_4 + 2H_2O \qquad \text{Equation 1-3}$$

Methanogens are generally colonized at the hindgut wall or inside the cells of several protists (Inoue *et al.*, 2008). In the hindgut of *R. flavipes*, methanogenic archaea are densely localized at the peripheral region of the hindgut and three strains of methanogens have been isolated: *Methanobrevibacter cuticularis*, *Methanobrevibacter curvatus* (Leadbetter and Breznak, 1996), and *Methanobrevibacter filiformis* (Leadbetter *et al.*, 1998). Methanogenic archaea requires H_2 and CO_2 to produce methane as shown in Equation 1-3. For this reason, they usually compete with acetogens for H_2. However, homoacetogens generally outcompete methanogens as the reasons mentioned above; thus, methane emission rate is only 10% of the reductive acetogenesis (Pester and Brune, 2007). By adding more exogenous H_2, the CH_4 emission rate was significantly increased, while the reductive acetogenesis from $^{14}CO_2$ was unchanged. This is an indication of H_2 limitation of the methanogens living at the gut pheripheral region (Ebert and Brune, 1997).

1.6.4 Acetate Formation from Lactate

$$2C_3H_6O_3 \rightarrow 3CH_3COOH + 2CO_2 + H_2 \qquad \text{Equation 1-4}$$

Various organic metabolites such as acetate, lactate, and formate are found in the termite hindgut. Among these, acetate is the major metabolite produced accounting for 86 mol% in the hindgut fluid of *R. flavipes* (Tholen and Brune, 2000). Besides acetate, lactate is one of the important intermediate by accounting for one-third of total carbon flux of hindgut metabolism (Tholen and Brune, 2000). In *R. flavipes*, lactate is produced from transformation of cellobiose and xylose into lactate by lactic acid bacteria (Tholen *et al.*, 1997). However, rapid turnover of lactate into acetate can contribute to low concentration of lactate in the hindgut. Influx of O_2 in the hindgut can affect on carbon metabolism and electron flow. In the presence of O_2 especially in the peripheral region, *Lactococci* and *Enterococci* transform lactate to acetate as shown in Equation 1-4 (Tholen *et al.*, 1997). On the other hand, lactate formation is favored than acetate in absence of O_2. For this reason, creating a microoxic condition in the hindgut is important for providing acetate to the host termite.

1.6.5 Nitrogen Metabolism

$$N_2 + 8H^+ + 8e- + 16ATP \rightarrow 2NH_3 + H_2 + 16ADP + 16\ P_i. \qquad \text{Equation 1-5}$$

Woody materials contain mostly carbon and a very small amount of nitrogen, so the termite's diet is generally a poor source containing a relatively high C/N ratio. However, termites can obtain nitrogen through bacterial nitrogen fixation. Free living nitrogen fixers and protist associated symbionts play significant roles in fixing nitrogen from the atmosphere for host termites in the form of NH_3 as shown in the Equation 1-5. This process requires high energy to transform N_2 to NH_3. For this reason, besides nitrogen fixation from atmosphere, gut microbes

have developed other methods to help host termite receive more N_2 by removing carbon to CO_2 and CH_4, recycling nitrogen from waste products, and by transforming poor nitrogenous compound like NH_3 into amino acids via glutamine synthetase (Noda *et al.*, 2005). Bacteria having ability to sequest nitrogen generally have genes encoding dinitrogenase reductase, an ammonium transporter, urease, and a urea transporter (Noda *et al.*, 2005).

The termite *R. flavipes* produces uric acid but it lacks the uricase enzyme to degrade uric acid. This product has to be transported to the hindgut and degraded by bacteria (Potrikus and Breznak, 1981). It is estimated that the gut of *R. flavipes* contains 6 X 10^4 cells of uricolytic bacteria, such as *Citrobacter* sp., *Bacteroides termitidis* (Potrikus and Breznak, 1980). Urea and ammonia are byproducts of uric acid degradation (Hongoh *et al.*, 2008b) and urea is further hydrolyzed by the urease enzyme producing more NH_3, which can be used for amino acid biosynthesis (Potrikus and Breznak, 1981). However, NH_3 is toxic and can cause damage to the cell, so the enzyme glutamate dehydrogenase (GDH) help detoxify excessive level of NH_3 by transforming NH_3 into glutamate (Sabree *et al.*, 2009). Genome sequences of endosymbionts of other insects consuming low nitrogen diet, such as the termite *C. formosanus*, and the cockroach *Dictyoptera*, also found both urease and GDH beside other nitrogen fixing genes (Hongoh *et al.*, 2008b; Sabree *et al.*, 2009).

1.7 The Phylum *Verrucomicrobia*

Verrucomicrobia was first proposed as a new bacterial division in 1997 (Hedlund *et al.*, 1997) and later was classified as a phylum. *Verrucomicrobia* means “warty” in Greek roots (Madigan *et al.*, 2008). This name represents physiological feature of most members that produce wart-like structure called prosthecae to help them attach to surrounding environments, take up nutrients, and expel wastes (Madigan, 2009). *Verrucomicrobia* were first discovered in freshwater in 1935 (Henrici and Johnson, 1935); then, members of this phylum were continuously found in various environments. They can be present in moderate temperature, cold areas (deep sea and Antarctica), hot environments at 75-95^{o}C (deep sea hydrothermal vent and hot spring), and

extreme environments (sulfide rich water, and soda lake) (Schlesner *et al.*, 2006). *Verrucomicrobia* are abundant in marine environment accounting for 98% of all analyzed samples; about 2% being present in marine water column and 1.4% found in sediment bacterial communities (Freitas *et al.*, 2012). In addition, they can be found in rice paddy soil, forest and agricultural soils, and eukaryotic hosts (Hengstmann *et al.*, 1999; Liesack and Stackebrandt, 1992; Suau *et al.*, 1999). The bacterial diversity in the gut of termite *Zootermopsis angusticollis* was studied by 16S rRNA gene cloning and sequencing. Among 91 clones, members of *Verrucomicrobia* were identified as genus *Verrucomicrobium* and *Opitutaceae* with 6 and 1 operational taxonomic units (OTUs), respectively (Rosengaus *et al.*, 2011).

Although abundant in various environments, the phylum *Verrucomicrobia* is still considered as one of the most obscure phyla because of unculturability of most members. In general, members of this group are identified by their 16S rRNA gene sequences from both culturable members and environmental clone sequences of unculturable members. All the isolates are Gram-negative and cells are present as coccoid or rod shaped bacteria. The presence of members in various habitats has indicated that they have the ability to utilize oxygen ranging from anaerobic, facultative anaerobic, and aerobic conditions. The oxygen requirement is the most selective physiological feature used to classify this group. Most members are chemoheterotroph favoring carbohydrates and complex polysaccharides such as hemicellulose and cellulose.

Currently, there are seven classes in the phylum *Verrucomicrobia*, which are *Verrucomicrobiae* (subdivision 1), *Spartobacteria* (subdivision 2), *Opitutae* (subdivision 4), subdivisions 3, 5, 6, and 7 (Hedlund, 2010b). Members of subdivision 5 and 7 are currently uncultured, so they are defined solely on their 16S rRNA gene sequences. The phylum *Verrucomicrobia* is most closely related to the phyla *Planctomycetes* and *Chlamydia* based on their 16S rRNA gene sequences (Hedlund *et al.*, 1996). A few genera that have been officially published are *Verrucomicrobium, Prosthecobacter, Opitutus, Alterococcus, Rubritalea,* and *Akkermansia* (Yoon); among these, *Verrucomicrobium spinosum* is most thoroughly studied.

1.8 Termite Associated *Verrucomicrobium* (TAV)

In order to gain insights into complex symbiosis in the termite hindgut, this research intends to work with Termite Associated *Verrucomicrobium* (TAV) microorganisms. Five TAV strains (TAV1 to TAV5) were obtained from the wood feeding termite, *Reticulitermes flavipes* (eastern subterranean termite). TAV strains were classified in the phylum *Verrucomicrobia*, class *Opitutae* based on their 16S rRNA gene sequences. The class *Opitutae* presently consists of two orders, *Puniceicoccales* and *Opitutales* (Yoon). The TAV strains or later named *Diplosphaera colitermitum* are classified in the order *Opitutales*. The order *Opitutales* comprises three genera, *Opitutus*, *Alterococus*, (Hedlund, 2010b), and *Diplosphaera*, while the order *Puniceicoccale*s comprises four genera, *Puniceicoccus*, *Cerasicoccus*, *Coraliomargarita*, and *Pelagicoccus*. All members of class *Opitutae* are Gram-negative and lack a peptidoglycan cell wall. Most members of the class *Opitutae* are found in soils, invertebrate digestive tracts, and marine environment; some are found in hypersaline, microbial mats, coastal hot springs and freshwater lakes (Hedlund, 2010b). Previously, members of the phylum *Verrucomicrobia* have been found in the termite hindgut based on the presence of their 16S rRNA gene but they were uncultured. This makes it extremely hard to study their functional, physiological and ecological roles in different ecosystems. Fortunately, five TAV strains are successfully isolated from the hindgut of wood-consuming worker termite, *R. flavipes* (Stevenson *et al.*, 2004). The TAV strains constitute a phenomenal opportunity to explore the obscured and limited information of the *Verrucomicrobia* living in the termite hindgut. The following information presents the physiological characteristics and genome sequences of the strain TAV2 and TAV1 that have been studied recently.

Diplosphaera colitermitum strain TAV2 is a Gram-negative, coccoid bacterium with a cell size of 0.5 µm in diameter. Cells are non-motile and usually present as pairs or groups. The strain produces 2-4 mm cream-colored colonies on R2A agar plates (Stevenson *et al.*, 2004). The strain TAV2 is microaerophile and could not grow under anaerobic condition. The optimum growth occurs between 2 and 8% O_2 with generation time of 7.7 – 8.3 hours. The strain can grow under

temperature from 15 to 35 ^{o}C with an optimum temperature at 30 ^{o}C. pH supporting growth ranges from 5.5 to 7.5 with optimum pH at 7.0. Cells can grow on monosaccharides (glucose and galactose), disaccharides (maltose), and cellulose derivatives (cellobiose, and starch). The strain has potential to fix nitrogen confirmed by presence of the *nif* genes and its ability to grow on nitrogen-free medium (Wertz *et al.*, 2012). The genome sequence of the strain TAV2 contains ~5.0 Mb nucleotides with 60.51% GC content (http://img.jgi.doe.gov/cgi-bin/pub/main.cgi). This sequence has revealed a total of 4,105 genes comprising 4,036 protein coding genes. The genome sequence revealed important genes in cellulose and hemicelluloses degradation (glycosyl hydrolase, xylanase, acetyl xylan esterase, 1,4-β-glucanase), nitrogen fixation (dinitrogenase reductase,) and an oxygen-binding enzyme (cbb_3–type cytochrome oxidase).

The genome sequence of the strain TAV1 contains 7.1 Mbp with 63.2% GC content. The draft genome contains 6,051 genes with 5,987 protein encoding genes. Based on carbohydrate-active enzyme database, TAV1 genome contains several genes encoding glycoside hydrolases (GH) which is responsible for breaking bonds of carbohydrates. These GH enzymes comprise both catalytic and carbohydrate-binding modules, such as glycoside hydrolase family 5, cellulase, endo-1,4-beta-xylanase, N-acetylglucosamine-6-phosphate deacetylase, and etc. Moreover, genes involving in nitrogen metabolism were determined, such as nitrogen iron reductase protein (*nifH*), nitrogenase molybdenum-iron (*nifD*), FeS assembly protein (*nifU*), and nitrogenase MoFe cofactor biosynthesis protein (*nifE)* genes. In addition, the cbb_3-type cytochrome oxidase, which has high affinity for oxygen, was also found in the strain TAV1 (Isanapong *et al.*, 2012).

1.9 Why the Termite Associated *Verrucomicrobium* (TAV) Strains?

1.9.1 Phylogenetic and Evolutionary Gradients in these Five TAV Strains Allow us to Study Genomic and Functional Variations among Closely Related Species.

Based on sequence identity of the 16S rRNA gene, TAV strains show an evolutionary gradient that can be used as a guide to study their genomic and functional variation. Sequence

analysis performed in the Rodrigues' laboratory revealed that TAV3 and TAV4 are the most closely related with 99.9% 16S rRNA identity. TAV2 shared 99.6% identity with TAV3 and TAV4. This level of identity indicates that these three isolates are considered to be strains of the same species. When the previous group is compared to TAV5, the percent identity is 97.5%. The last strain which is least identical to the previous strain is TAV1 with 95.3 – 95.8% identity.

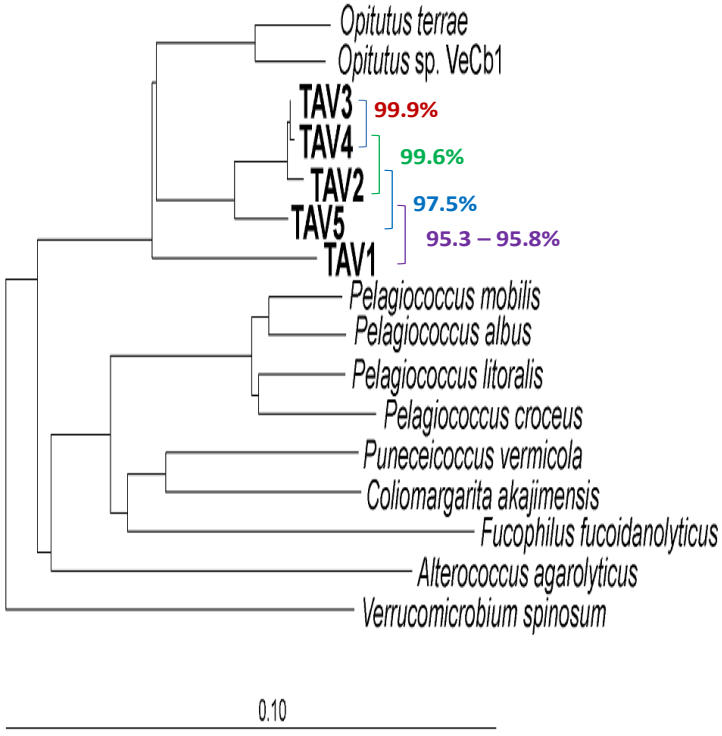

Figure 1.8 Phylogenetic tree of five TAV strains

Denef et al. (2010) found that the closely related strains may express distinct functional behavior in nature. This differentiation might depend on level of genetic variation among the natural population (Denef *et al.*, 2010). TAV1 produces numerous extracellular matrixes that are not present in other TAV strains (Stevenson *et al.*, 2004). According to preliminary result performed in the Rodrigues' laboratory, although TAV3 and TAV4 are the most closely related strains, their ability to grow in glucose varies substantially, with TAV3 presenting slower growth

rate than TAV4. Therefore, the study of their genomes will provide information about their genomic variations and functions of these closely related strains.

1.9.2 The Genome Sequence of TAV 1 and TAV2 Reveals Important Genes for Nitrogen Metabolism, Carbon Metabolisms, and Ability to Control Oxygen Levels in the Hindgut.

1.9.2.1 Nitrogen Metabolism

The termite's diet, mainly lignocellulose, is rich in carbon but poor in nitrogen; therefore, termites had devised ways of obtain and conserved nitrogen, such as symbiosis with nitrogen fixing bacteria (Hongoh *et al.*, 2003). Bacteria can fix nitrogen because they have enzymes involved in nitrogen fixation like dinitrogenase reductase (*nifH*). The presence of the genes encoding for dinitrogenase reductase and genes involving in recycling nitrogen from waste product in TAV genomes suggest an important functional role in the hindgut. Using variety of molecular techniques, this research aims to determine the ability of these strains in fixing nitrogen and the genes involved in this process.

1.9.2.2 Carbon Metabolism

Draft annotation of both TAV strains have revealed genes involved in carbon utilization. Among those genes, the genes involved in cellulose and hemicellulose degradation have been verified, such as the genes encoding for glycosyl hydrolases (cellulase family 5), xylanase and acetyl xylan esterase, 1,4-β-glucanase, and α-L-arabinofuranosidase. This research will investigate the capacity of the TAV strains in using various carbon sources. The main focus will be on ability of the strains to break down cellulose and hemicellulose.

1.9.2.3 Oxygen Controlling Ability

As moving from the foregut to the hindgut of termites, the level of oxygen decreases until it becomes anoxic at the hindgut core where bundles of microbial communities are present. Oxygen keeps circulating into the gut because the difference in concentrations between oxic and

anoxic interface. Oxygen is used up by microbes for their respiration; consequently, the center of the gut becomes anoxic region surrounded by microoxic periphery.

The TAV strains are believed to inhabit at the gut wall and might have potential role in establishing the microoxic periphery. Firstly, their genomes contain the gene coding for the cbb_3–type cytochrome oxidase which has high affinity to oxygen. Secondly, the TAV strains were isolated under hypoxic condition, which suggests the presence of these strains in microoxic region. Since their natural habitat is hypoxic, their growth rate and general performance in this condition are expected to be greater than those in high oxygen concentration. Willis (2009) indicated that TAV2 grown in R2B medium under 2% oxygen concentration had a higher growth rate than that observed in 20% oxygen concentration. The potential role of the TAV strains in maintaining oxygen levels in the termite hindgut is extremely important for microbial communities requiring anoxic environment to complete important mechanisms, such as cellulose degradations, nitrogen fixation, acetogenesis, and methanogenesis (Wertz and Breznak, 2007b).

The presence of genes associated with nitrogen fixation, lignocellulose degradation, and oxygen consumption in genome sequences of TAV1 and TAV2 implies significant roles of TAV strains in functioning of the hindgut ecosystem in maintaining of nitrogen inputs, funneling of oligosaccharides from celluloses and hemicelluloses, and maintaining hypoxic condition within the hindgut. Ultimately, this research intends to provide basic understanding of ecological, physiological, and functional roles of microbes capable of degrading cellulosic materials, helping to harness the microbes for biofuel production.

Chapter 2

High-Quality Draft Genome Sequence of the *Opitutaceae* Bacterium Strain TAV1, a Symbiont of the Wood-Feeding Termite *Reticulitermes flavipes*

2.1 Introduction

The *Opitutaceae* bacterium strain TAV1 was isolated from the hindgut of the wood-feeding termite *Reticulitermes flavipes* (Stevenson *et al.*, 2004). This termite-associated *Verrucomicrobia* (TAV) isolate is a Gram-negative, coccoid-shaped, microaerophilic bacterium. Owing to the importance of microbial symbionts in the termite hindgut for the degradation of cellulose and hemicellulose into acetate, hydrogen, and methane (Breznak and Brune, 1994; Brune and Ohkuma, 2011), we investigated the genetic potential of strain TAV1 for the degradation of lignocellulosic material and overall functional attributes associated with its ecological role.

2.2 Materials and Methods

The genomic DNA of strain TAV1 was isolated using a cetyltrimethylammonium bromide method (http://my.jgi.doe.gov/general/). Genome was sequenced to 30-fold coverage using a combination of Illumina and 454 pyrosequencing platforms. The individual reads were assembled with the Newbler assembler v. 2.3 (Roche) and generated 82 contigs with the largest being 590 kb and the smallest contig being 530 bp. All contigs span up to the length of 7.1 Mbp, and the average GC content of the genome is approximately 63.2%. Genes were identified using Prodigal (Hyatt *et al.*, 2010) as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline (Pati *et al.*, 2010). The predicted protein-coding genes (coding sequences [CDS]) were translated and used to search the National Center for Biotechnology

Information (NCBI) nonredundant database and the UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases. These data sources were combined to produce a product description for each predicted protein. Noncoding genes and miscellaneous features were predicted using tRNAscan-SE (Lowe and Eddy, 1997), RNAMMer (Lagesen *et al.*, 2007), Rfam (Griffiths-Jones *et al.*, 2003),TMHMM (Krogh *et al.*, 2001), and signalP (Bendtsen *et al.*, 2004).

2.3 Results

The draft genome contains 6,051 genes with 5,987 CDS. A total of 64 structural RNAs were identified in the genome, with the presence of one rRNA operon. Protein coding genes were classified according to the cluster of orthologous groups (COG categories). The top five functional groups were as follows: (i) general function prediction only (526 gene counts), (ii) transcription (476 gene counts), (iii) carbohydrate metabolism (406 gene counts), (iv) amino acid metabolism (403 gene counts), and (v) energy production and conversion (310 gene counts) (http://genome.ornl.gov/microbial/obac1/05feb11/fun.html). The last two groups in which genes were categorized were chromatin structure and dynamics (2 gene counts) and cytoskeleton (1 gene count). It was noteworthy to mention that catalase gene was not identified in the current result. This might be the case that the genome sequencing has not been closed yet and further investigation is required when the result is complete. However, there were the presence of alkyl hydroperoxide reductase and peroxiredoxin genes which can help scavenge H_2O_2. Further inspection using the carbohydrate active enzyme database (http://www.cazy.org) revealed that the TAV1 genome contains a large number of genes encoding glycoside hydrolases (GH), which are involved in the breakage of bonds between two or more carbohydrate moieties. These GH enzymes contained both catalytic and carbohydrate-binding modules, such as glycoside hydrolase family 5, cellulase, endo-1,4-beta-xylanase, *N*-acetylglucosamine-6-phosphate deacetylase, peptidoglycan glycosyltransferase, and 1,4-alpha-glucan branching enzyme. Moreover, 13 genes associated with nitrogen fixation were identified, such as nitrogen iron reductase protein (*nifH*), nitrogenase

molybdenum-iron (*nifD*), FeS assembly protein (*nifU*), and nitrogenase MoFe cofactor biosynthesis protein (*nifE*) genes, among others. In addition, the TAV1 genome contains the cbb_3- type cytochrome oxidase, which has high affinity for oxygen. Effective removal of O_2 is essential for the homoacetogenic and methanogenic process to occur. The presence of genes associated with lignocellulose degradation, nitrogen fixation, and oxygen consumption implies an important ecological role for strain TAV1 in the functioning of the hindgut ecosystem.

The high-quality draft genome sequence of the *Opitutaceae* bacterium was deposited in GenBank under the accession number AHKS00000000.

Chapter 3

Development of an Ecophysiological Model for *Diplosphaera colotermitum* TAV2, a Termite Hindgut *Verrucomicrobium*

3.1 Introduction

Termites have long been recognized for their ability to consume lignocellulosic plant material and soil (humus), converting it into substrates (primarily acetate) on which the termite depends for carbon and energy. These social insects are not only important for the global carbon cycling, but also for their biotechnological potential as efficient lignocellulose degraders (Brune, 1998). The success of termite feeding behavior is intimately associated with the presence of a diverse and abundant gut microbial community (Ohkuma and Brune, 2011). In the lower termite, *Reticulitermes flavipes*, the complexity of symbiosis spans three domains of life: methane-producing *Archaea*, cellulolytic *Eukarya* (protozoa), and *Bacteria* - all acting cooperatively to degrade lignocellulose, fix/recycle nitrogen, and remove oxygen, suggesting a division of metabolic activities among members of the community.

It has been long assumed that prokaryotic residents, in the hypoxic periphery of the termite gut, play the primary role of consuming O_2 and maintaining a steep radial gradient of this inwardly diffusing gas resulting in anoxia in the luminal portion of the gut (Brune *et al.*, 1995; Köhler *et al.*, 2012). However, this assumption has recently been challenged by the demonstration of a strong correlation between the O_2 consumption rate of extracted guts and the number of protozoans per gut (Wertz and Breznak, 2007a), suggesting a more modest role for bacteria in this process - at least in those termites that possess (putatively “strictly” anaerobic) cellulolytic protozoa in their hindguts. Although the relative importance of each group has yet to

be resolved, it is reasonable to expect that microbial community members have multiple ecological functions, owing to their ubiquitous presence in different termite species.

Among the Domain *Bacteria*, members of the phylum *Verrucomicrobia* have been consistently detected in molecular surveys of the 16S rRNA gene targeting the microbial community of various termite species (Berlanga *et al.*, 2011; Hongoh *et al.*, 2003; Köhler *et al.*, 2012; Nakajima *et al.*, 2005; Rosengaus *et al.*, 2011). Despite their ubiquity in higher and lower termite species, their functional role remains elusive. In an effort to understand the ecological functions of this microbial group, we previously isolated and characterized *Verrucomicrobia* strains from the gut of *Reticulitermes flavipes*, including the new species *Diplosphaera colotermitum* strain TAV2. We confirmed their autochthonous nature, estimated the population number, and unveiled the genomic properties (Isanapong *et al.*, 2012; Wertz *et al.*, 2012), suggesting new ecological roles for *Verrucomicrobia* within the termite gut.

