

- I. **Formate dehydrogenase gene diversity in  
lignocellulose-feeding insect gut microbial  
communities**
  
- II. **Metabolic impacts on the hydrogen isotope content of  
bacterial lipids**

Thesis by

Xinning Zhang

In Partial Fulfillment of the Requirements for the  
degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2010

(Defended April 15, 2010)

© 2010

Xinning Zhang

All Rights Reserved

*To my parents,*

*Xin and Jiajun*

## ACKNOWLEDGEMENTS

First, I would like to recognize my parents, Xin and Jiajun. This thesis is dedicated to them. There are really no words I can use to express the appreciation and admiration I have for them as parents and human beings. I credit my father for setting me on the path I am on today. From early on, he instilled in me the value of education and emphasized the importance of finding my talents and doing something meaningful with them. I thank my mother for being a great role model, for always encouraging me to follow my dreams, for never giving up on me (especially during those tumultuous teenage years), for always making time for me, and for so much more. I also want to thank my sister, Doris, who has been a consistent source of level-headed advice and encouragement throughout my life.

My mentors at Cornell deserve to be recognized for introducing me to environmental science research. Monroe Weber-Shirk taught me that a little sludge goes a long way, that research doesn't have to utilize fancy instruments (a pressure sensor, filter, and mud work quite well!), and that research doesn't have to cure cancer to be worthwhile. I thank Beth Ahner for believing in my abilities, for encouraging me to apply to Caltech, and for being a great role model for women in science.

I am also grateful to the people I've had the pleasure to learn from and get to know as colleagues and friends over the past 6 years at Caltech.

My advisor, Jared Leadbetter, has and continues to inspire me with his passion for science and his care for the people around him. I was very lucky to have been introduced to

microbiology by Jared. He truly has a gift for microbial enrichment and isolation; I would not be at all surprised if he were known someday as a “microbe whisperer.”

I give credit to Eric Matson for his superb mentorship at the bench over the past 6 years. I really can think of no other person whom I regard as highly with respect to work-ethic and skill at performing meticulous research. Besides being a great role model in science, Eric has also been a good friend over the years. I thank him for all this and more.

Adam Rosenthal is another lab member who deserves a lot of my thanks as he has, in the past year at Caltech, made important impacts on my life as a burgeoning scientist. Adam’s curiosity and enthusiasm for new molecular biology techniques pushed me (or dragged me?) forward to complete the microfluidic work presented in Chapter 4. This chapter contains our combined efforts on a “wouldn’t it be cool...” project thought up during regular coffee sessions.

And I shouldn’t forget Liz Ottesen, the master of the Biomark. Thanks to Liz, I may now be able to design primers for extraterrestrial DNA (or whatever they use to store information). Other members of the Leadbetter lab (past and present) whom I like thank include: Jean Huang, Yajuan Wang, Jian-Yuan Thum, Kasia Gora, Abbie Green, Nick Ballor, and Kaitlyn Lucey. Each has made a positive impact on my life.

I owe a debt of gratitude to numerous people outside the Leadbetter lab at Caltech and around the world. These people include (in no particular order): the Paulot family, Jessamay Howell, Ashley Jones, Gretchen Aleks, Xavier Levine, and Sky Rashby.

With respect to ESE faculty, I thank Alex Sessions for not giving up on our work together when I flooded his lab during my first year at Caltech and for always being available to chat. Dianne Newman is another faculty member who has made a big impact on my life, both scientifically and personally. I will always remember her thoughtfulness after my dad's passing. The rest of the ESE & GPS community has also been influential in my growth as a scientist and person. I couldn't imagine a better atmosphere for research and random musings on life.

Last, but definitely not least, un grand merci à Fabien Paulot. Had I not come to Caltech and had you not decided to venture abroad, we would have never met and I would not be the person I am today.

## ABSTRACT

I.) Symbiotic CO<sub>2</sub>-reducing acetogens are important bacterial members of lignocellulose-feeding termite and roach gut communities. Acetogens are the major consumers of H<sub>2</sub> derived from lignocellulose fermentation and can contribute up to 1/3 of the acetate that serves as fuel for the insect host. Many acetogens in wood-feeding termites belong to a diverse group of relatively unstudied, uncultured spirochetes within the genus *Treponema*. Here, we used the gene sequence for hydrogenase-linked formate dehydrogenase (FDH<sub>H</sub>), an enzyme utilized in sugar fermentation and the acetogenic metabolism of the isolate *T. primitia*, to investigate the diversity, evolution, and activity of uncultured acetogenic spirochetes in lignocellulose-feeding insect guts. To study diversity and evolution, we developed novel degenerate primers for FDH<sub>H</sub> genes and constructed gene inventories from the gut communities of taxonomically and nutritionally diverse termites and a wood-feeding roach. Phylogenetic analyses reveal that most genes group with those from *T. primitia*, forming two clades that encode selenium-dependent (Sec) and selenium-independent (Cys) enzymes, respectively. This result implies many uncultured acetogenic spirochetes encode FDH<sub>H</sub> genes. Phylogenetic patterns also imply FDH<sub>H</sub> gene pool composition between lower and higher termite taxa and termites with different lifestyles varies greatly. We interpret differences as shifts in acetogenic spirochete community structure that occurred during termite evolution. We then investigated activity by sequencing the gut community transcriptome of a termite using high-throughput sequencing techniques and mapping transcript reads to gene inventory and pure culture

data. We discover that  $FDH_H$  gene transcription is dominated by relatively few  $FDH_H$  phylotypes. Finally, we performed microfluidic digital PCR on gut bacteria to determine the specific 16S rRNA ribotypes of spirochetes that encode  $FDH_H$  genes. We report the ribotypes of transcriptionally active spirochetes herein. Our results have implications for the microbial ecology of uncultured acetogenic spirochetes. We suggest that (i) the trace element selenium has shaped the gene content of acetogenic spirochetes in gut communities over evolutionary time scales, (ii) acetogenic spirochete populations have undergone extinctions and radiations associated with an evolutionary bottleneck, convergent evolutions, and possibly even invasion during termite evolution, and (iii) termite gut acetogenesis is largely mediated by only a few spirochete species, which represent a small portion of total acetogenic spirochete diversity.

**II.)** The hydrogen-stable isotope compositions (D/H) of lipids in the environment vary greatly. All variations have been assumed to result from changes in the D/H of water, a source of lipid hydrogen. However, several studies suggest that water D/H may not be the only influential factor. In this study we report that lipid D/H values can vary by 500‰ in bacterial cultures despite constant water D/H. This indicates variations in lipid/water fractionation need to be considered when interpreting environmental data. More significantly, we demonstrate that lipid D/H values are systematically related to the utilization of different central metabolic pathways in bacteria. Our results suggest that different cellular mechanisms for NADPH synthesis result in lipids with characteristic D/H. We therefore propose that lipid D/H values may be useful isotopic markers of energy metabolism.

# TABLE OF CONTENTS

<b>Acknowledgements</b> .....	iv
<b>Abstract</b> .....	vii
<b>Table of Contents</b> .....	ix
<b>List of Figures</b> .....	xi
<b>List of Tables</b> .....	xiii
<b>Chapter 1: Introductory Background</b> .....	1-1
1.1 Preface to topics. ....	1-1
1.2 Introduction to formate dehydrogenase in insect guts.....	1-3
1.2.1 Early studies of acetogenesis .....	1-4
1.2.2 Acetogenesis pathway .....	1-5
1.2.3 Acetogenesis from H <sub>2</sub> and CO <sub>2</sub> .....	1-7
1.2.4 Acetogen phylogeny.....	1-8
1.2.5 Acetogen ecology.....	1-9
1.2.6 Ecological impacts of termite gut acetogenesis.....	1-10
1.2.7 Termite biology.....	1-11
1.2.8 Termite gut microbiology .....	1-15
1.2.9 Termite nutritional ecology.....	1-19
1.2.10 Termite gut acetogens.....	1-24
1.2.11 Formate dehydrogenase in a termite gut acetogen isolate .....	1-25
1.2.12 Overview of Chapters 2 - 4 .....	1-27
1.3 Introduction to bacterial lipid D/H and metabolism.....	1-29
Overview of Chapter 5.....	1-29
1.4 References.....	1-30
<b>Chapter 2: Formate dehydrogenase gene diversity in acetogenic gut communities of lower, wood-feeding termites and a wood-feeding roach</b> .....	2-1
Abstract .....	2-1
Introduction .....	2-3
Materials and Methods .....	2-6
Results.....	2-14
Discussion .....	2-30
Appendix .....	2-35
References .....	2-47
<b>Chapter 3: Formate dehydrogenase gene phylogeny in higher termites suggests gut microbial communities have undergone an evolutionary bottleneck, convergent evolution, and invasion</b> .....	3-1
Abstract .....	3-1
Introduction .....	3-3

Materials and Methods .....	3-6
Results .....	3-12
Discussion .....	3-34
Appendix .....	3-41
References .....	3-56
<b>Chapter 4: RNA-Seq and microfluidic digital PCR identification of transcriptionally active spirochetes in termite gut microbial communities.....</b>	<b>4-1</b>
Abstract .....	4-1
Introduction .....	4-3
Materials and Methods .....	4-6
Results .....	4-14
Discussion .....	4-27
Appendix .....	4-30
References .....	4-41
<b>Chapter 5: Large D/H variations in bacterial lipids reflect central metabolic pathways.....</b>	<b>5-1</b>
Abstract .....	5-1
Introduction .....	5-3
Results .....	5-5
Discussion .....	5-17
Conclusions .....	5-24
Materials and Methods .....	5-25
References .....	5-28
Appendix .....	5-33
<b>Chapter 6: Conclusions.....</b>	<b>6-1</b>

## LIST OF FIGURES

Number	Page
1.1. Wood-Ljungdahl pathway for acetogenesis.....	1-6
1.2. Termite family phylogeny .....	1-14
1.3. Carbon and energy flow in wood-feeding termites.....	1-21
2.1. Degenerate <i>fdhF</i> primer design.....	2-10
2.2. Phylogenetic analysis of FDH from bacterial isolates .....	2-17
2.3. Predicted SECIS-like elements for inventory sequences .....	2-22
2.4. Phylogenetic analysis of lower termite FDH <sub>H</sub> sequences.....	2-24
2.5. Phylogenetic analysis of Gut spirochete group FDH <sub>H</sub> : termites.....	2-26
2.6. Phylogenetic analysis of Gut spirochete group FDH <sub>H</sub> : termites and roach.....	2-29
2.7. Phylogenetic analysis of COII from lower termites and roaches .....	2-45
2.8. Rarefactions curves for lower termite and roach <i>fdhF</i> inventories .....	2-46
3.1. Phylogenetic analysis of COII from higher termites, lower termites, and roaches.....	3-17
3.2. Phylogenetic analysis of major FDH <sub>H</sub> sequence types.....	3-20
3.3. Phylogenetic analysis of Gut spirochete group FDH <sub>H</sub> . .....	3-23
3.4. Detailed phylogenetic analysis of higher termite spirochete group FDH <sub>H</sub> .....	3-26
3.5. Phylogenetic analysis of <i>Amitermes-Gnathamitermes-Rhynchotermes</i> clade FDH <sub>H</sub> .....	3-29
3.6. Phylogenetic analysis of Enteric <i>Proteobacteria</i> clade and Unclassified clade FDH <sub>H</sub> . .....	3-31
3.7. Principal component analysis of FDH <sub>H</sub> phylogeny. ....	3-33
3.8. Cys-clade specific PCR on insect gut DNA. ....	3-42
3.9. <i>Amitermes-Gnathamitermes-Rhynchotermes</i> clade specific PCR. ....	3-43

4.1. Gene inventory, RNA-Seq, and microfluidic PCR work-flow. ....	4-15
4.2. Microfluidic digital PCR: All Bacteria and Spirochete 16S rRNA. ....	4-22
4.3. Phylogenetic analysis of 16S rRNA, <i>fdhF</i> <sub>Sec</sub> , and <i>fdhF</i> <sub>Cys</sub> sequences amplified using microfluidic digital PCR. ....	4-26
5.1. Summary of D/H fractionations. ....	5-9
5.2. Regressions of $\delta D$ values versus water for <i>C. oxalaticus</i> ....	5-12
5.3. Fractionation factor curves for palmitic acid in <i>C. oxalaticus</i> ....	5-14
5.4. Representative bacterial growth curves. ....	5-43
5.5. Relationship between fatty acid and water $\delta D$ . ....	5-45
5.6. Fractionation curves for hypothetical sets of cultures ....	5-46
5.7. Fatty acid biosynthetic pathway ....	5-47
5.8. Summary of major central metabolic pathways highlighting important sources of NADPH ....	5-49
5.9. NADPH generation in the Embden–Meyerhoff–Parnas Pathway .....	5-51
5.10. NADPH generation in the oxidative pentose phosphate pathway and the Entner–Doudoroff pathway. ....	5-52
5.11. NADPH generation in the glycerate pathway. ....	5-52
5.12. NADPH generation in the tricarboxylic acid cycle and the glyoxylate shunt. ....	5-53
5.13. NADPH generation from transhydrogenase and the light reactions of photosynthesis. ....	5-54

## LIST OF TABLES

Number	Page
2.1.	PCR primers for <i>fdhF</i> -type formate dehydrogenases ..... 2-9
2.2.	Sequences used for <i>fdhF</i> primer design..... 2-36
2.3.	Primer combinations for PCR amplifications ..... 2-38
2.4.	Phylotype distribution in lower termite and roach FDH <sub>H</sub> libraries..... 2-39
2.5.	Summary of lower termite and roach FDH <sub>H</sub> libraries ..... 2-41
2.6.	Alignment of Gut spirochete group amino acid indel..... 2-42
3.1.	Description of higher termites ..... 3-13
3.2.	Summary of <i>fdhF</i> inventories from higher termites ..... 3-18
3.3.	Clone distribution amongst major clades..... 3-21
3.4.	PCR conditions for <i>fdhF</i> inventory construction..... 3-44
3.5.	Phylotype distribution in higher termite FDH <sub>H</sub> libraries..... 3-45
3.6.	Sequences used in COII and FDH <sub>H</sub> phylogenetic analyses..... 3-48
4.1.	Primer and probes used for microfluidics and inventory analyses ..... 4-13
4.2.	<i>Z. nevadensis</i> gut community FDH gene transcription: RNA-Seq, cDNA inventory, and qRT-PCR ..... 4-19
4.3.	Microfluidic chip retrieval frequencies for <i>fdhF</i> ..... 4-23
4.4.	RNA-Seq nucleotide scaffold data set ..... 4-31
4.5.	Microfluidic digital PCR experiment details ..... 4-40
5.1.	Summary of bacterial D/H culture experiments..... 5-7
5.2.	Relative abundances of fatty acids in bacterial cultures ..... 5-38
5.3.	$\delta D$ values of fatty acids and culture water..... 5-40
5.4.	Coefficients for regression of $R_l$ on $R_w$ ..... 5-42

## INTRODUCTORY BACKGROUND

### 1.1. Preface to topics

Microbes represent most of life's genetic and physiological diversity and are the major drivers of global biogeochemical cycles (57, 124). Yet the vast majority of microbes on Earth have not been successfully cultured (57). In the absence of pure cultures, many scientists rely on genetic and isotopic signatures in the environment to identify and investigate the ecological roles of uncultured microbes (1, 39). While such studies are commonly termed "culture-independent," interpretations ultimately rely on detailed pure culture investigations of phylogenetically or physiologically related microbes. My graduate work at Caltech has focused on two subjects in environmental science that highlight the synergism between environmental and pure culture studies.

*Topic I. Formate dehydrogenase gene diversity in lignocellulose-feeding insect gut microbial communities*

The majority of this thesis focuses on formate dehydrogenase gene diversity in the symbiotic gut microbial communities of lignocellulose-feeding termites and roaches. Formate dehydrogenase enzymes are crucial for autotrophic CO<sub>2</sub>-reductive acetogenesis, a bacterial process that significantly impacts insect nutrition, and, by way of host abundance, impacts the global carbon cycle. Chapters 2 – 4 describe studies of these functional genes in uncultured acetogenic bacteria, which relied on traditional and emerging molecular techniques in microbial ecology and were leveraged by pure culture

studies of a termite gut acetogen. The results shed light on the diversity, evolutionary biology, and activity of an important group of insect gut bacteria.

*Topic II. Metabolic impacts on the hydrogen isotope content of bacterial lipids*

The second topic of this thesis focuses on the biological determinants of bacterial lipid hydrogen stable isotope composition ( $H^2/H^1$ , D/H). Environmental measurements show lipid D/H values vary dramatically in ways that can not be explained by changes in the D/H of water, a source of lipid hydrogen. These data implicate biological processes as the sources of variation. However, such processes have remained almost completely unstudied. In Chapter 5 of this thesis, I describe studies on the relationship between energy metabolism and D/H of fatty acids in pure cultures of physiologically diverse bacteria. The results suggest lipid D/H may be a useful isotopic marker for energy metabolism.

The remainder of this chapter provides background information and a brief chapter outline for each topic.

## **1.2. Introduction to formate dehydrogenases in insect guts**

CO<sub>2</sub>-reductive acetogenesis is an anaerobic process used by certain microbes to gain energy, synthesize the key metabolic intermediate, acetyl-CoA, as well as cell carbon (34, 36). Of five known autotrophic mechanisms in nature, the reductive synthesis of one molecule of acetate from two molecules of CO<sub>2</sub> via acetogenesis is the only linear process for CO<sub>2</sub> fixation (34, 117). Its simplicity relative to other CO<sub>2</sub> fixation pathways has prompted some to consider it as the earliest evolved form of autotrophy (83). For these and other reasons, it has caught the interest of microbiologists, biochemists, environmental scientists, and geobiologists over the past ~80 years.

My work seeks to clarify the form, function, and evolution of microbes impacting CO<sub>2</sub> reductive acetogenesis in the guts of lignocellulose-feeding insects, where rates of acetate synthesis from H<sub>2</sub> + CO<sub>2</sub> are characteristically high and impact the global carbon cycle. Specifically, I use the gene for hydrogenase-linked formate dehydrogenase, a key enzyme in fermentative and acetogenic metabolism, as a “hook” to study uncultured acetogens inhabiting the guts of termites and roaches. In this introduction, I first present key microbiological, biochemical, and ecological aspects of acetogenesis. I then outline the biology of termites (and related insects) and termite gut microbial community composition before focusing on termite gut acetogenesis and acetogens. Finally, I introduce a genetic and transcriptional study of a specific termite gut acetogen, which formed the basis for the majority of my thesis work.

### 1.2.1. The discovery of acetogenesis: early microbiology and biochemistry

Studies of acetogenesis began in 1932 when Fischer *et al.* reported the H<sub>2</sub>-dependent synthesis of acetate from CO<sub>2</sub> in sewage sludge [(36) and references therein]. Four years later, the microbiologist Wieringa isolated the first acetogen (125), a Gram-negative spore forming bacterium of the phylum *Firmicutes*, which could grow via the complete synthesis of acetate (CH<sub>3</sub>COOH) from CO<sub>2</sub> and H<sub>2</sub> following the reaction:

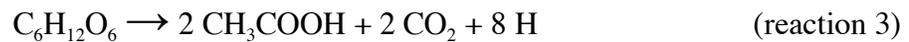


This organism, *Clostridium aceticum*, was lost, but a second *Firmicute* acetogen, *C. thermoaceticum* (renamed *Moorella thermoacetica*), was isolated a few years later from horse manure by Fontaine *et al.* (44). As no other acetogenic bacteria were isolated for several years, *M. thermoacetica* became the model bacterium for almost all biochemical and enzymological studies of acetogenesis.