In this study, we investigate the microaerophilic nature of TAV2 and report on novel ecological functions. We employed comprehensive and integrative transcriptomic and proteomic approaches to: (1) identify genes and proteins being expressed in response to different O_2 concentrations; (2) develop an experimentally tested metabolic map for TAV2; and (3) establish the first ecophysiological model for an isolate of the phylum *Verrucomicrobia*.

3.2 Materials and Methods

3.2.1 Culture Conditions

Diplosphaera colotermitum strain TAV2 (ATCC BAA-2264, DSMZ 25453) was grown in liquid R2A medium without agar (R2B) with a liquid-to-flask volume ratio of 40%. Four replicate cultures were incubated on an orbital shaker at 200 rpm inside a vinyl hypoxic chamber (Coy Laboratory Products, Glass Lake, MI) fitted with an oxygen sensor and automated controller set for 2% or 20% O_2 concentration. These O_2 concentrations were selected because 60% of the

termite hindgut is hypoxic (4% O_2 at the epithelial wall decreasing to anoxia approximately 100 µm inward; (Brune *et al.*, 1995)), whereas 20% O_2 concentration approximates atmospheric O_2 concentrations (20.9% at 1 atm). When cells reached an optical density (OD_{600}) of 0.3 (approximately 8.7×10^7 cells/ml), 400 ml cultures were harvested by centrifugation (17,644 × g) for 15 min at 4°C. Cells were washed twice with ice-cold 50 mM phosphate buffered saline (PBS; 10 mM phosphate buffer [pH 7.0] and 130 mM NaCl), pellets were frozen immediately in liquid nitrogen, and stored at -80 °C for further analysis. For the transcriptomic analysis, 5.0 ml of each culture was harvested by centrifugation for 5 min at 4°C and immediately processed for RNA isolation.

3.2.2 Microarray Design

The *D. colotermitum* microarray was designed based on genome sequence of TAV2 to represent 4,022 genes out of 4,105 genes total number (98% gene coverage) with coverage of three oligonucleotides per open reading frame. It contained 11,492 specific oligonucleotides (45-62 mer) replicated three times onto a glass slide (MYcroarray, Ann Arbor, MI) and 166 random sequence oligonucleotides were introduced randomly throughout the slide as internal hybridization controls.

3.2.3 Microarray Hybridization

Total RNA was extracted from cells grown under 2 and 20% O_2 conditions using the RiboPure™ bacterial RNA kit (Ambion, Austin, TX), followed by DNase I treatment according the manufacture's instructions. Transcripts were enriched with the MICROB*Express*™ bacterial mRNA enrichment kit (Ambion, Austin, TX) and samples (100 ng) were amplified and converted to anti-sense RNA (aRNA) with the MessageAmp II aRNA amplification kit. Samples were ethanol precipitated and quality was assessed with the Bioanalyzer 2100 (Agilent Tech., Santa Clara, CA). Fluorescent labeling was performed with either Alexa Flour® 555 or Alexa Flour® 647 dyes, following instructions of the manufacturer (Invitrogen Inc., Carlsbad, CA). Labeled samples were

purified using the RNeasy® kit (Qiagen Inc., Valencia, CA), vacuum dried, and stored at -20°C until hybridization.

Slides were pre-hybridized with buffer containing 5× SSPE (0.75 M NaCl, 0.05 M NaH_2PO_4, 0.005 M EDTA), 1% SDS, and 1 mg/ml acetylated bovine serum albumin (BSA) at 50 ^oC for 45 min and washed twice with 0.025× SSPE. Labeled samples (2.5 µg) were ressuspendend in 60µL of 1× hybridization buffer (25% SSPE, 25% formamide, 0.05% Tween 20, 0.01 mg/ml BSA) and transferred onto the slides. Single-slide rubber hybridization chambers were incubated in a 50 ^oC water bath for 16 h, followed by post-hybridization washes with 1× SSPE for 3 min and subsequently, with 0.1× SSPE at room temperature for 1 min. Each biological replicate was performed in three different runs (triplicate) and respective dye swaps were used. Slides were scanned at a 30-µm resolution using the Genepix® 4200A scanner (Axon Instruments, Sunnyvale, CA).

3.2.4 Transcriptomic Data Analysis

Signal intensities were normalized with the locally weighted scatterplot smoothing (LOWESS) algorithm, log_2-transformed, and filtered using a two-fold change cutoff (upper level set at 95% and lower level set at 10%) with the GeneSpring GX 11 software package (Agilent technologies, Santa Clara, CA). Statistical analysis of differentially expressed genes was performed using the Student's *T*-test with subsequent correction with the Benjamini-Hochberg multiple-hypothesis adjustment (Cui and Churchill, 2003).

3.2.5 Protein Extraction

Harvested *D. colotermitum* 2% O_2 and 20% O_2 growth samples were barocycled (Pressure Biosciences Inc., South Easton, MA) from ambient (10 sec) to 35kpsi (20 sec) for 10 times. Rapigest (Waters, Milford, MA) and dithiothreitol (Sigma, St. Louis, MO) were added to 0.1% and 5mM concentrations, respectively. Samples were incubated at 60 ˚C for 30 min with gentle shaking. Proteins were digested using sequencing-grade modified porcine trypsin

(Promega, Madison, WI) and peptide concentration measured according to established protocols (Callister *et al.*, 2006a).

3.2.6 Proteomic Data Analysis

Peptides were analyzed using the Accurate Mass and Time (AMT) tag proteomics approach (Smith *et al.*, 2002) in combination with a reference peptide database empirically generated for *D. colotermitum*. For database generation, a pooled *D. colotermitum* sample was fractionated (50 fractions total) using strong cation exchange high-pressure liquid chromatography (HPLC). Each fraction was further separated by reverse phase HPLC coupled to a linear ion trap mass spectrometer (LTQ, ThermoFisher Scientific Corp., San Jose, CA). Mass spectrometry (MS) instrument operating conditions and the HPLC separation method have previously been described (Sowell *et al.*, 2008). MS/MS spectra were analyzed using SEQUEST (Eng *et al.*, 1994) and the TAV2 genome (NCBI accession number ABEA00000000). Preliminary filtering of identified peptides was performed according to previously established criteria (Callister *et al.*, 2006b).

Proteome abundance information was obtained by LTQ-Orbitrap MS analysis of peptides separated using reverse phase HPLC. LC-MS spectra were de-isotoped (Jaitly *et al.*, 2009), mass and elution time features identified, then matched (Monroe *et al.*, 2007) to peptides stored in the reference peptide database filtered to exclude non-tryptic peptides, and peptides having a PeptideProphet probability (Keller *et al.*, 2002) <0.50 (false discovery rate of less than 4%; (Qian *et al.*, 2005)). Matched peptides having a mass error < 2 ppm and normalized elution time error < 0.06 were retained. Measured arbitrary abundance for a peptide was determined by integrating the area under each LC-MS peak for the detected feature matched to that peptide. Peptide abundances among technical replicates were combined, $\log_2$-transformed, normalized, and converted to protein abundances with the ZRollup method implemented in DAnTE (Polpitiya *et al.*, 2008). Only proteins with two or more unique peptides were retained for ANOVA analysis (p-

values < 0.05 and q-values to control the false discovery rate below 0.04 in multiple testing) (Storey, 2002).

3.2.7 Metabolic Reconstruction

A combined list of significantly expressed genes and detected proteins under both O_2 conditions was used for pathway prediction and metabolic overview of the strain TAV2 using the Oak Ridge National Laboratory annotation pipeline (http://genome.ornl.gov/microbial/verr/) and the Pathosystems Resource Integration Center genome viewer (http://patricbrc.vbi.vt.edu/).

3.2.8 Reverse-Transcription Quantitative PCR (RT-qPCR)

Eight differentially expressed genes or proteins were selected for RT-qPCR analysis (Appendix D). Specific primers were designed with the Oligoanalyzer software (www.idtdna.com). aRNA samples (5 μg) from the microarray analysis were reversed transcribed with the M-MuLV Taq RT-PCR kit (New England Biolabs, Ipswich, MA). cDNA (10 ng) was added to the SYBR Green master mix (BioRad, Hercule, CA) containing forward and reverse primers (150 nM each) and reactions were set according to manufacturer's instructions. Reactions were performed in triplicate for each biological replicate and results were normalized to the expression values of the housekeeping gene *gyrB* (β subunit of the DNA topoisomerase) to calculate fold induction values according to (Livak and Schmittgen, 2001). Statistical significance was calculated using the two-tailed Student's *T*-test.

3.3 Results

3.3.1 Confirmation of the Microaerophilic Phenotype of TAV2

Cells of *D. colotermitum* strain TAV2 were grown in liquid culture under controlled 2% or 20% O_2 concentrations. The growth rate during exponential phase for cells subjected to 2% O_2 was $0.0368 \cdot h^{-1}$, while a growth rate of $0.0313 \cdot h^{-1}$ was calculated for cells grown under 20% O_2

concentration ($P < 0.001$). We determined that this difference in growth rates was sufficient to reduce the population doubling time by approximately 4 h for the 2% O_2 culture.

3.3.2 Transcriptome Analysis

Microarray analysis allowed identification of 75 genes as differentially expressed ($P < 0.05$), with an up-regulation of 49 genes and a down-regulation of 26 genes, when cells were grown under an atmosphere of 2% O_2 as opposed to 20% O_2. When differentially expressed genes were classified by functional categories based on clusters of orthologous groups (COGs) (Tatusov *et al.*, 1997), hypothetical genes or proteins of unknown function comprised 42.6% of the data set (Appendix B). Functional categories representing energy production and conversion, cell cycle control, replication and recombination, and cellular trafficking and secretion were present in the up-regulated group for the 2% O_2 condition, while amino acid transport and metabolism and coenzyme transport were identified as down-regulated functional groups.

3.3.3 Proteome Analysis

A total of 36,109 peptides were identified in the LC-MS/MS analysis, with 18,044 and 18,065 peptides detected in samples derived from 2% O_2- and 20% O_2-grown cells, respectively. After data filtering to remove peptide and protein redundancies, 820 and 735 proteins were detected for each environmental condition, respectively. The above values account for 33% (2% O_2) and 30% (20% O_2), of the protein coding genes identified in the high draft genome of *D. colotermitum*. Among identified proteins, 665 (74.7%) were detected in both culture conditions.

A total of 96 proteins were identified as up-regulated ($P < 0.05$), with 79 having higher spectral count measurements under 2% O_2, whereas 17 proteins were more abundant in cells growing under 20% O_2 concentration as shown in Figure 3.1.

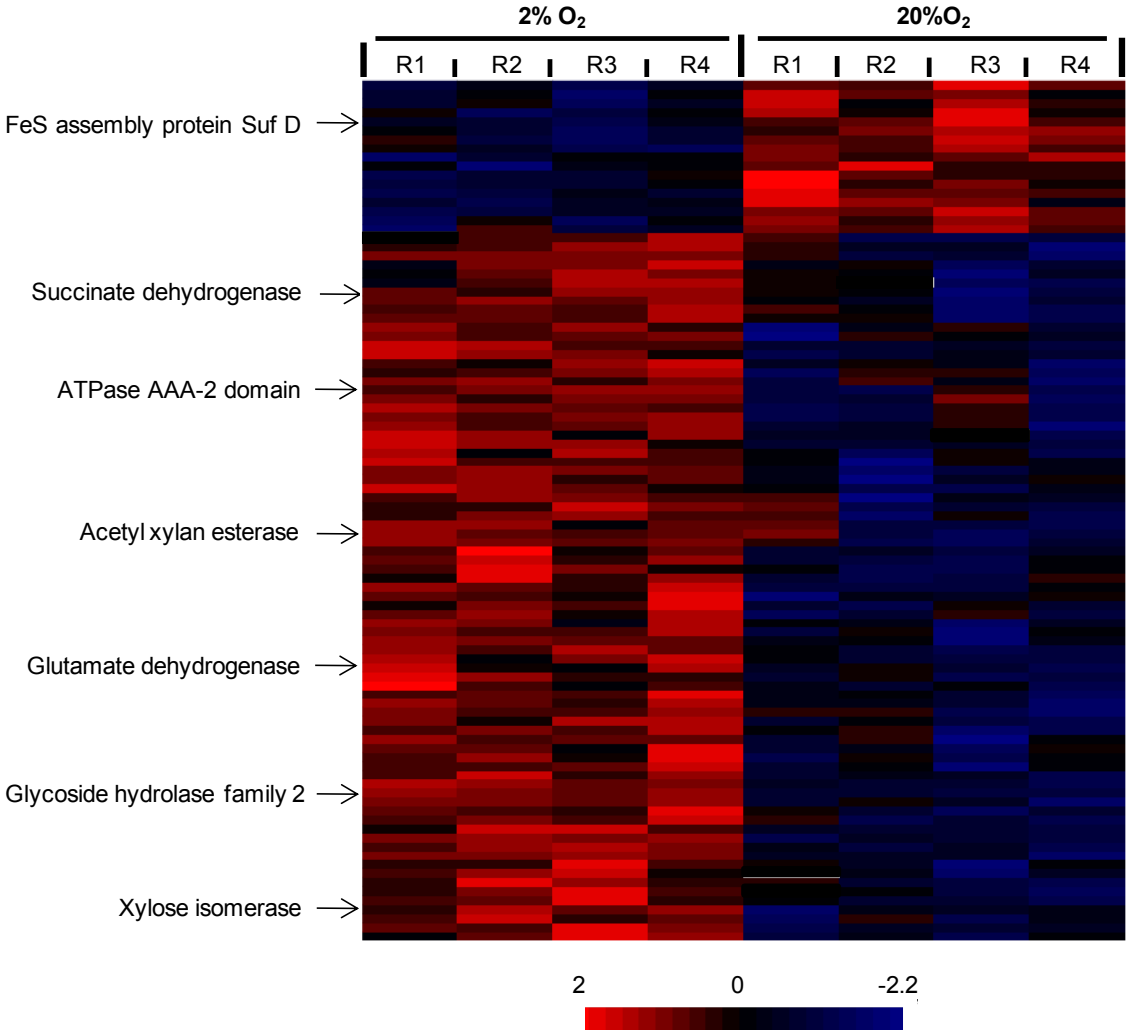

Figure 3.1 Heat map comparisons of calculated Z-scores for proteins differentially expressed by TAV2 cells grown under 2% or 20% O_2.

Note: Increasing intensity in the positive range (blue) represents abundances that are greater than the mean abundance derived from both conditions relative to the standard deviation associated with the mean. A few enzymes discussed in the text are illustrated here as indicated by the arrows)

Once assigned to functional categories, conserved hypothetical proteins or proteins encoded by genes previously classified as hypotheticals ranked as the largest group. The relative proportions of proteins classified in these two groups were 27.1% and 37.1% for 2% and 20% O_2, respectively. After removal of hypotheticals, the remaining proteins were placed in 22 COG

categories in order to describe changes in cellular functions in response to O_2 as shown in Figure 3.2.

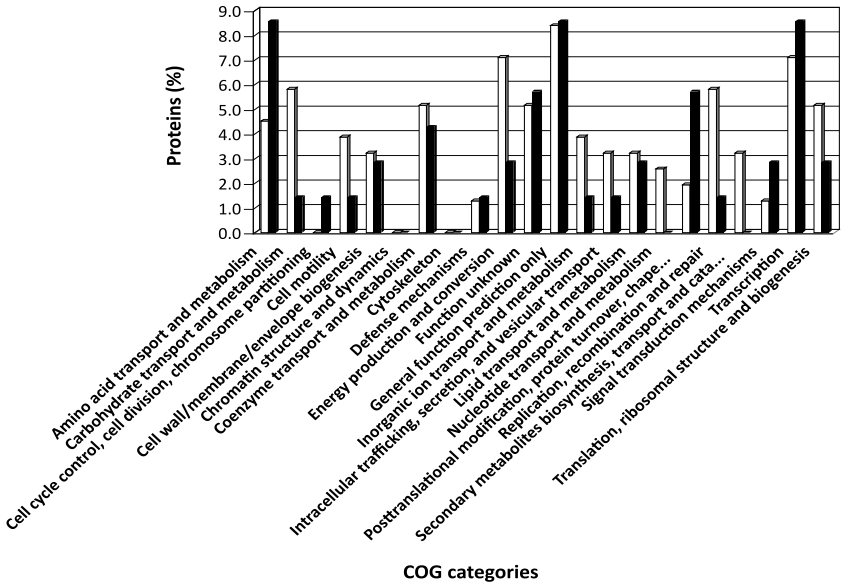

Figure 3.2 Relative proportions of unique proteins assigned to functional categories based on cluster of orthologous groups (COG).

Note: white bars represent proteins detected in cells grown under 2% O_2, while black bars represent proteins observed in cells grown under 20% O_2.

Eight COG categories had at least a two-fold higher protein expression for cells grown under 2% O_2 in comparison to those grown under 20% O_2 conditions. These included: energy production and conversion (7.1%), carbohydrate transport and metabolism (5.8%), replication and recombination (5.8%). Only three categories had a two times higher protein expression under 20% O_2 condition: amino acid transport and metabolism (8.4%), protein turnover and post-translational modifications (5.7%) and signal transduction mechanisms (2.9%). A principal component analysis of the proteomic results indicated that 72.34% of the variation could be explained by O_2 (55.9%), biological (9.67%) and technical (6.77%) replicates as shown in Figure 3.3.

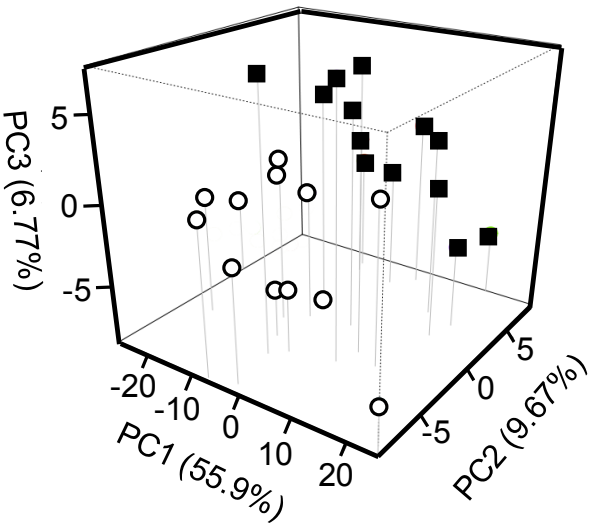

Figure 3.3 Principal component analysis of the proteome results for TAV2 cells grown under 2% or 20% O_2.

Components were: (1) O_2 concentration, (2) biological replicate, and (3) technical replicate. Open circles represent samples of cells grown under 2% O_2, while closed squares represent cells under 20% O_2.

3.3.4 Transcriptomic and Proteomic Data Comparison

We performed separate correlations for transcriptomic and proteomic data. The majority of the genes showed similar expression values in the two conditions tested (R^2 = 0.7909). The selection of differentially expressed genes modified the slope and lowered the coefficient of determination (R^2 = 0.3748). This is consistent with the larger number of genes that were up-regulated under 2% O_2 in comparison to 20% O_2 (Appendix Figure C – 1A). When the proteomic data were also examined in a similar manner, peptide abundances for proteins found in both conditions had a coefficient of determination of R^2 = 0.8641. The value was maintained approximately the same (R^2 = 0.8543) when only differentially detected proteins were selected for the calculation (Appendix Figure C – 1B).

Next, we compared differentially expressed genes and proteins within the different COG categories. Clusters that had a large number of genes expressed also displayed a larger number of proteins being detected: energy production and conversion, carbohydrate transport and

metabolism, and replication and recombination. An agreement between expressed transcripts and their corresponding proteins was only observed for a limited number of genes: ABC transporter, acetyl xylan esterase, ATP synthase, phosphofructokinase, ribosomal proteins, tetratricopeptide (TPR) repeat-containing protein, and a type II secretion system protein.

3.3.5 Metabolic Reconstruction

We combined enzyme commission numbers (EC numbers) of differentially expressed genes and proteins to construct a metabolic map for *D. colotermitum* as shown in Figure 3.4. Each EC number was assigned into metabolic pathways using metabolic pathway classification provided in the Oak Ridge National Laboratory annotation pipeline (http://genome.ornl.gov/microbial/verr/) and the Pathosystems Resource Integration Center genome viewer (http://patricbrc.vbi.vt.edu/).

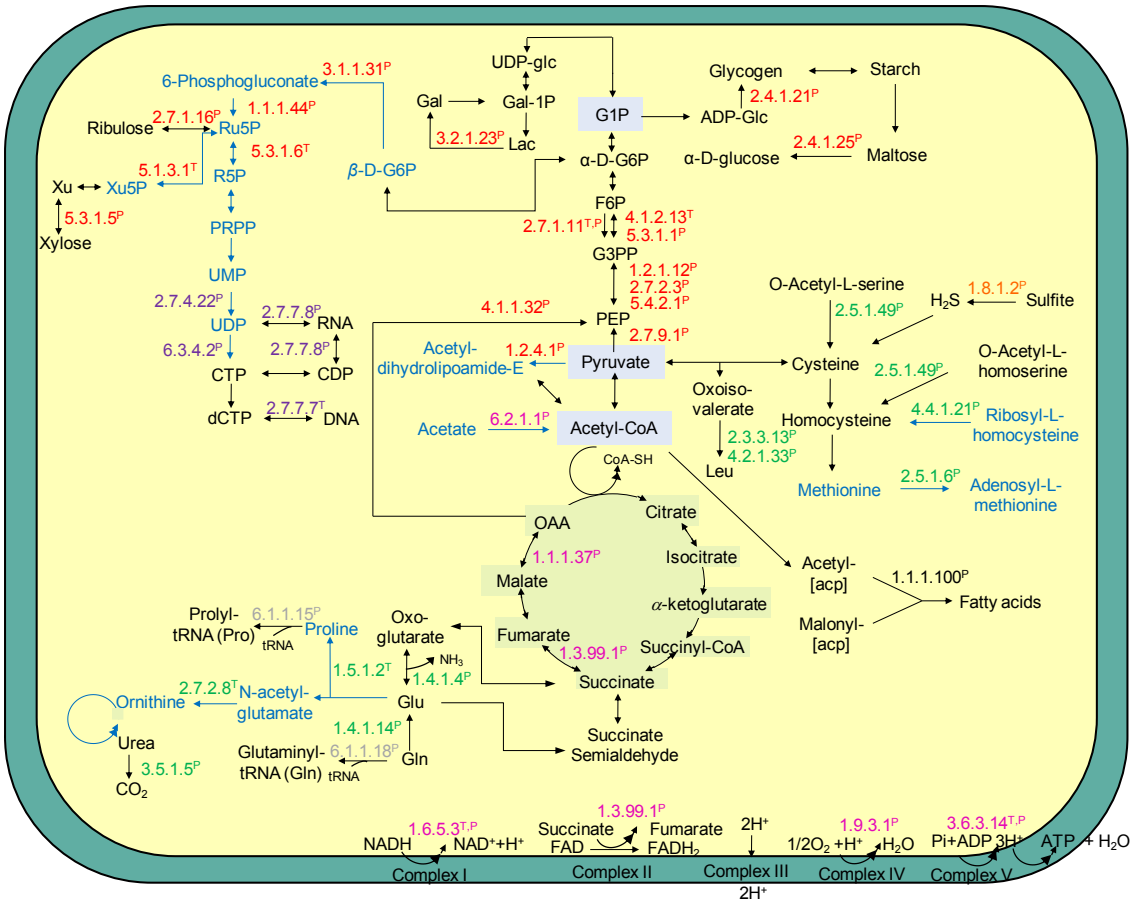

Figure 3.4 Metabolic map of *Diplosphaera colotermitum* strain TAV2 constructed with selected differentially expressed genes and proteins.

Note: Proteins (P) and/or their mRNA transcripts (T) are classified into six major metabolic classes: carbohydrate metabolism (red), energy generation (pink), amino acid biosynthesis (green), nucleotide metabolism (purple), fatty acid synthesis (black), and translation (grey). The enzyme sulfite reductase is shown in orange. Text marked in black represents up-regulated pathways under 2% O_2, while text marked in blue represents down-regulated pathways. Abbreviations: Ru5P, ribulose 5-phosphate; R5P, ribose 5-phosphate; PRPP, phosphoribosyl pyrophosphate; UMP, uridine monophosphate; UDP, uridine diphosphate; CTP, cytidine triphosphate; CDP, cytidine diphosphate; dCTP, deoxycytidine triphosphate; DNA, deoxyribonucleic acid; RNA, ribonucleic acid; UDP-glc, uridine diphosphate glucose; Gal-1P, galactose-1-phosphate; Lac, lactose; Gal, galactose; G1P, glucose-1-phosphate; G6P, glucose-6-phosphate; F6P, fructose-6-phosphate; G3PP, glyceraldehyde-3-phosphate; PEP, phosphoenolpyruvate; OAA, oxaloacetate; acp, acyl carrier protein; ADP-glc, adenosine diphosphoglucose; Leu, leucine; Glu, glutamate; Gln, glutamine, NADH, nicotinamide adenine dinucleotide; ADP, adenosine diphosphate; ATP, adenosine triphosphate.

Only enzymes assigned with an Enzyme Commission (EC) number were included in the reconstruction to provide an overall picture of TAV2 metabolism and determine the metabolic changes that occur in TAV2 at microoxic and atmospheric O_2 concentrations. The following functional categories were used for reconstruction:

3.3.5.1 Carbohydrate Metabolism

Two major carbohydrate pathways were derived from the combined data: glycolysis and the pentose phosphate pathway (PPP). Eight out of the 16 enzymes (50%) involved in glycolysis were up-regulated under 2% O_2. Among these enzymes, a positive correspondence between transcriptomic and proteomic analyses was found for phosphofructokinase (EC: $2.7.1.11^{T,P}$; superscript T and P denotes identification through transcriptome or proteome, respectively), the enzyme mediating the most important irreversible step from fructose 6-phosphate to fructose 1,6-bisphosphate. All enzymes associated with the PPP were down regulated under the same condition. It is noteworthy that the enzyme xylose isomerase ($E.C.5.3.1.5^{P}$) was up-regulated. This enzyme is responsible for providing D-xylulose entering the PPP.

Malate dehydrogenase ($EC:1.1.1.37^{P}$) and succinate dehydrogenase ($EC:1.3.99.1^{P}$), two enzymes of the tricarboxylic acid cycle, were significantly up-regulated under 2% O_2 concentration, while pyruvate dehydrogenase (acetyl-transferring) ($EC:1.2.4.1^{P}$), the enzyme mediating glucose catabolism from glycolysis to the TCA cycle was down-regulated. The former two enzymes have been found to increase their activities under microaerobic growth (Pauling *et al.*, 2001), while the latter enzyme has its expression repressed by low levels of O_2. Interestingly, the enzyme pyruvate phosphate-dikinase ($EC:2.7.9.1^{P}$), an important enzyme in reversing glycolysis, was up-regulated in cells maintained under microoxic conditions.

Up-regulated genes and proteins were searched against the catalytic module of the carbohydrate-active enzymes (CAZy) database. We identified enzymes in four groups, as follows: glycosyl hydrolases [family 2 TIM barrel, 4-α-glucanotransferase ($EC:2.4.1.25^{P}$)], carbohydrate

esterases [acetyl xylan esterase (EC:3.1.1.72^{P})], glycosyl transferases [glycogen/starch synthase, ADP-glucose type (EC:2.4.1.21^{P})], and miscellaneous [xylose isomerase (EC:5.3.1.5^{P})].

3.3.5.2 Amino Acid Biosynthesis

Our metabolic reconstruction analysis indicated that strain TAV2 is capable of synthesizing amino acids essential to the termite host (leucine, lysine, and tryptophan) and non-essential (alanine, glutamate, glutamine, cysteine, and serine) under 2% O_2, while enzymes involved in the synthesis of proline, ornithine, and methionine were down- regulated. Noteworthy, both enzymes glutamate dehydrogenase (EC:1.4.1.4^{P}) and glutamate synthase (EC:1.4.1.14^{P}), important for N metabolism, were up-regulated; while the enzymes involved in formation of proline, pyrroline-5-carboxylate reductase (EC:1.5.1.2^{T}) and ornithine, acetylglutamate kinase (EC:2.7.2.8^{T}), were down-regulated. The presence of an enzyme involved in the biosynthesis of ornithine and the up-regulated expression of the urease α-subunit domain protein (EC:3.5.1.5^{P}) denote the importance for N recycling in the termite gut.

3.3.5.3 Energy Transduction

TAV2 expressed several genes encoding for enzymes involved in the electron transport system (ETS) and ATP synthesis. Enzymes such as NADH dehydrogenase (EC:1.6.5.3^{T,P}), succinate dehydrogenase (EC:1.3.99.1^{P}), and cytochrome c oxidase (EC:1.9.3.1^{P}) were significantly up-regulated under microoxic conditions. A higher number of spectral counts was observed for the enzyme *cbb*$_3$ cytochrome oxidase for cells under 2% O_2 in comparison to those under 20% O_2. We observed a positive concurrence for the ATP synthase between proteomic and transcriptomic data, with both alpha (F_1) and beta (F_o) subunits of this large enzymatic complex (EC 3.6.3.14^{T,P}) significantly up-regulated under 2% O_2 concentration, indicating that ATP formation is carried out through oxidative phosphorylation. Proteomic results showed a higher number of superoxide dismutase peptides for cells grown in 20% O_2 concentration, although this increase was slight and not statistically different from those obtained for samples

grown under 2% O_2. This is consistent with the view that O_2 concentrations higher than those naturally found in the termite hindgut represent an oxidative stress for this microaerophile.

3.3.5.4 Nucleotide Metabolism and Post-Translational Modification

Genes and proteins involved in DNA replication, transcription, and translation were highly expressed at 2% O_2. TAV2 showed a significant expression of enzymes (above 3 spectral counts) responsible for syntheses of pyrimidine bases, RNA, and DNA. When combining proteomic and transcriptomic data, expressed genes and their detected proteins were observed for a subset of ribosomal proteins (L24^P, L9^P, S11^P, and L15^T) and aminoacyl-tRNA transferases, such as glutaminyl-tRNA synthetase (EC:6.1.1.18^P), tryptophanyl-tRNA synthetase (EC:6.1.1.2^P), and lysyl-tRNA synthetase (EC:6.1.1.6^T). TAV2 expressed several proteins responsible for proofreading and degradation of mismatched DNA base pairs, such as the DNA mismatch repair protein MutS^P, and the exodeoxyribonuclease VII small subunit^T. Moreover, we observed the presence of proteins involved in gene regulation such as GreA/GreB family elongation factor^P, two component transcriptional regulator LuxR family^P, regulatory proteins LacI^T and GntR HTH^T, and translational proteins, such as the initiation factor 3^P, the protein that attaches the mRNA into 30S ribosome subunit, and the translation elongation factor G^P. The cytoplasmatic proteins in charge of protein folding, chaperonin GroEL^P and chaperonin cpn10^P, were up-regulated under 2% O_2 condition.

3.3.5.5 Secretion and Transport Systems

Our analysis showed that secretion and transport proteins had significantly higher spectral counts under 2% O_2 with significant expression of an ATP-binding cassette protein (ABC transporter), type II secretion system protein, and tetratricopeptide (TPR) repeat-containing protein in both transcriptomic and proteomic analyses. The strain TAV2 expressed the enzyme SecA wing and scaffold^P, involved in the Sec translocase system. SecA is responsible for hydrolyzing ATP during protein export to the periplasm (Economou and Wickner, 1994). Under

2% O_2, TAV2 cells expressed solute-binding proteins family 1^P and family 3^P. These carrier-mediated transport proteins help to increase the rate of solute uptake and accumulate solute inside the cells.

3.3.6 Validation of Expressed Genes through RT-qPCR

To confirm the transcriptomic results, induction values were calculated for eight selected genes using reverse transcription followed by quantitative PCR analysis. All up-regulated genes selected for comparison were confirmed to be expressed at significantly higher levels. Owing to our particular interest in carbon metabolism, one acetyl xylan esterases ($P < 0.05$) and three xylose isomerases (ObacDRAFT_3012, $P < 0.01$; ObacDRAFT_1973 and ObacDRAFT_0419, $P < 0.05$) that were not observed to be up-regulated in our transcriptomic results, but showed ≥ 2 fold peptide change were included in this analysis. Up-regulated expressions for all four genes were observed when measurements were carried out through RT-qPCR.

3.4 Discussion

The results of this study confirm the microaerophilic nature of strain TAV2 (Wertz *et al.*, 2012), with a higher growth rate observed for cells grown under 2% O_2 than those grown under 20% O_2. The long lag phase shown for cells grown in 20% O_2 suggests the need of an acclimation period and the possibility of cells being under oxidative stress. Similarly, longer lag phases have been observed for the termite gut microaerophile *Stenoxybacter acetivorans* grown under increasing O_2 concentrations (Wertz and Breznak, 2007b). When TAV2 cells are maintained under 2% O_2, a condition associated with the physicochemical characteristics of the peripheral zone of the termite hindgut (30 mbar O_2; (Brune *et al.*, 1995)), genes associated with functional categories responsible for energy production and conversion, carbohydrate transport and metabolism, and replication and recombination are up-regulated as evidenced in both transcriptomic and proteomic data. It is noteworthy to observe that the COG category cell cycle control showed a divergent pattern between proteomic and transcriptomic approaches. We

attribute this difference to the small number of genes (18 or 0.66% of the genome) belonging to cell cycle control in comparison to other categories. Refinements on gene identification and annotation methods will help to mitigate possible differences. In addition, several factors may explain transcriptome and proteome differences, such as post-translational modifications, half-lives of mRNA and proteins, and detection limits for both technologies (Zhang *et al.*, 2010).

TAV2 cells show signs of oxidative stress when grown under 20% O_2 condition. First, even the superoxide dismutase was not significantly up-regulated in both transcriptomic and proteomic analyses, our proteomic approach detected the presence of peptides derived from superoxide dismutase. This result provides an explanation on how TAV2 manages reactive oxygen species (ROS) and helps to conciliate our previous experimental work, in which the activity of superoxide dismutase was not observed (Wertz *et al.*, 2012). Second, TAV2 cells promote a shift in protein expression with a substantial increase of categories associated with amino acid transport and metabolism, protein turnover and post-translational modifications, and signal transduction mechanisms. Specifically, proteins encoded by the *suf*C- and *suf*D-like genes in *Escherichia coli*, and predicted to be an ATPase and a complex-stabilizing protein (Nachin *et al.*, 2001; Nachin *et al.*, 2003), respectively, were highly abundant. These proteins are associated with the formation of [Fe-S] clusters and involved in repair during oxidative stress. Considering the importance of the initial microbial colonization of recently hatched termite larvae or frequent trophallactic transfers between individuals (Brune and Ohkuma, 2011), the presence of different mechanisms to cope with ROS is an important feature for any termite microaerophile. Adaptations to O_2 toxicity might be a widespread attribute in the termite hindgut. (Leadbetter and Breznak, 1996) showed that methanogens, usually thought to be strict anaerobes, possess catalase activity. It is noteworthy that their isolates were *in situ* localized as present on or near the hindgut wall.

TAV2 cells employ the glycolytic pathway for processing carbohydrates to pyruvate and exert a tight control over the tricarboxylic acid (TCA) cycle. The expression of pyruvate dehydrogenase, the enzyme responsible for providing acetyl-CoA to the TCA cycle, was down-

regulated under 2% O_2. Previously, Pauling and co-workers (2001) demonstrated that *Azorhizobium caulinodulans* cells exhibit a dramatic increase in NADH:NAD^+ ratio when cultures are shifted from aerobic to microaerobic conditions. It is possible that hypoxia in TAV2 also resulted in increase of NADH, a well known inhibitor of the enzymes pyruvate dehydrogenase, citrate synthase, isocitrate dehydrogenase, and α-ketoglutarate dehydrogenase (Barton, 2004), which slows down the overall activity of the TCA cycle. The adaptive consequence of a slow TCA cycle turnover remains to be investigated in depth for microaerophiles, but it suggests a controlled rate of substrate conversion. We hypothesize this metabolic design might be a consequence of the O_2 limitation, preventing NADH and $FADH_2$ from being readily oxidized. This is similar to the strategy used by fermentative bacteria operating a reductive TCA cycle (Ludwig, 2004). Although one might expect that the electron transport system and the ATP synthesis machinery may be affected as well, enzymes such as NADH dehydrogenase, cytochrome c oxidase, and ATP synthase alpha (F_1) and beta (f_0) subunits were highly expressed under the low O_2 condition. This is consistent with previously observed increase in ATP formation with decreasing O_2 concentrations in other microaerophiles (Bergersen and Turner, 1975; Jackson and Dawes, 1976). Interestingly, (Graber and Breznak, 2004) observed a high molar ratio conversion for the termite spirochete *Treponema primitia* without substantial increase in biomass. The authors suggested that this mechanism would prevent overpopulation of the termite hindgut, destabilizing the symbiosis.

One of the well-known features of wood-feeding termites is that their diet is low in combined nitrogen (Brune and Ohkuma, 2011; Potrikus and Breznak, 1981). The TAV2 cell seems to be well adapted to its environment by carrying genes for biosynthesis of 20 of the most common amino acids plus ornithine. Our combined transcriptomic and proteomic data made possible for the identification of gene transcripts and proteins for biosynthesis of eight amino acids. This is a surprising finding considering that the growth medium was selected to elicit a broad metabolic response and contained yeast extract and casamino acids, albeit in low concentrations. The potential for biosynthesis and the expression of genes related to amino acid

formation have been shown for two protozoa endosymbionts (Hongoh *et al.*, 2008a; Hongoh *et al.*, 2008b) and two co-cultivated termite gut spirochete species (Rosenthal *et al.*, 2011). Previously, (Sabree *et al.*, 2012) proposed that the nutrient provisioning in termite-bacterial interaction was altered as the intracellular microorganism *Blattabacterium*, a symbiont in cockroaches, was replaced by hindgut protozoa and/or bacteria in almost all termite species. The presence of biosynthetic genes for all 20 common amino acids in the TAV2 genome and the subsequent expression of many of these genes suggest an array of possible interactions between the *Verrucomicrobia* population and its host as well as bacterial and protozoan members of the gut community.

Our proteomic results indicated the presence of specific peptides corresponding to the dinitrogenase reductase (NifH) only when cells were grown with 2% O_2, corroborating our previous findings for the presence of nitrogen fixation genes (*nif*) and growth on N-free medium (Wertz *et al.*, 2012). Finally, the expression of a urea-binding signal peptide protein and the α subunit of urease may indicate a role for TAV2 in the recycling of uric acid excreted by the host. Previously, (Potrikus and Breznak, 1981) determined uricolysis to be a process mediated by gut bacteria occurring under anoxic conditions. The possibility that uric acid degradation occurs under microaerophilic conditions remains an interesting aspect of symbiosis, yet to be studied. Taken together, the above results imply that TAV2 can contribute to the nitrogen requirements of the termite. We remain cautious about establishing such a link based on the low *Verrucomicrobia* population size (1.2×10^3 cells) residing in the hindgut (Köhler *et al.*, 2012; Wertz *et al.*, 2012), but even a small contribution could be important in an N-limited ecosystem.

Lately, a renewed interest in the members of the phylum *Verrucomicrobia* has arisen because of their potential role in complex polysaccharide degradation (Flint *et al.*, 2012; Martinez-Garcia *et al.*, 2012) and importance in the biogeochemical cycling of carbon (Bergmann *et al.*, 2011; Freitas *et al.*, 2012). Although, we have not yet to identify specific conditions allowing strain TAV2 to grow in xylan as sole C source (Wertz *et al.*, 2012), our integrated approach identified genes and proteins involved in lignocellulose degradation, specifically the enzymes acetyl-xylan

esterase, xylan α-1,2-glucuronosidase, and xylose isomerase, the later being responsible for converting xylose into xylulose, which enters the PPP (White, 2007). Because much of investigation in termite guts have targeted cellulose degradation through biochemical characterization of host celullases (Watanabe and Tokuda, 2010), gene expression of glycosyl hydrolases by protists (Todaka *et al.*, 2010; Todaka *et al.*, 2007), and metagenomic (Warnecke *et al.*, 2007) and proteomic (Burnum *et al.*, 2011) approaches, the understanding of xylan hydrolysis remains virtually unexplored. Our results reveal an important and perhaps compartmentalized, ecological role for the TAV2 population on the debranching of xylan residues. Owing to the heteropolymeric structure of xylans, their degradation requires a coordinated action of a microorganism's transcriptional apparatus and a substantial metabolic investment on the production of many xylanolytic enzymes. We were able to identify different acetyl xylan esterases and xylose isomerases being expressed in TAV2 cells (Appendix D). It appears that the degradation of the xylans might be a wide spread function among members of the phylum in different environments. Besides the strain TAV2, which belongs to the class *Opitutae*, soil isolates from classes *Verrucomicrobiae, Spartobacteria*, and subdivision 3 were obtained with diluted medium containing xylan as sole C source (Sangwan *et al.*, 2005). Furthermore, experiments on single cell genomics combined with fluorescently-labeled xylan and laminarin in marine and fresh waters yielded *Verrucomicrobia* genomes particularly enriched in glycoside hydrolases (0.91% of total number of genes) in comparison to other sequenced bacterial genomes (0.2%) (Martinez-Garcia *et al.*, 2012). Together, these independent results not only indicate an important ecological role in biopolymer recycling for the phylum, but also raise new questions on environmental conditions necessary for a coordinated response.

In this study we asked whether O_2 was a relevant factor for the proteomic profile differences observed for strain TAV2. Approximately 55.9% of the variance is explained solely by changes in O_2 concentrations, suggesting a strong adaptive evolution of TAV2 within the wall-associated, oxygen-consuming termite gut community. Certainly, the identification of other environmental parameters and their combined effects on cell physiology will increase our

understanding of the functional role of this species and other members of the phylum *Verrucomicrobia*.

3.4.1 A working model for Diplosphaera colotermitum strain TAV2 in the termite hindgut

Our work provides the first integrated omics approach for understanding the ecological role of a member of the phylum *Verrucomicrobia*. We found that the TAV2 strain can contribute to the metabolism of the termite gut microbial community through biological N_2 fixation, amino acid production, degradation of xylans, and removal of free O_2. A conceptual model for the different functions carried out by the *Diplosphaera colotermitum* strain TAV2 is proposed in Figure 3.5.

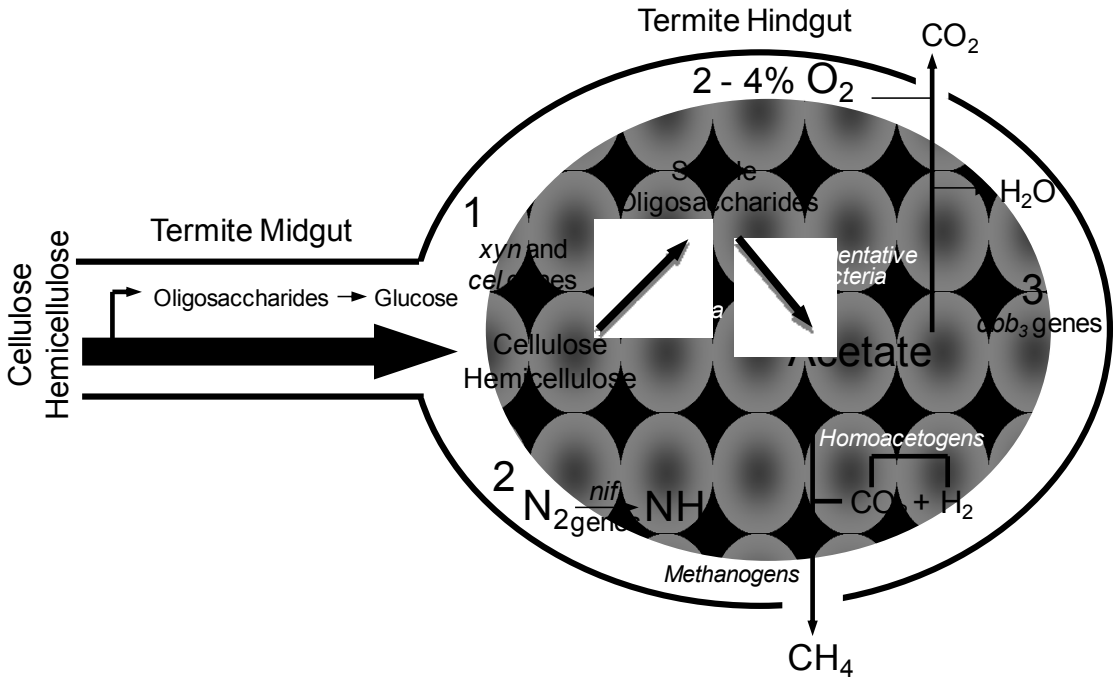

Figure 3.5 Model representation of possible functional roles for the *Diplosphaera colotermitum* strain TAV2 in the hindgut of the termite *Reticulitermes flavipes*.

Note: Strain TAV2 expressed genes and/or proteins associated with (1) hemicellulose degradation, (2) biological nitrogen fixation, and (3) controlled oxygen consumption. Shaded area represents anoxic core and drawing is not to scale. Other important microbial groups are also represented in the hindgut ecosystem. For detailed morphological descriptions of microorganisms

and the termite gut epithelium through transmission electron- and scanning electron- microscopy see (Breznak and Pankratz, 1977).

Finally, our results underscore the importance of O_2 for whole genome expression and, ultimately, the physiological state of a cell. Understanding the factors that regulate the microaerophilic physiology will broaden our view of an evolutionarily adapted group, sometimes mistakenly regarded as slow-growing aerobes.

Chapter 4

Cloning, Expression and Purification of Acetyl Xylan Esterase from the *Diplosphaera colotermitum* Strain TAV2

4.1 Introduction

Plant cell wall is composed of cellulose, hemicellulose, and lignin. After the cellulose, hemicellulose is the second most abundant natural polysaccharides making up to 20-30% of plant dry weight (Eriksson *et al.*, 1990). Hemicelluloses are heterogeneous polysaccharides consisting of various sugars including pentoses (xylose, arabinose), hexoses (glucose, galactose, mannose), sugar acids (glucuronic acid), acetic acid, ferulic acid, etc. (Saha, 2003). The most abundant hemicellulose commonly found in hardwood is xylan, while glucomannans are found in softwood (McMillan James, 1994). Typically, xylan contains a *β*-1,4-linked xylopyranosyl backbone attached to various side chains of acetyl, arabinofuranosyl, and 4-*O*-methyl glucuronyl groups (Thomson, 1993). Xylans can be categorized into four groups which are homoxylans, arabinoxylans, glucuronoxylans, and arabinoglucuronoxylans (Dodd and Cann, 2009). Homoxylans are composed of recurring units of xylose molecules connected by *β*-1,4- and *β*-1,3 linkage which are commonly found in cell walls of red seaweeds. Xylans found in higher plants have a structure of *β*-1,4-linked xylopyranose sugars substituted with acetyl groups. Arabinoxylans, usually found in cell walls of grasses, cereal grains and wheat, consist of arabinose residues linked to the *O*-2 or *O*-3 of the xylose backbone. Glucuronoxylans consist of 4-*O*-methyl-*α*-D-glucuronic acid side groups linked with the *O*-2 of xylose backbone. These xylans are found in hardwoods, herbs, and woody plants. Arabinoglucuronoxylans have arabinofuranosyl, 4-*O*-methyl-*α*-D-glucuronic acid, and acetyl side chains linked to the xylose backbone. This type of xylan is found in lignocellulose isolated from grasses and is considered as a potential source for biofuel production (Dodd and Cann, 2009).

Composition of xylan in hardwood, softwood, cereal, and grasses has different constituents. Birchwood xylan contains relatively high numbers of xylose at 89.3%, and the

remaining is composed of anhydrouronic acid (8.3%), glucose (1.4%) and arabinose (1%) (Kormelink and Voragen, 1993). Xylan of rice bran contains xylose (46%), arabinose (44.9%), galactose (6.1%), glucose (1.9%), and anhydrouronic acid (1.1%) (Shibuya and Iwasaki, 1985). Wheat arabinoxylan contains mostly xylose and arabinose about 65.8% and 33.5%, respectively, and small percentages of glucose, mannose, and galactose (Gruppen *et al.*, 1993). From the breakdown of xylan, xylose and arabinose will be the mainly constituent sugars.

In order to breakdown polysaccharides of xylan into monosaccharides, various enzymes in a xylanolytic system are required to complete hydrolysis process. Backbone-degrading enzymes are *β*-D-xylosidases (EC 3.2.1.37), endo-1,4-*β*-xylanases (EC 3.2.1.8), and side group-cutting enzymes are acetyl xylan esterases (EC 3.1.1.72), *α*-L-arabinofuranosidases (EC 3.2.1.55), *α*-glucuronidases (EC 3.2.1.139), and ferulic/coumaric acid esterases (EC 3.1.1.73) (Zhang *et al.*, 2011) . Acetyl xylan esterases (AXEs) are enzymes that hydrolyse ester linkages of acetyl side groups especially in positions 2 and/or 3 of the xylose backbone (Dupont *et al.*, 1996). Typically, the acetyl ester groups of xylan are found in hardwoods and cereals, but are generally deficient in softwoods (Dupont *et al.*, 1996). The AXEs were present in various bacteria, such as *Fibrobacter succinogenes* (McDermid *et al.*, 1990), *Bacillus pumilus* (Degrassi *et al.*, 1998), *Streptomyces lividans* (Shareck *et al.*, 1995), and also present in fungi such as the anaerobic fungus *Orpinomyces* sp. Strain PC-2 (Blum *et al.*, 1999), and *Trichoderma reesei* (Poutanen *et al.*, 1990).