In contrast to *C. aceticum*, acetogenesis in *M. thermoacetica* was ironically identified in the context of its heterotrophic metabolism of glucose (C<sub>6</sub>H<sub>12</sub>O<sub>6</sub>). In 1942, Fontaine and colleagues (44) noted that this bacterium's metabolism of glucose to acetate yielded a novel stoichiometry (reaction 2), which precluded the typical 3-3 split of glucose typical in glycolysis:



Later, in 1944, Barker (3) employed fermentation balances to propose that reaction 2 was actually a sum of two reactions, whereby the reducing equivalents from glucose oxidation (reaction 3, where “2H” represents a reducing equivalent such as H<sub>2</sub>) are used to generate an additional molecule of acetate via CO<sub>2</sub> reductive acetogenesis (reaction 4):

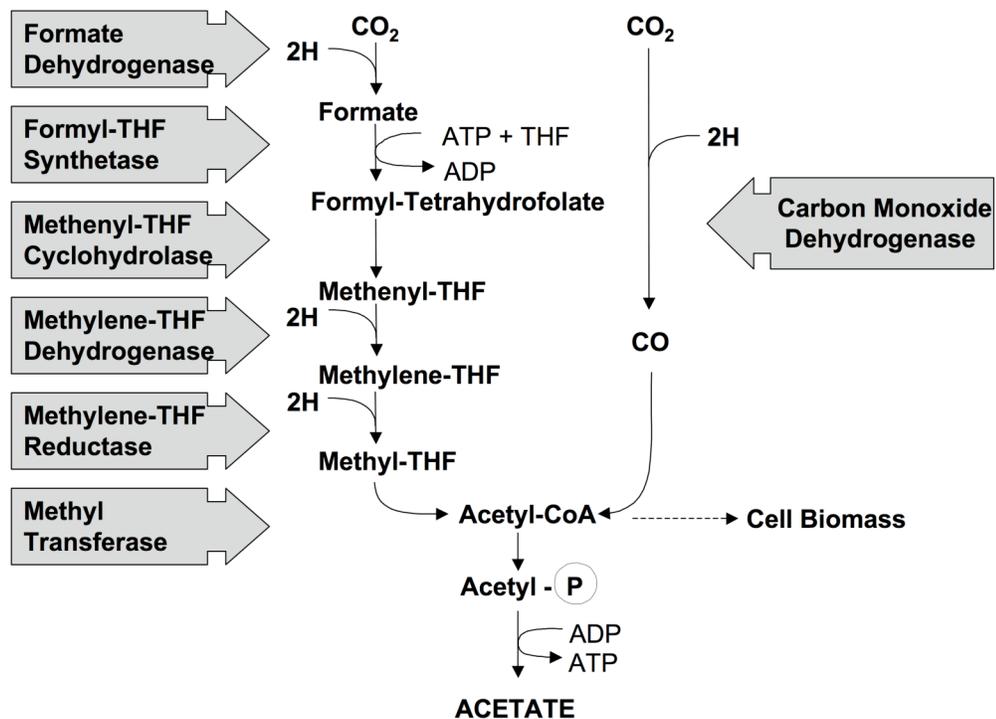


Barker and Kamen (4) confirmed this reaction scheme by demonstrating that *M. thermoacetica* produces acetate labeled in both carbon positions when grown with <sup>14</sup>CO<sub>2</sub>. This landmark study not only refined the concepts of autotrophy and heterotrophy, it was also the first biological study performed which utilized <sup>14</sup>C as a metabolic tracer. H. G. Wood, in a mass-balance study using <sup>13</sup>C, provided definitive isotopic evidence for the complete synthesis of acetate from CO<sub>2</sub> via reaction 4 (126).

### 1.2.2. The Wood-Ljungdahl pathway for acetogenesis

In the next 40 years, Wood and his student L. G. Ljungdahl led efforts to elucidate the biochemistry and enzymology of CO<sub>2</sub> reductive acetogenesis. As a result, the enzymatic steps underlying CO<sub>2</sub> reductive acetogenesis have been collectively termed the “Wood-Ljungdahl” pathway (Figure 1.1). The pathway, also known as the acetyl-CoA pathway, was first proposed in 1969 by Wood and Ljungdahl and later summarized by the same

authors in 1991, after its resolution in *M. thermoacetica* — and the demonstration that this organism could, in fact, grow as an autotroph on  $H_2 + CO_2$  (31, 79, 127). From here on, I will use the term “acetogen” to describe microbes that utilize the Wood-Ljungdahl pathway for energy conservation, acetyl-CoA production, and assimilation of  $CO_2$  into cell carbon, either during heterotrophic metabolism (e.g., *M. thermoacetica*) or autotrophic metabolism (e.g., *C. aceticum*).



**Figure 1.1.** Wood-Ljungdahl pathway for  $CO_2$ -reductive acetogenesis (adapted from a personal communication by J. R. Leadbetter and (96)).  $CO_2$  is the terminal electron acceptor for reducing equivalents,  $2H$ , which are typically  $H_2$  or derived from it (32). Carbohydrates,  $CO$ , methanol, and other incompletely oxidized one-carbon and two-carbon compounds can also serve as sources of  $2H$  (32, 33). The major site of chemiosmotic energy conservation is thought to be the methyl-transferase reaction (32). THF, tetrahydrofolate; Acetyl-P, acetyl-phosphate. Dashed line indicates intermediate drawn off for biosynthetic processes.

The heterotrophic capacity of acetogens is quite diverse. Sources of reductant include carbohydrates, one-carbon (C1), and even two-carbon compounds (36). These substrates interact with the Wood-Ljungdahl pathway in two conceptually different ways. As observed in *M. thermoacetica*, carbohydrate fermentation provides reducing equivalents but no reduced carbon skeletons for acetogenesis. In contrast, acetogens using C1 compounds obtain reducing equivalents, as well as incompletely oxidized carbon substrates, by fully oxidizing a portion of their C1 substrates to CO<sub>2</sub> for reductant, and then channeling the remaining C1 molecules into the acetogenesis pathway at intermediates of similar redox state.

### **1.2.3. Acetogenesis from H<sub>2</sub> and CO<sub>2</sub>**

The autotrophic growth of acetogens depends on chemiosmotic energy generation, since there is no net substrate level phosphorylation of ATP (Figure 1.1) when H<sub>2</sub> and CO<sub>2</sub> are substrates for acetogenesis. Two types of chemiosmotic mechanisms have been identified in acetogens. Organisms like *M. thermoacetica*, which possess membranous electron transport proteins, perform proton-based chemiosmosis [(34) and references therein] when they catalyze the methyl transferase reaction in acetogenesis. The autotrophic acetogen, *Acidobacterium woodii*, lacks such membrane features and instead uses the methyl transferase reaction (Figure 1.1) to create a sodium ion-based membrane gradient that drives chemiosmotic ATP production [(34) and references therein].

Formate dehydrogenase and carbon monoxide dehydrogenase are critical enzymes for the

pathway's CO<sub>2</sub> assimilating capacity (Figure 1.1). Formate dehydrogenase from *M. thermoacetica* was identified as an NADPH-linked cytoplasmic protein containing catalytic tungsten and selenium moieties (128). These latter aspects are notable as they represent discoveries of the first tungsten protein and one of the first selenoproteins in nature [(127) and references therein].

Lastly, autotrophic growth of acetogens depends on H<sub>2</sub>-oxidizing hydrogenases. These enzymes either “activate” H<sub>2</sub> into physiological forms (e.g., ferredoxin, NADPH) that are the direct reductants for acetogenesis, or transduce H<sub>2</sub> to a reducing enzyme, like formate dehydrogenase, when they are physically linked to other enzymes in a multi-enzyme complex (36). The latter has not been biochemically established in any acetogen to date, but may be relevant in H<sub>2</sub> rich environments. Hydrogenases are not explicitly shown in Figure 1.1 but are nonetheless just as important as CO<sub>2</sub> “activating” enzymes like formate dehydrogenase for acetogenesis from H<sub>2</sub> + CO<sub>2</sub>.

#### **1.2.4. Phylogeny of acetogenic bacteria**

Acetogens, as a group, represent over 20 bacterial genera [(34) and references therein]. Almost all acetogens belong to the phylum *Firmicutes*, but many of these genera are not monophyletic (i.e., sister taxa of acetogens may not be acetogenic) (101). The majority of acetogen isolates belong to the *Firmicute* genera *Acetobacterium* and *Clostridium*. Acetogen isolates outside the *Firmicutes* include two species of spirochete that belong to the genus *Treponema*, within the phylum *Spirochaetes* (34);  $\delta$ -*Proteobacteria* like the sulfate reducer, *Desulfotignum phosphitoxidans*, which grows as an acetogen in the

absence of sulfate (107); and *Holophaga foetida*, in the phylum *Acidobacteria* (76).

No known acetogenic *Archaea* have been identified, although the use of a major portion of the Wood-Ljungdahl pathway (i.e., all enzymes except formate dehydrogenase and formyl-tetrahydrofolate synthetase) for synthesis of cell carbon in autotrophic methanogenic *Archaea* is well-established (36). Theoretically, a bonified *Archaeal* acetogen should possess better energetics (i.e., conserve more ATP per acetate formed) than bacterial acetogens, which burn additional ATP to active THF with formyl-tetrahydrofolate synthetase (Figure 1.1). The Wood-Ljungdahl pathway is also employed by anaerobes like the sulfate reducing bacterium *Desulfobacterium autotrophicum* for anabolism. In addition, acetate oxidizing methanogens (e.g., *Methanosarcina barkeri*) and sulfate reducers (e.g., *Desulfotomaculum acetoxidans*) can use the Wood-Ljungdahl pathway in reverse to fuel catabolism. However, it is worth emphasizing that the aforementioned microbes are not true “acetogens” even though they may employ all or parts of the Wood-Ljungdahl pathway in their metabolism.

#### **1.2.5. Ecology of acetogenic bacteria**

Many acetogens are able to grow in a non-acetogenic capacity; for example, *M. thermoacetica* can use nitrate as a terminal electron acceptor instead of CO<sub>2</sub> (37). This physiological diversity translates into a wide distribution in nature (35, 38). Bacteria with acetogenic capabilities have been isolated from diverse anaerobic environments, including sewage, sediments, animal waste, hot springs, rumen fluid, and termite hindguts (36), but our knowledge of the ecological role such bacteria play as acetogens

within their respective environments remains fairly limited. As a consequence, there are few estimates for the impacts of acetogenesis outside environments such as certain sediments (49, 81) and termite hindguts (20). Nevertheless, these estimates, which credit acetogens for generating ~ 10% of sediment acetate production (127) and ~  $10^{12}$  kg of acetate annually in termite guts (20, 36), indicate acetogenesis plays an important role in the global carbon cycle.

#### **1.2.6. Ecological impacts of termite gut acetogenesis**

The global significance of acetogenesis is clearest in termites. These insects are best known for their ability to consume cellulose and lignocellulose, the most abundant biopolymers on land. These, and other activities, confer termites with the status of ecologically important arthropods that mediate carbon turnover and maintain soil fertility in terrestrial ecosystems (7, 9). Globally, termites consume ~ 4% of terrestrial plant biomass and account for 2 – 5% of terrestrial CO<sub>2</sub> and 3 – 5% of methane emitted to the atmosphere (10, 15, 67, 84, 102, 104, 116). Regionally, termites may be responsible for as much as 20% of carbon mineralization (9). These ecological roles are associated with one of nature's most striking nutritional mutualisms, wherein a complex community of obligate symbiotic microbes inhabiting termite hindguts degrades lignocellulose and other recalcitrant food substrates into carbon forms like acetate which fuel the host termite's metabolism. Among the most important termite gut symbionts are CO<sub>2</sub> reductive acetogens, which generate nearly 1/3 of the host's fuel (92). Globally termite gut acetogens generate ~  $1.22 \times 10^{12}$  kg acetate annually (20); this is ~ 10% of all acetate (~  $10^{13}$  kg) metabolized in anaerobic environments (36, 127).

Acetogenesis in termite guts also has implications for global climate. Acetogenesis is typically outcompeted by methanogenesis for reductant generated from anaerobic degradation (e.g., H<sub>2</sub>) in environments that are poor in electron acceptors other than CO<sub>2</sub>. The best example can be found in animal rumens where methanogenic *Archaea* dominate H<sub>2</sub> consumption and generate enough methane, a potent greenhouse gas, to account for 13 – 15% of global emissions (114). In contrast, the dominant H<sub>2</sub> consuming process in termite guts is bacterial CO<sub>2</sub>-reductive acetogenesis (21). As a result, termite emissions account for 3 – 5% of total methane emissions (114), rather than upwards of ~ 10%, as predicted based on ruminant-like carbon flows in termite guts (14).

### **1.2.7. Termite Biology**

Termites are eusocial animals that belong to the insect order *Isoptera*, which means “equal winged” and describes the fact that fore- and hind-wings are of approximately equal size in reproductive castes (48). They constitute one of our planet’s most diverse and abundant animal groups, with ~ 3000 extant species (41) representing at least 10<sup>18</sup> individuals (7). The majority of termites inhabit tropical environments (9), where they account for as much as 95% of insect biomass in soils (9). Termite abundance, biomass, and species diversity tend to decline with distance from the equator (40).

Termites belong to the detritivorous insect superorder *Dictyoptera*, which also includes cockroaches and mantids (48, 64). The general consensus is that termites descend from a wood-feeding cockroach, whose extant representatives belong to the sub-social, wood-

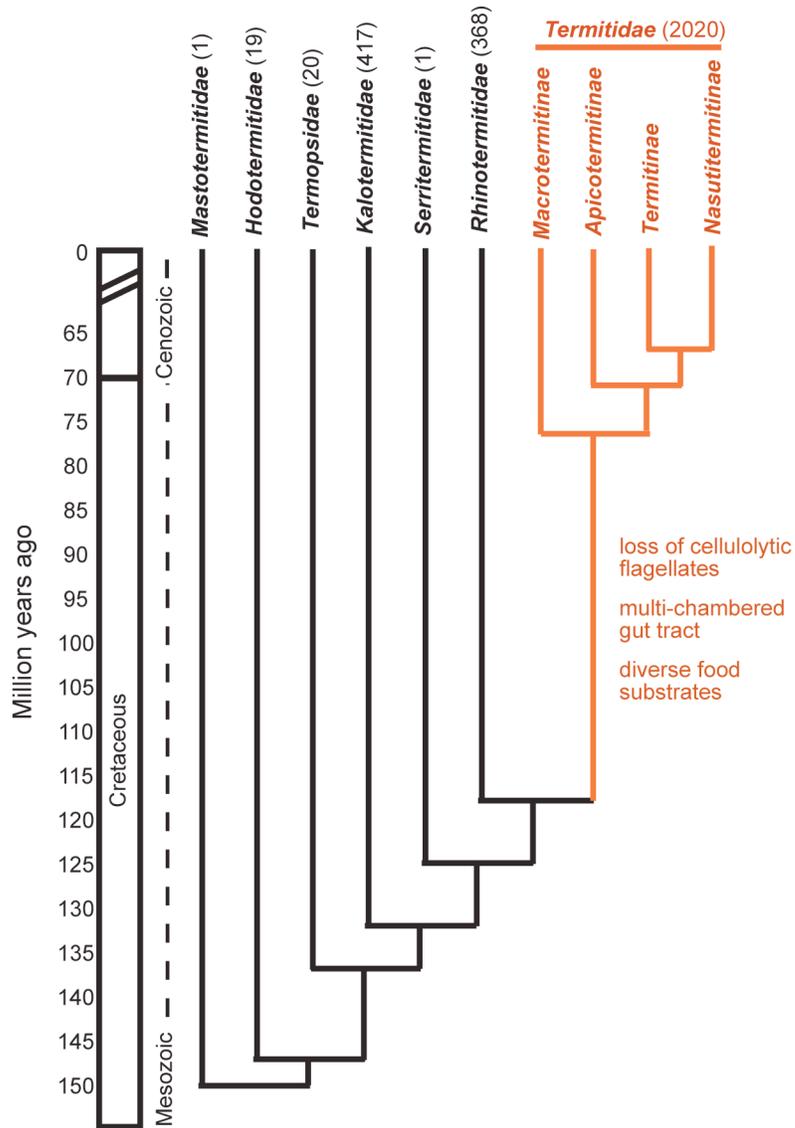
feeding cockroach genus *Cryptocercidae* (64, 80). The split between cockroach and termite lineages is estimated to have occurred ~ 140 million years ago (mya) in the early Cretaceous (48, 119). Within *Isoptera*, termites are traditionally classified into seven families (Figure 1.2) based on morphology and molecular data (64, 69).

Termites are also classified into two broad taxonomic groups, termed “lower” and “higher” termites. Members of six families are considered “lower” termites: *Mastotermitidae*, *Hodotermitidae*, *Termopsidae*, *Kalotermitidae*, *Rhinotermitidae*, and *Serritermitidae*. Members of the seventh family, *Termitidae*, are “higher” termites. This single family comprises the most termite individuals and > 85% of all termite genera. It is further classified into four sub-families: *Macrotermitinae*, *Apicotermitinae*, *Termitinae*, and *Nasutitermitinae*. The relationships between the subfamilies are still being debated, but taxonomists generally agree that fungus-cultivating *Macrotermitinae* are basal to other higher termite lineages and that *Nasutitermitinae* and *Termitinae* are the most derived (41, 64). Fossil remains date higher termite evolution to ~ 50 mya in the Eocene (48).

Apart from molecular phylogeny, important characteristics related to diet, gut structure, and gut microbial community composition also distinguish lower termites from their higher termite relatives. The diet of lower termites is comprised mainly of wood but can also include grass (64). The gut tracts of lower termites are relatively simple, consisting of foregut (crop and gizzard), midgut (site of digestive enzyme secretion), hindgut (largest compartment where most microbes are located and where digestion occurs),

colon, and rectum [(16, 89), see diagrams in (25)]. Microelectrode measurements have revealed that the hindgut is circumneutral and features steep radial gradients of O<sub>2</sub> (microoxic at the epithelial surface, anoxic in the luminal center), H<sub>2</sub> (~ 70 kPa in the center, decreasing outwards), and redox potential (-130 to -290 mV in the center, increasing outwards) (8, 26, 100). Axial variations in pH and redox have also been documented (6, 8). Physicochemical variations correlate with host activity (i.e., host-controlled tracheal gas exchange) and microbial distribution within the hindgut (77). Overall, the hindgut harbors a dense microbial community composed of morphologically diverse prokaryotes and cellulose fermenting flagellate protozoa that together enable lower termites to thrive on wood (16, 24).

Higher termites, on the other hand, are able to eat a wider range of substrates. These include wood in various stages of decay, grass, dung, leaf-litter, soil rich in humus, and fungi (9). They also possess multi-chambered hindguts with varying levels of segmentation (i.e., segments P1-P5) (24, 90). Degree of hindgut segmentation varies among higher termite species; soil-feeding termites possess the most highly differentiated hindguts, whereas fungus-cultivating termites have the least complex gut tracts. Physicochemical parameters such as pH and redox potential vary between and within hindgut chambers (26), and have been related to host factors and prokaryotic community composition along axial and radial axes of the gut tract (109, 111). Higher termites harbor *Bacteria* and *Archaea* in their hindguts (like lower termites) but do not possess any gut flagellates. The complete lack of flagellate protozoa is perhaps the most striking feature of the higher termite gut microbial community.



**Figure 1.2.** Phylogenetic reconstruction of termite family and subfamily evolution using extant and fossil insect data [adapted from Figure 7.88 in (48)]. Species number and key characteristics distinguishing higher (in red) from lower termites (in black) are listed.

### **1.2.8. Termite gut microbiota**

All termites harbor a complex community of microbes in their gut tracts. Wet mount preparations of hindgut contents reveal an environment densely packed with microbes that have diverse morphologies and exhibit varying levels of motility. Microscopy counts indicate at least  $10^9$  cells are present in every milliliter of hindgut fluid [(16) and references therein]. This is 3-orders of magnitude greater than the microbial density of seawater (124).

Lower termite gut microbiota include members of all three domains of life. Higher termites, as previously described, have robust prokaryotic gut populations but do not possess eukaryotic gut protists. In the following section, I provide a summary of current knowledge for each microbial group.

#### *Eukaryotic Flagellates*

Gut flagellates are perhaps the most visible members of the gut community in lower termites and wood-feeding roaches. They dominate  $\sim 90\%$  of hindgut volume, have striking morphologies, and interesting patterns of movement [(16) and references therein]. In addition, flagellates which are morphologically and phylogenetically similar to those in termites and roaches do not appear to exist anywhere else in nature (23, 55). As such, they have been the focus of over a century of study (74).

Taxonomic studies indicate gut flagellates represent over 400 different species of *Oxymonadida* and *Parabasalia* (23, 55, 62, 132). All but two species, *Trichonympha*

*termosidis* and *Trichonympha sphaerica* isolated by Yamin in the 1980s (129, 130), remain uncultured. Difficulties in isolation result from incomplete knowledge of the nutritional requirements for different protists — it is likely many obtain unidentified nutrients from other microbes. Indeed, prokaryotes are known to colonize protist cell surfaces as ectosymbionts and even exist inside protozoa as endosymbionts (27, 61, 62, 105). The recently published genome sequences of two endosymbionts, a putative nitrogen fixing *Bacteroidete* (53) and a member of the *Endomicrobia* (previously known as phylum TG1) thought to produce amino acids for their host protist (52), support the concept that termite gut protists form nutritional symbioses with other gut microbes. The associations between endosymbionts and their hosts appear to be fairly stable based on recent phylogenetic evidence indicating protist-endosymbiont cospeciation (88, 122).