Xylose backbone can be substituted with acetyl group at position *O*-2 and *O*-3 about 22-50% (Blum *et al.*, 1999). Increasing degrees of side chain acetylation can obstruct xylanase activity in hydrolysis of xylan (Poutanen *et al.*, 1987), and solubility of xylan can be improved by removal of these substituted side groups like acetyl, arabinosyl, etc. It was observed that the more acetyl groups have been removed, the more xylose released from hydrolysis of corn stover substrates by synergistic activity of endoxylanase and acetyl xylan esterase (Selig *et al.*, 2009). Furthermore, removal of xylan can enhance performance of cellulase in enzymatic hydrolysis of cellulose. This is because xylans are distributing between cellulose microfibrils, so they can

occasionally obstruct accessibility of cellulase to hydrolyze cellulose (Saha, 2003). By addition of xylanases, the ability of cellulase in degrading cellulose is significantly increased (García-Aparicio *et al.*, 2007).

In addition to the function of enzyme in removing acetyl group from xylan backbone, this enzyme is found to be up-regulated in our transcriptomic and proteomic analyses of *D. colotermitum* strain TAV2. Moreover, most researches have studied properties of AXE mainly produced by fungi but the AXE produced by bacteria is not well characterized. In this study, we perform gene cloning and expression of acetyl xylan esterase of strain TAV2 as well as purification of the AXE in order to characterize the enzyme properties.

4.2 Materials and Methods

4.2.1 Plasmid, Bacterial Strains, and Culture Conditions

Plasmid pET-28a (5369 bp), *Escherichia coli* TOP10, and *E. coli* BL21 (DE3) strains were used. Growth media for gene cloning were Luria Bertani (LB) agar (HIMEDIA) and LB broth miller (EMD, Billerica, MA) supplemented with 50 µg/ml kanamycin. Incubating temperature was set at 37 °C, 200 rpm in case of growing on the broth media.

4.2.2 Preparation of axe Gene Clones

TAV2 genomic DNA was isolated using a cetyltrimethylammonium bromide method (http://my.jgi.doe.gov/general/) and was used as a template for amplifying 1317 bp *of axe* gene (gene ID 641180470, ObacDRAFT_0993, http://img.jgi.doe.gov). Two oligonucleotide primers including restriction enzyme sequences (Sigma-Aldrich, St. Louis, MO) which are *Nde*I (underline) in AXEup (5'-GGAATTCCATATGATGAAATCCCGCATCCG-3') and *Xho*I in AXEdown 5'-GCCCTCGAGTCATTTGTTTTTAACTACAA-3') were used to amplify the *axe* gene. Reagents used for PCR reaction were the FastStart High Fidelity PCR system (Roche, Indianapolis, IN). PCR reaction was run in C1000 Thermal Cycle (Bio-Rad, Hercules, CA) with following parameters: 95^{o}C for 2 min, 35 cycles of 95^{o}C for 30 sec, 55^{o}C for 30 sec, and 72^{o}C for 1.30 min,

and 72^{o}C for 5 min. PCR product was purified with the QIAquick PCR purification kit (Qiagen Sciences, Maryland, MD). The correct band size of amplified product was verified by loading the PCR product onto a 1% agarose in 1X Tris-Acetate EDTA (TAE) gel, running the gel at 80 V for 40 minutes using a PowerPac^{TM} Basic (Biorad, Hercules, CA) and visualizing the desired band under UV light using GelLogic 212 PRO Carestream (Carestream Health, New Haven, CT). The amplified PCR product was digested with *Nde*I and *Xho*I restriction enzymes (New England Biolabs, Ipswich, MA BioLabs) at 37^{o}C for 17 h and the reaction was inactivated at 65^{o}C for 20 min. Ligation was performed in 20 µl reaction by using 0.0577 pmol of PCR product, 0.0192 pmol of linearized pET-28a plasmid, 2 µl of 10X T4 DNA ligase buffer, and 1 µl (400 U/µl) of T4 DNA ligase (New England Biolabs, Ipswich, MA) and incubated for 20 min at room temperature.

4.2.3 Transformation of axe-pET-28a into E. coli

Two µl of axe-ligated plasmid from ligation reaction was transformed into 50 µl of competent *E.coli* TOP10 by electroporation (2 KV). One hundred µl of transformed *E. coli* TOP10 cells were spread on Luria–Bertani (LB) agar (HIMEDIA) containing 50 µg/ml kanamycin and incubated at 37^{o}C overnight. Plasmids were extracted from white colonies using the QIAprep spin miniprep kit (Qiagen Sciences, Maryland, MD) and screened for the presence of insert using primers specific to *axe* gene as described above. The axe-pET-28a expression plasmid was extracted from *E. coli* TOP10 and then transformed to the host *E. coli* BL21 (DE3) strain for protein expression by electroporation (2 KV).Transformed *E. coli* BL21 (DE3) cells were plated on LB agar containing 50 µg/ml kanamycin and incubated at 37^{o}C overnight.

4.2.4 Confirmation of Cloned Insert and the Presence of Histidine Tag

Plasmid was extracted from a transformant *E. coli* BL21 (DE3). Confirmation for the presence of axe insert was performed in 50 µl of enzyme digestion reaction by using 5 µg of plasmid, 5 µl of 10X NEBuffer 4, 0.5 µl of 100X purified BSA (10 mg/ml), 25 U of *Nde*I and *Xho*I restriction enzymes (New England Biolabs, Ipswich, MA BioLabs). The reaction was incubated at

37^{o}C for 17 hrs, followed by restriction enzyme inactivation at 65^{o}C for 20 min. The sample was loaded onto 1% agarose gel as the method described above to check for the expected band size.

DNA sequencing was another method used to confirm the presence of cloned insert. Colony PCR of a transformant *E. coli* BL21 (DE3) was prepared for DNA sequencing. Primers used for gene amplification were *Nde*I:AXEup and *Xho*I:AXEdown, and the primers for T7 promoter (5'-TAATACGACTCACTATAGGG-3') and *Xho*I:AXEdown were used to amplify the promoter region of the pET-28a plasmid. The PCR reagents were the FastStart High Fidelity PCR system (Roche) and followed the thermal cycle condition as stated above. Expected band size was verified by loading PCR product onto 1% agarose gel. Nine µl of the resulting PCR products were cleaned by mixing with 1 µl of 10% ExoSap-IT (Affymetrix, Santa Clara, CA), and the mixture was run in the C1000 Thermal Cycle (Biorad, Hercules, CA) at the followings; 37°C for 20 min, 80°C for 20 min, and 4°C. The reaction was centrifuge at 16,000 x g for 2 min, and then 1 µl of supernatant was transferred to a clean tube containing 9 µl of primers (primers concentration were set to 0.4 pmol/µl for gene sequencing and was set at 1.6 pmol/µl for plasmid sequencing). The samples were sequenced using the Sanger method at the Genomics Core Facility (University of Texas at Arlington). The sequences were assembled using Sequencer 4.10.1 (Gene Codes Corporation, Ann Arbor, MI), aligned and analyzed using basic local alignment search tools (BLAST) in the NCBI website (http://blast.ncbi.nlm.nih.gov/).

4.2.5 Time-Course Analysis for Optimal Protein Expression

A single colony of *E. coli* BL21 (DE3) harboring axe-pET-28a was grown overnight at 37^{o}C, 242 rpm in 10 ml ZYM-5052 starter medium supplemented with 50 µg/ml kanamycin as described in Appendix E. One ml of uninduced culture was harvested at 16,000 x g for 1 min, resuspended in ice-cold PBS, pH 7.3, and the pellet was stored in -20^{o}C. The remaining culture was diluted to an optical density (OD_{600}) of 1.0 using ZYP-5052 rich medium containing lactose for auto-induction. The cell culture was further incubated and 1 ml of the culture was harvested periodically at 5 h, 8 h and 24 h after adding the rich media. The uninduced and induced samples

were resuspend in 200 µl of ice-cold PBS, pH 7.3 supplemented with 1 mg/ml lysozyme, undergone five freeze-thaw cycles (-80 and 37^{o}C, 10 min each), and centrifuged at 16,000 x g for 2 min to remove the cellular debris. Four µl of lysates was mixed with 5X sample buffer, heated to 95°C for 10 min. Optimal protein expression was identified by loading the lysate samples onto 8-12% SDS-PAGE gel (Nusep, Bogart, GA) and running at 120 V for 1 hour using the VWR power source 300 V (VWR International, Radnor, PA).

4.2.6 Culture Condition for Protein Expression

A single colony of *E. coli* BL21 (DE3) harboring *axe*-pET-28a was grown overnight at 37^{o}C, 242 rpm in 25 ml ZYM-5052 starter media supplemented with 50 µg/ml kanamycin. Then, the cell culture was diluted to an optical density (OD_{600}) of 1.0 using ZYP-5052 rich medium and the culture was further incubated for 24 h after adding the rich media. Cell culture (45 ml) was harvested by centrifugation at 10,967 x g for 5 min at 4^{o}C (Sorvall RC 6 plus Thermo electron corporation), and washed with ice-cold 50 mM phosphate buffered saline (PBS; pH 7.3). Pellets were stored at -80 ^{o}C for next step of purification.

4.2.7 Histidine-Tagged Protein Purification

Pellet harvested from 45 ml cell culture was thawed on ice for 15 min and resuspended in 10 ml lysis buffer (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0). The cell suspension was supplemented with 1X protease inhibitor and 1 mg/ml lysozyme and stirred overnight at 4^{o}C. The cell suspension was further disrupted by five freeze-thaw cycles (-80 and 37^{o}C, 10 min, each). The cell extracts were centrifuged at 17,146 x g for 20 min to remove cellular debris. The clear lysate was filtered through 0.45 µl cellulose membrane, and then 4 ml lysate was mixed with 1 ml of Ni-nitrilotriacetic acid agarose slurry (Ni-NTA, Qiagen Sciences, Maryland, MD). The mixture was incubated at 4^{o}C for 2 h using end-to-end rotator. AXE protein was purified by gravitational flow purification by applying the incubated mixture into a 15 ml column and collecting the flow-through. The Ni-resin was washed three times with 15 ml wash

buffers 1, 12 ml wash buffer 2, and 10 ml wash buffer 3 (10 mM Tris, 100 mM NaH_2PO_4, pH 8.0, 6.3, and 5.9, respectively). AXE protein was eluted four times with 500 µl of elution buffer (10 mM Tris, 100 mM NaH_2PO_4, pH 4.0). 400 µl of eluate was applied to a strong cation exchange mini spin column (Pierce, Rockford, IL). The protein was eluted by 150 µl of elution buffer (25 mM CH_3COONa, 500 mM NaCl). The eluted proteins were concentrated using Amicon ultra-0.5 centrifugal filter with 30K cut-off (Millipore, Billerica, Massachusetts) and the eluate was analyzed by SDS-PAGE. Protein concentrations were determined with a bicinchoninic acid protein assay kit (Pierce, Rockford, IL) with bovine serum albumin as the standard.

4.2.8 SDS-PAGE Analysis

After purification steps, the protein samples were mixed with 5X sample buffer (in 10 ml buffer; 250 mM Tris-HCl pH 6.8, 500 mM DTT, 10% SDS, 0.1% bromophenol blue, 50% glycerol), heated to 95^{o}C for 10 min, run through 8-12% SDS-PAGE (Nusep, Bogart, GA) in 1X Tris-glycine electrophoresis running buffer (3.02 g Tris base, 14.4 g glycine, 1 g SDS in 1L dH_2O) at 120 V for 1 hour using the VWR power source 300 V (VWR International, Radnor, PA). After that, the gel was stained overnight in 15 ml coomassie brilliant blue NuBlu Express stain (Nusep). The protein marker was the PageRuler broad range unstained protein ladder 10-200 kDa (Thermo scientific, Waltham, MA). The gel was visualized under white light in the GelLogic 212 PRO Carestream (Carestream Health, New Haven, CT).

4.2.9 Western Blot Analysis

Eluted protein from gravitational flow purification and spectra multicolor broad range protein ladder (Thermo Scientific, Rockford, IL) were first separate in 8-12% SDS-PAGE at 70 volt for 60 min and then the voltage was increased to 100 and run until bromophenol blue reached the end of gel. Next, the gel was equilibrated for 30 min in transfer buffer (25 mM Tris base, 192 mM glycine, 10% methanol, per 1L of dH_2O) with gentle shaking. For electroblotting, a cassette of the gel, polyvinylidene difluoride (PVDF) membrane, filter papers, and scour pads

were assembled and placed in a blotting tank (Bio-Rad, Hercules, CA) containing ice packs and ice-cold 1X Tris-buffer saline (TBS; per 1L, 10mM Tris HCl (pH 7.5) and 154 mM NaCl) and run at 100 volt for 90 min. Upon completion, the membrane was washed in 1X TBS at 50 rpm for 2 min. The membrane was blocked in 5% non-fat dry milk for 1 h, and then the membrane was removed from the milk solution and probed with a 1:1000 dilution of anti-histidine tagged antibody, clone HIS.H8 (Millipore, Billerica, Massachusetts) in 5% non-fat dry milk for 1 h. To remove unbound antibody, the membrane was washed three times, 15 min each, in 50 ml of 1X Tween-20 in Tris Buffered Saline (TTBS; per 1L, 10 mM Tris HCl (pH 7.5), 154 mM NaCl, 0.01% Tween 20) and washed one time in 50 ml 1X TBS for 15 min. The membrane was probed overnight with a 1:1000 dilution of secondary antibody (Goat anti-Mouse IgG, (H+L) HRP conjugate, Millipore, Billerica, Massachusetts) in 5% non-fat dry milk. Then, the blot was washed with 1X TTBS and 1X TBS as described above. To detect the probe bound to his-tagged protein, the membrane was developed with 4 ml of 1:1 ratio of chemiluminescence reagents (SuperSignal West Pico**;** Pierce, Rockford, IL). The membrane was blotted to dry, wrapped in clear plastic, and exposed to film in dark room.

4.2.10 Enzyme Assay

Acetylated birchwood xylan preparation and enzyme assay methods were performed by the method described by Johnson et al. 1988. For optimum temperature characterization, 100 µl of 100 mM potassium phosphate pH 6.0 was added, 4 µg of AXE enzyme, and distilled water added up to total volume of 300 µl. Then, 200 µl of 10 % acetylated xylan in distilled water was added to initiate the reaction. The reaction was incubated at 25, 37, 50, 60, and 70^{o}C for 25 min and stopped the reaction by adding 10 µl of 1N of sulfuric acid. The tube was centrifuged at 8000 g for 2 min and the clear supernatant was used for measuring released acetic acid using acetic acid assay kit (K-ACET, Megazyme International Ireland, Wicklow, Ireland) For optimum pH characterization, the procedure was the same as optimum temperature assays except buffers used in each pH range were

different; 100 mM citrate buffer from pH 3.0 to 5.0, 100 mM phosphate buffer from pH 6.0 to 7.0, and 100 mM tris buffer from pH 8.0 to 9.0 and the reactions were incubated at 50^{o}C for 25 min.

4.3 Results and Discussion

4.3.1 Nucleotide and Amino Acid Sequences of Acetyl Xylan Esterase

Nucleotides sequencing of *axe* gene (ObacDRAFT_0993) contains 1,317 bp encoding 438 amino acids with 47.85 kilodalton of calculated molecular weight. The sequence has a predicted isoelectric point at 9.66 and charge of 13.5 (http://img.jgi.doe.gov). Deduced amino acid sequence of AXE lacks of signal peptide when searching with SignalP4.0 (Petersen *et al.*, 2011), while some enzymes in xylanolytic system such as xylanase of *Clostridium cellulovorans* (Kosugi *et al.*, 2002) and AXE from the anaerobic fungus *Orpinomyces* sp. strain PC-2 contain signal peptide and catalytic domain (Blum *et al.*, 1999). Comparison of the deduced amino acid sequence of AXE with amino acid sequences database deposited in NCBI using blastp (http://www.ncbi.nlm.nih.gov/blast) showed that AXE contains sequence similar to AXE of the strain TAV2 encoding from different locus tags with 85% maximum identity. Moreover, the sequence has 36-77% identity with several AXEs of their closely related species, strain TAV1 and TAV 5. Outside the TAV strains, the AXE sequence reveals 37% identity to cephalosporin-C deacetylase, [*Rhodopirellula maiorica* SM1]. The presence of AXE in a group of TAV strains and some other enzymes which are homologous to AXE could be a result of horizontal gene transfer from one organism to other organisms, and then the gene has subsequently been lost or added some domains. Blum *et.al.,* 1999 found that AXE of anaerobic fungus *Orpinomyces* sp. strain PC-2 has 56% amino acid sequence identity with BnaA from *Neocallimastix patriciarum* but the *N. patriciarum* contains both catalytic and dockerin domains, while *Orpinomyces* sp. has only catalytic domain. When searching conserved domain using CD-search (Marchler-Bauer and Bryant, 2004) in NCBI, functions of the deduced amino acid could not be defined with high confidence because it doesn't have a specific hit with the conserved domain. However, the

sequence has a statistically significant hit (E-value cut off of 0.01) with some domain models, so in this case, the general function can be inferred as superfamily but specific function is uncertain. The sequence of AXE revealed that it is classified in esterases and lipases (cl2031) superfamily. Enzymes of this group (EC: 3.1.1.-) function on carboxylic esters in which the carbonyl carbon atom of the ester bond is broken down by catalytic residues. Furthermore, the sequence revealed that AXE is classified in multifamily domain of alpha/beta hydrolase family (pfam12695) containing several single domains, which are, acetyl xylan esterase (AXE1; pfam05448), acetyl esterase (deacetylase) [COG3458, secondary metabolites biosynthesis, transport, and catabolism], dipeptidyl aminopeptidases/acylaminoacyl-peptidases [COG1506, amino acid transport and metabolism], dienelactone hydrolase [COG0412, secondary metabolites biosynthesis, transport, and catabolism] (Marchler-Bauer *et al.*, 2013). Comparison of expect values (E-value) among these domains, acetyl xylan esterase (AXE1; pfam05448) and acetyl esterase (deacetylase) [COG3458] have the lowest E-value at 3.19×10^{-22} and 1.36×10^{-18}, respectively, so it is more confident to infer functions of this sequence as acetyl xylan esterase and acetyl esterase.

4.3.2 Confirmation of Cloned Insert

From genome sequence of the strain TAV2, *axe* gene has a nucleotide sequence of 1317 bp. The gene is inserted into a polylinker site of the expression vector pET-28a downstream from the T7 RNA polymerase promoter. In order to comfirm insertion of the gene, pET-28a plasmid was extracted from *E.coli* BL21 (DE3) and digested by *Nde*I and *Xho*I restriction enzymes. Figure 4.1 shows the presence of *axe* gene with the size of 1317 bp which is digested from 5369 bp of pET-28a plasmid.

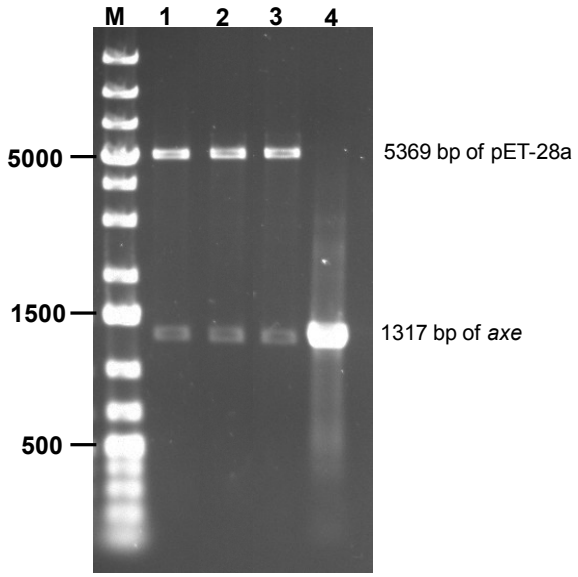

Figure 4.1 Confirmation of cloned insert, *axe* gene in pET-28a, using restriction enzyme digestion.

M, 1 kb DNA ladder (bp); lane1-3 triplicate samples of *axe* digested from pET-28a plasmid; lane 4, positive control of axe gene.

In addition to restriction enzyme digestion, DNA sequence is also used to confirm the presence of *axe* gene fused into pET-28a as well as histidine tag. When aligning two DNA sequences of the *axe* gene (ObacDRAFT_0993) obtained from genome sequences (Joint Genome Institute, http://img.jgi.doe.gov/) and the consensus sequence of the cloned insert of this experiment, both sequences have 96% maximum identity with 76% query coverage (E value = 0.0) (Zhang *et al.*, 2000). Moreover, the sequence of cloned insert shows gene encoding for 6x histidine-tag (6xHis-tag) when using T7 promoter primer to amplify the insert. The presence of *axe* gene in pET-28a confirms that *axe* is inserted into the vector and It is expected that when the *lac* promoter is induced by lactose, the 6xHis-tag fused to its N-terminus will be expressed as a his-tagged target protein. The his-tagged protein facilitates protein purification by binding to Ni-resins and all other proteins are removed.

4.3.3 Determination of Optimal Protein Production

Optimal protein induction is performed by growing an overnight culture of *E.coli* BL21 (DE3) containing axe-pET-28a in a starter medium. At this time point, a light band of AXE protein at approximately 48 kDa is present as shown in figure 4.2 A (lane 1). After adding the rich media containing lactose to induce the *lac* operon, the band size of the AXE protein is increasing with time after 5, 8, and 24 h of incubation (lane 2-4), and the maximum band size appears at 24 hours. For this reason, the cells for protein purification are collected at this time frame in order to get the maximum protein yield. In order to compare between uninduced and induced AXE protein, *E.coli* BL21 (DE3) containing only pET-28a without *axe* insert is grown in the same condition for 24 h, and the result shows that there is no band of induced AXE protein at 48 kDa in comparison to the one containing *axe* insert as shown in figure 4.2 B.

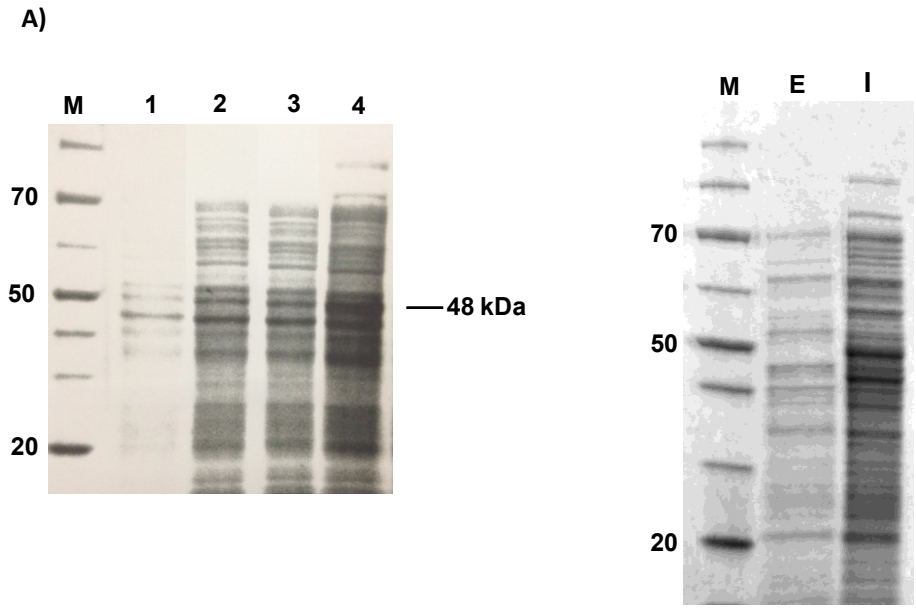

Figure 4.2 SDS-PAGE [8-12% (w/v)] of time course for optimal protein induction (A), and comparison of induced proteins from the *E. coli* BL21 (DE3) containing empty plasmid and *axe*-inserted plasmid (B).

M, standard protein marker (kDa); Lane1, lysate of cells grown overnight in starter media; Lane2-4, lysate of cells grown 5, 8 and 24 hours after adding rich media; E, cell lysate of *E. coli* containing empty pET-28a without *axe* insert; I, cell lysate of *E. coli* containing pET-28a with *axe* insert.