### *Archaea*

Termite guts harbor significantly fewer Archaea than Bacteria (i.e., ~ 5% versus 95%) (13). In addition, Archaeal populations in wood- and litter-feeding termites are consistently lower than in soil-feeding termites (13). Phylogenetic analyses have indicated that Archaea are methanogens that belong to the genus *Methanobrevibacter* in the order *Methanobacteriales* (24). Only two termite gut methanogens (both *Methanobrevibacter sp.*) have been isolated to date (71, 72). These appear to be attached or in close proximity, to the gut wall of the lower termite *Reticulitermes* and are tolerant of microoxic levels of O<sub>2</sub> (71, 72). The latter finding is significant as all methanogens were once considered strict anaerobes. Non-methanogenic Archaea exist in termites (e.g., *Thermoplasmales* and *Crenarchaeota*) but have been much less studied (45, 113).

A recent review highlights developments in our understanding of these gut community members (24).

### *Bacteria*

Bacteria are, by far, the most abundant prokaryotes in termite guts (13). 16S rRNA gene studies have greatly aided efforts to define this population. Extensive surveys from the wood-feeding lower termite *Reticulitermes speratus* (50, 54) have shown that members of the phylum *Spirochaetes* are the most dominant (~ 50% of clones), followed by *Cytophaga-Flexibacter-Bacteroides* (CFB group, ~ 20%), low G+C *Firmicutes* (~ 15%), and *Endomicrobia* (~ 10%). The remaining ~ 5% of clones in these studies affiliated with *Proteobacteria*, *Actinobacteria*, *Mycoplasma* and other phyla. Practically all phylotypes recovered in 16S rRNA surveys represent uncultured species; many appear to be unique to termite gut environments.

Wood-feeding higher termite *Nasutitermes* (subfamily *Nasutitermitinae*) and *Microcerotermes* (subfamily *Termitinae*) bacterial communities have also been investigated with 16S rRNA gene inventory methods (51, 86). The prevalent gut bacteria are *Spirochaetes* (~ 60%), *Firmicutes* (~ 10%), and CFB group (~ 10%) bacteria, similar to lower termites. But two differences are worth noting: Higher termite guts lack *Endomicrobia* and harbor a new group of bacteria, the *Fibrobacters* (phylum TG3, ~10%). The absence of *Endomicrobia* is expected given the loss of cellulolytic gut protists by higher termites. The presence of fibrobacters in higher termites is more intriguing as the ruminant isolate *Fibrobacter succinogenes* is a well-known cellulolytic

bacterium (115). A recent metagenomic analysis of the gut community in a wood-feeding higher termite not only confirmed previous gene inventory studies, but also implicated fibrobacters and, surprisingly, spirochetes as functional replacements for cellulolytic flagellate protozoa (123).

Coevolution of termites with their gut bacterial communities has been another focus of exploration. Hongoh *et al.* (51) analysed bacterial 16S rRNA genes from 8 different species of Japanese *Reticulitermes* and *Microcerotermes* and showed that communities from termites of the same species are more similar to each other than to communities from termites of different species. This suggests some degree of host-symbiont coevolution (51); the extent of coevolution remains under debate. Lower and higher termite gut 16S rRNA sequences appear interspersed within bacteria phyla (e.g., *Spirochaetes*) suggesting symbiont phylogeny deviates from termite host phylogeny at family scales (5, 78, 93-95, 97). However, Berlanga *et al.* (5) argued against this interpretation with data that indicates 16S rRNA sequences from termites of the same lower termite family (*Kalotermitidae*) are more closely related to each other than to sequences from termites of a different lower termite family (*Rhinotermitidae*). A recent publication concluded the evolutionary history between ectosymbiotic *Bacteroidales* and gut protists has involved multiple instances of symbiont acquisition, suggesting the evolutionary history of termites and their gut symbionts is influenced by additional factors besides coevolution (87). More taxon sampling of bacteria and their hosts is needed for clarification.

Other factors like diet (e.g., soil versus wood) and gut physicochemistry (e.g., pH) also

impact community composition (12, 13, 86). With respect to diet, low G+C *Firmicutes* (~ 70%) are the dominant bacteria in soil-feeding termites but not wood-feeding higher termites, whose bacterial communities are instead dominated by spirochetes (24, 111). With respect to physicochemistry, the highly alkaline anterior gut sections (pH ~ 11, P1) of both wood- and soil- feeding termites are consistently dominated by firmicutes (110, 118); posterior hindgut sections (P3, P4) of soil-feeders are more circumneutral (pH ~ 7 – 10) and harbor CFB group, proteobacteria and spirochetes (111). Circumneutral P3 sections of wood-feeders are dominated by spirochetes. More details on termite gut microbial community structure can be found in several reviews (16, 17, 19, 24, 62).

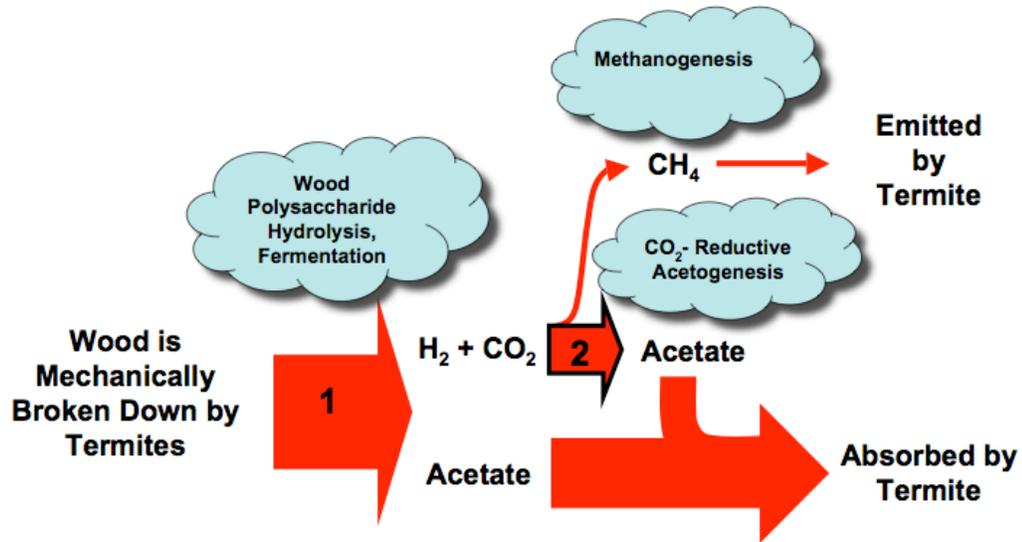
### **1.2.9. Termite gut nutritional ecology**

All termites engage in obligate nutritional symbioses with a dense and complex community of microbes in their hindguts (16, 19, 24, 62). Symbioses supporting the carbon, energy, and nitrogen metabolism of lignocellulose-feeding termites have been identified (24). These symbioses are not unexpected given the nutritional paucity of the host diet – lignocellulose lacks essential nutrients like amino acids and vitamins and has a C:N ratio 100-fold higher than insect tissue (70). Much less is known about the symbiotic relationships in soil-feeding termites, whose food substrates are very ill-defined (12).

As this thesis focuses on acetogenesis, a process mediating carbon and electron flow, I will only discuss organisms and processes related to lignocellulose degradation in termites. Details on nitrogen symbioses (e.g., N<sub>2</sub> fixation and uric acid recycling) can be found in references (16, 17, 19, 24, 62).

*Model of lignocellulose degradation*

Lignocellulose is a complex matrix comprised of three biopolymers: cellulose, hemicellulose, and lignin. The relative contribution of each polymer may vary with plant species and tissue, but lignocellulose generally contains 20 – 35% cellulose, 30% lignin, and 5% hemicellulose (82). Several studies have shown that termites and their gut communities only metabolize the cellulose and hemicellulosic fractions (16, 24, 56). Figure 1.3 depicts a schematic of the current model for wood degradation in termites. This degradation is stepwise: (1) termites increase wood particle surface area to volume ratio by maceration; (2) polysaccharides are hydrolyzed and then fermented, yielding the fermentation by-products  $H_2$ ,  $CO_2$ , and acetate; (3)  $H_2$  and  $CO_2$  are converted to additional acetate by acetogenic bacteria. Very little carbon and energy is lost from the system as methane; this has prompted some to consider termite guts as the most efficient bioreactors in nature (100).



**Figure 1.3.** General scheme outlining carbon and energy flow in wood-feeding termites. (Adapted from schematic by J. R. Leadbetter, personal communication). Step 1 results from the combined activities of host insect and cellulolytic protozoa in lower termites; spirochete and fibrobacter bacteria are implicated in step 1 in higher termites (123). Spirochetes dominate acetogenesis (step 2) in lower and higher wood-feeding termites (99, 103, 123). Acetogenesis generates up to  $\sim 1/3$  of gut acetate (21), which can accumulate up to 80 mM in the gut and support up to 100% of termite respiratory metabolism (92).

### *Role of termites*

Termites contribute to lignocellulose degradation by providing finely macerated wood particles with increased surface area to aid their symbionts in their degradative activities. However, the role of termite-derived cellulase enzymes is more debatable. Termites encode endoglucanase genes which, when expressed, have hydrolyzing activity on crystalline cellulose (120). However, the site of expression (salivary glands and midgut) argues against a driving role for the termite host in cellulose hydrolysis and fermentation, since these processes are thought to occur in the hindgut paunch (121, 133).

*Role of flagellate protozoa*

The role of flagellate protozoa in termite nutrition has been the subject of study since the 1920s. The first studies, performed by Cleveland (28-30), demonstrated that the selective removal of anaerobic protozoa with hyperbaric O<sub>2</sub> correlates with low survival rates of termites fed wood and cellulose. This was the first hard evidence that flagellate protozoa play a fundamental role in the wood-feeding ability of termites and roaches. Later, Hungate reported the biochemical basis for the termite's dependence on protozoa is protozoal depolymerization of cellulose into glycosyl units and the subsequent fermentation of these units into short chain fatty acids, which could be absorbed by the host for its metabolism (58-60). Kovoov then identified acetate as the major short chain fatty acid in termite guts [(68), reviewed by (15)]. Measurements of cellulose fermentation stoichiometries in axenic cultures of termite gut protists by Yamin and colleagues (91, 129-132) were consistent with Hungate's calculations.

Odelsen and Breznak's study of volatile fatty acid (VFA) production in wood-feeding termite gut homogenates using gas chromatography/mass spectroscopy confirmed acetate as the dominant gut short chain fatty acid and the major fuel for termite respiration (acetate supported 77-100% of *Reticulitermes flavipes* respiration) (92). Their results supported the idea that protists are the major gut acetate producers. But Odelsen and Breznak's most striking finding was that the concentration of acetate was lower in the guts of antibiotic dosed termites than the control group. This was the first evidence implicating bacteria in the carbon and energy nutrition of their host.

*Role of acetogenic bacteria*

A landmark study by Breznak and Switzer in 1986 (21) firmly supported Odelsen and Breznak's hypothesis that bacteria contribute to acetate production (and, hence, termite nutrition) via acetogenesis from  $H_2 + CO_2$ . More significantly, Breznak and Switzer showed acetogenesis rates in termite guts are high enough to impact the global carbon cycle. Using  $^{14}C$  to trace carbon and electron flows in gut homogenates, they demonstrated that the dominant reductant consuming process in the guts of wood-feeding termites is  $CO_2$ -reductive acetogenesis, rather than methanogenesis. A later study by Brauman *et al.* (14), which reports acetogenesis and methanogenesis rates in termites with different feeding habits, confirmed Breznak and Switzer's observations and extended the dominance of acetogenesis to grass-feeding termites. This latter data set indicated that acetogenesis rates could be 15 to 20-fold higher in wood- and grass-feeding termites than soil- and fungus-feeding termites.

Recently, Pester *et al.* (98) showed that  $H_2$  is the central free intermediate in lignocellulose degradation using hydrogen microsensors to infer  $H_2$  flux in three lower termite species. Their measurements indicated termite guts are characterized by high concentrations of  $H_2$  ( $\sim 70$  kPa) and rapid turnover, with little  $H_2$  loss from the system. In addition, microinjections of  $NaH^{14}CO_2$  into intact hindguts revealed methane emission accounts for only 4% of respiratory electron flow, whereas acetogenesis corresponds to  $\sim 20\%$ . Their results were consistent with previous estimates (based on gut homogenates) that  $CO_2$ -reductive acetogenesis can fuel up to  $\sim 30\%$  of termite respiration (21).

The finding that acetogens outcompete methanogens for H<sub>2</sub> in termite guts remains curious. Methanogenesis is predicted to dominate acetogenesis as an electron sink in anaerobic environments, based on its energetic favorability (methanogenesis  $\Delta G^{0'}$  = -136 kJ/mol, acetogenesis  $\Delta G^{0'}$  = -105 kJ/mol), and typically does so in sulfate depleted sediments and the ruminant gut. More study is needed to elucidate reasons underlying robust acetogen and meager methanogen populations in termite guts.

### 1.2.10. Termite gut acetogens

#### *Firmicutes and Spirochaetes isolates*

Only seven termite gut acetogens have been isolated to date. Almost all isolates, at the time, were considered new bacterial species. The first gut acetogen isolate, *Sporomusa termitidae*, was obtained by Breznak and colleagues (22) from the wood-feeding higher termite *Nasutitermes nigriceps*. Later isolates (and their origins) include: *Acetonema longum* (lower termite) (66), *Clostridium mayombei* (soil-feeding higher termite)(65), *Sporomusa termitidis* (wood-feeding higher termite) (47), *Sporomusa aerivorans* (soil-feeding higher termite) (11), *Treponema primitia* str. ZAS-1 (lower termite) (73), and *T. primitia* str. ZAS-2 (lower termite) (73). Most of these isolates can utilize carbohydrate-derived reductants for acetogenesis in addition to H<sub>2</sub>; some are capable of mixotrophic growth wherein H<sub>2</sub> + CO<sub>2</sub> and organic substrates are utilized simultaneously for catabolism and anabolism (18, 46). Of the 7 isolates, strains ZAS-1 and ZAS-2 are the only bacteria that are not *Firmicutes*. To this day, they remain the sole examples of chemolithoautotrophy in the phylum *Spirochaetes*.

*“Culture-independent” surveys of acetogens implicate spirochetes*

The abundance and diversity of spirochetes in termite guts led investigators to speculate that spirochetes might be the dominant acetogens in these environments. Leadbetter’s isolation of two acetogenic spirochetes provided the first concrete evidence moving this speculation into the realm of hypothesis (73). However, the in situ relevance of microbial isolates is typically questionable, since major cultivation biases limit isolate studies to those bacteria that manage to thrive in a given enrichment medium. The predictive capacity of pure culture studies, therefore, needs to be corroborated with “culture-independent” molecular profiling techniques, which can survey the entire gut community. Salmassi *et al.* (103) used such an approach to survey the acetogen population inhabiting the guts of the wood-feeding lower termite, *Zootermopsis nevadensis*. They analyzed clone libraries of the Wood-Ljungdahl pathway marker gene, *fhs*, which encodes formyl-tetrahydrofolate synthetase (FTHFS) (Figure 1.2), and discovered that the majority of sequences phylogenetically affiliated with the FTHFS from the spirochete, *T. primitia*. This finding was the first solid molecular evidence supporting the hypothesis that spirochetes are the dominant acetogens in termite guts. Pester and Brune’s survey of expressed *fhs* genes in lower termite guts lent additional support to the concept that spirochetes are the major acetogens in wood-feeding termites (99).

**1.2.11. *Treponema primitia* str. ZAS-2 formate dehydrogenases**

Gene surveys of 16S rRNA and FTHFS prompted a recent study by Matson *et al.* (85) of the Wood-Ljungdahl pathway in *T. primitia* str. ZAS-2, in which they report the

discovery of two genes for the CO<sub>2</sub>-fixing Wood-Ljungdahl enzyme, formate dehydrogenase (FDH). This finding is significant for two reasons:

First, FDH genes in *T. primitia* strain ZAS-2 are only distantly related to FDH in the model acetogen, *M. thermoacetica*. FDH genes in *T. primitia* phylogenetically group with genes for hydrogenase-linked FDH (FDH<sub>H</sub>, *fdhF*) in enteric *Gammaproteobacteria*, which utilize FDH<sub>H</sub> for carbohydrate fermentation (43). This is highly unexpected as no Wood-Ljungdahl pathway enzymes in *M. thermoacetica* (or any other acetogen) are known to be hydrogenase-linked (i.e., associated in a multi-enzyme complex like FDH<sub>H</sub> in the *Escherichia coli* formate hydrogenase lyase complex) (36). Indeed, the canonical FDH in *Moorella* is a NADPH-linked enzyme that only interacts indirectly with hydrogenase (128). The identification of hydrogenase-linked FDH genes in *T. primitia* therefore suggests that the current model of acetogenesis based *M. thermoacetica* may not have relevance in termite guts, where acetogenesis rates translate into globally relevant carbon fluxes.

Second, the two genes encode selenium-dependent (Sec) and selenium-independent (Cys) FDH<sub>H</sub> homologs. This finding is the first indication that the redox active trace element selenium may influence *T. primitia*'s energy metabolism and physiological ecology. The selenium-dependant FDH<sub>H</sub> gene contains an in-frame TGA stop codon encoding the non-canonical amino acid, selenocysteine, at the enzyme active site. In contrast the other gene encodes the amino acid cysteine at the corresponding catalytically relevant position. Studies on selenocysteine enzymes show that they are more catalytically active than their

cysteine homologs (2). This led Matson *et al.* (85) to hypothesize that *T. primitia* preferentially utilizes its selenium-dependent FDH<sub>H</sub> when selenium is replete and switches to its selenium independent variant as a “back up” when selenium is scarce. They tested this hypothesis with quantitative RT-PCR and confirmed that transcription of the two gene variants varied with selenium concentration in the predicted directions (i.e., Sec variant transcription increased and Cys variant transcription decreased with the addition of selenium). Matson *et al.* (85) therefore posited that selenium may influence the genome content and physiological ecology of uncultured termite gut acetogens. My work on uncultured termite gut acetogens stems from this hypothesis and is presented in Chapters 2 – 4 of this thesis.

#### **1.2.12. Overview of chapters 2 – 4**

My work aims at understanding the diversity, evolution, and activity of termite gut acetogens at a functional gene level. I have employed traditional gene inventory, novel sequencing, and single cell techniques to:

- (i) assess whether hydrogenase-linked FDH genes (*fdhF*) are relevant for CO<sub>2</sub>-reductive acetogenesis in the guts of taxonomically and nutritionally diverse termites,
- (ii) determine whether Sec and Cys forms of *fdhF* have relevance in uncultured termite gut microbes,
- (iii) discover uncultured acetogenic spirochetes which encode both Sec and Cys *fdhF* like *T. primitia*, and

- (iv) identify microbes whose activity dominates *fdhF* transcription in gut communities.

In Chapter 2, I present work related to degenerate *fdhF* primer design and gene inventory analysis from three species of phylogenetically lower termite and a wood-feeding roach (objectives i and ii).

In Chapter 3, I discuss surveys of *fdhF* diversity in taxonomically and nutritionally diverse higher termites and compare *fdhF* phylogeny in higher and lower termites (objectives i and ii).

In Chapter 4, I demonstrate that high-throughput sequencing of community mRNA can be leveraged by gene inventory and pure culture data to identify the major transcriptionally active *fdhF* phylotypes in termite gut microbial communities (objective iv). I also show that microfluidic multiplex digital PCR can be used to discover the 16S rRNA identity of transcriptionally active uncultured microbes as well as those that encode both Sec and Cys gene variants (objective iii).

### **1.3. Introduction to bacterial lipid D/H and metabolism**

Compound specific isotope analysis is a powerful tool for identifying sources and inferring processes in the environment (42). Accordingly, compound-specific approaches have gained popularity in a wide variety of fields; these include organic geochemistry, paleoclimate, bioremediation science, and archeology [(42, 108) and references therein]. Many of these studies have relied on stable isotopes of carbon, but recent instrumental developments now allow the stable isotopes of hydrogen to be measured easily and accurately (112).

Compound specific stable hydrogen isotope ratios (D/H) have already proved useful in environmental science. For example, D/H of lacustrine sedimentary lipids have been used to infer D/H of environmental water and reconstruct the geochemistry of past environments (106). Measurements of lipid D/H in marine sediments have also been made but are more difficult to interpret (63, 75). This primarily stems from our limited mechanistic understanding of the factors and processes underlying D/H signals in organic matter. My work seeks to clarify the biological determinants of lipid D/H in microbes, organisms which account for ~ 50% of all living biomass on the planet (124).