4.3.4 Purification of Acetyl Xylan Esterase

AXE enzyme was produced as cell associated and was detected in both soluble protein and in inclusion body after cell lysis. The enzyme was not found in culture supernatant. To perform protein purification, four ml of soluble protein lysate was loaded in to a gravitational flow column containing 1 ml Ni-NTA slurry, then, the protein was washed three times in washing buffers and eluted four times in elution buffer. The concept of protein purification used in this experiment was to decrease pH of the washing buffers from 8.0, 6.3, to 5.9, respectively and pH of the elution buffer was lowered to 4.0. The protein could be eluted at low pH because histidine residues in 6xHis tag have a pKa of 6.0. By lowering pH to 4.0, the histidine residues become protonated, so the his-tagged protein cannot bind to Ni resin anymore and will desorb from the resin. Figure 4.3A shows that most of non his-tagged proteins that do not bind to the resins are removed first in flow through and washing buffers, while his-tagged protein binds firmly to the Ni resins so it is eluted later in elution buffer. Since other proteins are eluted together with AXE as shown in lane E1, the eluate was further purified using ion exchange column. Since the protein has predicted charge of 13.5, it is a positively charged protein and is purified using a cation exchange column until the protein band of AXE with a molecular weight of 48 kDa is present as shown in figure 4.3B.

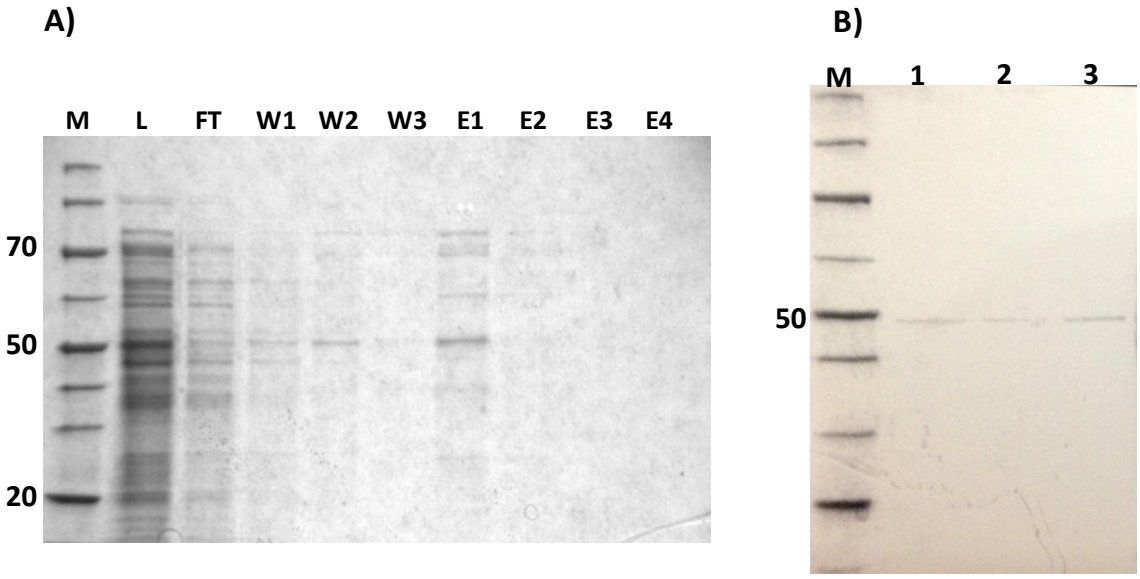

Figure 4.3 Gravitational flow purification of AXE protein using Ni-NTA resin (A), and purified AXE eluted from cation exchange column (B).

Note: M, standard protein marker; L, cell lysate; FT, flow through; W1-W3, wash 1 to 3; E1-4, elution 1 to 4

A Western blot analysis was performed in order to detect the his-tagged AXE. The eluted protein was tested against anti-histidine tagged antibody specific to the his-tagged protein. The result shows the his-tagged protein of approximately 48 kDa, which matches to the molecular weight of AXE protein.

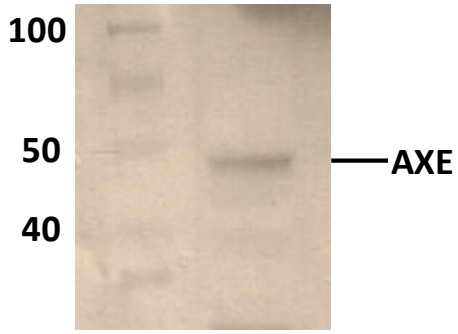

Figure 4.4 Western blot analysis of AXE.

4.3.5 Characterization of Acetyl Xylan Esterase

The activity of acetyl xylan esterase was tested against acetylated birchwood xylan. The purified AXE enzyme showed esterase activity toward actylated birchwood xylan by releasing acetic acid from the xylan backbone. The optimum temperature of the enzyme which was incubated with the substrate for 25 min was at 50 ^{o}C and the AXE maintained its activity of 75% and 72% at 25 and 35 ^{o}C, respectively (Figure 4.5). The activity of AXE was drastically decreased to 37% at 60 ^{o}C and 70 ^{o}C. Increasing temperature of the reaction can increase kinetic energy of the molecules as well as movement of the molecules, so catalyzed reaction rate is increased. However, at high temperature (e.g., 70 ^{o}C) can weaken and destabilize bonds in different areas of the enzyme, and thus can deform active site of the enzyme in binding with substrate in some extent.

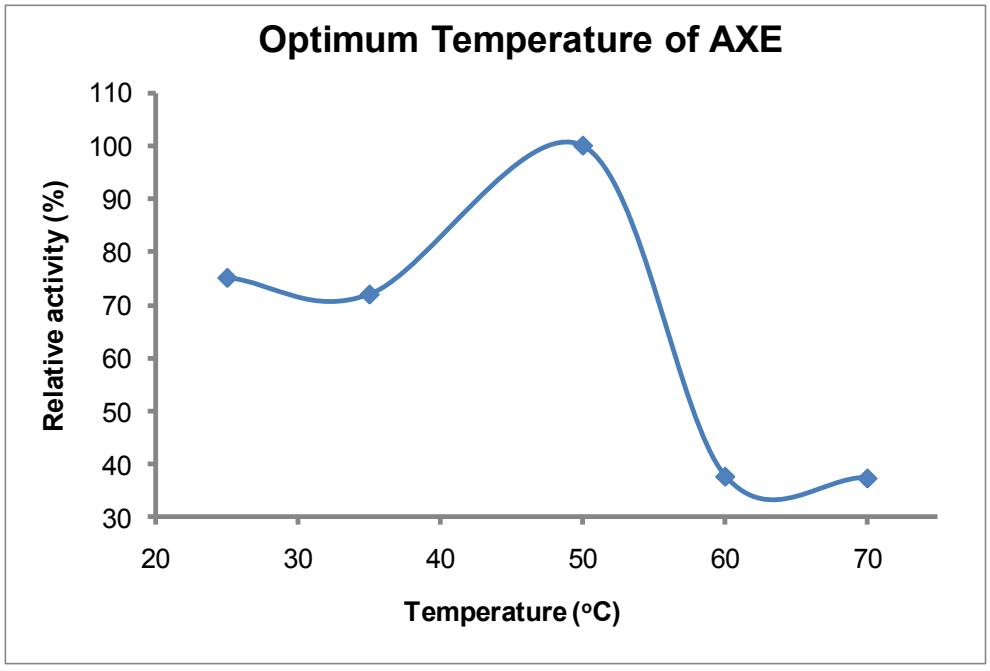

Figure 4.5 Optimum temperature of AXE toward acetylated xylan

Assays were tested in 100 mM potassium phosphate buffer, pH 6.0.

When comparing the optimum temperature of AXE produced from bacteria, *D. colotermitum* strain TAV2, *Fibrobacter succinogenes* S85 (the cellulose-degrading bacterium of the rumen) (McDermid *et al.*, 1990), and *Bacillus pumilus* PS213 (the aerobic soil bacterium) (Degrassi *et al.*, 2000), they revealed relatively close optimum temperature at 50, 45, and 45 ^{o}C respectively. In addition, the purified AXEs of these bacteria have similar molecular mass of 48, 55, and 40 kDa, respectively. Other organisms expressed activity toward acetylated xylan including *Clostridium cellulovorans*, anaerobic fungus *Orpinomyces* sp. strain PC-2, and fungus *Trichoderma reesei*. Poutanen and Sundberg (1988) found that activity of purified acetyl esterase from *T. reesei* in liberating acetic acid was stimulated by adding enzymes such as xylanase and xylosidase together with the acetyl esterase (Poutanen and Sundberg, 1988). This suggested that deacetylation by *T. reesei* might require synergism of enzymes in xylanolytic system and may produce multiple esterases specific to different substrate sizes such as short or long chain of substrates. The recombinant AXE cloned from *D. colotermitum* strain TAV2 alone was able to cleave acetyl group from the xylose backbone without other xylanolytic enzymes. Furthermore, the enzyme can still maintain its activity up to 75% at room temperature (25 ^{o}C); this level is close to the AXE activity at 25 ^{o}C of the cellulose-degrading bacterium of the rumen *F. succinogenes* S85 (McDermid *et al.*, 1990). In addition, the optimum temperature pattern of AXE of strain TAV2 is silmilar to AXE of *F. succinogenes*, which is the enzyme still maintains the activity of at least 60% below its optimum temperature, but the activity is dramatically decreased at temperature greater than 50 ^{o}C.

The optimum pH of AXE cloned from *D. colotermitum* strain TAV2 was assayed from pH 3.0 to 9.0 at 50 ^{o}C which is the temperature at which AXE has 100% relative activity. The optimum pH of the AXE which was incubated for 25 min with acetylated xylan was at pH 7.0. The AXE maintained its activity of 93% and 90% at pH 6.0 and 8.0, respectively. The activity of AXE was decreased to 71% at pH 9.0 and the enzyme had decreased activity when tested at low pH such as 3.0, 4.0, and 5.0 (Figure 4.6).

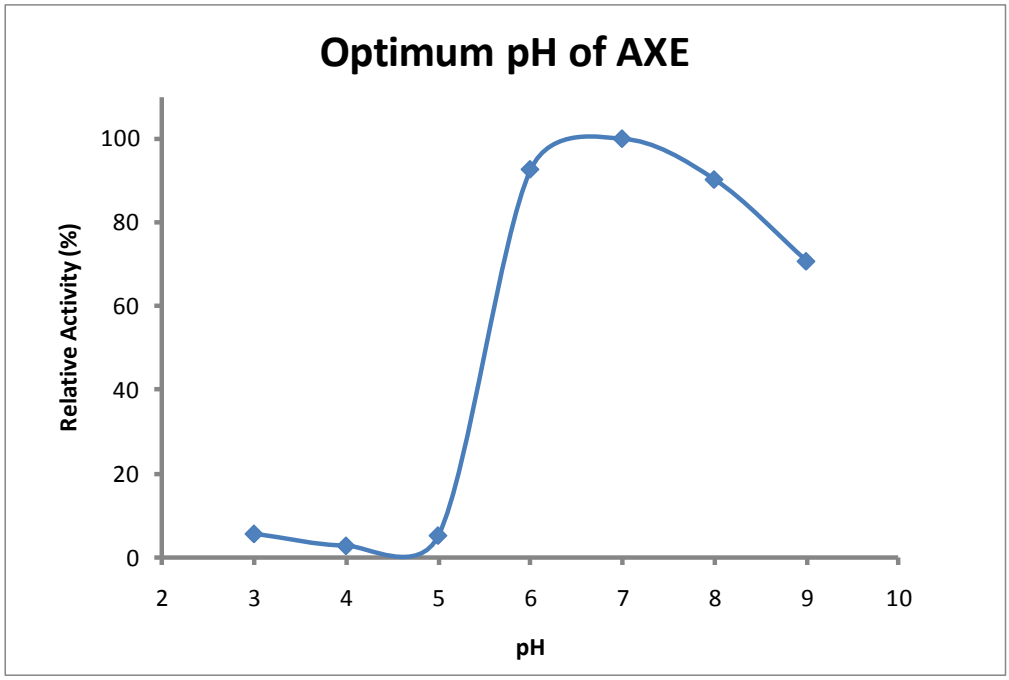

Figure 4.6 Optimum pH of AXE toward acetylated xylan

Assays were performed at 50°C in 100 mM citrate buffer from pH 3.0 to 5.0, 100 mM phosphate buffer from pH 6.0 to 7.0, and 100 mM tris buffer from pH 8.0 to 9.0.

At low pH, both acidic and basic amino acid side chains which have negatively charged and positively charged are exposed to high concentration of H^+, so breakdown of the ionic interactions holding the active site structure, and tertiary and quaternary structures of the enzyme is occurred, thus the enzyme is incapable of binding properly with the substrate and producing the products. At high pH, there is low concentration of H^+, so the negatively charged and positively charged amino acid side chains are likely to liberate a bound H^+ which again results in alteration of chemical properties and structure of the active site. The precise degree of H^+ binding and dissociation can be determined from the optimum pH of the enzyme to catalyze the reaction and generate the products. The effect of pH on AXE enzyme reveals that AXE loses most of its activity at pH 3.0 – 5.0 and still has some activity (71%) at pH 9.0. From the optimum temperature and pH of the AXE, it shows that the optimum temperature pattern of *D. colotermitum* strain TAV2 is relatively similar to the *F. succinogenes*, and the optimum pH at 7.0 is identical. Interestingly,

both bacteria come from the same habitat of rumen and involve in cellulose degradation. For the optimum pH of fungi, *T. reesei* had optimum pH at 5.5 and there was decreased to 10% at pH 7.0 (Poutanen and Sundberg, 1988) and *Orpinomyces* sp. Strain PC-2 had optimum pH at 9.0 (Blum *et al.*, 1999).

When testing the activity of enzyme using p-nitrophenyl acetate as a substrate, the AXE enzyme was unable to catalyze the reaction and release p-nitrophenol as a product. This might be because the AXE is very specific to xylan substrate and can hydrolyze short chain aliphatic esters but cannot hydrolyze acetyl ester of phenols. The ability of the AXE in hydrolyzing acetylated xylan suggests the roles of AXE in liberating acetyl group from xylan backbone, and thus increasing accessibility of other enzymes like xylanase, xylosidase, and cellulase to degrade hemicellulose and cellulose.

4.4 Future Direction

The acetyl xylan esterase will be tested its activity on broad ranges of substrates such as short chain fatty acid esters (galactose pentaacetate and xylose tetraacetate), napthyl fatty acid esters (napthyl acetate and napthyl butyrate), then substrate specificity of acetylaed xylan and other substrates will be calculated. Moreover, synergistic activity of AXE and other enzymes in xylanolytic system will be studied to verify if the efficiency in degrading hemicellulose is increasing or not. In addition, other property of the enzyme such as enzyme stability will be tested.

Research Conclusions

A group of bacteria, Termite Associated *Verrucomicrobium* (TAV) strains TAV1 - TAV5, were isolated from the hindgut of lower termite *Reticulitermes flavipes*. Based on 16S rRNA gene sequences, they belong to the phylum *Verrucomicrobia*, family *Opitutaceae*. As we know that termite hindgut is not completely anaerobic, but there is hypoxic region containing 2-4% of O_2 surrounding at the gut wall. This tightly controlled oxygen gradient is very important for spatial distribution of microorganisms living in the hindgut to complete their metabolic activities such as cellulose degradation, nitrogen fixation, acetogenesis, and methanogenesis. For example, strictly anaerobic microorganisms like protists, could not survive in the presence of O_2 and they require anaerobic condition to complete cellulose degradation. Whereas most study focus on microbial diversity in the termite hindgut, this study revealed ecophysiological roles of TAV in the hindgut using high throughput methods.

The genome sequences of the strain TAV1, TAV2, and TAV5 revealed several important genes: 1) genes involved in cellulose and hemicellulose degradation such as glycoside hydrolase family 5, endo-1,4-beta-xylanase, and 1,4-alpha-glucan branching enzyme, 2) nitrogen fixation genes such as nitrogen iron reductase protein (*nifH*), nitrogenase molybdenum-iron (*nifD*), FeS assembly protein (*nifU*), and nitrogenase MoFe cofactor biosynthesis protein (*nifE*), and 3) the gene encoding the enzyme cbb_3- type cytochrome oxidase, which is involved in O_2 consumption having high affinity to oxygen and helping to maintain hypoxic condition in the termite hindgut. The presence of genes associated with lignocellulose degradation, nitrogen fixation, and oxygen consumption implied ecological roles of TAV strains in the hindgut ecosystem.

Transcriptomic and proteomic analyses of the strain TAV2 revealed statistically significant genes and proteins in response to 2% and 20% O_2 concentrations. Among these, we found positive correlation between transcriptomics and proteomics in acetyl xylan esterase, an enzyme involving in deacetylation of hemicellulose, which was up-regulated under 2% O_2. Moreover, the

presence of up-regulated alpha-urease enzyme under 2% O_2 implied functional role of the strain in recycling of nitrogenous waste excreted from termite host. The *cbb*$_3$- type cytochrome oxidase was not significantly up-regulated under 2% O_2, however, we found high spectral count of this enzyme under 2% O_2. Moreover, a metabolic map was constructed with differentially expressed genes and proteins, providing information on pathways that are up- or down- regulated under 2% O_2. To our knowledge, this is the first experimentally tested metabolic map for a microorganism of the phylum *Verrucomicrobia* and can serve as a reference for researches interested in microbial physiology/biochemistry and comparative studies among members of the *Planctomycetes-Verrucomicrobia-Chlamydia* (PVC) superphylum. In addition, study of genomics, transcriptomics, and proteomics can be used as a guide to study their physiology based on what we found on high throughput analyses. For example, ability of the TAV in fixing nitrogen, or ability to use a variety of carbon sources such as xylan as a substrate will be tested.

Gene cloning, expression, and purification of the acetyl xylan esterase (AXE) from the strain TAV2 allowed us to describe the properties of this enzyme. To our knowledge, this is the first study of AXE purified from *Verrucomicrobium* sp., whereas most of previous studies focused on AXE purified from fungi such as *Orpinomyces* sp., and *Trichoderma reesei.* We found that purified AXE was able to cleave acetyl groups from xylan backbone. The activity of the enzyme was highest at 50^{o}C and optimum pH was at 7.0, which is the pH of its natural habitat. Besides cellulose degradation, study of AXE in removal of acetyl side groups would advance our knowledge and pave the way for hemicellulose degradation.

Appendix A

Family *Opitutaceae*

Family *Opitutaceae*

Jorge L. M. Rodrigues and Jantiya Isanapong

Department of Biology, University of Texas, Arlington, TX, 76019, USA

Abstract

The family *Opitutaceae,* within the phylum *Verrucomicrobia*, is currently comprised of isolated microbial members of three different genera: *Opitutus*, *Alterococcus*, and *Diplosphaera.* This grouping is solely based on the phylogenetic analysis of their 16S rRNA gene sequences. Culture-independent studies indicate that family members can be found in freshwater and marine environments, hot springs, soils, and termite hindguts. All isolates are Gram-negative and characterized for their coccus shape, cell size varying between 0.4 and 0.9 μm, and colorless or white colony formation on agar plates. Successful isolation has been associated with use of moderately recalcitrant heteropolysaccharides such as xylan and pectin. Sequenced genomes of four isolates revealed an average G+C content of 63.1 mol % and genome sizes ranging from 5.22 to 7.41 Mb. The ecological contributions of this family to different ecosystems remain to be understood.

Taxonomy, historical and current

O.pi.tu.ta.ce.a.e. N. L. masc. n. *Opitutus* type genus of the family; -*aceae* ending to denote a family; N. L. fem. pl. n. *Opitutaceae* the *Opitutus* family.

The family *Opitutaceae*, formally proposed by (Choo *et al.*, 2007), is based on a comparative analysis of the 16S rRNA gene of few isolated microorganisms and a large number of sequences that have been obtained from different environmental samples. Phylogenetically, this group of sequences forms a distinct clade within the class *Opitutae.* Molecular retrieval of sequences of the 16S rRNA gene indicated that members of this family are found in termite and earthworm guts (Köhler *et al.*, 2012; Rosengaus *et al.*, 2011), soils (Chin *et al.*, 2001), and lakes (Burkert *et al.*, 2003; Shieh and Jean, 1998). Currently, the *Opitutaceae* family is composed of

three described genera with phylogenetic support based on their 16S rRNA sequences. However, only two genera, *Opitutus* (Chin *et al.*, 2001) and *Alterococcus* (Shieh and Jean, 1998) have been recognized with standing in the literature, while the third genus *Diplosphaera* (Wertz *et al.*, 2012) has yet to be validated. All genera have multiple isolates that are characterized as Gram-negative, coccus or diplococcus, and with a GC content varying between 60.5 and 65.8 mol %. Colonies on agar vary from clear to white, opaque, and of circular shape. Formation of spores has not been observed for members of this family.

Molecular Analyses

Phylogenetically, the family *Opitutaceae* is closely positioned with the *Puniceicoccaceae*, a family within the class *Opitutae,* containing isolates obtained from samples of marine origin. While the *Puniceicoccaceae* appears to be of a polyphyletic origin, the family *Opitutaceae* is grouped as a single distinct clade (Figure A - 1). However, one should remain cautious about using this phylogenetic tree to suggest a clear distinction between these two families. As new members of the phylum *Verrucomicrobia* continue to be successfully isolated and characterized, the phylogenetic coverage for members of the above families will also increase.

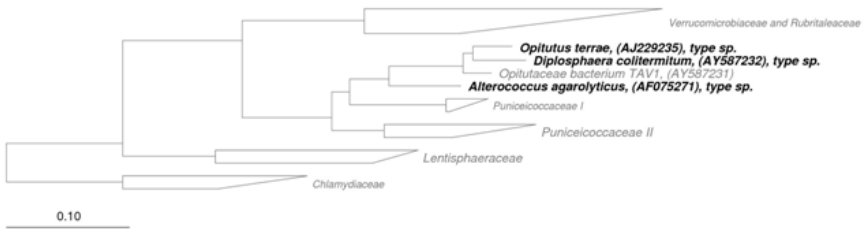

Figure A - 1 Maximum likelihood genealogy reconstruction based on the 16S rRNA gene from isolated members of the family *Opitutaceae*.

A phylogenetic analysis based on 16S rRNA gene suggests that the family is composed of at least three described genera, *Opitutus*, *Alterococcus*, and *Diplosphaera.* Another termite hindgut microorganism of the *Opitutaceae* family, strain TAV1 (NCBI accession number AY587231) has been isolated (Stevenson *et al.*, 2004) and its genome has been sequenced (Isanapong *et al.*, 2012), but a formal description of the genus is yet to be provided.

Genome Analyses

Isolated members of the *Opitutaceae* family have a higher G+C content (average of 63.1 mol %) in comparison to those belonging to the sister family *Puniceicoccaceae* (52.9 mol %). Among the four sequenced *Opitutaceae* members (Table A - 1), the genome sequence belonging to *Opitutus terrae* strain PB90-1 is the only one fully assembled (van Passel *et al.*, 2011). It maintains a circular chromosome of 5,957,605 bp with 3 rRNA operons and 4,632 protein-coding sequences. A total of 67% of the proteins had assigned functions and 30.8% of them were predicted to contain signal peptides. The other three genomes are derived from termite hindgut isolates, strains TAV1, TAV2, and TAV5. Their genomes were sequenced to high-quality drafts, but long stretches of GC content prevented closure. The genome of *Diplosphaera colotermitum* strain TAV2 is of approximately 5.22 Mb, calculated by the sum of all contigs. It contains 4,826 protein-coding sequences and 1 rRNA operon, the latter confirmed by Southern blot hybridizations of the 16S rRNA gene from enzymatic digestions of total genomic DNA (Wertz *et al.*, 2012). *Opitutaceae* bacterium strains TAV1 and TAV5 have similar genome sizes with 7.07 and 7.41 Mb, respectively. Only one ribosomal operon was found in each genome and the numbers of protein-coding sequences were 5,987 (TAV1) and 6,006 (TAV5).

Table A - 1 Genomic properties of members of the family *Opitutaceae*

Species	Size (Mb)	GC %	Number of Genes	Number of Proteins	NCBI accession
Opitutus terrae	5.96	65.3	4,704	4,632	NC_010571
Diplosphaera colotermitum	5.22	60.8	4,896	4,826	NZ_ABEA00000000
Alterococcus agarolyticus	ND[a]	65.8	ND	ND	-
Opitutaceae bacterium strain TAV1	7.07	63.4	6,051	5,987	NZ_AHKS02000000
Opitutaceae bacterium strain TAV5	7.41	63.9	6,056	6,006	NZ_AGJF00000000

[a] Not determined.