#### **Overview of chapter 5**

In the last chapter of this thesis, I present my investigation of the relationship between metabolism and lipid D/H in physiologically diverse bacteria. I show evidence that lipid D/H varies systematically with energy metabolism and propose a biological basis for lipid D/H variations.

## 1.4. References

1. **Allen, E. E., and J. F. Banfield.** 2005. Community genomics in microbial ecology and evolution. *Nature Rev Microbiol* **3**:489-498.
2. **Axley, M. J., A. Böck, and T. C. Stadtman.** 1991. Catalytic properties of an *Escherichia coli* formate dehydrogenase mutant in which sulfur replaces selenium. *Proc Natl Acad Sci U S A* **88**:8450-8454.
3. **Barker, H. A.** 1944. On the role of carbon dioxide in the metabolism of *Clostridium thermoaceticum*. *Proc Natl Acad Sci U S A* **30**:88-90.
4. **Barker, H. A., and M. D. Kamen.** 1945. Carbon dioxide utilization in the synthesis of acetic acid by *Clostridium thermoaceticum*. *Proc Natl Acad Sci U S A* **31**:219-225.
5. **Berlanga, M., B. J. Paster, and R. Guerrero.** 2007. Coevolution of symbiotic spirochete diversity in lower termites. *Int Microbiol* **10**:133-9.
6. **Bignell, D. E.** 1984. Direct potentiometric determination of redox potentials of the gut contents in the termites *Zootermopsis nevadensis* and *Cubitermes severus* and in three other arthropods. *J Insect Physiol* **30**:167-174.
7. **Bignell, D. E.** 2006. Termites as soil engineers and soil processors, p. 183-220. *In* H. König and A. Varma (ed.), *Intestinal microorganisms of termites and other invertebrates*. Springer, Heidelberg, Germany.
8. **Bignell, D. E., and J. M. Anderson.** 1980. Determination of pH and oxygen status in the guts of lower and higher termites. *J Insect Physiol* **26**:183-188.
9. **Bignell, D. E., and P. Eggleton.** 2000. Termites in Ecosystems, p. 363-387. *In* T. Abe, D. E. Bignell, and M. Higashi (ed.), *Termites: Evolution, Sociality, Symbioses, Ecology*. Springer, Dordrecht, The Netherlands.
10. **Bignell, D. E., P. Eggleton, K. Thomas, and L. Nunes.** 1997. Termites as mediators of carbon fluxes in tropical forest: budgets for carbon dioxide and methane emissions, p. 119-134. *In* A. Watt, N. E. Stork, and M. Hunter (ed.), *Insects and Forests*. Chapman and Hall, London.
11. **Boga, H. I., W. Ludwig, and A. Brune.** 2003. *Sporomusa aerivorans* sp. nov., an oxygen-reducing homoacetogenic bacterium from the gut of a soil-feeding termite. *Int J Syst Evol Microbiol* **53**:1397-1404.
12. **Brauman, A., D. E. Bignell, and I. Tayasu.** 2000. Soil-feeding termites: biology, microbial associations and digestive mechanisms, p. 233-259. *In* T. Abe, D. E. Bignell, and M. Higashi (ed.), *Termites: Evolution, Sociality, Symbioses, Ecology*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
13. **Brauman, A., J. Dore, P. Eggleton, D. Bignell, J. A. Breznak, and M. D. Kane.** 2001. Molecular phylogenetic profiling of prokaryotic communities in guts of termites with different feeding habits. *FEMS Microbiol Ecol* **35**:27-36.
14. **Brauman, A., M. D. Kane, M. Labat, and J. A. Breznak.** 1992. Genesis of Acetate and Methane by Gut Bacteria of Nutritionally Diverse Termites. *Science* **257**:1384-1387.
15. **Breznak, J. A.** 2000. Acetogenesis from carbon dioxide in termite guts, p. 303-330. *In* H. L. Drake (ed.), *Acetogenesis*. Chapman and Hall, New York, NY.
16. **Breznak, J. A.** 2000. Ecology of prokaryotic microbes in the guts of wood-and litter-feeding termites, p. 209-231. *In* T. Abe, D. E. Bignell, and M. Higashi (ed.),

- Termites: Evolution, Sociality, Symbiosis, Ecology Kluwer Academic Publishers Dordrecht, The Netherlands.
17. **Breznak, J. A.** 1982. Intestinal microbiota of termites and other xylophagous insects. *Annu Rev Microbiol* **36**:323-43.
  18. **Breznak, J. A., and J. S. Blum.** 1991. Mixotrophy in the termite gut acetogen, *Sporomusa termitida*. *Arch Microbiol* **156**:105-110.
  19. **Breznak, J. A., and A. Brune.** 1994. Role of microorganisms in the digestion of lignocellulose by termites. *Ann Rev Entomol* **39**:453-487.
  20. **Breznak, J. A., and M. D. Kane.** 1990. Microbial H<sub>2</sub>/CO<sub>2</sub> acetogenesis in animal guts: nature and nutritional significance. *FEMS Microbiol Rev* **7**:309-13.
  21. **Breznak, J. A., and J. M. Switzer.** 1986. Acetate synthesis from H<sub>2</sub> plus CO<sub>2</sub> by termite gut microbes. *Appl Environ Microbiol* **52**:623-630.
  22. **Breznak, J. A., J. M. Switzer, and H. J. Seitz.** 1988. *Sporomusa termitida* sp. nov., an H<sub>2</sub> CO<sub>2</sub> utilizing acetogen isolated from termites. *Arch Microbiol* **150**:282-288.
  23. **Brugerolle, G., and R. Radek.** 2006. Symbiotic Protozoa of Termites, p. 244-269. *In* H. König and A. Varma (ed.), *Intestinal microorganisms of termites and other invertebrates*. Springer, Heidelberg, Germany.
  24. **Brune, A.** 2006. Symbiotic associations between termites and prokaryotes, p. 439-474. *In* M. Dworkin, S. Falkow, E. Rosenber, K. H. Schleifer, and E. Stackebrandt (ed.), *The Prokaryotes*, 3 ed, vol. 1. Springer, New York.
  25. **Brune, A., D. Emerson, and J. A. Breznak.** 1995. The termite gut microflora as an oxygen sink: microelectrode determination of oxygen and pH gradients in guts of lower and higher termites. *Appl Environ Microbiol* **61**:2681-2687.
  26. **Brune, A., and M. W. Friedrich.** 2000. Microecology of the termite gut: Structure and function on a microscale. *Curr Opin Microbiol* **3**:263-269.
  27. **Brune, A., and U. Stingl.** 2006. Prokaryotic symbionts of termite gut flagellates: phylogenetic and metabolic implications of a tripartite symbiosis. *Prog Mol Subcell Biol* **41**:39-60.
  28. **Cleveland, L. R.** 1923. Correlation between the food and morphology of termites and the presence of intestinal protozoa. *Am J Epidemiol* **3**:444-461.
  29. **Cleveland, L. R.** 1925. The effects of oxygenation and starvation on the symbiosis between the termite, *Termopsis*, and its intestinal flagellates. *Biol Bull* **48**:309-312.
  30. **Cleveland, L. R.** 1925. Toxicity of oxygen for protozoa in vivo and in vitro: Animals defaunated without injury. *Biol Bull* **48**:455-468.
  31. **Daniel, S. L., T. Hsu, S. I. Dean, and H. L. Drake.** 1990. Characterization of the H<sub>2</sub>- and CO-dependent chemolithotrophic potentials of the acetogens *Clostridium thermoaceticum* and *Acetogenium kivui*. *J Bacteriol* **172**:4464-4471.
  32. **Diekert, G., and G. Wohlfarth.** 1994. Energetics of acetogenesis from C1 units, p. 157-179. *In* H. L. Drake (ed.), *Acetogenesis*. Chapman and Hall, New York, NY.
  33. **Diekert, G., and G. Wohlfarth.** 1994. Metabolism of homocetogens. *Antonie van Leeuwenhoek* **66**:209-21.
  34. **Drake, H., A. Gossner, and S. Daniel.** 2007. Old Acetogens, New Light. *Ann N Y Acad Sci* **1125**:100-128.

35. **Drake, H., K. Küsel, and C. Matthies.** 2002. Ecological consequences of the phylogenetic and physiological diversities of acetogens. *Antonie van Leeuwenhoek* **81**:203-213.
36. **Drake, H. L.** 1994. Introduction to Acetogenesis, p. 3-60. *In* H. L. Drake (ed.), *Acetogenesis*. Chapman and Hall, New York.
37. **Drake, H. L.** 2004. Physiology of the thermophilic acetogen *Moorella thermoacetica*. *Res Microbiol* **155**:422-436.
38. **Drake, H. L., K. Küsel, and C. Matthies.** 2006. Acetogenic prokaryotes, p. 354-420. *In* M. Dworkin, S. Falkow, E. Rosenber, K. H. Schleifer, and E. Stackebrandt (ed.), *The Prokaryotes*, 3 ed. Springer.
39. **Dumont, M. G., and J. C. Murrell.** 2005. Stable isotope probing - linking microbial identity to function. *Nature Rev Microbiol* **3**:499-504.
40. **Eggleton, P.** 2000. Global patterns of termite diversity, p. 25-51. *In* T. Abe, D. E. Bignell, and M. Higashi (ed.), *Termites: Evolution, Sociality, Symbioses, Ecology*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
41. **Engel, M. S., D. A. Grimaldi, and K. Krishna.** 2009. Termites (*Isoptera*): Their Phylogeny, Classification, and Rise to Ecological Dominance. *American Museum Novitates* **3650**:1-27.
42. **Evershed, R. P., I. D. Bull, J. T. Corr, Z. M. Crossman, B. E. Van Dongen, C. J. Evans, S. Jim, H. R. Mottram, A. J. Mukherjee, and R. D. Pancost.** 2007. Compound-specific stable isotope analysis in ecology and paleoecology, p. 480-540. *In* R. Michener and K. Lajtha (ed.), *Stable Isotopes in Ecology and Environmental Science*. Blackwell Publishing, Malden, MA.
43. **Ferry, J. G.** 1990. Formate Dehydrogenase: Microbiology, Biochemistry and Genetics, p. 117-141. *In* G. A. Codd, L. Dijkhuizen, and F. R. Tabita (ed.), *Autotrophic Microbiology and One Carbon Metabolism*. Kluwer Academic Publishers, Dordrecht.
44. **Fontaine, F. E., W. H. Peterson, E. McCoy, M. J. Johnson, and G. J. Ritter.** 1942. A new type of glucose fermentation by *Clostridium thermoaceticum*. *J Bacteriol* **43**:701-715.
45. **Friedrich, M. W., D. Schmitt-Wagner, T. Lueders, and A. Brune.** 2001. Axial differences in community structure of *Crenarchaeota* and *Euryarchaeota* in the highly compartmentalized gut of the soil-feeding termite *Cubitermes orthognathus*. *Appl Environ Microbiol* **67**:4880-4890.
46. **Graber, J. R., and J. A. Breznak.** 2004. Physiology and nutrition of *Treponema primitia*, an H<sub>2</sub>/CO<sub>2</sub>-acetogenic spirochete from termite hindguts. *Appl Environ Microbiol* **70**:1307-1314.
47. **Grech-Mora, I., M.-L. Fardeau, B. K. C. Patel, B. Ollivier, A. Rimbault, G. Prensier, J.-L. Garcia, and E. Garnier-Sillam.** 1996. Isolation and Characterization of *Sporobacter termitidis* gen. nov., sp. nov., from the Digestive Tract of the Wood-Feeding Termite *Nasutitermes lujae*. *Int J Syst Bacteriol* **46**:512-518.
48. **Grimaldi, D., and M. S. Engel.** 2005. *Evolution of the Insects*. Cambridge University Press, New York, NY.

49. **Hoehler, T. M., D. B. Albert, M. J. Alperin, and C. S. Martens.** 1999. Acetogenesis from CO<sub>2</sub> in an anoxic marine sediment. *Limnol Oceanogr* **44**:662-667.
50. **Hongoh, Y.** 2003. Molecular analysis of bacterial microbiota in the gut of the termite *Reticulitermes speratus* (Isoptera; *Rhinotermitidae*). *FEMS Microbiol Ecol* **44**:231-242.
51. **Hongoh, Y., P. Deevong, T. Inoue, S. Moriya, S. Trakulnaleamsai, M. Ohkuma, C. Vongkaluang, N. Noparatnaraporn, and T. Kudo.** 2005. Intra- and Interspecific Comparisons of Bacterial Diversity and Community Structure Support Coevolution of Gut Microbiota and Termite Host. *Appl Environ Microbiol* **71**:6590-6599.
52. **Hongoh, Y., V. K. Sharma, T. Prakash, S. Noda, T. D. Taylor, T. Kudo, Y. Sakaki, A. Toyoda, M. Hattori, and M. Ohkuma.** 2008. Complete genome of the uncultured Termite Group 1 bacteria in a single host protist cell. *Proc Natl Acad Sci USA* **105**:5555-5560.
53. **Hongoh, Y., V. K. Sharma, T. Prakash, S. Noda, H. Toh, T. D. Taylor, T. Kudo, Y. Sakaki, A. Toyoda, M. Hattori, and M. Ohkuma.** 2008. Genome of an Endosymbiont Coupling N<sub>2</sub> Fixation to Cellulolysis Within Protist Cells in Termite Gut. *Science* **322**:1108-1109.
54. **Hongoh, Y., H. Yuzawa, M. Ohkuma, and T. Kudo.** 2003. Evaluation of primers and PCR conditions for the analysis of 16S rRNA genes from a natural environment. *FEMS Microbiol Lett* **221**:299-304.
55. **Honigberg, B. M.** 1970. Protozoa associated with termites and their role in digestion, p. 1-36. *In* K. Krishna and F. M. Weesner (ed.), *Biology of Termites*, vol. 2. Academic Press, New York.
56. **Hopkins, D. W., J. A. Chudek, D. E. Bignell, J. Frouz, E. A. Webster, and T. Lawson.** 1998. Application of <sup>13</sup>C NMR to investigate the transformations and biodegradation of organic materials by wood- and soil-feeding termites, and a coprophagous litter-dwelling dipteran larva. *Biodegradation* **9**:423-431.
57. **Hugenholtz, P., B. M. Goebel, and N. Pace.** 1998. Impact of Culture-Independent Studies on the Emerging Phylogenetic View of Bacterial Diversity. *J Bacteriol* **180**: 4765-4774.
58. **Hungate, R. E.** 1939. Experiments on the nutrition of *Zootermopsis*. III: The anaerobic carbohydrate dissimilation by the intestinal protozoa. *Ecology* **20**:230-245.
59. **Hungate, R. E.** 1955. Mutualistic intestinal protozoa, p. 159-199. *In* S. H. Hutner and A. Lwoff (ed.), *Biochemistry and Physiology of Protozoa*. Academic Press New York, NY.
60. **Hungate, R. E.** 1943. Quantitative analyses of the cellulose fermentation by termite protozoa. *Annals of the Entomological Society of America* **36**:730-739.
61. **Inoue, J. I., S. Noda, Y. Hongoh, S. Ui, and M. Ohkuma.** 2008. Identification of Endosymbiotic Methanogen and Ectosymbiotic Spirochetes of Gut Protists of the Termite *Coptotermes formosanus*. *Microbes Environment* **23**:94-97.
62. **Inoue, T., O. Kitade, T. Yoshimura, and I. Yamaoka.** 2000. Symbiotic association with protists, p. 275-288. *In* T. Abe, D. E. Bignell, and M. Higashi

- (ed.), *Termites: Evolution, Sociality, Symbioses*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
63. **Jones, A., A. L. Sessions, B. Campbell, C. Li, and D. Valentine.** 2008. D/H ratios of fatty acids from marine particulate organic matter in the California Borderland Basins. *Org Geochem* **39**:485-500.
  64. **Kambhampati, S., and P. Eggleton.** 2000. Taxonomy and phylogeny of termites, p. 1-24. *In* T. Abe, D. E. Bignell, and M. Higashi (ed.), *Termites: Evolution, Sociality, Symbioses, Ecology*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
  65. **Kane, M. D., A. Brauman, and J. A. Breznak.** 1991. *Clostridium mayombei* sp. nov., an H<sub>2</sub>/CO<sub>2</sub> acetogenic bacterium from the gut of the African soil-feeding termite, *Cubitermes speciosus*. *Arch Microbiol* **156**:99-104.
  66. **Kane, M. D., and J. A. Breznak.** 1991. *Acetonema longum* gen. nov. sp. nov., an H<sub>2</sub>/CO<sub>2</sub> acetogenic bacterium from the termite, *Pterotermes occidentis*. *Arch Microbiol* **156**:91-8.
  67. **Khalil, M. A. K., R. A. Rasmussen, J. R. J. French, and J. A. Holt.** 1989. The Influence of Termites on Atmospheric Trace Gases: CH<sub>4</sub>, CO<sub>2</sub>, CHCl<sub>3</sub>, N<sub>2</sub>O, CO, H<sub>2</sub>, and Light Hydrocarbons. *J Geophys Res* **95**:3619–3634.
  68. **Kovoor, J.** 1968. L'intestin d'un termite supérieur (*Microcerotermes edentatus*, Wasman, *Amitermitinae*). *Histophysiologie et flore bacterienne symbiotique*. *Bull Biol Fr Belg* **102**:45–84.
  69. **Krishna, K.** 1970. Taxonomy, phylogeny, and distribution of termites, p. 127-152. *In* K. Krishna and F. M. Weesner (ed.), *Biology of Termites*. Academic Press, New York.
  70. **LaFage, J. P., and W. L. Nutting.** 1978. Nutrient dynamics of termites. *In* M. V. Brian (ed.), *Production ecology of ants and termites*. Cambridge University Press.
  71. **Leadbetter, J. R., and J. A. Breznak.** 1996. Physiological ecology of *Methanobrevibacter cuticularis* sp. nov. and *Methanobrevibacter curvatus* sp. nov., isolated from the hindgut of the termite *Reticulitermes flavipes*. *Appl Environ Microbiol* **62**:3620–3631.
  72. **Leadbetter, J. R., L. D. Crosby, and J. A. Breznak.** 1998. *Methanobrevibacter filiformis* sp. nov., A filamentous methanogen from termite hindguts. *Arch Microbiol* **169**:287-92.
  73. **Leadbetter, J. R., T. M. Schmidt, J. R. Graber, and J. A. Breznak.** 1999. Acetogenesis from H<sub>2</sub> plus CO<sub>2</sub> by spirochetes from termite guts. *Science* **283**:686-689.
  74. **Leidy, J.** 1881. The parasites of the termites. *J Acad Nat Sci (Phila) 2nd Ser* **8**:425–447.
  75. **Li, C., A. L. Sessions, F. S. Kinnaman, and D. L. Valentine.** 2009. Hydrogen-isotopic variability in lipids from Santa Barbara Basin sediments. *Geochim Cosmochim Acta* **73**:4803-4823.
  76. **Liesack, W., F. Bak, J. U. Kreft, and E. Stackebrandt.** 1994. *Holophaga foetida* gen. nov., sp. nov., a new, homoacetogenic bacterium degrading methoxylated aromatic compounds. *Arch Microbiol* **162**:85-90.

77. **Lighton, J. R. B., and E. A. Ottesen.** 2005. To DGC or not to DGC: oxygen guarding in the termite *Zootermopsis nevadensis* (Isoptera: *Termopsidae*). *J Exp Biol* **208**:4671-4678.
78. **Lilburn, T. G., T. M. Schmidt, and J. A. Breznak.** 1999. Phylogenetic diversity of termite gut spirochaetes. *Environ Microbiol* **1**:331-45.
79. **Ljungdahl, L. G., and H. G. Wood.** 1969. Total Synthesis of Acetate From CO<sub>2</sub> by Heterotrophic Bacteria. *Annu Rev Microbiol* **23**:515-538.
80. **Lo, N., G. Tokuda, H. Watanabe, H. Rose, M. Slaytor, K. Maekawa, C. Bandi, and H. Noda.** 2000. Evidence from multiple gene sequences indicates that termites evolved from wood-feeding cockroaches. *Curr Biol* **10**:801-804
81. **Lovley, D. R., and M. J. Klug.** 1983. Methanogenesis from Methanol and Methylamines and Acetogenesis from Hydrogen and Carbon Dioxide in the Sediments of a Eutrophic Lake. *Appl Environ Microbiol* **45**:1310-1315.
82. **Lynd, L. R., P. J. Weimer, W. H. van Zyl, and I. S. Pretorius.** 2002. Microbial Cellulose Utilization: Fundamentals and Biotechnology. *Microbiol Mol Biol Rev* **66**:506-577.
83. **Martin, W., J. Baross, D. Kelley, and M. J. Russell.** 2008. Hydrothermal vents and the origin of life. *Nat Rev Micro* **6**:805-814.
84. **Martius, C.** 1994. Diversity and ecology of termites in Amazonian forests. *Pedobiologia* **38**:407-428.
85. **Matson, E. G., X. Zhang, and J. R. Leadbetter.** 2010. Selenium controls expression of paralogous formate dehydrogenases in the termite gut homoacetogen *Treponema primitia*. *Environ Microbiol* *Accepted*.
86. **Miyata, R., N. Noda, H. Tamaki, K. Kinjyo, H. Aoyagi, H. Uchiyama, and H. Tanaka.** 2007. Influence of Feed Components on Symbiotic Bacterial Community Structure in the Gut of the Wood-Feeding Higher Termite *Nasutitermes takasagoensis*. *Biosci Biotechnol Biochem* **71**:1244-1251.
87. **Noda, S., Y. Hongoh, T. Sato, and M. Ohkuma.** 2009. Complex coevolutionary history of symbiotic *Bacteroidales* bacteria of various protists in the gut of termites. *BMC Evolutionary Biology* **9**:158.
88. **Noda, S., O. Kitade, T. Inoue, M. Kawai, and M. Kanuka.** 2007. Cospeciation in the triplex symbiosis of termite gut protists (*Pseudotriconympha* spp.), their hosts, and their bacterial endosymbionts. *Molecular Ecology* **16**:1257 - 1266.
89. **Noirot, C.** 1995. The gut of termites (*Isoptera*) comparative anatomy, systematics, phylogeny. I. Lower termites. *Ann Soc Entomol Fr* **31**:197-226.
90. **Noirot, C.** 2001. The gut of termites (*Isoptera*), comparative anatomy, systematics, phylogeny. II: Higher termites (*termitidae*). *Ann Soc Entomol Fr* **37**:431-471.
91. **Odelson, D. A., and J. A. Breznak.** 1985. Nutrition and growth characteristics of *Trichomitopsis termopsidis*, a cellulolytic protozoan from termites. *Appl Environ Microbiol* **49**:614-621.
92. **Odelson, D. A., and J. A. Breznak.** 1983. Volatile fatty acid production by the hindgut microbiota of xylophagous termites. *Appl Environ Microbiol* **45**:1602-1613.
93. **Ohkuma, M., Y. Hongoh, and T. Kudo.** 2006. Diversity and Molecular Analyses of Yet-Uncultivated Microorganisms. *In* H. König and A. Varma (ed.),

- Intestinal microorganisms of termites and other invertebrates. Springer, Heidelberg, Germany.
94. **Ohkuma, M., T. Iida, and T. Kudo.** 1999. Phylogenetic relationships of symbiotic spirochetes in the gut of diverse termites. *FEMS Microbiol Lett* **181**:123–129.
  95. **Ohkuma, M., and T. Kudo.** 1998. Phylogenetic analysis of the symbiotic intestinal microflora of the termite *Cryptotermes domesticus*. *FEMS Microbiol Lett* **164**:389–395.
  96. **Ottesen, E. A.** 2008. The Biology and Community Structure of CO<sub>2</sub>-Reducing Acetogens in the Termite Hindgut. Ph.D. dissertation. California Institute of Technology, Pasadena.
  97. **Paster, B. J., F. E. Dewhirst, S. M. Cooke, V. Fusing, L. K. Poulsen, and J. A. Breznak.** 1996. Phylogeny of not-yet-cultured spirochetes from termite guts. *Appl Environ Microbiol* **62**:347-52.
  98. **Pester, M.** 2006. Hydrogen metabolism in the hindgut of lower termites: Fluxes of hydrogen-dependent and related processes and identification of the homoacetogenic microbiota. Ph.D. dissertation. Philipps-Universität Marburg, Marburg, Germany.
  99. **Pester, M., and A. Brune.** 2006. Expression profiles of *fhs* (FTHFS) genes support the hypothesis that spirochaetes dominate reductive acetogenesis in the hindgut of lower termites. *Environ Microbiol* **8**:1261-1270.
  100. **Pester, M., and A. Brune.** 2007. Hydrogen is the central free intermediate during lignocellulose degradation by termite gut symbionts. *ISME J* **1**:551-65.
  101. **Ragsdale, S. W., J. E. Clark, L. G. Ljungdahl, L. L. Lundie, and H. L. Drake.** 1983. Properties of purified carbon monoxide dehydrogenase from *Clostridium thermoaceticum*, a nickel, iron-sulfur protein. *J Biol Chem* **258**:2364-9.
  102. **Rasmussen, R. A., and M. A. K. Khalil.** 1983. Global production of methane by termites. *Nature* **301**:704–705.
  103. **Salmassi, T. M., and J. R. Leadbetter.** 2003. Analysis of genes of tetrahydrofolate-dependent metabolism from cultivated spirochaetes and the gut community of the termite *Zootermopsis angusticollis*. *Microbiology* **149**:2529–2537.
  104. **Sanderson, M. G.** 1996. Biomass of termites and their emissions of methane and carbon dioxide: A global database. *Global Biogeochem Cycles* **10**:543–557.
  105. **Sato, T., Y. Hongoh, S. Noda, S. Hattori, S. Ui, and M. Ohkuma.** 2009. *Candidatus Desulfovibrio trichonymphae*, a novel intracellular symbiont of the flagellate *Trichonympha agilis* in termite gut. *Environmental Microbiology* **11**:1007-1015.
  106. **Sauer, P. E., T. I. Eglinton, J. M. Hayes, A. Schimmelmann, and A. L. Sessions.** 2001. Compound specific D/H ratios of lipid biomarkers from sediments as a proxy for environmental and climatic conditions. *Geochim Cosmochim Acta* **65**:213-222.
  107. **Schink, B., V. Thiemann, H. Laue, and M. Friedrich.** 2002. *Desulfotignum phosphitoxidans* sp. nov., a new marine sulfate reducer that oxidizes phosphite to phosphate. *Arch Microbiol* **177**:381-391.

108. **Schmidt, T., L. Zwank, M. Elsner, M. Berg, R. Meckenstock, and S. Haderlein.** 2004. Compound-specific stable isotope analysis of organic contaminants in natural environments: a critical review of the state of the art, prospects, and future challenges. *Anal Bioanal Chem* **378**:283-300.
109. **Schmitt-Wagner, D., and A. Brune.** 1999. Hydrogen profiles and localization of methanogenic activities in the highly compartmentalized hindgut of soil-feeding higher termites (*Cubitermes* spp.). *Appl Environ Microbiol* **65**:4490-4496.
110. **Schmitt-Wagner, D., M. W. Friedrich, B. Wagner, and A. Brune.** 2003. Axial dynamics, stability, and interspecies similarity of bacterial community structure in the highly compartmentalized gut of soil-feeding termites (*Cubitermes* spp.). *Appl Environ Microbiol* **69**:6018–6024
111. **Schmitt-Wagner, D., M. W. Friedrich, B. Wagner, and A. Brune.** 2003. Phylogenetic diversity, abundance, and axial distribution of bacteria in the intestinal tract of two soil-feeding termites (*Cubitermes* spp.). *Appl Environ Microbiol* **69**:6007-17.
112. **Sessions, A. L., T. W. Burgoyne, A. Schimmelmann, and J. M. Hayes.** 1999. Fractionation of hydrogen isotopes in lipid biosynthesis. *Org Geochem* **30**:1193-1200.
113. **Shinzato, N., T. Matsumoto, I. Yamaoka, T. Oshima, and A. Yamagishi.** 1999. Phylogenetic diversity of symbiotic methanogens living in the hindgut of the lower termite *Reticulitermes speratus* analyzed by PCR and in situ hybridization. *Appl Environ Microbiol* **65**:837–840.
114. **Solomon, S., D. Qin, M. Manning, Z. Chen, M. Marquis, K. Averyt, M. Tignor, and H. Miller.** 2008. *Climate change 2007: the physical science basis.* Cambridge University Press Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, S, Paulo, Delhi.
115. **Stewart, C. S., and H. J. Flint.** 1989. *Bacteroides (Fibrobacter) succinogenes*, a cellulolytic anaerobic bacterium from the gastrointestinal tract. *Appl Microbiol Biotechnol* **30**:433-439.
116. **Sugimoto, A., D. E. Bignell, and J. A. MacDonald.** 2000. Global impact of termites on the carbon cycle and atmospheric trace gases, p. 409-435. *In* T. Abe, D. E. Bignell, and M. Higashi (ed.), *Termites: Evolution, Sociality, Symbioses.* Kluwer Academic Publishers, Dordrecht, The Netherlands.
117. **Thauer, R. K.** 2007. A fifth pathway of carbon fixation. *Science* **318**:1732-1733.
118. **Thongaram, T., Y. Hongoh, S. Kosono, M. Ohkuma, S. Trakulnaleamsai, N. Noparatnaraporn, and T. Kudo.** 2005. Comparison of bacterial communities in the alkaline gut segment among various species of higher termites. *Extremophiles* **9**:229-238.
119. **Thorne, B. L., D. Grimaldi, and K. Krishna.** 2000. Early fossil history of termites, p. 77-93. *In* T. Abe, D. E. Bignell, and M. Higashi (ed.), *Termites: Evolution, Sociality, Symbioses, Ecology.* Springer, Dordrecht, The Netherlands.
120. **Tokuda, G., N. Lo, and H. Watanabe.** 2005. Marked variations in patterns of cellulase activity against crystalline- vs. carboxymethyl-cellulose in the digestive systems of diverse, wood-feeding termites. *Physiol Entomol* **30**: 372 - 380.
121. **Tokuda, G., and H. Watanabe.** 2007. Hidden cellulases in termites: revision of an old hypothesis. *Biology Letters* **3**:336-339.

122. **Wakako, I., and A. Brune.** 2009. Cospeciation of termite gut flagellates and their bacterial endosymbionts: *Trichonympha* species and Candidatus *Endomicrobium trichonymphae*. *Molecular Ecology* **18**:332-342.
123. **Warnecke, F., P. Luginbühl, N. Ivanova, M. Ghassemian, T. Richardson, J. Stege, M. Cayouette, A. Mchardy, G. Djordjevic, N. Aboushadi, R. Sorek, S. Tringe, M. Podar, H. Martin, V. Kunin, D. Dalevi, J. Madejska, E. Kirton, D. Platt, E. Szeto, A. Salamov, K. Barry, N. Mikhailova, N. Kyrpides, E. Matson, E. Ottesen, X. Zhang, M. Hernández, C. Murillo, L. Acosta, I. Rigoutsos, G. Tamayo, B. Green, C. Chang, E. Rubin, E. Mathur, D. Robertson, P. Hugenholtz, and J. Leadbetter.** 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**:560-565.
124. **Whitman, W. B., D. C. Coleman, and W. J. Wiebe.** 1998. Prokaryotes: The unseen majority. *Proc Natl Acad Sci USA* **95**:6578 – 6583.
125. **Wieringa, K.** 1939. The formation of acetic acid from carbon dioxide and hydrogen by anaerobic spore-forming bacteria. *Antonie van Leeuwenhoek* **6**:251-262.
126. **Wood, H. G.** 1952. A study of carbon dioxide fixation by mass determination on the types of <sup>13</sup>C-acetate. *J Biol Chem* **194**.
127. **Wood, H. G., and L. G. Ljungdahl.** 1991. Autotrophic character of the acetogenic bacteria, p. 201-250. *In* J. M. Shively and L. L. Barton (ed.), *Variations in Autotrophic Life*. Academic Press, New York.
128. **Yamamoto, I., T. Saiki, S. M. Liu, and L. G. Ljungdahl.** 1983. Purification and properties of NADP-dependent formate dehydrogenase from *Clostridium thermoaceticum*, a tungsten-selenium-iron protein. *J Biol Chem* **258**:1826-32.
129. **Yamin, M. A.** 1978. Axenic cultivation of the cellulolytic flagellate *Trichomitopsis termopsidis* (Cleveland) from the termite *Zootermopsis*. *J Protozool* **25**:535-538.
130. **Yamin, M. A.** 1981. Cellulose metabolism by the flagellate *Trichonympha* from a termite is independent of endosymbiotic bacteria. *Science* **211**:58-59.
131. **Yamin, M. A.** 1980. Cellulose metabolism by the termite flagellate *Trichomitopsis termopsidis*. *Appl Environ Microbiol* **39**:859-863.
132. **Yamin, M. A.** 1979. Termite flagellates. *Sociobiology* **4**:1-119.
133. **Zhou, X., J. Smith, F. Oi, P. Koehler, G. Bennett, and M. Scharf.** 2007. Correlation of cellulase gene expression and cellulolytic activity throughout the gut of the termite *Reticulitermes flavipes*. *Gene* **395**:29-39.

# Formate dehydrogenase gene diversity in acetogenic gut communities of lower, wood-feeding termites and a wood-feeding roach

## Abstract

The bacterial Wood-Ljungdahl pathway for CO<sub>2</sub>-reductive acetogenesis is important for the nutritional mutualism occurring between wood-feeding insects and their hindgut microbiota. A key step in this pathway is the reduction of CO<sub>2</sub> to formate, catalyzed by the enzyme formate dehydrogenase (FDH). Putative selenocysteine- (Sec) and cysteine- (Cys) containing paralogs of hydrogenase-linked FDH (FDH<sub>H</sub>) have been identified in the termite gut acetogenic spirochete, *Treponema primitia*, but knowledge of their relevance in the termite gut environment remains limited. In this study, we designed degenerate PCR primers for FDH<sub>H</sub> genes (*fdhF*) and assessed *fdhF* diversity in insect gut bacterial isolates and the gut microbial communities of termites and roaches. The insects examined herein represent the wood-feeding termite families *Termopsidae*, *Kalotermitidae*, and *Rhinotermitidae* (phylogenetically “lower” termite taxa), the wood-feeding roach family *Cryptocercidae* (the sister taxon to termites), and the omnivorous roach family *Blattidae*. Sec and Cys FDH<sub>H</sub> variants were identified in every wood-feeding insect but not the omnivorous roach. Of 68 novel phlotypes obtained from inventories, 66 affiliated phylogenetically with enzymes from *T. primitia*. These formed two sub-clades (37 and 29 phlotypes) almost completely comprised of Sec-containing and Cys-containing enzymes, respectively. A gut cDNA

inventory showed transcription of both variants in the termite *Zootermopsis nevadensis* (family *Termopsidae*). The results suggest  $\text{FDH}_\text{H}$  enzymes are important for the  $\text{CO}_2$ -reductive metabolism of uncultured acetogenic treponemes and imply that the trace element selenium has shaped the gene content of gut microbial communities in wood-feeding insects.

## Introduction

Xylophagy, the ability to feed exclusively on lignocellulose, in termites and wood-feeding roaches results from obligate nutritional mutualisms with their hindgut microbes (Breznak, 1982; Breznak and Brune, 1994; Inoue *et al.*, 2000). Studies on phylogenetically lower wood-feeding termites and wood-feeding roaches in the genus *Cryptocercus* have pointed to commonalities in their gut microbiota, in particular the presence of unique cellulolytic protozoa, as the basis for similarities in nutritional physiology (Honigberg, 1970; Inoue *et al.*, 2000). The process of lignocellulose degradation, elucidated in lower termites, is stepwise (Odelson and Breznak, 1983; Breznak and Switzer, 1986; Pester and Brune, 2007), and is applicable to *Cryptocercus* based on radiotracer measurements of hindgut carbon flow (Breznak and Switzer, 1986). Lignocellulose-derived polysaccharides are first hydrolyzed to glycosyl units and then fermented into acetate, H<sub>2</sub>, and CO<sub>2</sub> by cellulolytic protozoa. The H<sub>2</sub> generated by this activity can approach saturation levels (Ebert and Brune, 1997; Pester and Brune, 2007). The final step in hindgut fermentation is CO<sub>2</sub>-reductive acetogenesis, which outcompetes methanogenesis as the H<sub>2</sub> sink in the guts of these wood-feeding insects (Odelson and Breznak, 1983; Breznak and Switzer, 1986; Brauman *et al.*, 1992; Pester and Brune, 2007).

Acetogenesis in wood-feeding termites is mediated by anaerobic bacteria (Odelson and Breznak, 1983; Breznak and Switzer, 1986) and is estimated to contribute up to 1/3 of the acetate used by the insect host as its carbon and energy source (Odelson and Breznak, 1983). H<sub>2</sub> flux measurements in termite guts have shown that acetogens catalyze rapid and efficient turnover of H<sub>2</sub> in a system that is essentially optimized for acetate production (Pester and

Brune, 2007). The pathway by which these bacteria metabolize H<sub>2</sub> and CO<sub>2</sub> to acetate is the Wood-Ljungdahl pathway for CO<sub>2</sub> reductive acetogenesis (Ljungdahl, 1986). The enzymology underlying the four reductions culminating in the reductive fixation of CO<sub>2</sub> to acetate has been largely elucidated in the model actogen, *Moorella thermoacetica*, a member of the phylum *Firmicutes*, which comprises the majority of acetogens (Wood and Ljungdahl, 1991; Drake, 1994; Drake *et al.*, 2006; Drake *et al.*, 2007). The enzyme formate dehydrogenase (FDH) is one of two enzymes in the pathway critical for both H<sub>2</sub> turnover and autotrophic carbon fixation in acetogens (Drake, 1994; Drake *et al.*, 2002; Vorholt and Thauer, 2002). FDH catalyzes the reduction of CO<sub>2</sub> to formate with H<sub>2</sub> (or its equivalent) in the first step within the methyl branch of the Wood-Ljungdahl pathway (Drake *et al.*, 2006).

Culture and gene-inventory studies of termite gut acetogens indicate spirochetes, rather than firmicutes, are the predominant acetogenic bacteria in wood-feeding termite guts (Leadbetter *et al.*, 1999; Graber and Breznak, 2004; Graber *et al.*, 2004; Pester and Brune, 2006; Warnecke *et al.*, 2007). Despite the importance of acetogenic spirochetes in termite nutrition, only two isolates have ever been obtained; they remain, to this day, the sole examples of chemolithoautotrophy in the phylum *Spirochaetes* (Leadbetter *et al.*, 1999). The H<sub>2</sub>-utilizing acetogenic spirochete *Treponema primitia* str. ZAS-2, isolated from the hindgut of the lower wood-feeding termite *Zootermopsis angusticollis*, is one these isolates.

A recent study of FDH in *T. primitia* reported the identification of two FDH genes (Matson *et al.*, 2010). Sequence comparisons with structurally characterized FDH enzymes (Axley *et al.*, 1991; Gladyshev *et al.*, 1994; Boyington *et al.*, 1997; Jormakka *et al.*, 2002;

Raaijmakers *et al.*, 2002; Jormakka *et al.*, 2003) indicated the non-canonical amino acid selenocysteine (Sec) is likely encoded within the active site of one FDH variant whereas the amino acid cysteine (Cys) is encoded at the corresponding catalytic position in the other. Despite the catalytic advantages of selenoproteins over their selenium-free counterparts (Axley *et al.*, 1991; Berry *et al.*, 1992; Lee *et al.*, 2000; Gromer *et al.*, 2003; Kim and Gladyshev, 2005), several studies have demonstrated that Cys variants may be useful, if not required, when selenium is scarce (Jones and Stadtman, 1981; Berghöfer *et al.*, 1994; Vorholt *et al.*, 1997; Valente *et al.*, 2006). Consistent with these previous studies was the finding that selenium controls transcription of genes for both selenium- and selenium-independent FDH enzymes in *T. primitia* (Matson *et al.*, 2010). Taken together, these results implied *T. primitia* may be challenged by changing selenium availability in the termite gut.

Phylogenetic analysis of the FDH genes in *T. primitia* indicated they are *fdhF* paralogs that encode hydrogenase-linked FDH enzymes (FDH<sub>H</sub>), similar to those used for formate oxidation during sugar fermentation in *Gammaproteobacteria* such as *Escherichia coli* (Pecher *et al.*, 1985; Zinoni *et al.*, 1986). The result was noteworthy as *T. primitia* FDH<sub>H</sub> enzymes were expected to group with the well-characterized FDH for acetogenesis, an NADPH-linked tungsten containing selenoprotein from the classic acetogen, *Moorella thermoacetica* (Thauer, 1972; Yamamoto *et al.*, 1983; Pierce *et al.*, 2008). This led to the suggestion that the use of hydrogenase-linked FDH enzymes to directly access H<sub>2</sub> for CO<sub>2</sub>-reductive metabolism may be an adaptation of *T. primitia* to life its H<sub>2</sub>-rich gut environment.