The frequency of specific tetranucleotides in genomic sequences is known to exhibit species-specific signals (Teeling *et al.*, 2004) and can be used to draw distinction between two species. When a pairwise comparative analysis for tetranucleotide usage patterns was performed with the above genomes (Table A - 1), a correlation coefficient for tetranucleotide presence between TAV1 and TAV5 was of 0.9951 (Table A - 2**,** upper triangle), suggesting that these two isolates belong to the same species. They are immediately followed by strain TAV2 with correlation coefficients of 0.8949 and 0.8879 for TAV1 and TAV5, respectively. Finally, the comparison between the genome of *O. terrae* and the other three genomes yielded correlation coefficients below 0.80, suggesting a distant relationship with the other strains. These results are corroborated by average nucleotide identity (ANI) values calculated between any two genomes (Konstantinidis and Tiedje, 2005). Strains TAV1 and TAV5 share a mean ANI of 95.75%, a value above the 95% threshold used for the definition of a microbial species. When the TAV2 genome was compared to the other termite hindgut isolate genomes, the ANI values were 79.62 and 79.89% for TAV1 and TAV5, respectively (Table A - 2**,** lower triangle). This finding indicates that different *Opitutaceae* populations co-exist in the termite hindgut, but the ecological significance of those differences remains unknown.

Table A - 2 Correlation coefficients of tetranucleotide usage patterns (upper triangle) and % average nucleotide identity (lower triangle) calculated between *Opitutaceae* genomes.

Strain	*Opitutus terrae*	TAV1	TAV2	TAV5
Opitutus terrae		0.7834	0.6587	0.7887
TAV1	71.96		0.8949	0.9951
TAV2	71.08	79.62		0.8878
TAV5	71.79	95.75	79.89	

Phenotypic analyses

OPITUTUS CHIN, LIESACK, JANSSEN 2001, 1967.

O.pi.tu'tus. L. fem. n. *Ops, Opis* a Roman Earth and harvest goddess; L. part. Adj. *tutus* protected; N.L. masc. n. *Opitutus* the one protected by Ops).

The type strain of *Opitutus terrae* PB90-1^T is a Gram-negative, coccoid bacterium with a cell size of 0.4 – 0.6 µm in diameter. A cell of PB90-1^T is motile by use of one flagellum. Diplococci with only one flagellum are frequently observed (Chin *et al.*, 2001). A few connected cells can be observed using phase contrast microscopy, causing them to appear similar to short rod bacteria. Spores are not formed in fresh cultures containing 50 mM glucose or in cultures containing a low concentration of glucose (4 mM) and high concentrations of $CaCl_2 \cdot 2H_2O$ (0.3 g/l), $MnCl_2 \cdot 4H_2O$ (50 mg/l), thiamine chloride hydrochloride (30 mg/l), and soil extract (10% v/v). The type strain shows negative results for catalase and oxidase activities. The aminopeptidase reaction and KOH test yield positive results, confirming the type strain as Gram-negative. The strain PB90-1^T is an obligate anaerobe and no growth is observed for cells in the upper 20 mm of agar when 20 kPa of O_2 is used. Colonies appearing in agar deeps are colorless and granular. Growth occurs at temperatures ranging from 10 to 37 ^oC and no growth has been observed at low (4 ^oC) or high (40 ^oC) temperatures. The optimum pH for supporting growth is 7.5 – 8, but growth occurs in a range of pH units (5.5 – 9). Growth will occur in liquid media containing 30 g NaCl/l, but increasing salinity to 40 g NaCl/l causes growth inhibition. The strain is able to utilize the following substrate: glucose, fructose, galactose, mannose, galacturonic acid, mannitol,

arabinose, cellobiose, maltose, sucrose, lactose, melibiose, xylan, pectin, and starch (Table A - 3). Propionate and acetate are major fermentation products for this strain, while a minor production of succinate, lactate, ethanol, and H_2 can also occur (Chin and Janssen, 2002). The type strain grows much faster in glucose (μ = 2.8 day^{-1}) in comparison to growth when pectin is used (μ = 0.17 day^{-1}). In both case, propionate is formed in high concentrations. The strain PB90-1^T uses nitrate as a major electron acceptor, but do not use sulfur, sulfate, sulfite, thiosulfate, or fumarate. The following substrates were not utilized by the strain under tested conditions: xylose, ribose, sorbose, methyl-α-glucopyranoside, cellulose, chitin, arabinogalactan, pyruvate, lactate, fumarate, malate, tartrate, citrate, crotonate, glycerol (with or without acetate), aspartate, alanine, serine, leucine, isoleucine, glutamate, proline, and lysine.

ALTEROCOCCUS SHIEH AND JEAN, 1998, 664.

Al.te.ro.coc'cus. L. *alter* another, Gr. N. *coccus* a grain or berry; M.L. masc. n. *Alterocococcus* another coccus (Shieh and Jean, 1998).

The species *Alterococcus agarolyticus* strain $ADT3^T$ is a Gram-negative, spherical bacterium with a cell size of 0.8 – 0.9 μm in diameter. Cells appear as single or in pairs, being motile by means of a single flagellum. There are five isolates designated as ADT1 to ADT5. All strains produce white circular opaque colonies on peptone-yeast (PY) agar. These microorganisms have the ability to degrade and liquify agar. Shallow holes around colonies appear after two days of incubation, indicating their capability of dissolving agar. Within one week, deep holes are formed as the colony mass increases. Agarase activity under aerobic condition is of approximately 74-100 μg galactose per ml of culture per day. The most abundant fatty acid (46.0–51.5 % mol) measured for these strains is the saturated anteiso-15-carbon acid (anteiso-C15:0). All strains show catalase and oxidase activities. These agar-degrading strains are heterotrophic, halophillic, facultatively anaerobic, and moderately thermophilic bacteria. Substrates supporting growth are glucose, galactose, lactose, cellobiose, trehalose, sucrose, and xylose (Table A - 3). Under aerobic condition, growth occurs at temperatures ranging from 40 to

56 ^{o}C, with an optimum at 48 ^{o}C. No growth is observed at temperatures lower than 36 ^{o}C or higher than 60 ^{o}C. Growth occurs at narrow pH 7.0–8.5, but not at pHs below 6.5 or above 9.0. The type strain grows at NaCl concentrations of 0.17–0.60 M (ca. 1.0–3.5%) with optimum growth at 0.34–0.43 M (ca. 2.0–2.5%), whereas no growth is observed at very low concentrations (0–0.09 M) or high concentrations (0.68–0.85 M). When KCl (0.17–0.51 M) is used as a substitute for NaCl, growth is not observed.

In peptone–yeast–glucose (PYG) broth, strain ADT3^{T} is able to grow under both aerobic and anaerobic conditions. Once cultures reach stationary phase, large and rapid decreases in OD_{600} are observed as well as decreases in the medium pH of approximately 2.1 and 1.5 units for anaerobic and aerobic cultures, respectively. During growth under both conditions, major organic acids generated are butyrate, propionate, and formate. Lactate and acetate can also be detected. Cells can be grown aerobically or anaerobically in PY semisolid medium with subsequent formation of butyrate and propionate.

DIPLOSPHAERA WERTZ ET AL. 2012, 1553.

Di.plo.spha.e'ra. Gr. Adj. *diploos*, double; L. fem. n. *sphaera*, globe, sphere; N.L. fem. n. *diplosphaera,* double sphere or diplococcus (Wertz *et al.*, 2012).

Diplosphaera colotermitum strain TAV2 is Gram-negative, coccoid bacteria with a cell size of 0.5 µm in diameter. Cells are non-motile and usually stay in pairs. The strain forms cream-colored colonies with 2-4 mm in diameter on R2A agar plates (Stevenson et al. 2004). The colonies have mucoid appearance, an entire rim, and a low convex elevation. Cells show no catalase and peroxidase activities. The strain has the ability to fix nitrogen confirmed by presence of the *nif* genes and its ability to grow on nitrogen-free medium. The microorganism is microaerophile and no growth is observed under anaerobic condition when supplied with a variety of terminal electron acceptors. The generation time during exponential growth phase is of approximately 7.7 – 8.3 hours and 9.2 – 11.7 hours under 2 to 8% O_2 (v/v) and 12 to 20% O_2, respectively. The optimum growth occurs between 2 and 8% O_2. There is no detectable

formation of acetate or other organic acids produced during the growth on glucose under 2% and 21% O_2. Temperatures ranging from 15 to 35 ^oC support growth, with an optimum at 30 ^oC. No growth is observed at 4 ^oC or 37 ^oC. Growth is observed between pH units of 5.5 – 7.5 with optimum pH at 7.0, but no growth at pH 5.0 or 8.0. Substrates supporting growth are the monosaccharides glucose and galactose, the disaccharides maltose, the cellulose derivative cellobiose, and starch (Table A - 3). Under the conditions tested, cells are not able to utilize fructose, mannose, ribose, xylose, arabinose, trehalose, sucrose, raffinose, mannitol, sorbitol, lactate, sodium pyruvate, sodium fumarate, sodium acetate, allantoin, glucuronate, galacturonate, gluconic acid, xanthine, tannic acid, resourcinol, vanillic acid, sodium benzoate, and trimethylbenzoate.

Table A - 3 Comparison of selected phenotypic characteristics of the family *Opitutaceae*.

Characteristics	***O. terrae***	***A. agarolyticus***	***D. colotermitum***
Habitat	Soil	Hot spring	Termite hindgut
Gram-stain	Negative	Negative	Negative
Cell shape	Coccus	Coccus	Diplococcus
Cell size (μm)	0.4 - 0.6	0.8 - 0.9	0.5 - 0.6
Colony color	Colorless	White opaque	Cream
Metabolism	Anaerobe	Facultative anaerobe	Microaerophile
Oxidase	-	+	-
Catalase	-	+	-
Motility	Flagellum	Flagellum	Non-motile
Temp range (^{o}C)	10-37	44-52	15-35
pH range	5.5-9	7.0–8.5	5.5–7.5
Growth in NaCl (%)	3	1.0–3.5	1.5
Carbon sources			
Glucose	+	+	+
Galactose	+	+	+
Cellobiose	+	+	+
Fructose	+	ND	-
Xylose	-	+	-
Arabinose	+	ND	-
Maltose	+	ND	+
Lactose	+	+	-
Sucrose	+	+	-
Starch	+	ND	+

Note: +, present or utilized; - , absent or not utilized; ND, not determined.

Isolation, enrichment and maintenance procedures

Opitutus terrae strain PB90-1^T was isolated from anoxic rice paddy soil microcosms as previously described by Chin et al. (1999). Anaerobic dry soil cores obtained from 90 days old rice cultures were thoroughly resuspended in degassed sterile-deionized water (1:1, wt/wt). Soil slurries were ten-fold serially diluted using sulfide-reduced, bicarbonate-buffered mineral medium (SdM) supplied with vitamins as described by (Chin *et al.*, 1998). The mineral salt of SdM contains (per liter): 0.1 g of $MgCl_2 \cdot 6H_2O$, 0.1 g of NH_4Cl, 0.07 g of KH_2PO_4, and 0.1 g of $CaCl_2 \cdot 2H_2O$, and 1 ml of riboflavin solution (50 mg/l). The pH of the medium was adjusted to 7.2-7.3 (with NaOH or HCl). Different growth substrates (L-isomers of organic and amino acids, D-isomers of sugars, amorphous cellulose, and pectin) with concentration of 0.2-2 M or 5% (w/v) were added to the medium just prior to inoculation. The tubes were secured with black rubber stoppers, and the gas headspace was supplied with N_2 and CO_2 (80:20 [v/v]). Sodium dithionite (100 µM) was used as an additional reducing agent and the tubes were incubated in the dark at 25^{o}C. Growth results were considered positive when the cells were present and fermentation products such as propionate, acetate, and lactate were produced at a concentration greater than 1 mM. Results considered as negatives had fermentation product concentrations less than 0.05 mM. The strain PB90-1^T was successfully isolated from pectin-containing medium as the growth substrate.

Alterococcus agarolyticus was isolated by a method described in (Shieh and Jean, 1998). Water samples were collected from coastal hot springs with temperatures of 39–51°C and ten-fold serially diluted using sterile NaCl–MOPSO buffer (25 g NaCl and 0.45 g 3-(*N*-morpholino)-2-hydroxypropanesulfonic acid (MOPSO) per liter of deionized water, pH 7.0). Each diluted sample was spread on the plate-containing peptone-yeast agar (PY; pH 7.0) of the following composition (per liter of deionized water): 4 g peptone, 2 g yeast extract, 20 g NaCl; 0.5 g $MgSO_4 \cdot 7H_2O$, 0.01 g $CaCl_2$, 4.5 g MOPSO, and 15 g agar. All plates were aerobically incubated in the dark at 50°C for 3–7 days. Agar-dissolving colonies present on the plates were consecutively transferred to fresh PY agar plates until pure cultures were obtained. The cells can be maintained every 3

months at 45°C by inoculating 0.25 ml of early stationary phase cells grown in PY broth into 5 ml of 60% seawater. For long-term storage, cells can be maintained in NaCl/glycerol stock (2 g NaCl and 15 g glycerol in 100 ml of deionized water). Cell suspension should be pre-cooled in -4^{o}C for 1 h and then kept in -80^{o}C (Hedlund, 2010a).

Diplosphaera colotermitum strain TAV2 and its closely related strains (TAV1, TAV3, and TAV4) were isolated from the hindgut of the worker termite *Reticulitermes flavipes* (Stevenson *et al.*, 2004). The hindgut homogenates were prepared under hypoxia (2% O_2, 5% CO_2, 93% N_2) and were serially diluted in basal salts solution containing the following composition (per liter): 0.2 g KH_2PO_4, 0.25 g NH_4Cl, 0.5 g KCl, 0.15 g $CaCl_2 \cdot 2H_2O$, 1.0 g NaCl, 0.62 g $MgCl_2 \cdot 6H_2O$, 2.84 g Na_2SO_4, and 10 mM morpholinepropanesulfonic acid (MOPS), pH 7.0. Each dilution was spread onto basal medium agar containing sodium acetate (2.46 g/l) and/or a mixture of peptone and yeast extract (0.1 g/l each). The basal medium agar was modified from the basal salts solution described above, but 10 mM HEPES (pH 6.8) was added instead of MOPS. The medium agar was supplied with 1 ml trace element solution, 1 ml vitamin B12 (50 mg/l), 1 ml of mixed vitamin solution (Widdel and Bak, 1991), and 15 g agar. The plates were incubated under hypoxia (2% O_2, 5% CO_2, and 93% N_2) and low light at room temperature for at least 30 days. All colonies growing on agar surface were used for a PCR-based surveillance method known as plate wash PCR (Stevenson *et al.*, 2004). Selected plates were flooded with buffer and colonies were resuspendend for DNA extraction, followed by PCR with specific primers targeting the 16S rRNA gene of the phyla *Verrucomicrobia.* Remaining colonies-containing plates were continually monitored by PCR with specific primers and colonies were re-streaked onto fresh basal medium agar until pure cultures were obtained. TAV1 and TAV2 were isolated from basal medium supplemented with yeast extract, peptone, and acetate, while TAV3 and TAV4 were isolated from the same medium but lacking acetate.

Ecology

Culture-independent studies based on the retrieval of the 16S rRNA gene sequences have indicated that members of the *Opitutaceae* family are present in soils, hot springs, waste water treatment plants, lakes, peat bogs, and inside the digestive track of invertebrates (Chin *et al.*, 1999; Hongoh *et al.*, 2003; Köhler *et al.*, 2012; Parveen *et al.*, 2013; Rosengaus *et al.*, 2011; Wertz *et al.*, 2012). In comparison to other *Verrucomicrobia* families, the number of sequences classified as *Opitutaceae* can be considered small. *Opitutaceae* represents only 2.4% of all sequences retrieved for the phylum (Cole *et al.*, 2009). Successful isolation of members belonging to this family appears to be associated with a provision of carbon substrates that are moderately recalcitrant such as plant heteropolysaccharides (xylan and pectin) adequately supplied in low concentrations (Chin *et al.*, 2001; Shieh and Jean, 1998; Stevenson *et al.*, 2004). (Chin *et al.*, 1999) suggested that, because polymers need to be hydrolyzed first, these carbon compounds eliminate sudden exposure to high concentrations of a growth substrate, preventing substrate-accelerated death.

The ecological functions of the *Opitutaceae* family members are largely unknown. However, some specific information on their ecology is emerging. For instance, the type strain *O. terrae* was isolated from a rice paddy soil (Chin *et al.*, 1999), a system known to be an important source of methane. Propionate-producing bacteria such as *O. terrae* are dominant in rice fields as propionate is an important intermediate for methanogenesis. Another ecological function for microorganisms isolated from the hindgut of the lower termite *Reticulitermes flavipes* is their potential for biological nitrogen fixation. Genomic analysis of the TAV strains indicate that these *Verrucomicrobia* are capable of biological nitrogen fixation (Wertz *et al.*, 2012), an important attribute for a very N-limited environment owing to the termite diet (Brune and Ohkuma, 2011).

Application

There is no information available on the clinical importance of the *Opitutaceae* family.

The species *A. agarolyticus* is resistant to gentamicin, kanamycin, nalidixic acid, polymyxin B, and streptomycin as determined by inhibition zone with the use of method of standardized disks. The antibiotics ampicillin, chloramphenicol, erythromycin, penicillin G, and tetracycline are suitable antimicrobial agents for this species (Shieh and Jean, 1998).

Pathogenicity, clinical relevance

The species *Alterococcus agarolyticus* is capable of degrading agar through the production of an extracellular agarase. This is the first thermophile isolate known to degrade agar (Shieh and Jean, 1998).

Although no other industrial applications have been reported for members of the *Opitutaceae* family, the utilization of *Verrucomicrobia* for biotechnological purposes of biofuel production and bioremediation has been suggested (Martinez-Garcia *et al.*, 2012). There is a significant enrichment of genes encoding for polysaccharide degradation enzymes in verrucomicrobial genomes, making them important source for gene discovery and future applications in biotechnology.

Appendix B

List of Genes and Proteins that were Significantly Up- or Down-Regulated during Growth of *Diplosphaera colotermitum* Strain TAV2 Cells under 2% O_2 as Compared to Cells Grown under 20% O_2

Table B - 1 List of genes and proteins that were significantly up- or down-regulated under 2% and 20% O_2

Locus Tag	Gene Object ID	ID Description	EC Number	COG	Group	Functional Categories
ObacDRAFT_1002	641180481	ATP synthase F1, alpha subunit	3.6.3.14	0056	C	Energy production and conversion
ObacDRAFT_2008	641183125	NADH-ubiquinone/plastoquinone oxidoreductase	1.6.5.3	0839	C	Energy production and conversion
ObacDRAFT_4074	641180385	Quinone oxidoreductase putative PIG3	1.6.5.5	0604	CR	Energy production and conversion/General function prediction only
ObacDRAFT_3725	641183128	putative exopolysaccharide biosynthesis protein	N/A	0489	D	Cell cycle control, cell division, chromosome partitioning
ObacDRAFT_1474	641182918	ABC transporter related	N/A	0410	E	Amino acid transport and metabolism
ObacDRAFT_3647	641182410	D-isomer specific 2-hydroxyacid dehydrogenase	1.1.1.95	0111	E	Amino acid transport and metabolism
ObacDRAFT_1038	641183406	phosphofructokinase	2.7.1.11	0205	G	Carbohydrate transport and metabolism
ObacDRAFT_1367	641181180	Fructose-bisphosphate aldolase	4.1.2.13	0191	G	Carbohydrate transport and metabolism
ObacDRAFT_3927	641181225	Alpha/beta hydrolase fold-3 domain protein	3.7.1.-	0657	I	Lipid transport and metabolism
ObacDRAFT_2695	641182545	ribosomal protein L15	N/A	0200	J	Translation, ribosomal structure and biogenesis
ObacDRAFT_1498	641183640	lysyl-tRNA synthetase	6.1.1.6	1190	J	Translation, ribosomal structure and biogenesis
ObacDRAFT_3158	641182322	regulatory protein LacI	N/A	1609	K	Transcription
ObacDRAFT_1903	641180891	regulatory protein GntR HTH	N/A	1609	K	Transcription
ObacDRAFT_3386	641180148	helix-turn-helix- domain containing protein AraC	N/A	2207	K	Transcription
ObacDRAFT_3083	641183638	DNA-directed DNA polymerase	2.7.7.7	0389	L	Replication, recombination and repair
ObacDRAFT_1380	641183673	exodeoxyribonuclease VII, small subunit	3.1.11.6	1722	L	Replication, recombination and repair
ObacDRAFT_3711	641182838	integrase family protein	N/A	4974	L	Replication, recombination and repair
ObacDRAFT_3566	641180943	major facilitator superfamily MFS_1	N/A	2223	P	Inorganic ion transport and metabolism
ObacDRAFT_0993	641180470	Acetyl xylan esterase	N/A	3458	Q	Secondary metabolites biosynthesis, transport and catabolism
ObacDRAFT_0729	641183105	integral membrane sensor signal transduction	2.7.13.3	0642	T	Signal transduction mechanisms

Table B - 1 List of genes and proteins that were significantly up- or down-regulated under 2% and 20% O_2 - *Continued*

Locus Tag	Gene Object ID	ID Description	EC Number	COG	Group	Functional Categories
ObacDRAFT_1492	641181088	type II and III secretion system protein	N/A	1450	U	Intracellular trafficking, secretion, and vesicular transport
ObacDRAFT_0510	641180714	type II secretion system protein	N/A	1459	U	Intracellular trafficking, secretion, and vesicular transport
ObacDRAFT_2682	641181536	export protein FliQ family 3	N/A	1987	U	Intracellular trafficking, secretion, and vesicular transport
ObacDRAFT_2604	641182131	WD40 domain protein beta Propeller	N/A	0823	U	Intracellular trafficking, secretion, and vesicular transport
ObacDRAFT_2168	641183701	TPR repeat-containing protein	N/A	N/A	N/A	
ObacDRAFT_1988	641181411	Inner membrane CreD family protein	N/A	4452	V	Defense mechanisms
ObacDRAFT_2832	641180795	ISRSO5-transposase protein	N/A	N/A		
ObacDRAFT_1479	641182388	regulatory protein LacI				
ObacDRAFT_2532	641181909	regulatory protein LacI				
ObacDRAFT_3485	641183617	helix-turn-helix- domain containing protein AraC				
ObacDRAFT_2379	641179917	hypothetical protein				
ObacDRAFT_3770	641180523	hypothetical protein				
ObacDRAFT_4115	641182554	hypothetical protein				
ObacDRAFT_2467	641180041	hypothetical protein				
ObacDRAFT_0304	641180335	protein of unknown function DUF147				
ObacDRAFT_0671	641182958	conserved hypothetical protein				
ObacDRAFT_0198	641179730	hypothetical protein				
ObacDRAFT_2291	641180546	hypothetical protein				
ObacDRAFT_2505	641181167	hypothetical protein				
ObacDRAFT_0172	641182653	hypothetical protein				
ObacDRAFT_0891	641181281	hypothetical protein				
ObacDRAFT_1818	641182755	hypothetical protein				
ObacDRAFT_2387	641179929	hypothetical protein				
ObacDRAFT_2703	641182547	hypothetical protein				
ObacDRAFT_2603	641182130	protein of unknown function DUF59				
ObacDRAFT_3299	641183335	hypothetical protein				
ObacDRAFT_0149	641180431	hypothetical protein				

Table B - 1 List of genes and proteins that were significantly up- or down-regulated under 2% and 20% O_2 - *Continued*