Study of FDH<sub>H</sub> in *T. primitia* (Matson *et al.*, 2010) imply that it has both evolved mechanisms to deal with changing selenium availability and adapted its metabolism to take advantage of high H<sub>2</sub> levels in the gut of its host termite, *Z. angusticollis*. Yet the extent to which *T. primitia* reflects general characteristics of the gut microbial community in *Zootermopsis* and other wood-feeding insects remains unknown. Here, using novel degenerate *fdhF* primers, we investigated FDH<sub>H</sub> diversity in insect-gut isolates, the gut microbial communities of three wood-feeding lower termite species (*Zootermopsis nevadensis*, *Reticulitermes hesperus*, *Incisitermes minor*), wood-feeding roaches (*C. punctulatus*), and an omnivorous roach (*Periplaneta americana*). Together these insect species represent 4 of 5 wood-feeding basal phylogenetic taxa within the insect superorder *Dictyoptera*, comprised of termites, roaches, and mantids (Grimaldi and Engel, 2005). Insights into Sec/Cys FDH variant evolution in gut communities are highlighted and the likely importance of FDH<sub>H</sub> enzymes to CO<sub>2</sub>-reductive acetogenesis is discussed.

## Materials and Methods

### Microbial strains

Microbial isolates *Treponema primitia* str. ZAS-1 (DSM 12426), *Treponema primitia* str. ZAS-2 (DSM 12427), and *Acetonema longum* str. APO-1 (DSM 6540) were grown in anaerobic YACo medium under a headspace of 80% H<sub>2</sub> + 20% CO<sub>2</sub> as described previously (Kane and Breznak, 1991; Leadbetter *et al.*, 1999; Lilburn *et al.*, 2001). *Treponema azotonutricium* str. ZAS-9 (DSM 13862) was cultivated in a similar medium (Leadbetter *et al.*, 1999; Lilburn *et al.*, 2001). *Buttiauxiella* sp. str. SN-1 was isolated from the homogenized gut of a common garden snail (*Helix aspersa*), collected at the California

Institute of Technology (E. Matson, personal communication). This and other microbial isolates *Citrobacter sp.* str. TSA-1, *Escherichia coli* K12 str. MG1655, *Serratia grimesii* str. ZFX-1, and *Pantoea stewartii* subsp. *stewartii* (DSM 30176), were grown in LB shaking at 250 rpm, 30°C (*P. stewartii*, *Citrobacter*) or 37°C (*E. coli*, *S. grimesii*, *Buttiauxiella*). Cultures were harvested during exponential phase for DNA extraction. DNA was extracted from pure culture microbial isolates using a DNeasy extraction kit (QIAGEN, Valencia, CA).

### **Insect collection**

*Z. nevadensis* collection ChiA1, *Z. nevadensis* collection ChiB, and *R. hesperus* collection ChiB worker termites were obtained from fallen *Pinus ponderosa* (Ponderosa pine) in the San Gabriel Mountains of Southern California. *I. minor* collection Pas1 worker termites were collected from a decaying chaparral oak pile in Pasadena, CA. Nymph specimens of the wood roach *C. punctulatus* were collected in the South Mountains of North Carolina and made available for this study by C. A. Nalepa (North Carolina Department of Agriculture, North Carolina State University). Specimens of the omnivorous cockroach *Periplaneta americana* were collected on the Caltech campus.

### **Hindgut nucleic acid extraction**

Hindguts of 30 *Z. nevadensis* ChiA1, 180 *R. hesperus* ChiA2, 7 *I. minor* Pas1 worker termites, and 1 *P. americana* roach were extracted within 24 h of collection and pooled by collection into 1X Tris-EDTA (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) for DNA analyses. The guts of 3 *C. punctulatus* nymphs were similarly extracted within 1 week of receipt.

Whole gut community DNA was obtained using the method described by Matson *et al.* (Matson *et al.*, 2007). For gut community RNA analyses, the hindguts of 8 *Z. nevadensis* ChiB worker termites were extracted and suspended in 100  $\mu\text{l}$  of RNA stabilization buffer (RNA Protect Bacteria Reagent, QIAGEN) immediately after collection in the field. Crude lysate containing RNA and DNA was extracted using the bead-beating/phenol procedure described for gut DNA extraction (Matson *et al.*, 2007). Total RNA was purified from crude lysate using RNeasy Mini columns with an on-column DNase I treatment (QIAGEN), followed by 30 min of off-column DNA digestion at 37°C using RQ1 DNase enzyme (0.1 U  $\cdot$   $\mu\text{l}^{-1}$ ) in 1X DNase buffer (Promega Corp., Madison, WI), finishing with a final RNeasy Mini column purification for further template purification and enzyme removal. RNA purity and yield (900 ng  $\cdot$   $\mu\text{l}^{-1}$ ) were evaluated spectrophotometrically and by agarose gel electrophoresis. Gut community RNA (450 ng) was converted to single strand cDNA by randomly-primed reverse transcription (1<sup>st</sup> Strand Synthesis Kit for RT-PCR, Roche Applied Science, Indianapolis, IN).

### ***fdhF* primer design**

The translated amino acid sequences of *fdhF* genes in *T. primitia* str. ZAS-2 were used as queries in BLAST (Altschul *et al.*, 1997) searches of NCBI databases to identify a set of homologous sequences (>70% similarity) for which *fdhF* primers could be designed. Oligonucleotide primers (Table 2.1) target conserved regions within the molybdopterin oxidoreductase Fe4S4 and molybdopterin dinucleotide binding domains (Fig. 2.1A) using an alignment of *T. primitia* and other *fdhF* nucleotide sequences (Fig. 2.1B). Together, forward and reverse primers span the entire molybdopterin oxidoreductase domain, which harbors

catalytic selenocysteine and cysteine amino acids and is the largest protein domain in FDH<sub>H</sub> enzymes. The primers to yield a ca. 1.8 kb amplicon from a typical 2.2 kb *fdhF* gene. Confirmation of *fdhF* amplification from pure culture templates using non-degenerate primer combinations led to the modification of primers into a degenerate “universal” *fdhF* primer set: EntfdhFunv-F1, TgfdhFunv-F1, and fdhFunv-R1.

**Table 2.1.** PCR primers for *fdhF* type formate dehydrogenase genes.

Primer <sup>1</sup>	Sequence <sup>2</sup>
fdhF-F1	5'– GCT GGT ACG GCT GGG ATT –3'
fdhF-F2	5'– GTT ATT ATG GCT GGG ACT –3'
fdhF-F3	5'– GCT ACT ACG GCT GGG ATT –3'
fdhF-R1	5'– ACC CAC CAC TGG TAG GTC AT –3'
fdhF-R2	5'– ATC CAC CAC TGG TAG GTC AT –3'
TgfdhFunv-F1	5'– TGG TAY GGI TGG GAY T –3'
EntfdhFunv-F1	5'– GIT AYT AYG GIT GGG AYT –3'
fdhFunv-R1	5'– CCA CCA YTG RTA IGT CAT –3'

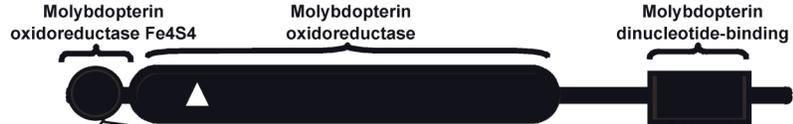
<sup>1</sup>Forward primers indicated by monikers, ‘F1’, ‘F2’, or ‘F3’ in primer name; reverse primers by ‘R1’, ‘R2’ in primer name.

<sup>2</sup>Canonical nucleotides symbols are used. I, inosine.

**Figure 2.1.** Degenerate *fdhF* PCR primer design.

(A) Hydrogenase-linked formate dehydrogenase (FDH<sub>H</sub>) enzymes, encoded by *fdhF*, have three characteristic domains: molybdopterin oxidoreductase Fe4S4 (PFAM ID: PF04879, Interpro ID: IPR006963), molybdopterin oxidoreductase (PFAM ID: PF00384, Interpro ID: IPR006656) containing catalytic selenocysteine or cysteine amino acids (triangle), and the molybdopterin dinucleotide binding domain (PFAM ID: PF01568, Interpro ID: IPR006657). (B) ClustalW (Larkin *et al.*, 2007) alignment of conserved regions of DNA within the molybdopterin oxidoreductase Fe4S4 domain and the molybdopterin dinucleotide binding domain used for primer design. The nucleotide positions corresponding to each primer in the *fdhF* of *E. coli* are listed above the alignment. The sequence targets of each primer are indicated to the right of each alignment (F1=fdhF-F1, F2=fdhF-F2, F3=fdhF-F3, EU=EntfdhFunv-F1, TU=TgfdhFunv-F1, R1=fdhF-R1, R2=fdhF-R2, and RU=fdhFunv-R1). Mismatches to universal primer EntfdhF-unvF1 are highlighted and multiple copies of *fdhF* are indicated. Accession numbers are listed in Appendix, Table 2.2.

A



B

	Forward primer		Reverse primer	
	G--Y--Y--G--W--D--	(AA)	-M--T--Y--Q--W--W--	(AA)
	134	151	1973	1954
<i>Aggregatibacter aphrophilus</i> NJ8700	GATATTATGGCTGGGACT	EU	ATGACCTATCAATGGTGGAT	RU
<i>Citrobacter koseri</i> ATCC BAA-895 copy 1	GCTACTA TGGCTGGGACT	EU	ATGACCTATCAATGGTGGAT	RU
<i>Citrobacter koseri</i> ATCC BAA-895 copy 2	GCTACTATGGCTGGGATT	EU	ATGACCTACCAGTGGTGGAT	R2 RU
<i>Citrobacter</i> sp. 30_2 copy 1	GCTACTATGGCTGGGACT	EU	ATGACCTACCAGTGGTGGAT	R2 RU
<i>Citrobacter</i> sp. 30_2 copy 2	GCTATTATGGTTGGGACT	EU	ATGACCTATCAATGGTGGAT	RU
<i>Citrobacter rodentium</i> ICC168 copy 1	GCTATTATGGCTGGGATT	EU	ATGACCTACCAGTGGTGGAT	R2 RU
<i>Citrobacter rodentium</i> ICC168 copy 2	GCTACTACGGCTGGGATT	F3 EU	ATGACTTACCAGTGGTGGAT	RU
<i>Citrobacter youngae</i> ATCC 29220 copy 1	GCTACTATGGCTGGGATT	EU	ATGACCTACCAGTGGTGGAT	R2 RU
<i>Citrobacter youngae</i> ATCC 29220 copy 2	GCTACTATGGTTGGGACT	EU	ATGACCTATCAATGGTGGAT	RU
<i>Cronobacter turicensis</i> copy 1	GCTATTACGGCTGGGATT	EU	ATGACGTACCAGTGGTGGAT	RU
<i>Cronobacter turicensis</i> copy 2	GCTATTACGGCTGGGATT	EU	ATGACCTACCAGTGGTGGAT	R2 RU
<i>Dickeya zeae</i> Ech1591	GCTATTATGGCTGGGACT	EU	ATGACCTACCAGTGGTGGAT	R2 RU
<i>Dickeya dadantii</i> Ech586	GCTATTACGGTTGGGATT	EU	ATGACTTATCAGTGGTGGAT	RU
<i>Edwardsiella ictaluri</i> 93-146	GGTACTATGGCTGGGACT	EU	ATGACCTACCAGTGGTGGAT	R2 RU
<i>Edwardsiella tarda</i> EIB202	GGTACTACGGCTGGGATT	EU	ATGACCTACCAGTGGTGGAT	R2 RU
<i>Enterobacter</i> sp. 638 copy 1	GCTATTACGGCTGGGATT	EU	ATGACTTATCAGTGGTGGAT	RU
<i>Enterobacter</i> sp. 638 copy 2	GTTTCTACGGCTGGGATT	EU	ATGACCTATCAGTGGTGGAT	RU
<i>Enterobacter cancerogenus</i> ATCC 35316	GCTACTATGGCTGGGATT	EU	ATGACCTACCAGTGGTGGAT	R2 RU
<i>Cronobacter sakazakii</i> ATCC BAA-894	GCTATTACGGCTGGGATT	EU	ATGACCTACCAGTGGTGGAT	R3 RU
<i>Escherichia coli</i> K-12 substr. MG1655	GTTATTATGGCTGGGACT	F2 EU	ATGACCTACCAGTGGTGGAT	R4 RU
<i>Escherichia fergusonii</i> ATCC 35469	GGTATTACGGCTGGGATT	EU	ATGACCTACCAGTGGTGGAT	R5 RU
<i>Pantoea</i> sp. At-9b	GCTACTATGGCTGGGACT	EU	ATGACTTACCAGTGGTGGAT	RU
<i>Pectobacterium atrosepticum</i> SCRI1043 copy 1	GCTATTACGGTTGGGATT	EU	ATGACTTACCAGTGGTGGAT	RU
<i>Pectobacterium atrosepticum</i> SCRI1043 copy 2	GGTATTACGGCTGGGATT	EU	ATGACCTACCAGTGGTGGAT	R2 RU
<i>Pectobacterium carotovorum</i> WPP14	GCTATTACGGCTGGGATT	EU	ATGACCTACCAGTGGTGGAT	R2 RU
<i>Pectobacterium wasabiae</i> WPP163	GCTACTACGGTTGGGATT	EU	ATGACTTACCAATGGTGGAT	RU
<i>Providencia alcalifaciens</i> DSM 30120	GTTATTATGGCTGGGACT	F2 EU	ATGACCTATCAGTGGTGGAT	RU
<i>Proteus mirabilis</i> HI4320 copy 1	GGTATTATGGATGGGATT	EU	ATGACTTACCAATGGTGGAT	RU
<i>Proteus mirabilis</i> HI4320 copy 2	GTCTCTATGGATGGGATT	EU	ATGACTTATCAATGGTGGAT	RU
<i>Providencia rustigianii</i> DSM 4541	GCTATTATGGCTGGGACT	EU	ATGACGTATCAGTGGTGGAT	RU
<i>Shigella</i> sp. D9	GTTATTATGGCTGGGATT	EU	ATGACCTACCAGTGGTGGAT	R2 RU
<i>Salmonella typhimurium</i> LT2	GCTACTACGGCTGGGATT	F3 EU	ATGACCTACCAGTGGTGGAT	R2 RU
<i>Salmonella enterica</i> Typhi CT18	GCTACTATGGCTGGGATT	EU	ATGACCTACCAGTGGTGGAT	R2 RU
<i>Yersinia aldovae</i> ATCC 35236	GCTATTATGGCTGGGATT	EU	ATGACTTATCAGTGGTGGAT	RU
<i>Yersinia bercovieri</i> ATCC 43970	GCTACTACGGCTGGGATT	F3 EU	ATGACTTACCAGTGGTGGAT	RU
<i>Yersinia enterocolitica</i> 8081	GCTATTATGGCTGGGATT	EU	ATGACTTATCAGTGGTGGAT	RU
<i>Yersinia frederiksenii</i> ATCC 33641 copy 1	GCTACTATGGCTGGGATT	EU	ATGACTTATCAGTGGTGGAT	RU
<i>Yersinia frederiksenii</i> ATCC 33641 copy 2	GTTACTATGGTTGGGATT	EU	ATGACTTATCAGTGGTGGAT	RU
<i>Yersinia mollaretii</i> ATCC 43969 copy 1	GCTACTACGGCTGGGATT	F3 EU	ATGACCTATCAGTGGTGGAT	RU
<i>Yersinia mollaretii</i> ATCC 43969 copy 2	GCTACTACGGCTGGGATT	F3 EU	ATGACTTACCAGTGGTGGAT	RU
<i>Yersinia rohdei</i> ATCC 43380	GCTATTATGGCTGGGATT	EU	ATGACTTATCAGTGGTGGAT	RU
<i>Yersinia ruckeri</i> ATCC 29473	GCTATTACGGCTGGGATT	EU	ATGACTTATCAATGGTGGAT	RU
<i>Klebsiella pneumoniae</i> NTUH-K2044 copy 1	GCTATTATGGCTGGGACT	EU	ATGACCTACCAGTGGTGGAT	R2 RU
<i>Klebsiella pneumoniae</i> NTUH-K2044 copy 2	GCTACTACGGCTGGGATT	F3 EU	ATGACCTATCAGTGGTGGAT	RU
<i>Serratia proteamaculans</i> 568	GCTATTACGGCTGGGACT	EU	ATGACCTACCAGTGGTGGAT	R2 RU
<i>Aeromonas salmonicida</i> A449	GTTATTACGGTTGGGATT	EU	ATGACCTACCAGTGGTGGAT	R2 RU
<i>Psychromonas</i> sp. CNPT3	GATATTATGGTTGGGATT	EU	ATGACTTATCAATGGTGGAT	RU
<i>Photobacterium profundum</i> 3TCK	GCTACTATGGTTGGGATT	EU	ATGACCTACCAGTGGTGGAT	R2 RU
<i>Vibrio angustum</i> S14	GTTACTATGGCTGGGATT	EU	ATGACTTACCAATGGTGGAT	RU
<i>Clostridium beijerinckii</i> NCIMB 8052	GATATTATGGCTGGGATT	EU	ATGACATATCAATGGTGGAT	RU
<i>Clostridium carboxidivorans</i> P7 copy 1	GACTACTATGGATGGGATT	EU	ATGACATACCAATGGTGGAT	RU
<i>Clostridium carboxidivorans</i> P7 copy 2	GATACTACGGATGGGATT	EU	ATGACTTATCAATGGTGGAT	RU
<i>Clostridium bartlettii</i> DSM 16795	GACTACTATGGATGGGACT	EU	ATGACTTACCAATGGTGGAT	RU
<i>Clostridium difficile</i> 630	GTCTTATGGATGGGATT	EU	ATGACTTATCAGTGGTGGAT	RU
	G--W--Y--G--W--D--	(AA)	-M--T--Y--Q--W--W--	(AA)
<i>Treponema primitia</i> str. ZAS-2 <i>fdhF</i> sec (copy 1)	GCTGGTACGGCTGGGATT	F1 TU	ATGACCTACCAGTGGTGGAT	R1 RU
<i>Treponema primitia</i> str. ZAS-2 <i>fdhF</i> cys (copy 2)	GCTGGTACGGCTGGGATT	F1 TU	ATGACCTACCAGTGGTGGAT	R2 RU

***fdhF* amplification and cloning**

PCR amplifications of *fdhF* genes were performed with oligonucleotide primers listed in Table 1 on nucleic acids from microbial isolates and insect hindgut communities in the combinations and concentrations described in Appendix, Table 2.3. Primers with 5' phosphate groups were synthesized by Integrated DNA Technologies (Coralville, IA). Reactions with pure culture ( $1 \text{ ng} \cdot \mu\text{l}^{-1}$ ) or termite gut templates ( $1 \text{ ng} \cdot \mu\text{l}^{-1}$  DNA,  $5 \text{ ng} \cdot \mu\text{l}^{-1}$  cDNA) were assembled with 1X FAILSAFE Premix D (EPICENTRE Biotechnologies, Madison, WI) and  $0.07 \text{ U} \cdot \mu\text{l}^{-1}$  EXPAND High Fidelity polymerase (Roche Applied Science). *C. punctulatus* and *P. americana* reactions contained more enzyme ( $0.28 \text{ U} \cdot \mu\text{l}^{-1}$ ) and less DNA ( $0.5 \text{ ng} \cdot \mu\text{l}^{-1}$  and  $0.1 \text{ ng} \cdot \mu\text{l}^{-1}$ , respectively) due to the presence of PCR inhibitors in the template. Cycling conditions on a Mastercycler Model 5331 thermocycler (Eppendorf, Westbury, NY) were initial denaturation at  $94^{\circ}\text{C}$  for 2 min, followed by cycles of denaturation at  $94^{\circ}\text{C}$  for 30 sec, annealing for 1 min, and extension at  $68^{\circ}\text{C}$  for 2 min 30 sec, and finishing with a final extension at  $68^{\circ}\text{C}$  for 10 min. 30 cycles were used for pure culture DNA and gut cDNA templates, 23 cycles for termite and wood-roach gut DNA, and 35 cycles for *P. americana* DNA. Annealing temperature for PCR reactions containing non-degenerate primers and pure culture template (*T. primitia* str. ZAS-1, *Buttiauxiella* sp. str. SN-1, *Serratia grimesii* str. ZFX-1) was  $56^{\circ}\text{C}$  based on PCR optimization. Annealing temperature for amplifications with degenerate primers (*Citrobacter* sp. str. TSA-1, *Acetoneama longum* str. APO-1, gut DNA, and cDNA) was  $51^{\circ}\text{C}$ . PCR at  $51^{\circ}\text{C}$  with *C. punctulatus* DNA yielded products of different sizes, necessitating gel purification of the correct sized band (QIAquick Gel Extraction Kit, QIAGEN). A second PCR amplification at an annealing temperature of  $57^{\circ}\text{C}$  was performed on *C. punctulatus* DNA and yielded a

single band of the correct size. Both sets of *C. punctulatus* PCR products along with products from pure culture and other gut template amplifications were cloned using the TOPO-TA cloning kit (Invitrogen, Carlsbad, CA).

### **COII amplification**

Mitochondrial cytochrome oxidase subunit II (COII) gene fragments were amplified from termites using primers A-tLEU and B-tLYS (Lo *et al.*, 2000) and from *C. punctulatus* using primers described by Park *et al.* (Park *et al.*, 2004). Whole gut community DNA containing host insect DNA was used as template for each amplification. PCR products were purified using a QIAquick PCR purification kit (QIAGEN), sequenced, and analyzed to verify the species identity of the insect specimens.

### **RFLP analysis, sequencing, and diversity assessment**

Preliminary assessment of diversity was accomplished by digesting ca. 80–140 clones with restriction enzyme *RsaI* (New England Biolabs, Beverly, MA), followed by visualization of restriction fragment length polymorphisms (RFLPs) by gel electrophoresis using 2% (w/v) agarose (Invitrogen). Plasmids from all clones with unique RFLP patterns were purified using a QIAprep Spin Miniprep Kit (QIAGEN) and sequenced with T3 and T7 primers at Laragen, Inc. (Los Angeles, CA) using an Applied Biosystems Incorporated ABI3730 automated sequencer. Lasergene (DNASTAR, Inc, Madison, WI) software was used to assemble and edit sequences. Sequences were confirmed to be of *fdhF* type by (i) comparison to *γ-Proteobacterial fdhF* sequences in public databases at the National Center for Biotechnology using BLAST methods (expect value  $< e^{-100}$ ) (Altschul *et al.*, 1997) and

(ii) identification of key amino acid residues required for catalytic activity (Romão, 2009). Multiple sequence alignments of nucleotide and their deduced translated amino acid sequences were constructed using the program ClustalW (Larkin *et al.*, 2007) and manually adjusted. Sequences were grouped into operational taxonomic units at a 97% protein similarity level based on distance calculations (Phylip Distance Matrix using a Jones-Thorton-Taylor correction) and DOTUR (Schloss and Handelsman, 2005). The program EstimateS v8.2.0 (Colwell, 2009) was used to assess *fdhF* inventory diversity. Sec and Cys FDH<sub>H</sub> abundance statistics were calculated for each inventory using the exact binomial test for goodness-of-fit.

### **Phylogenetic analysis**

The ARB software package v.09.08.29 (Ludwig *et al.*, 2004) was used for phylogenetic analysis of protein and nucleotide sequences. Details of tree construction can be found in figure legends. COII DNA phylogeny was generated with the AxML method (Stamatakis *et al.*, 2004) whereas FDH protein phylogenies were calculated with the Phylip protein maximum likelihood (PROTML) algorithm (Felsenstein, 1989). The same filter and alignments were employed when additional tree algorithms (Fitch distance, Phylip protein parsimony) were used to infer node robustness (Felsenstein, 1989). All phylogenetic inference models were run assuming a uniform rate of change for each nucleotide or amino acid position.

### **RNA secondary structure prediction**

Selenocysteine Insertion Sequences (SECIS) elements downstream of in-frame TGA stop codons within selenoprotein encoding *fdhF* were inferred using bSECIS, a webserver for bacterial SECIS prediction (Zhang and Gladyshev, 2005), and mFOLD (Zuker, 2003). In some cases output RNA structures were vastly different from that previously proposed for *T. primitia* (Matson *et al.*, 2010). Manual inspection and adjustment were performed to find the structure with closest fit to the SECIS in ZAS-2 for the purpose of determining a minimum consensus set of structures. All RNA secondary structures free energies were calculated using mFOLD's user defined structure prediction function, available at <http://mfold.bioinfo.rpi.edu/cgi-bin/efn-form1.cgi>.

### **Accession numbers**

Sequences recovered in this study were deposited in GenBank under accession numbers GQ922348-GQ922450, GU563432-GU563485, HM208259, and HM208251.

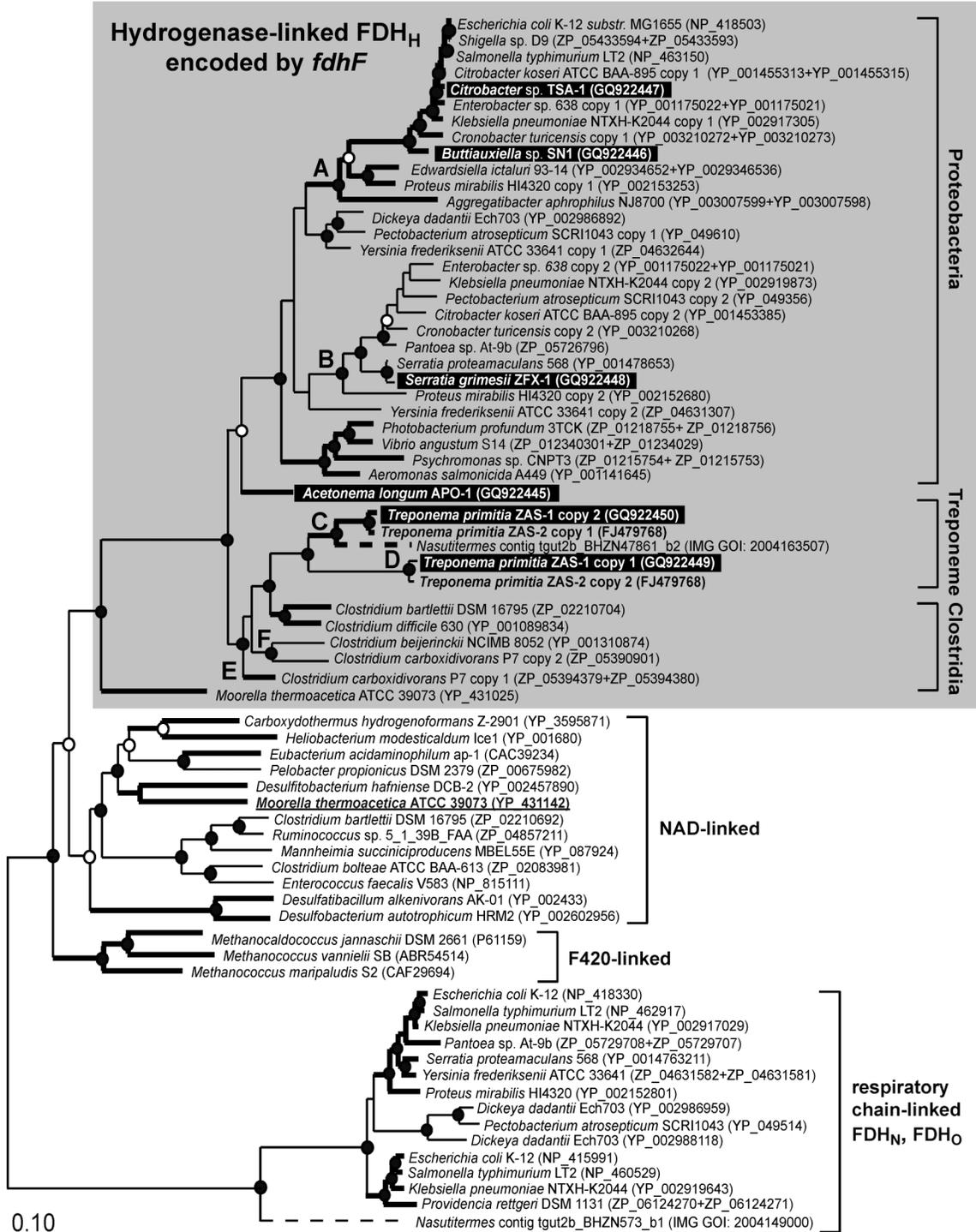
## **Results**

### ***fdhF* primers amplify phylogenetically diverse FDH<sub>H</sub> genes from pure cultures**

Primers for *fdhF* (Table 2.1) were designed using an alignment of *T. primitia* str. ZAS-2,  $\gamma$ -*Proteobacteria*, and *Firmicutes* sequences (Fig. 2.1B, Appendix, Table 2.2) to amplify ca. 1.8 kb of a typical 2.2 kb *fdhF* gene (Fig. 2.1A). Non-degenerate *fdhF* primer combinations (fdhF-F1 + fdhF-R1, fdhF-F1 + fdhF-R2, fdhF-F2 + fdhF-R1, fdhF-F3 + fdhF-R2) were tested on DNA from a second termite gut acetogenic spirochete isolate and two invertebrate gut-associated enteric  $\gamma$ -*Proteobacteria* in which *fdhF* had yet to be identified. This yielded

one or more *fdhF* homologs, confirmed by BLAST (expect value  $< e^{-100}$ ), from the spirochete *T. primitia* str. ZAS-1, termite hindgut enteric *Serratia grimesii* str. ZFX-1 (Graber and Breznak, 2005), and snail gut isolate *Buttiauxiella sp.* str. SN-1 (Table S2). Thereafter, the primers were modified into degenerate “universal” *fdhF* primers: EntfdhFunv-F1, TgfdhFunv-F1, and fdhFunv-R1 (Table 2.1, Fig. 2.1B). Universal forward primers, EntfdhFunv-F1 and TgfdhFunv-F1, target *Proteobacteria/Firmicute* and termite spirochete *fdhF*, respectively. The universal reverse primer (fdhFunv-R1) targets all *fdhF* variants. This degenerate primer set recovered *fdhF* from the termite hindgut enteric *Citrobacter sp.* str. TSA-1 (Schultz and Breznak, 1978) but did not amplify DNA from *Pantoea stewartii* subsp. *stewartii*, a  $\gamma$ -*Proteobacterium* that lacks *fdhF* and consequently ferments carbohydrates without gas production (Appendix, Table 2.3). Universal primers were also able to recover a *fdhF* homolog from the termite gut acetogenic *Firmicute*, *Acetonema longum* str. APO-1 (Kane and Breznak, 1991). Primer set specificity for *fdhF*-like genes is supported by phylogenetic analysis of novel pure culture sequences (Fig. 2.2). Each sequence falls within a cohesive clade (Fig. 2.2, grey box) that clusters to the exclusion of other FDH types (i.e., NAD-linked, coenzyme F420, or respiratory chain-linked) (Vorholt and Thauer, 2002) and contains the well-studied hydrogenase-linked FDH of *E. coli* (Sawers, 1994). This clade has been previously described by Matson *et al.* (Matson *et al.*, 2010) as encoding FDH<sub>H</sub>-like enzymes. The distribution of novel pure culture sequences throughout the FDH<sub>H</sub> clade illustrates the breadth of target range of *fdhF* primers.

**Figure 2.2.** Protein phylogeny of formate dehydrogenases from pure microbial cultures and the gut community metagenome of a phylogenetically higher termite. FDH may be coupled to hydrogenase (FDH<sub>H</sub>), coenzymes NADH/NADPH, coenzyme F420, or respiratory chains (FDH<sub>N</sub>, FDH<sub>O</sub>). The FDH<sub>H</sub> clade is highlighted by a grey box. Paralogous FDH<sub>H</sub> enzymes are indicated (copy 1 or copy 2). Sequences recovered with *fdhF* primers are highlighted by black boxes. Branches in bold indicate Sec-containing FDH<sub>H</sub>. Nodes A and B define Sec and Cys clades within the *Proteobacteria* FDH<sub>H</sub> lineage, nodes C and D for Treponeme FDH<sub>H</sub>, nodes E and F for Clostridium FDH<sub>H</sub>. The NAD-linked FDH derived from studies of acetogenesis in the model acetogen *Moorella thermoacetica* is underlined. The tree was constructed with 562 aligned amino acids using a protein maximum likelihood algorithm (Phylip PROTML). The length of dashed lines for *Nasutitermes* metagenome sequence contigs are not comparable to other branches as these were short (253 and 255 amino acids) and added by parsimony to tree. Filled circles at branch nodes denote support by distance (Fitch), parsimony (Phylip PROTPARS), and maximum likelihood (Phylip PROTML) tree construction algorithms. Unfilled circles denote support from two of these algorithms. Scale bar represents 0.1 amino acid changes per alignment position. Multiple protein accession numbers for a sequence refer to truncated portions of a selenocysteine encoding FDH. These were manually assembled into a full selenocysteine encoding open reading frame based on nucleotide sequence.



### **Convergent evolution of *fdhF*<sub>Sec</sub> and *fdhF*<sub>Cys</sub> in *Proteobacteria*, *Treponemes*, and *Firmicutes***

The recovery of genes for both Sec and Cys variants of FDH<sub>H</sub> from the acetogenic spirochete isolate, *T. primitia* str. ZAS-1, was noteworthy as this proved the presence of both genes, hereafter referred to as *fdhF*<sub>Sec</sub> and *fdhF*<sub>Cys</sub>, in all acetogenic spirochete isolates to date (Leadbetter *et al.*, 1999). Search of the NCBI database resulted in the discovery of several other distantly related organisms possessing dual *fdhF*<sub>Sec</sub> and *fdhF*<sub>Cys</sub> variants. These were *Enterobacteriaceae* belonging to the phylum  $\gamma$ -*Proteobacteria* (i.e., isolates in the genera *Citrobacter*, *Cronobacter*, *Enterobacter*, *Klebsiella*, *Pantoea*, and *Proteus*) and a *Firmicute*, *Clostridium carboxidovorans*. FDH<sub>H</sub> in enteric bacteria such as *E. coli* is used in the direction of H<sub>2</sub> production from formate oxidation during sugar fermentation (Sawers, 1994). In contrast, the FDH<sub>H</sub> in the solvent producing acetogen *C. carboxidovorans* may catalyze the opposite reaction (i.e., formate production from H<sub>2</sub> + CO<sub>2</sub>, as has been implicated for *T. primitia* str. ZAS-2 during acetogenic growth) (Liou *et al.*, 2005; Matson *et al.*, 2010).

Phylogenetic analysis of dual Sec and Cys FDH<sub>H</sub> sequences in *Proteobacteria*, *Treponemes*, and *Firmicutes* (Fig. 2.2) indicates sequences first group based on similarities in organism descent rather than the Sec or Cys character, resulting in clades comprised of only *Proteobacteria*, *Treponeme*, or *Firmicute* sequences. This phylogenetic pattern supports that the evolution of paralogous *fdhF*<sub>Sec</sub> and *fdhF*<sub>Cys</sub> in *T. primitia* str. ZAS-2 was independent from that of other dual *fdhF* genes in *Proteobacteria* and *Firmicutes*. Indeed the observation that *fdhF*<sub>Sec</sub> and *fdhF*<sub>Cys</sub> genes cluster into Sec and Cys sub-clades within each major FDH<sub>H</sub>

lineage (Fig. 2.2, defined by nodes A and B in the *Proteobacteria* FDH<sub>H</sub> clade, C and D in the *Treponeme* clade, E and F in *Clostridia*) strongly implies the occurrence of at least 3 independent *fdhF* gene duplications, one in each of these FDH lineages. Effects of diversification after gene duplication are most prominent in the *Proteobacteria* FDH<sub>H</sub> clade as dual *fdhF*<sub>Sec</sub> and *fdhF*<sub>Cys</sub> variants are present in organisms belonging to different genera rather than different strains (which is the case for the *Treponeme* clade). Other than sequences from the treponeme isolates, the *Treponeme* FDH<sub>H</sub> cluster also contains a single truncated *fdhF*<sub>Sec</sub> sequence derived from a metagenomic analysis of gut contents in a phylogenetically higher wood-feeding termite (Breznak and Warnecke, 2008). This result pointed to the possibility of greater, unexplored *fdhF* diversity occurring within the guts of wood-feeding insects.

### **Wood-feeding insect gut microbial communities harbor a diversity of *fdhF* homologs**

Degenerate *fdhF* primer sets were used to investigate *fdhF* diversity in the gut microbial communities of lower wood-feeding termite species (*Z. nevadensis*, *R. hesperus*, and *I. minor*), a wood-feeding roach species (*C. punctulatus*), and an omnivorous roach species (*P. americana*). The lower termites (Appendix, Fig. 2.7) examined in this study represent 3 of 6 major termite families. When considered with *C. punctulatus*, a member of the wood-feeding roach family *Cryptocercidae* - considered the sister taxon of termites (Appendix, Fig. 2.7), the insects studied herein represent half of all wood-feeding families in the detritivorous insect superorder *Dictyoptera*, comprised of orders *Isoptera* (termites), *Blattidae* (roaches), and *Mantodea* (mantids) (Kambhampati and Eggleton, 2000; Grimaldi and Engel, 2005).

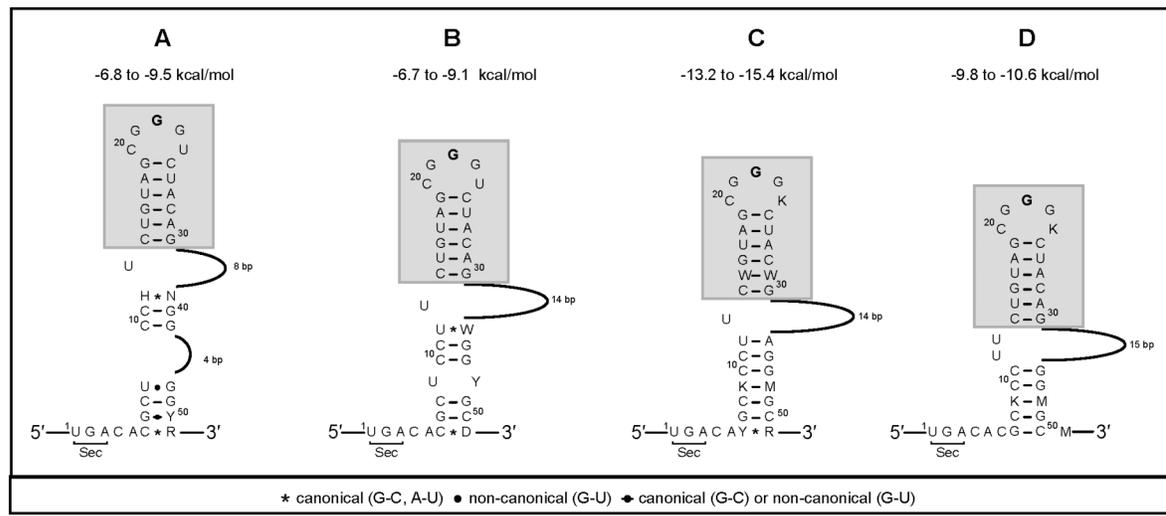
Polymerase chain reactions (PCR) containing gut community DNA from termites or the wood-feeding roach consistently yielded *fdhF* amplicons. In contrast, PCR with gut DNA from the *P. americana* did not amplify *fdhF* after repeated attempts. Clone inventories for *fdhF*, each comprised of 80-136 clones, were constructed from every termite species and *C. punctulatus* (Appendix, Tables 2.3 and 2.4). Termite inventories contained 21-32 *fdhF* genotypes or 7-15 phylotypes (based on an operational taxonomic unit definition of 97% amino acid similarity). Mean Chao1 non-parametric estimates of sequence diversity (7.53-14.96, Appendix, Table 2.5) and rarefaction analyses (Appendix, Fig. 2.8) indicate sampling efforts captured the majority of diversity present in a sample. Two inventories were generated from *Z. nevadensis* collection ChiA (family *Termopsidae*) for the purpose of comparing the target breadth of different primer sets (primer set 1, *fdhF*-F1, *fdhF*-F2, *fdhF*-F3, *fdhF*-R1, *fdhF*-R2 versus primer set 2, *EntfdhFunv*-F1, *TgfdhFunv*-F1, *fdhFunv*-R1). Universal primers (set 2) increased the number of recovered operational taxonomic units from 7 to 15. Phylotype diversity in the subterranean termite *R. hesperus* (family *Rhinotermitidae*) was on par with that in *Z. nevadensis*, a dampwood termite collected from the same mountainous region (13 versus 15 phylotypes recovered; mean Chao1  $\pm 1\sigma$ , 13.66  $\pm 3.86$  versus 14.96  $\pm 2.78$ ). Fewer phylotypes (11, 10.92  $\pm 0.62$ ) were recovered from the drywood termite *I. minor* (family *Kalotermitidae*), a result supported by 95% confidence intervals for the mean Chao1 (10.69 - 13.6) (Chao, 1987). Gut DNA from the wood-roach, *C. punctulatus*, yielded more *fdhF* diversity than any termite gut (64 genotypes, 24 phylotypes). While this result may reflect the greater sampling effort used for the *C. punctulatus* inventory, mean Chao1 estimates (21.52  $\pm 2.97$ ) suggest the wood-roach gut likely harbors greater sequence diversity than the termite gut.