Locus Tag	Gene Object ID	ID Description	EC Number	COG	Group	Functional Categories
ObacDRAFT_2383	641179922	hypothetical protein				
ObacDRAFT_1681	641182784	Uncharacterized protein-like protein				
DOWN-regulated						
ObacDRAFT_3488	641183430	Beta-glucosidase	3.2.1.21	2723	G	Carbohydrate transport and metabolism
ObacDRAFT_2780	641182885	acetylglutamate kinase	2.7.2.8	0548	E	Amino acid transport and metabolism
ObacDRAFT_1952	641183714	Pyrroline-5-carboxylate reductase	1.5.1.2	0345	E	Amino acid transport and metabolism
ObacDRAFT_1865	641180361	Ribulose-phosphate 3-epimerase	5.1.3.1	0036	G	Carbohydrate transport and metabolism
ObacDRAFT_3182	641183345	sugar-phosphate isomerase, RpiB/LacA/LacB	5.3.1.6	0698	G	Carbohydrate transport and metabolism
ObacDRAFT_3215	641183152	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase	2.7.6.3	0801	H	Coenzyme transport and metabolism
ObacDRAFT_3157	641182327	biotin--acetyl-CoA-carboxylase ligase	6.3.4.15	0340	H	Coenzyme transport and metabolism
ObacDRAFT_2424	641182974	GCN5-related N-acetyltransferase	N/A	1670	J	Translation, ribosomal structure and biogenesis
ObacDRAFT_1590	641182699	translation elongation factor Ts	N/A	0264	J	Translation, ribosomal structure and biogenesis
ObacDRAFT_0988	641181375	helix-turn-helix- domain containing protein AraC	N/A	2207	K	Transcription
ObacDRAFT_2322	641183407	transcriptional regulator, LysR family	N/A	0583	K	Transcription
ObacDRAFT_2549	641182511	DNA polymerase III, beta subunit	2.7.7.7	0847	K	
ObacDRAFT_1223	641181018	Methyltransferase type 12	N/A	N/A		
ObacDRAFT_3212	641183546	hypothetical protein				
ObacDRAFT_2769	641181319	hypothetical protein				
ObacDRAFT_2328	641182819	hypothetical protein				
ObacDRAFT_2044	641179849	hypothetical protein				
ObacDRAFT_0797	641181333	hypothetical protein				
ObacDRAFT_0711	641181999	conserved hypothetical protein				
ObacDRAFT_3441	641181152	hypothetical protein				
ObacDRAFT_2130	641180866	hypothetical protein				
ObacDRAFT_0819	641179783	conserved hypothetical protein				

Table B - 1 List of genes and proteins that were significantly up- or down-regulated under 2% and 20% O_2 - *Continued*

Locus Tag	Gene Object ID	ID Description	EC Number	COG	Group	Functional Categories
ObacDRAFT_2414	641181788	hypothetical protein				
ObacDRAFT_0265	641180554	hypothetical protein				
ObacDRAFT_4102	641180409	hypothetical protein				
ObacDRAFT_0241	641179761	hypothetical protein				

PROTEOMICS

UP-regulated

Locus Tag	Gene Object ID	ID Description	EC Number	COG	Group	Functional Categories
ObacDRAFT_3061	641182673	aldo/keto reductase	N/A	0667	C	Energy production and conversion
ObacDRAFT_1943	641182117	D-lactate dehydrogenase (cytochrome)	1.1.3.15	0277	C	Energy production and conversion
ObacDRAFT_2903	641182424	FGGY-family pentulose kinase	2.7.1.16	1069	C	Energy production and conversion
ObacDRAFT_1327	641181096	Phosphoenolpyruvate carboxykinase (GTP)	4.1.1.32	1274	C	Energy production and conversion
ObacDRAFT_2731	641180635	succinate dehydrogenase and fumarate reductase iron-sulfur protein	1.3.99.1	0479	C	Energy production and conversion
ObacDRAFT_2730	641180634	succinate dehydrogenase or fumarate reductase, flavoprotein subunit	1.3.5.4	1053	C	Energy production and conversion
ObacDRAFT_2013	641181659	ferredoxin	1.6.5.3	1143	C	Energy production and conversion
ObacDRAFT_2017	641181663	NADH dehydrogenase (ubiquinone) 30 kDa subunit	1.6.5.3	0852	C	Energy production and conversion
ObacDRAFT_2887	641182650	Alanine--glyoxylate transaminase	2.6.1.-	0075	E	Amino acid transport and metabolism
ObacDRAFT_0053	641181674	Glutamate dehydrogenase (NADP(+))	1.4.1.4	0334	E	Amino acid transport and metabolism
ObacDRAFT_1864	641180360	lipolytic protein G-D-S-L family		2755	E	Amino acid transport and metabolism
ObacDRAFT_0053	641181674	glutamate synthase, NADH/NADPH, small subunit	1.4.1.13	0493	E	Amino acid transport and metabolism
ObacDRAFT_0782	641181325	ROK family protein	2.7.1.63	1940	G	Carbohydrate transport and metabolism
ObacDRAFT_1414	641181838	glyceraldehyde-3-phosphate dehydrogenase, type I	1.2.1.12	0057	G	Carbohydrate transport and metabolism
ObacDRAFT_1969	641182248	glycogen/starch synthase, ADP-glucose type	2.4.1.21	0297	G	Carbohydrate transport and metabolism

Table B - 1 List of genes and proteins that were significantly up- or down-regulated under 2% and 20% O_2 - *Continued*

Locus Tag	Gene Object ID	ID Description	EC Number	COG	Group	Functional Categories
ObacDRAFT_2426	641181843	glycoside hydrolase family 2 TIM barrel	3.2.1.23	3250	G	Carbohydrate transport and metabolism
ObacDRAFT_3988	641181601	phosphofructokinase	2.7.1.11	0205	G	Carbohydrate transport and metabolism
ObacDRAFT_0714	641182004	pyruvate, phosphate dikinase	2.7.9.1	0574	G	Carbohydrate transport and metabolism
ObacDRAFT_1412	641181836	Triose-phosphate isomerase	5.3.1.1	149	G	Carbohydrate transport and metabolism
ObacDRAFT_3012	641180254	Xylose isomerase domain protein TIM barrel	5.3.1.5	1082	G	Carbohydrate transport and metabolism
ObacDRAFT_1973	641182252	Xylose isomerase domain protein TIM barrel				
ObacDRAFT_0419	641182902	Xylose isomerase domain protein TIM barrel				
ObacDRAFT_2850	641182487	TonB-dependent receptor plug	N/A	4206	H	Coenzyme transport and metabolism
ObacDRAFT_3600	641182497	biotin/lipoyl attachment domain-containing protein	N/A	0511	I	Lipid transport and metabolism
ObacDRAFT_2918	641181729	glutaminyl-tRNA synthetase	6.1.1.18	0008	J	Translation, ribosomal structure and biogenesis
ObacDRAFT_2699	641182550	ribosomal protein S11	N/A	0100	J	Translation, ribosomal structure and biogenesis
ObacDRAFT_1703	641181809	translation elongation factor G	3.6.5.3	0480	J	Translation, ribosomal structure and biogenesis
ObacDRAFT_2586	641183246	tryptophanyl-tRNA synthetase	6.1.1.2	0180	J	Translation, ribosomal structure and biogenesis
ObacDRAFT_3342	641180098	two component transcriptional regulator, LuxR family	N/A	2197	K	
ObacDRAFT_3586	641182162	DNA mismatch repair protein MutS domain protein	N/A	0249	L	Replication, recombination and repair
ObacDRAFT_0959	641182396	twitching motility protein	N/A	2805	N	Cell motility
ObacDRAFT_1758	641181629	ATPase AAA-2 domain protein	N/A	0542	O	Posttranslational modification, protein turnover, chaperones
ObacDRAFT_1710	641182761	ferric uptake regulator, Fur family	N/A	0735	P	Inorganic ion transport and metabolism
ObacDRAFT_1220	641181014	phosphopantetheine-binding	N/A	0236	Q	Secondary metabolites biosynthesis, transport and catabolism
ObacDRAFT_1706	641181812	2,5-didehydrogluconate reductase	N/A	0656	R	General function prediction only
ObacDRAFT_0445	641181248	band 7 protein	N/A	2268	S	Function unknown
ObacDRAFT_0605	641181912	putative CRISPR-associated protein, APE2256 family		4006	S	Function unknown
ObacDRAFT_1966	641183420	SecA Wing and Scaffold	N/A	0653	U	Intracellular trafficking, secretion, and

Table B - 1 List of genes and proteins that were significantly up- or down-regulated under 2% and 20% O_2 - *Continued*

Locus Tag	Gene Object ID	ID Description	EC Number	COG	Group	Functional Categories
						vesicular transport
ObacDRAFT_3531	641180213	Tetratricopeptide TPR_2 repeat protein		5010	U	Intracellular trafficking, secretion, and vesicular transport
ObacDRAFT_2613	641183326	TPR repeat-containing protein				
ObacDRAFT_1558	641181288	Nucleotidyl transferase	2.7.7.8	N/A		
ObacDRAFT_3338	641180094	hypothetical protein ObacDRAFT_3338				
ObacDRAFT_2602	641182983	conserved hypothetical protein				
ObacDRAFT_3954	641181565	conserved hypothetical protein				
ObacDRAFT_1126	641180760	hypothetical protein ObacDRAFT_1126				
ObacDRAFT_0266	641180556	hypothetical protein ObacDRAFT_0266				
ObacDRAFT_4041	641183491	hypothetical protein ObacDRAFT_4041				
ObacDRAFT_3315	641181266	conserved hypothetical protein				
ObacDRAFT_1965	641183567	conserved hypothetical protein-signal peptide and transmembrane prediction				
ObacDRAFT_3500	641182172	protein of unknown function DUF303 acetylesterase putative				
ObacDRAFT_1489	641181085	hypothetical protein ObacDRAFT_1489				
ObacDRAFT_3940	641181463	hypothetical protein ObacDRAFT_3940				
ObacDRAFT_1245	641182420	conserved hypothetical protein				
ObacDRAFT_1015	641183098	protein of unknown function DUF303 acetylesterase putative				
ObacDRAFT_1491	641181087	hypothetical protein ObacDRAFT_1491				
DOWN-regulated						
ObacDRAFT_3556	641183484	6-phosphogluconate dehydrogenase NAD-binding	1.1.1.44	2084	I	Lipid transport and metabolism
ObacDRAFT_0038	641183259	uridylate kinase	2.7.4.22	0528	F	Nucleotide transport and metabolism
ObacDRAFT_1019	641181412	Adenosylhomocysteinase	4.4.1.21	0499	H	Coenzyme transport and metabolism

Table B - 1 List of genes and proteins that were significantly up- or down-regulated under 2% and 20% O_2 - *Continued*

Locus Tag	ID	ID Description	Number	COG	Group	Functional Categories
ObacDRAFT_1018	641183101	Methionine adenosyltransferase	2.5.1.6	0192	H	Coenzyme transport and metabolism
ObacDRAFT_2307	641183495	acetate--CoA ligase	6.2.1.1	0365	I	Lipid transport and metabolism
ObacDRAFT_4010	641180594	FeS assembly protein SufB	N/A	0719	O	Posttranslational modification, protein turnover, chaperones
ObacDRAFT_2287	641180534	Hpt protein	N/A	2198	T	Signal transduction mechanisms
ObacDRAFT_1073	641181950	cell surface receptor IPT/TIG domain protein	N/A	N/A		
ObacDRAFT_2744	641182035	Uncharacterized protein-like protein				
ObacDRAFT_1288	641183076	conserved repeat domain				
ObacDRAFT_1612	641180499	hypothetical protein ObacDRAFT_1612				

Appendix C

Scatter Plots Comparisons of Transcriptomic and Proteomic Whole Cell Profiles of *Diplosphaera colotermitum* under 2% and 20% O_2 Concentrations.

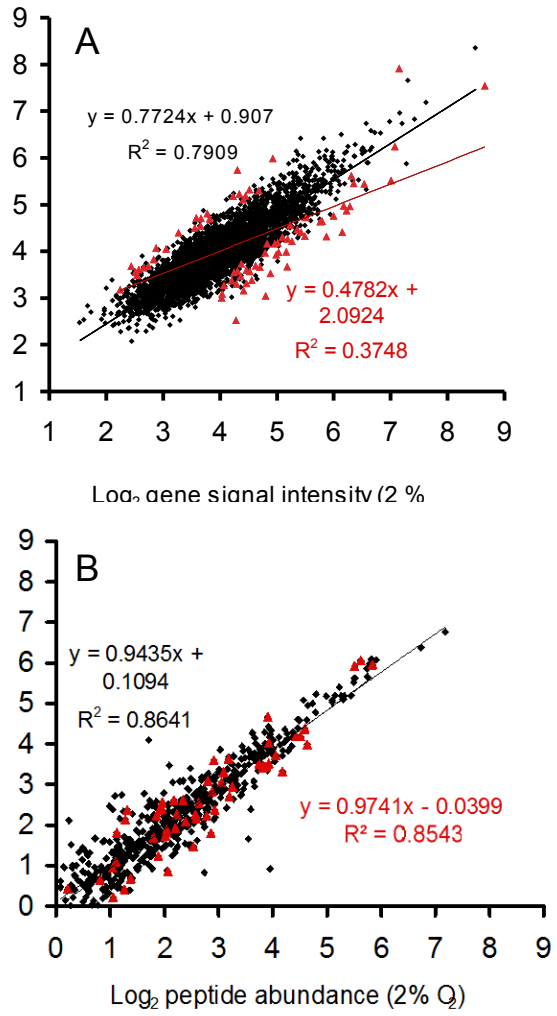

Figure C- 1 Log_2 transformed normalized scatter plots comparisons of (A) transcriptomic and (B) proteomic whole cell profiles of *Diplosphaera colotermitum* under 2% and 20% O_2 concentrations.

Red dots represent genes ($P < 0.01$) or proteins ($P < 0.05$) found to be differentially regulated or expressed between the two O_2 concentrations. Black dots represent all other genes or proteins found under both conditions.

Appendix D

Fold-Change Results for Selected Genes or Proteins with RT-qPCR, Microarray, and Log_2-Transformed Peptide Numbers for the *Diplosphaera colotermitum* Strain TAV2

Table D- 1 Fold-change results for selected genes or proteins with RT-qPCR, microarray, and log_2-transformed peptide numbers for the *Diplosphaera colotermitum* strain TAV2. Locus IDs were selected based on fold induction or number of peptides detected.

Target Gene Locus Tag	Gene Name	NCBI reference sequence	Microarray fold change	Peptide fold change[a]	RT-qPCR fold change
ObacDRAFT_0993	Acetyl xylan esterase	ZP_02011263	2.42	ND	10.27^{+}
ObacDRAFT_1002	ATP synthase F1, alpha subunit	ZP_02011249	3.61	4.087	4.47^{+++}
ObacDRAFT_2695	ribosomal protein L15	ZP_02013299	3.14	2.322	1.01^{+}
ObacDRAFT_1939	Acetyl xylan esterase	ZP_02012874	NS	2.807	3.91^{+}
ObacDRAFT_3012	Xylose isomerase TIM barrel	ZP_02011037	NS	2	1.98^{++}
ObacDRAFT_1973	Xylose isomerase domain protein TIM barrel	ZP_02013011	NS	2.807	9.00^{+}
ObacDRAFT_0419	Xylose isomerase domain protein TIM barrel	ZP_02013650	NS	2.585	4.40^{+}
ObacDRAFT_2549	DNA polymerase III, beta subunit	ZP_02011517	2.25	4.087	1.27^{++}

[a] Log_2-transformed peptide counts
ND: Not detected.
NS: Not significant.
$^{+}$ $P < 0.05$, $^{++}$ $P < 0.01$, $^{+++}$ $P < 0.001$ (two tailed Student's *T*-test)

Appendix E

Auto-Induction Reagent Lists

Auto-Induction Reagent Lists

ZY (1 L)

- 10 g Tryptone or peptone
- 5 g yeast extract
- Distilled H_2O

50x MPS (500 mL)

- 112 g Na_2HPO_4 (1.25 M)
- 85 g KH_2PO_4 (1.25 M)
- 67 g NH_4Cl (2.5 M)
- Distilled H_2O

50x 5052 (1 L)

- 250 g Glycerol (weigh in beaker; 0.5%)
- 25 g Glucose (0.05%)
- 100 g α-lactose (0.2%)
- 730 mL distilled H_2O

1 M $MgSO_4$(100 mL)

- 24.65 g $MgSO_4 \bullet 7H_2O$
- Distilled H_2O

1 M Na_2SO_4 (100 mL)

- 14.2 g Na_2SO_4
- Distilled H_2O

40% Glucose (w/v) (100 mL)

- 40 g Glucose
- 74 mL distilled H_2O

• ZYM-5052 starter media

Component	Volume	Final conc.
ZY	Up to 100 mL	-
1 M Na_2SO_4	500 µL	5 mM
1 M $MgSO_4$	200 µL	2 mM
40% Glucose	125 µL	0.05%
50x MPS	2 mL	1X
Antibiotics, as needed from the list below		
Kanamycin(50 mg/mL)	50 µL	25 µg/mL

•ZYP-5052 rich medium for auto-induction

Component	Volume	Final conc.
ZY	Up to 650 mL	-
1 M Na_2SO_4	3.25 mL	5 mM
1 M $MgSO_4$	1.3 mL	2 mM
50x 5052	13 mL	1X
50x MPS	13 mL	1X
Antibiotics, as needed from the list below		
Kanamycin (50 mg/mL)	325 µL	25 µg/mL

References

Badger PC. 2002. Ethanol from cellulose: A general review. Trends in new crops and new uses. Alexandria, VA: ASHS Press. p. 17-21.

Barton LL. 2004. Structural and functional relationships in prokaryotes. 2005 ed. New York, NY: Springer. p. 844.

Bendtsen JD, Nielsen H, von Heijne G, Brunak S. 2004. Improved prediction of signal peptides: SignalP 3.0. Journal of Molecular Biology. 340(4):783-795.

Bergersen FJ, Turner GL. 1975. Leghaemoglobin and the supply of O_2 to nitrogen-fixing root nodule bacteroids: presence of two oxidase systems and ATP production at low free O_2 concentration. Journal of General Microbiology 91(2):345-354.

Bergmann GT, Bates ST, Eilers KG, Lauber CL, Caporaso JG, Walters WA, Knight R, Fierer N. 2011. The under-recognized dominance of *Verrucomicrobia* in soil bacterial communities. Soil Biology & Biochemistry 43(7):1450-1455.

Berlanga M, Paster BJ, Grandcolas P, Guerrero R. 2011. Comparison of the gut microbiota from soldier and worker castes of the termite *Reticulitermes grassei*. International Microbiology 14(2):83-93.

Blum DL, Li X-L, Chen H, Ljungdahl LG. 1999. Characterization of an acetyl xylan esterase from the anaerobic fungus *Orpinomyces* sp. Strain PC-2. Applied and Environmental Microbiology 65(9):3990-3995.

Boden TA, G. Marland, and R.J. Andres. 2010. Global, regional, and national fossil-fuel CO_2 emissions. Carbon Dioxide Information Analysis Center. Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A.

Breznak JA, Brill WJ, Mertins JW, Coppel HC. 1973. Nitrogen-fixation in termites. Nature 244(5418):577-579.

Breznak JA, Brune A. 1994. Role of microorganisms in the digestion of lignocellulose by termites. Annual Review of Entomology. 39:453-487.

Breznak JA, Pankratz HS. 1977. *In situ* morphology of gut microbiota of wood-eating termites *Reticulitermes flavipes* (Kollar) and *Coptotermes formosanus* (Shiraki). Applied and Environmental Microbiology 33(2):406-426.

Brune A. 1998. Termite guts: the world's smallest bioreactors. Trends in Biotechnology 16(1):16-21.

Brune A. 2006. Symbiotic associations between termites and prokaryotes. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E, editors. The Prokaryotes. Springer New York. p. 439-474.

Brune A, Emerson D, Breznak JA. 1995. The termite gut microflora as an oxygen sink-microelectrode determination of oxygen and pH gradients in guts of lower and higher termites. Applied and Environmental Microbiology 61(7):2681-2687.

Brune A, Ohkuma M. 2011. Role of the termite gut microbiota in symbiotic digestion. In: Bignell DE, Roisin Y, Lo N, editors. Biology of Termites: a Modern Synthesis. Springer Netherlands. p. 439-475.

Burkert U, Warnecke F, Babenzien D, Zwirnmann E, Pernthaler J. 2003. Members of a readily enriched beta-proteobacterial clade are common in surface waters of a humic lake. Applied and Environmental Microbiology 69(11):6550-6559.

Burnum KE, Callister SJ, Nicora CD, Purvine SO, Hugenholtz P, Warnecke F, Scheffrahn RH, Smith RD, Lipton MS. 2011. Proteome insights into the symbiotic relationship between a captive colony of *Nasutitermes corniger* and its hindgut microbiome. The ISME journal 5(1):161-164.

Callister SJ, Dominguez MA, Nicora CD, Zeng X, Tavano CL, Kaplan S, Donohue TJ, Smith RD, Lipton MS. 2006a. Application of the accurate mass and time tag approach to the proteome analysis of sub-cellular fractions obtained from *Rhodobacter sphaeroides* 2.4.1. aerobic and photosynthetic cell cultures. Journal of Proteome Research 5(8):1940-1947.

Callister SJ, Nicora CD, Zeng X, Roh JH, Dominguez MA, Tavano CL, Monroe ME, Kaplan S, Donohue TJ, Smith RD, Lipton MS. 2006b. Comparison of aerobic and photosynthetic *Rhodobacter sphaeroides* 2.4.1 proteomes. Journal of Microbiological Methods 67(3):424-436.

Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Research 37:233-238.

Chin K-J, Janssen PH. 2002. Propionate formation by *Opitutus terrae* in pure culture and in mixed culture with a hydrogenotrophic methanogen and implications for carbon fluxes in anoxic rice paddy soil. Applied and Environmental Microbiology 68(4):2089-2092.

Chin K-J, Rainey FA, Janssen PH, Conrad R. 1998. Methanogenic degradation of polysaccharides and the characterization of polysaccharolytic *Clostridia* from anoxic rice field soil. Systematic and Applied Microbiology 21(2):185-200.

Chin KJ, Hahn D, Hengstmann U, Liesack W, Janssen PH. 1999. Characterization and identification of numerically abundant culturable bacteria from the anoxic bulk soil of rice paddy microcosms. Applied and Environmental Microbiology 65(11):5042-5049.

Chin KJ, Liesack W, Janssen PH. 2001. *Opitutus terrae* gen. nov., sp. nov., to accommodate novel strains of the division '*Verrucomicrobia*' isolated from rice paddy soil. International Journal of Systematic and Evolutionary Microbiology 51(Pt 6):1965-1968.

Choo Y-J, Lee K, Song J, Cho J-C. 2007. *Puniceicoccus vermicola* gen. nov., sp. nov., a novel marine bacterium, and description of *Puniceicoccaceae* fam. nov., *Puniceicoccales* ord. nov., *Opitutaceae* fam. nov., *Opitutales* ord. nov. and *Opitutae* classis nov. in the phylum '*Verrucomicrobia*'. International Journal of Systematic and Evolutionary Microbiology 57(Pt 3):532-537.

Ciais P, Tans PP, Denning AS, Francey RJ, Trolier M, Meijer HAJ, White JWC, Berry JA, Randall DA, Collatz GJ, Sellers PJ *et al.* . 1997. A three-dimensional synthesis study of $\delta^{18}O$ in atmospheric CO_2: 2. Simulations with the TM2 transport model. J. Geophys. Res. 102(D5):5873-5883.

Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ, Kulam-Syed-Mohideen AS, McGarrell DM, Marsh T, Garrity GM, Tiedje JM. 2009. The ribosomal database project: improved alignments and new tools for rRNA analysis. Nucleic Acids Research 37(Database issue):D141-145.

Collins NM, Wood TG. 1984. Termites and atmospheric gas-production. Science 224(4644):84-86.

Cui X, Churchill GA. 2003. Statistical tests for differential expression in cDNA microarray experiments. Genome Biology 4(4).

Degrassi G, Kojic M, Ljubijankic G, Venturi V. 2000. The acetyl xylan esterase of *Bacillus pumilus* belongs to a family of esterases with broad substrate specificity. Microbiology 146(7):1585-1591.

Degrassi G, Okeke BC, Bruschi CV, Venturi V. 1998. Purification and characterization of an acetyl xylan esterase from *Bacillus pumilus*. Applied and Environmental Microbiology 64(2):789-792.

Denef VJ, Kalnejais LH, Mueller RS, Wilmes P, Baker BJ, Thomas BC, VerBerkmoes NC, Hettich RL, Banfield JF. 2010. Proteogenomic basis for ecological divergence of closely related bacteria in natural acidophilic microbial communities. Proceedings of the National Academy of Sciences of the United States of America 107(6):2383-2390.

Dodd D, Cann IKO. 2009. Enzymatic deconstruction of xylan for biofuel production. Global Change Biology. Bioenergy 1(1):2-17.

Dupont C, Daigneault N, Shareck F, Morosoli R, Kluepfel D. 1996. Purification and characterization of an acetyl xylan esterase produced by *Streptomyces lividans*. Biochemical Journal 319(Pt 3):881-886.

Ebert A, Brune A. 1997. Hydrogen concentration profiles at the oxic-anoxic interface: a microsensor study of the hindgut of the wood-feeding lower termite *Reticulitermes flavipes* (Kollar). Applied and Environmental Microbiology 63(10):4039-4046.

Economou A, Wickner W. 1994. SecA promotes preprotein translocation by undergoing ATP-driven cycles of membrane insertion and deinsertion. Cell 78(5):835-843.

Eng JK, McCormack AL, Yates Iii JR. 1994. An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. Journal of the American Society for Mass Spectrometry 5(11):976-989.

Engel MS, Krishna K. 2004. Family-group names for termites (Isoptera). American Museum Novitates:1-9.

Eriksson K-EL, Blanchette RA, Ander P. 1990. Microbial and enzymatic degradation of wood and wood components. Berlin, Germany: Springer-Verlag.

Farrell AE, Plevin RJ, Turner BT, Jones AD, O'Hare M, Kammen DM. 2006. Ethanol can contribute to energy and environmental goals. Science 311(5760):506-508.

Faulon J-L, Carlson GA, Hatcher PG. 1994. A three-dimensional model for lignocellulose from gymnospermous wood. Organic Geochemistry 21(12):1169-1179.

Flint HJ, Scott KP, Duncan SH, Louis P, Forano E. 2012. Microbial degradation of complex carbohydrates in the gut. Gut Microbes 3(4):289-306.

Freitas S, Hatosy S, Fuhrman JA, Huse SM, Welch DBM, Sogin ML, Martiny AC. 2012. Global distribution and diversity of marine *Verrucomicrobia*. ISME Journal 6(8):1499-1505.

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

García-Aparicio MP, Ballesteros M, Manzanares P, Ballesteros I, González A, Negro MJ. 2007. Xylanase contribution to the efficiency of cellulose enzymatic hydrolysis of barley straw. In: Mielenz JR, Klasson KT, Adney WS, McMillan JD, editors. Applied Biochemistry and Biotecnology. Humana Press. p. 353-365.

Graber JR, Breznak JA. 2004. Physiology and nutrition of *Treponema Primitia*, an H_2/CO_2 Acetogenic spirochete from termite hindguts. Applied and Environmental Microbiology 70(3):1307-1314.

Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. 2003. Rfam: an RNA family database. Nucleic Acids Research. 31(1):439-441.

Gruppen H, Kormelink FJM, Voragen AGJ. 1993. Water-unextractable cell wall material from wheat flour. 3. A structural model for arabinoxylans. Journal of Cereal Science 18(2):111-128.

Hedlund BP. 2010a. Phylum XXIII. *Verrucomicrobia* phyl. nov. In: Krieg NR, Staley JT, Brown DR, Hedlund BP, Paster BJ, Ward NL, Ludwig W, Whitman WB, editors. Bergey's Manual® of Systematic Bacteriology. Springer New York. p. 795-841.

Hedlund BP. 2010b. Phylum XXIII. *Verrucomicrobia* phyl. nov. In: Krieg NR, Staley JT, Brown DR, Hedlund BP, Paster BJ, Ward NL, Ludwig W, Whitman WB, editors. Bergey's Manual® of Systematic Bacteriology. Springer New York. p. 795-841.

Hedlund BP, Gosink JJ, Staley JT. 1996. Phylogeny of *Prosthecobacter*, the fusiform caulobacters: Members of a recently discovered division of the Bacteria. International Journal of Systematic Bacteriology 46(4):960-966.

Hedlund BP, Gosink JJ, Staley JT. 1997. *Verrucomicrobia* div. nov., a new division of the bacteria containing three new species of *Prosthecobacter*. Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology 72(1):29-38.

Hengstmann U, Chin KJ, Janssen PH, Liesack W. 1999. Comparative phylogenetic assignment of environmental sequences of genes encoding 16S rRNA and numerically abundant culturable bacteria from an anoxic rice paddy soil. Applied and Environmental Microbiology 65(11):5050-5058.

Henrici AT, Johnson DE. 1935. Studies of freshwater bacteria II stalked bacteria, a new order of *schizomycetes*. Journal of Bacteriology 30(1):61-93.

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

Hongoh Y, Ohkuma M, Kudo T. 2003. Molecular analysis of bacterial microbiota in the gut of the termite *Reticulitermes speratus* (Isoptera; Rhinotermitidae). FEMS Microbiology Ecology 44(2):231-242.

Hongoh Y, Sharma VK, Prakash T, Noda S, Taylor TD, Kudo T, Sakaki Y, Toyoda A, Hattori M, Ohkuma M. 2008a. Complete genome of the uncultured Termite Group 1 bacteria in a single host protist cell. Proceedings of the National Academy of Sciences of the United States of America 105(14):5555-5560.

Hongoh Y, Sharma VK, Prakash T, Noda S, Toh H, Taylor TD, Kudo T, Sakaki Y, Toyoda A, Hattori M, Ohkuma M. 2008b. Genome of an endosymbiont coupling N_2 fixation to cellulolysis within protist cells in termite gut. Science (New York, N.Y.) 322(5904):1108-1109.

Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 11:119.

Inoue JI, Noda S, Hongoh Y, Ui S, Ohkuma M. 2008. Identification of endosymbiotic methanogen and ectosymbiotic spirochetes of gut protists of the termite *Coptotermes formosanus*. Microbes and Environments 23(1):94-97.

Inoue JI, Saita K, Kudo T, Ui S, Ohkuma M. 2007. Hydrogen production by termite gut protists: Characterization of iron hydrogenases of parabasalian symbionts of the termite *Coptotermes formosanus*. Eukaryotic Cell 6(10):1925-1932.

Isanapong J, Goodwin L, Bruce D, Chen A, Detter C, Han J, Han CS, Held B, Huntemann M, Ivanova N, Land ML *et al.* . 2012. High-quality draft genome sequence of the *Opitutaceae* Bacterium Strain TAV1, a symbiont of the wood-feeding termite *Reticulitermes flavipes*. Journal of Bacteriology 194(10):2744-2745.

Jackson FA, Dawes EA. 1976. Regulation of the Tricarboxylic Acid Cycle and poly-β-hydroxybutyrate metabolism in *Azotobacter beijerinckii* grown under nitrogen or oxygen limitation. Journal of General Microbiology 97(2):303-312.

Jaitly N, Mayampurath A, Littlefield K, Adkins JN, Anderson GA, Smith RD. 2009. Decon2LS: An open-source software package for automated processing and visualization of high resolution mass spectrometry data. BMC bioinformatics 10. doi 10.1186/1471-2105-10-87.

Keller A, Nesvizhskii AI, Kolker E, Aebersold R. 2002. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. Analytical chemistry 74(20):5383-5392.

Khalil MAK, Rasmussen RA, French JRJ, Holt JA. 1990. The influence of termites on atmospheric trace gases- CH_4, CO_2, $CHCl_3$, N_2O, CO, H_2, and light-hydrocarbons. Journal of Geophysical Research-Atmospheres 95(D4):3619-3634.

Köhler T, Dietrich C, Scheffrahn RH, Brune A. 2012. High-resolution analysis of gut environment and bacterial microbiota reveals functional compartmentation of the gut in wood-feeding higher termites (*Nasutitermes* spp.). Applied and Environmental Microbiology 78(13):4691-4701.

Konstantinidis KT, Tiedje JM. 2005. Towards a genome-based taxonomy for prokaryotes. Journal of Bacteriology 187(18):6258-6264.

Kormelink FJM, Voragen AGJ. 1993. Degradation of different [(glucurono)arabino]xylans by a combination of purified xylan-degrading enzymes. Applied Microbiology and Biotechnology 38(5):688-695.

Kosugi A, Murashima K, Doi RH. 2002. Xylanase and acetyl xylan esterase activities of XynA, a key subunit of the *Clostridium cellulovorans* cellulosome for xylan degradation. Applied and Environmental Microbiology 68(12):6399-6402.

Krogh A, Larsson B, von Heijne G, Sonnhammer EL. 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. Journal of Molecular Biology. 305(3):567-580.

Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Research. 35(9):3100-3108.

Leadbetter JR, Breznak JA. 1996. Physiological ecology of *Methanobrevibacter cuticularis* sp. nov. and *Methanobrevibacter curvatus* sp. nov., isolated from the hindgut of the termite *Reticulitermes flavipes*. Applied and Environmental Microbiology 62(10):3620-3631.

Leadbetter JR, Crosby LD, Breznak JA. 1998. *Methanobrevibacter filiformis* sp. nov., a filamentous methanogen from termite hindguts. Archives of Microbiology 169(4):287-292.

Leadbetter JR, Schmidt TM, Graber JR, Breznak JA. 1999. Acetogenesis from H_2 plus CO_2 by spirochetes from termite guts. Science 283(5402):686-689.

Liesack W, Stackebrandt E. 1992. Occurrence of novel groups of the domain bacteria as revealed by analysis of genetic material isolated from an australian terrestrial environment. Journal of Bacteriology 174(15):5072-5078.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods (San Diego, Calif.) 25(4):402-408.

Ljungdahl LG, Eriksson KE. 1985. Ecology of microbial cellulose degradation. Advances in Microbial Ecology 8:237-299.

Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Research. 25(5):955-964.

Ludwig RA. 2004. Microaerophilic bacteria transduce energy via oxidative metabolic gearing. Research in Microbiology 155(2):61-70.

Madigan M. 2009. Brock Biology of Microorganisms. Pearson International Edition.

Madigan MT, Martinko JM, Dunlap PV, Clark DP. 2008. Brock Biology of Microorganisms. Benjamin Cummings.

Marchler-Bauer A, Bryant SH. 2004. CD-Search: protein domain annotations on the fly. Nucleic acids research 32(Web Server issue):W327-W331.

Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F *et al.* . 2013. CDD: conserved domains and protein three-dimensional structure. Nucleic acids research 41(Database issue):D348-352.

Martinez-Garcia M, Brazel DM, Swan BK, Arnosti C, Chain PSG, Reitenga KG, Xie G, Poulton NJ, Gomez ML, Masland DED, Thompson B *et al.* . 2012. Capturing single cell genomes of active polysaccharide degraders: An unexpected contribution of *Verrucomicrobia*. PLoS ONE 7(4).

McDermid KP, Forsberg CW, MacKenzie CR. 1990. Purification and properties of an acetylxylan esterase from *Fibrobacter succinogenes* S85. Applied and Environmental Microbiology 56(12):3805-3810.

McMillan James D. 1994. Pretreatment of Lignocellulosic Biomass. Enzymatic Conversion of Biomass for Fuels Production. American Chemical Society. p. 292-324.

Metz B, Davidson OR, Bosch PR, Dave R. 2007. Contribution of working group III to the fourth assessment report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Monroe ME, Tolić N, Jaitly N, Shaw JL, Adkins JN, Smith RD. 2007. VIPER: an advanced software package to support high-throughput LC-MS peptide identification. Bioinformatics 23(15):2021-2023.

Nachin L, El Hassouni M, Loiseau L, Expert D, Barras F. 2001. SoxR-dependent response to oxidative stress and virulence of *Erwinia chrysanthemi*: the key role of SufC, an orphan ABC ATPase. Molecular Microbiology 39(4):960-972.

Nachin L, Loiseau L, Expert D, Barras F. 2003. SufC: an unorthodox cytoplasmic ABC/ATPase required for [Fe-S] biogenesis under oxidative stress. The EMBO Journal 22(3):427-437.

Nadin E. 2007. For the love of termites. Engineering and Science 2:24 - 31.

Nakajima H, Hongoh Y, Usami R, Kudo T, Ohkuma M. 2005. Spatial distribution of bacterial phylotypes in the gut of the termite *Reticulitermes speratus* and the bacterial community colonizing the gut epithelium. FEMS Microbiology Ecology 54(2):247-255.

Noda S, Iida T, Kitade S, Nakajima H, Kudo T, Ohkuma M. 2005. Endosymbiotic *Bacteroidales* bacteria of the flagellated protist *Pseudotrichonympha grassii* in the gut of the termite *Coptotermes formosanus*. Applied and Environmental Microbiology 71(12):8811-8817.

Ohkuma M. 2003. Termite symbiotic systems: efficient bio-recycling of lignocellulose. Applied Microbiology and Biotechnology 61(1):1-9.

Ohkuma M. 2008. Symbioses of flagellates and prokaryotes in the gut of lower termites. Trends in Microbiology 16(7):345-352.

Ohkuma M, Brune A. 2011. Diversity, structure, and evolution of the termite gut microbial community. In: Bignell DE, Roisin Y, Lo N, editors. Biology of Termites: a Modern Synthesis. Springer Netherlands. p. 413-438.

Ohkuma M, Kudo T. 1998. Phylogenetic analysis of the symbiotic intestinal microflora of the termite *Cryptotermes domesticus*. FEMS Microbiology Letters 164(2):389-395.

Ophardt CE. [cited 2012 09/22] Cellulose. Elmhurst College. Available from: http://www.elmhurst.edu/~chm/vchembook/547cellulose.html

Parveen B, Mary I, Vellet A, Ravet V, Debroas D. 2013. Temporal dynamics and phylogenetic diversity of free-living and particle-associated *Verrucomicrobia* communities in relation to environmental variables in a mesotrophic lake. FEMS Microbiology Ecology 83(1):189-201.

Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC. 2010. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. Nature Methods. 7(6):455-457.

Pauling DC, Lapointe JP, Paris CM, Ludwig RA. 2001. *Azorhizobium caulinodans* pyruvate dehydrogenase activity is dispensable for aerobic but required for microaerobic growth. Microbiology (Reading, England) 147(Pt 8):2233-2245.

Pester M, Brune A. 2006. Expression profiles of fhs (FTHFS) genes support the hypothesis that spirochaetes dominate reductive acetogenesis in the hindgut of lower termites. Environmental Microbiology 8(7):1261-1270.

Pester M, Brune A. 2007. Hydrogen is the central free intermediate during lignocellulose degradation by termite gut symbionts. ISME Journal 1(6):551-565.

Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nature methods 8(10):785-786.

Polpitiya AD, Qian WJ, Jaitly N, Petyuk VA, Adkins JN, Camp DG, Anderson GA, Smith RD. 2008. DAnTE: a statistical tool for quantitative analysis of -omics data. Bioinformatics 24(13):1556-1558.

Potrikus CJ, Breznak JA. 1980. Uric-acid degrading bacteria in guts of termites *Reticulitermes flavipes* (Kollar). Applied and Environmental Microbiology 40(1):117-124.

Potrikus CJ, Breznak JA. 1981. Gut bacteria recycle uric acid nitrogen in termites: A strategy for nutrient conservation. Proceedings of the National Academy of Sciences of the United States of America 78(7):4601-4605.

Poutanen K, Rättö M, Puls J, Viikari L. 1987. Evaluation of different microbial xylanolytic systems. Journal of Biotechnology 6(1):49-60.

Poutanen K, Sundberg M. 1988. An acetyl esterase of *Trichoderma reesei* and its role in the hydrolysis of acetyl xylans. Applied Microbiology and Biotechnology 28(4-5):419-424.

Poutanen K, Sundberg M, Korte H, Puls J. 1990. Deacetylation of xylans by acetyl esterases of *Trichoderma reesei*. Applied Microbiology and Biotechnology 33(5):506-510.

Qian W-J, Liu T, Monroe ME, Strittmatter EF, Jacobs JM, Kangas LJ, Petritis K, Camp DG, 2nd, Smith RD. 2005. Probability-based evaluation of peptide and protein identifications from tandem mass spectrometry and SEQUEST analysis: the human proteome. Journal of Proteome Research 4(1):53-62.

Rosengaus RB, Zecher CN, Schultheis KF, Brucker RM, Bordenstein SR. 2011. Disruption of the termite gut microbiota and its prolonged consequences for fitness. Applied and Environmental Microbiology 77(13):4303-4312.

Rosenthal AZ, Matson EG, Eldar A, Leadbetter JR. 2011. RNA-seq reveals cooperative metabolic interactions between two termite-gut spirochete species in co-culture. The ISME journal 5(7):1133-1142.

Sabree ZL, Huang CY, Arakawa G, Tokuda G, Lo N, Watanabe H, Moran NA. 2012. Genome shrinkage and loss of nutrient-providing potential in the obligate symbiont of the primitive termite *Mastotermes darwiniensis*. Applied and Environmental Microbiology 78(1):204-210.

Sabree ZL, Kambhampati S, Moran NA. 2009. Nitrogen recycling and nutritional provisioning by *Blattabacterium*, the cockroach endosymbiont. Proceedings of the National Academy of Sciences of the United States of America 106(46):19521-19526.

Saha BC. 2003. Hemicellulose bioconversion. Journal of Industrial Microbiology & Biotechnology 30(5):279-291.

Salmassi TM, Leadbetter JR. 2003. Analysis of genes of tetrahydrofolate-dependent metabolism from cultivated spirochaetes and the gut community of the termite *Zootermopsis angusticollis*. Microbiology-Sgm 149:2529-2537.

Sangwan P, Kovac S, Davis KER, Sait M, Janssen PH. 2005. Detection and cultivation of soil *Verrucomicrobia*. Applied and Environmental Microbiology 71(12):8402-8410.

Schlesner H, Jenkins C, Staley J. 2006. The Phylum *Verrucomicrobia*: A phylogenetically heterogeneous bacterial group. In: Dworkin M, Falkow S, Rosenberg E, Schleifer K-H, Stackebrandt E, editors. The Prokaryotes. Springer New York. p. 881-896.

Schultz JE, Breznak JA. 1978. Heterotrophic bacteria present in hindguts of wood-eating termites *Reticulitermes flavipes* (Kollar). Applied and Environmental Microbiology 35(5):930-936.

Selig MJ, Adney WS, Himmel ME, Decker SR. 2009. The impact of cell wall acetylation on corn stover hydrolysis by cellulolytic and xylanolytic enzymes. Cellulose 16(4):711-722.

Shareck F, Biely P, Morosoli R, Kluepfel D. 1995. Analysis of DNA flanking the *xlnB* locus of *Streptomyces lividans* reveals genes encoding acetyl xylan esterase and the RNA component of ribonuclease P. Gene 153(1):105-109.

Shibuya N, Iwasaki T. 1985. Structural features of rice bran hemicellulose. Phytochemistry 24(2):285-289.

Shieh WY, Jean WD. 1998. *Alterococcus agarolyticus*, gen.nov., sp.nov., a halophilic thermophilic bacterium capable of agar degradation. Canadian Journal of Microbiology 44(7):637-645.

Smith RD, Anderson GA, Lipton MS, Pasa-Tolic L, Shen Y, Conrads TP, Veenstra TD, Udseth HR. 2002. An accurate mass tag strategy for quantitative and high-throughput proteome measurements. Proteomics 2(5):513-523.

Somerville C. 2007. Biofuels. Current Biology: CB 17(4):R115-119.

Sowell SM, Norbeck AD, Lipton MS, Nicora CD, Callister SJ, Smith RD, Barofsky DF, Giovannoni SJ. 2008. Proteomic analysis of stationary phase in the marine bacterium "*Candidatus* Pelagibacter ubique". Applied and Environmental Microbiology 74(13):4091-4100.

Stevenson BS, Eichorst SA, Wertz JT, Schmidt TM, Breznak JA. 2004. New strategies for cultivation and detection of previously uncultured microbes. Applied and Environmental Microbiology. 70(8):4748-4755.

Storey JD. 2002. A direct approach to false discovery rates. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64(3):479-498.

Suau A, Bonnet R, Sutren M, Godon JJ, Gibson GR, Collins MD, Dore J. 1999. Direct analysis of genes encoding 16S rRNA from complex communities reveals many novel molecular species within the human gut. Applied and Environmental Microbiology 65(11):4799-4807.

Tatusov RL, Koonin EV, Lipman DJ. 1997. A genomic perspective on protein families. Science (New York, N.Y.) 278(5338):631-637.

Teeling H, Meyerdierks A, Bauer M, Amann R, Glöckner FO. 2004. Application of tetranucleotide frequencies for the assignment of genomic fragments. Environmental Microbiology 6(9):938-947.

Tholen A, Brune A. 2000. Impact of oxygen on metabolic fluxes and in situ rates of reductive acetogenesis in the hindgut of the wood-feeding termite *Reticulitermes flavipes*. Environmental Microbiology 2(4):436-449.

Tholen A, Schink B, Brune A. 1997. The gut microflora of *Reticulitermes flavipes*, its relation to oxygen, and evidence for oxygen-dependent acetogenesis by the most abundant *Enterococcus* sp. FEMS Microbiology Ecology 24(2):137-149.

Thompson NS. 1983. Hemicellulose as a biomass resource. In: Soltes E, editor. Wood and Agricultural Residues. Research on Use for Feed, Fuels, and Chemicals. New York: Academic. p. pp. 101-119.

Thomson JA. 1993. Molecular biology of xylan degradation. FEMS Microbiology Reviews 10(1-2):65-82.

Todaka N, Inoue T, Saita K, Ohkuma M, Nalepa CA, Lenz M, Kudo T, Moriya S. 2010. Phylogenetic analysis of cellulolytic enzyme genes from representative lineages of termites and a related cockroach. PLOS ONE 5(1).

Todaka N, Moriya S, Saita K, Hondo T, Kiuchi I, Takasu H, Ohkuma M, Piero C, Hayashizaki Y, Kudo T. 2007. Environmental cDNA analysis of the genes involved in lignocellulose digestion in the symbiotic protist community of *Reticulitermes speratus*. FEMS Microbiology Ecology 59(3):592-599.

Tokuda G, Watanabe H. 2007. Hidden cellulases in termites: revision of an old hypothesis. Biology Letters 3(3):336-339.

van Passel MWJ, Kant R, Palva A, Copeland A, Lucas S, Lapidus A, Glavina del Rio T, Pitluck S, Goltsman E, Clum A, Sun H *et al.* . 2011. Genome sequence of the *Verrucomicrobium Opitutus terrae* PB90-1, an abundant inhabitant of rice paddy soil ecosystems. Journal of Bacteriology 193(9):2367-2368.

Warnecke F, Luginbuhl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, Cayouette M, McHardy AC, Djordjevic G, Aboushadi N, Sorek R *et al.* . 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. Nature 450(7169):560-U17.

Watanabe H, Tokuda G. 2010. Cellulolytic systems in insects. Annual Review of Entomology 55:609-632.

Wertz JT, Breznak JA. 2007a. Physiological ecology of *Stenoxybacter acetivorans*, an obligate microaerophile in termite guts. Applied and Environmental Microbiology 73(21):6829-6841.

Wertz JT, Breznak JA. 2007b. *Stenoxybacter acetivorans* gen. nov., sp. nov., an acetate-oxidizing obligate microaerophile among diverse O_2-consuming bacteria from termite guts. Applied and Environmental Microbiology 73(21):6819-6828.

Wertz JT, Kim E, Breznak JA, Schmidt TM, Rodrigues JLM. 2012. Genomic and physiological characterization of the *Verrucomicrobia* Isolate *Diplosphaera colitermitum* gen. nov., sp nov., reveals microaerophily and nitrogen fixation genes. Applied and Environmental Microbiology 78(5):1544-1555.

White D. 2007. The physiology and biochemistry of prokaryotes. Oxford University Press.

Widdel F, Bak F. 1991. Gram negative mesophilic sulfate-reducing bacteria. In: Balows A, Trüper HG, Dworkin M, Harder W, Schleifer K-H, editors. The Prokaryotes. 2 ed. New York, NY: Springer-Verlag. p. 3352-3378.

Yoon J. 2011. Phylogenetic studies on the bacterial phylum '*Verrucomicrobia*'. Microbiol. Cult. Coll 27(2):61-65.

Zhang J, Siika-Aho M, Tenkanen M, Viikari L. 2011. The role of acetyl xylan esterase in the solubilization of xylan and enzymatic hydrolysis of wheat straw and giant reed. Biotechnology for Biofuels 4(1).

Zhang W, Li F, Nie L. 2010. Integrating multiple 'omics' analysis for microbial biology: application and methodologies. Microbiology 156:287-301.

Zhang Z, Schwartz S, Wagner L, Miller W. 2000. A greedy algorithm for aligning DNA sequences. Journal of Computational Biology: A Journal of Computational Molecular Cell Biology 7(1-2):203-214.

Biographical information

Jantiya Isanapong completed her Bachelor degree in Sanitary Science in 1998 from KhonKaen University, Thailand. She received a scholarship from the Royal Thai Government to pursue her Master and Doctoral degrees in the United States. She completed her Master's degree in Environmental Engineering in 2008 from Lehigh University, Pennsylvania. Then, she got accepted to the doctoral program of Earth and Environmental Sciences at the University of Texas at Arlington in spring of 2009 and completed her research on environmental microbiology under supervision of Dr. Jorge Rodrigues at Department of Biology. After completing the doctoral program, she will work as a lecturer at the Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Thailand.