To investigate whether diverse *fdhF* sequence types were also utilized within the gut community, we constructed a gut cDNA inventory from a second collection of *Z. nevadensis* (collection ChiB). The cDNA inventory yielded 15 phlotypes, the same number of phlotypes recovered from the *Z. nevadensis* gut DNA inventory. Altogether, surveys of *fdhF* in gut DNA and cDNA from wood-feeding termites and roaches resulted in 68 new *fdhF* phlotypes.

***fdhF* genes in gut microbial communities encode both Sec and Cys FDH<sub>H</sub> variants.**

Alleles for both Sec and Cys FDH<sub>H</sub> variants were identified in the gut communities of the four wood-feeding insects. The abundances of unique *fdhF*<sub>Sec</sub> and *fdhF*<sub>Cys</sub> phlotypes within each DNA inventory (4 Sec, 3 Cys and 6 Sec, 9 Cys in *Z. nevadensis*; 7 Sec, 6 Cys in *R. hesperus*; 7 Sec, 4 Cys for *I. minor*; 15 Sec, 9 Cys in *C. punctulatus*; Appendix, Table 2.5) were not statistically different (exact binomial test of goodness-of-fit p-value > 0.30). The classification of *fdhF*<sub>Sec</sub> sequences as encoding the non-canonical amino acid Sec was supported by the presence of an in-frame TGA codon followed immediately by a SECIS-like mRNA secondary structure (Fig. 2.3) identified using the programs bSECIS (Zhang and Gladyshev, 2005) and mFOLD (Zuker, 2003). These secondary structures in mRNA, along with GTP, a specialized elongation factor SelB, tRNA-Sec, and ribosome, are required for the proper insertion of Sec at the stop codon TGA (Böck, 2000). The SECIS-like structures predicted for gut sequences and *T. primitia* str. ZAS-2 (Matson *et al.*, 2010) were almost exactly matched in their apical stem loop regions, although there was a surprising amount of variability in the lower stem regions (Fig. 2.3).

**Figure 2.3.** Predicted SECIS-like elements in the mRNA of *T. primitia* (Matson *et al.*, 2010) and gut inventory *fdhF*<sub>Sec</sub> sequences can be classified into one of four consensus categories (A, B, C, D). The elements in *T. primitia* str. ZAS-1 and ZAS-2 *fdhF*<sub>Sec</sub> fall in category A. The structure of the apical stem and loop (grey box) as well as the apical loop guanine (highlighted in bold) predicted to interact with elongation factor SelB are conserved in all predicted SECIS-like elements. Free energies ranges for each mRNA structure are comparable to that for the SECIS of *fdhF* in *E. coli* (-10.7 kcal/mol). Canonical single letter coding is used for positions corresponding to more than one nucleotide (i.e., R = A or G, W = A or U, M = A or C, Y = U or C, H = A or C or U, D = A or G or U, N = any of the four nucleotides). Base pairing constraints are listed at the bottom.



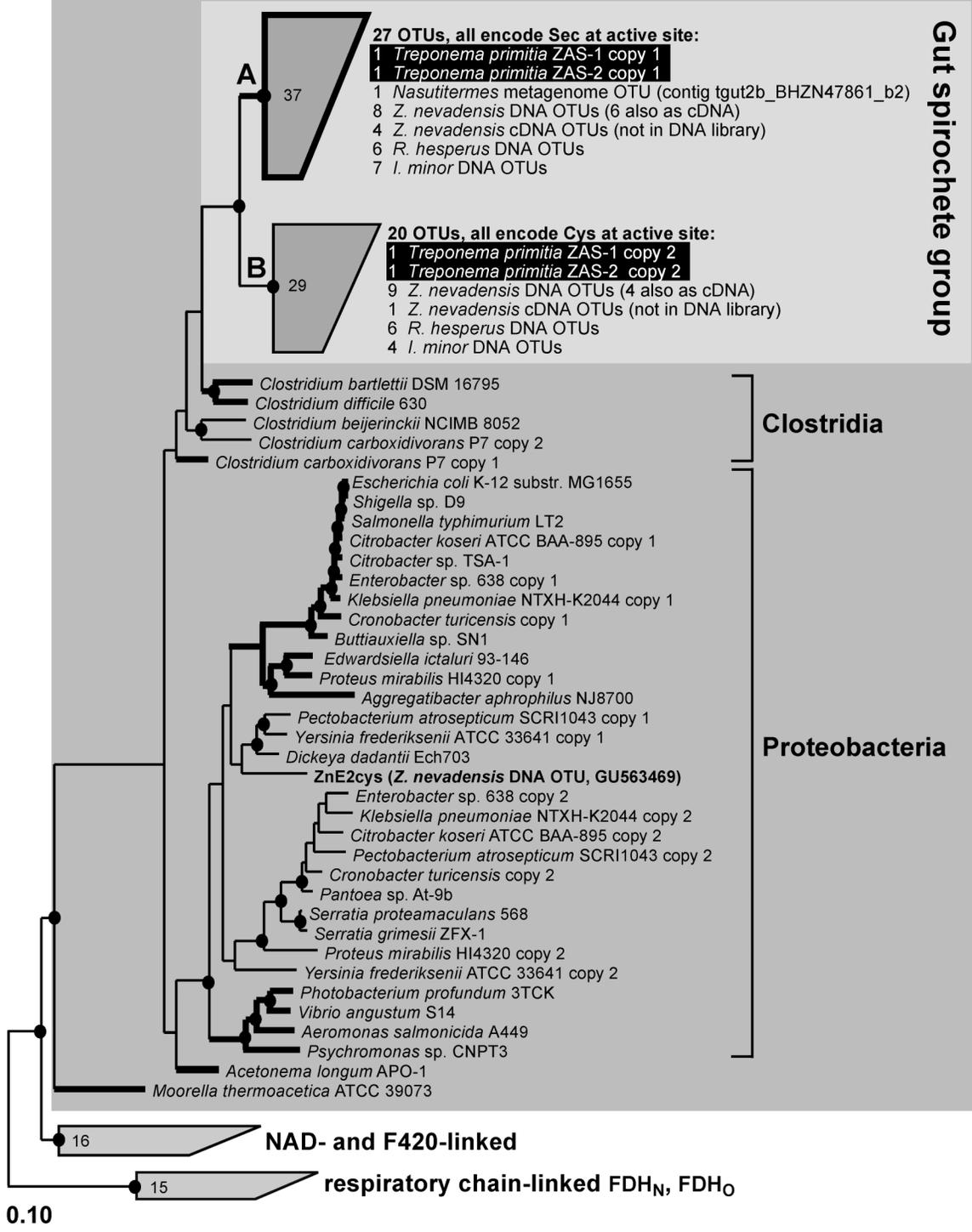
### Phylogenetic analysis of termite gut FDH<sub>H</sub> sequences

Phylogenetic analysis of termite gut FDH<sub>H</sub> sequences (Fig. 2.4) grouped all but one phylotype from gut DNA and all phylotypes from cDNA with sequences from cultured acetogenic treponemes. Based on phylogenetic inference, we have designated the clade formed by these *Treponeme*-like FDH<sub>H</sub> sequences as the “Gut spirochete group.” A polypeptide character (4-5 residues in length) was identified in all termite gut and *T. primitia* sequences (Appendix, Table 2.6) but was absent in sequences outside the Gut spirochete group. This character, omitted from phylogenetic analysis, thus serves as independent support for the observed phylogenetic patterns.

**Figure 2.4.** Phylogenetic analysis of lower termite DNA and cDNA FDH<sub>H</sub> sequences in relationship to other FDH<sub>H</sub>, NAD-linked, and respiratory-chain linked (FDH<sub>N</sub>, FDH<sub>O</sub>) sequence types. Gut inventory sequences group with other FDH<sub>H</sub> sequences (i.e., within the dark grey box). All but one gut sequence fall into clades A and B, together forming the “Gut spirochete group” (light grey box). Clade A contains 37 sequences forming 27 operational taxonomic units; all encode selenocysteine (Sec) at the catalytic active site. Clade B contains 29 sequences, forming 20 operational taxonomic units; all encode cysteine (Cys) at the catalytic active site. Tree was constructed with the maximum likelihood algorithm Phylip PROTML based on 539 aligned amino acids. Higher termite gut metagenome sequences were added in by parsimony using 253 and 255 amino acids, respectively. Branches in bold indicate Sec-encoding FDHs. Filled circles denote nodes supported by maximum likelihood, protein parsimony, and neighbor joining methods. The scale bar corresponds to 0.1 amino acid changes per alignment position. Accession numbers for pure cultures are found in the legend of Fig. 2.2.

# Hydrogenase-linked FDH<sub>H</sub>

Gut spirochete group

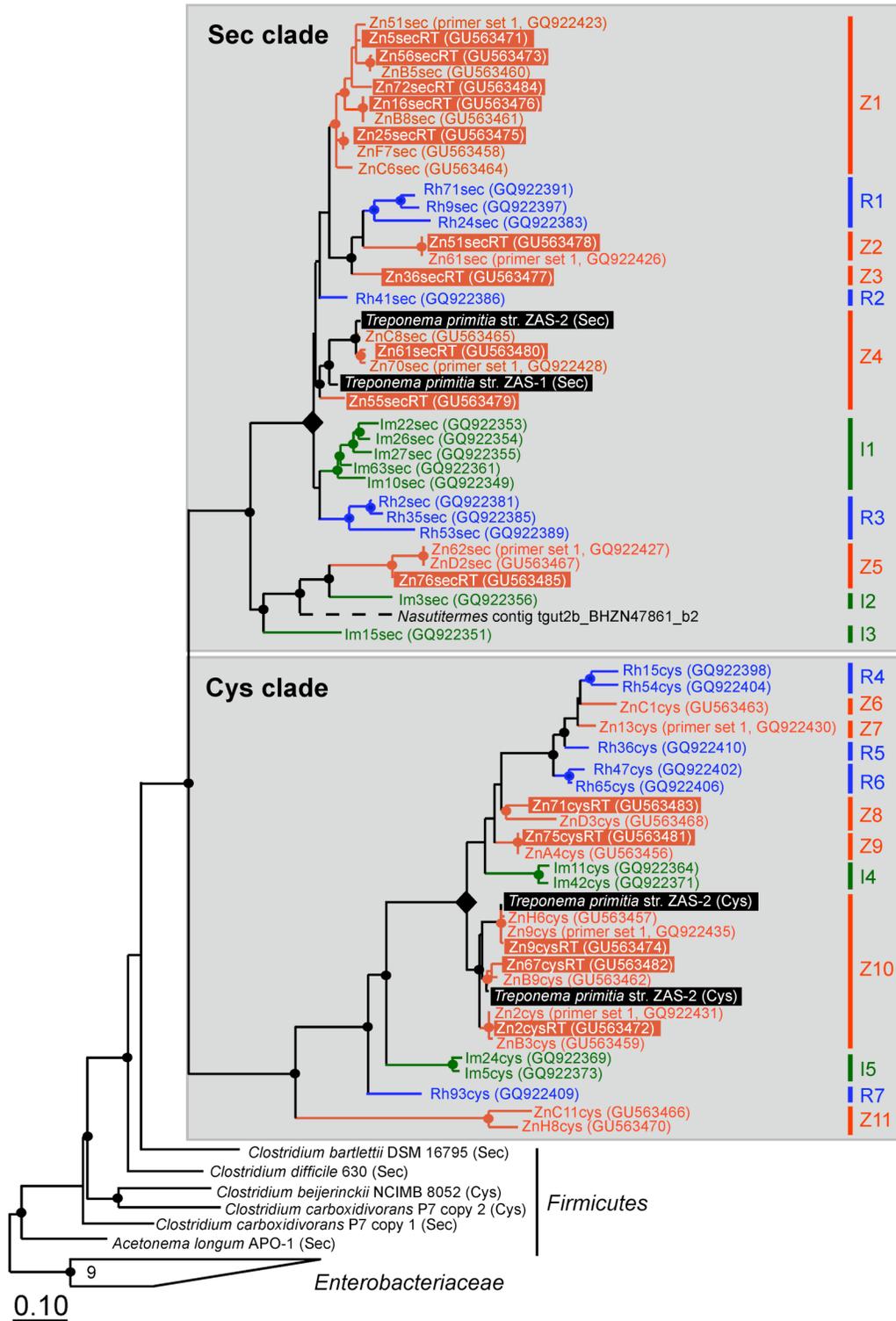


Within the Gut spirochete group, DNA and cDNA sequences could be consistently sub-grouped into one of two clades (Fig. 2.4, grouped clades A and B) based on whether a Sec or Cys amino acid is encoded at the position corresponding to the catalytically active Sec140 residue within the FDH<sub>H</sub> from *E. coli* (Axley *et al.*, 1991; Gladyshev *et al.*, 1994; Romão, 2009). Clade A is solely comprised of Sec encoding sequences whereas clade B contains only Cys encoding sequences. Hereafter, we refer to clades A and B as the “Sec” and “Cys clade.” Local phylogenetic topologies within Sec and Cys clades are shown in Fig. 2.5.

Comparison of *Z. nevadensis* gut DNA and cDNA sequence phylogeny in Fig. 2.5 indicates that several *fdhF* alleles recovered in the DNA inventory are also transcribed. The majority of the transcribed alleles (77%) encode Sec FDH<sub>H</sub>. These sequences primarily group within the monophyletic Sec sub-clades Z1 and Z4. Most of the remaining cDNA clones (23%), which encode Cys FDH<sub>H</sub>, group within the Cys sub-clade Z9 which features *T. primitia* Cys FDH<sub>H</sub>, shown to be transcribed under selenium limited conditions in pure culture (Matson *et al.*, 2010). This result implies uncultured acetogenic treponemes closely related to *T. primitia* experience selenium limitation in the termite gut and respond by transcribing genes for the selenium independent FDH<sub>H</sub>.

**Figure 2.5.** Phylogeny of termite gut Sec and Cys clade FDH<sub>H</sub> sequences (see clades A and B in Fig. 2.4). Sequences from different termite species are indicated by the following monikers and colors: ‘Zn’ and red for *Z. nevadensis*, ‘Rh’ and blue for *R. hesperus*, ‘Im’ and green for *I. minor*. Sequence names containing the moniker ‘RT’ are derived from *Z. nevadensis* gut cDNA and are highlighted in orange. Selenocysteine-encoding FDH<sub>H</sub> sequences are denoted by ‘sec’ in the sequence name; cysteine-encoding FDH<sub>H</sub> are denoted by ‘cys’. Monophyletic groups are indicated on the right side of the figure (*Z. nevadensis* clades Z1-11; *R. hesperus*, R1-R7; *I. minor*, I1-I5). Filled diamonds denote a node in Sec and Cys clades from which monophyletic groups representing each termite radiate. Tree was constructed with the maximum likelihood algorithm Phylip PROTML based on 563 aligned amino acids; a metagenome sequence fragment (dashed branch) was added in by parsimony

using 243 amino acids. Circles denote nodes supported by maximum likelihood, protein parsimony, and neighbor joining methods. The scale bar corresponds to 0.1 amino acid changes per alignment position.



Comparison of FDH<sub>H</sub> phylogeny for three termite species in Fig. 2.5 indicates sequences tend to cluster by termite of origin after grouping based on the Sec/Cys character. The phylogeny of sequences from *T. primitia*, isolated from the gut of the termite *Zootermopsis angusticollis* (Leadbetter *et al.*, 1999), is consistent with this interpretation. The grouping of *Z. nevadensis* sequences ZnHcys and Zn13cys within a clade of *R. hesperus* sequences represents a notable instance of intermingling between sequences associated with different termites. At a broader scale, a level of phylogenetic congruence between Sec and Cys clades is suggested by the radiation of approximately equal numbers of monophyletic groups (i.e., comprised of sequences from one termite species) from a robustly supported internal node (diamond shaped node in Fig. 2.5) within both Sec and Cys clades for every termite examined.

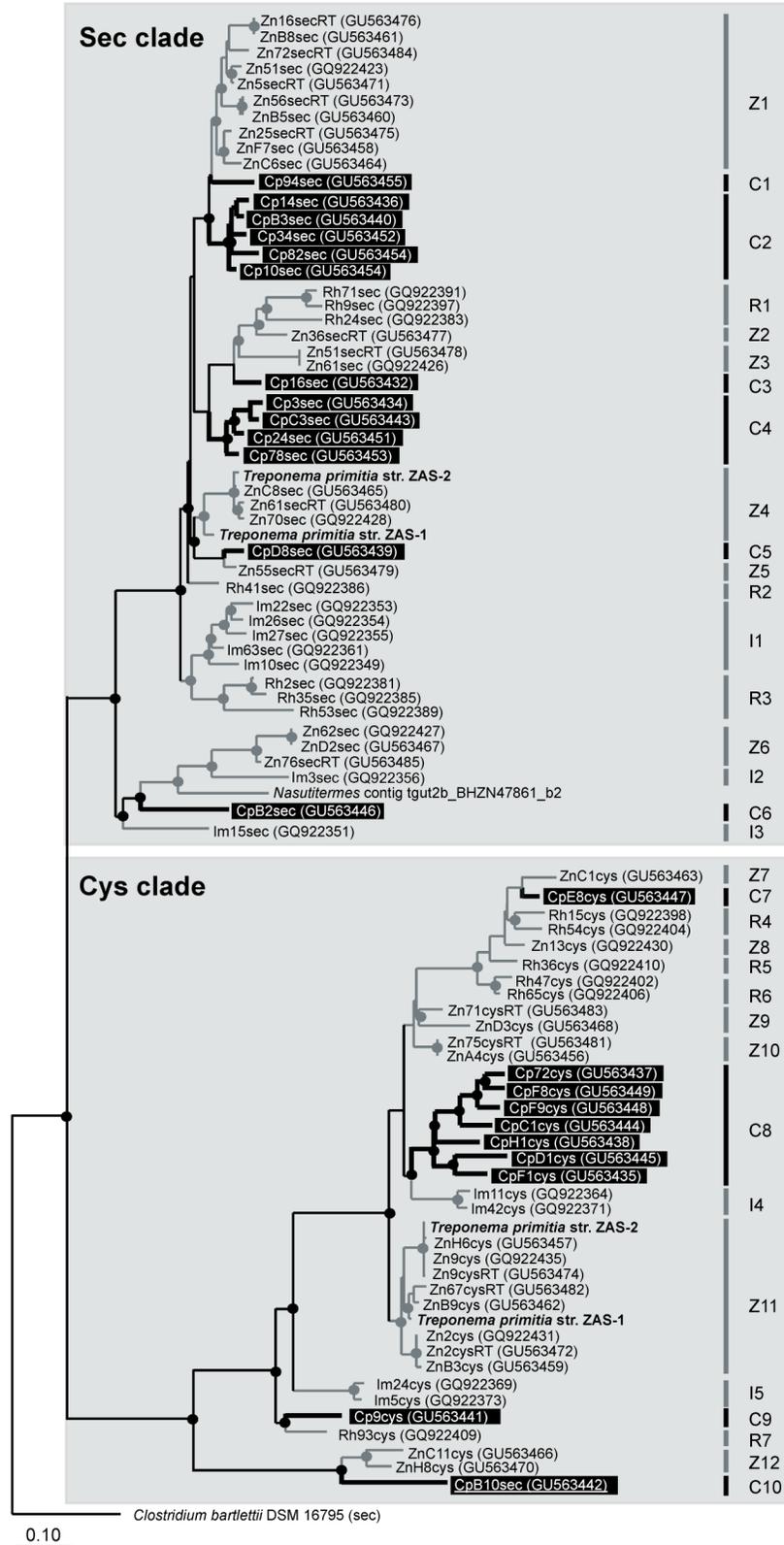
### **Phylogenetic analysis of wood-roach FDH<sub>H</sub> sequences**

The phylogeny of *C. punctulatus* FDH<sub>H</sub> sequences relative to termite FDH<sub>H</sub> sequences is shown in Fig. 2.6. Only one phylotype from wood-roach guts falls outside the Gut spirochete group (phylotype Cp28sec, Genbank no. GU563450, groups with enteric Proteobacteria FDH<sub>H</sub> sequences). This result is also supported by the lack of a polypeptide character signature (Appendix, Table 2.6). Of the 23 of 24 phylotypes clustering in the Gut spirochete group, 13 phylotypes encode Sec FDH<sub>H</sub>. These form 6 monophyletic wood-roach clades within the Sec clade of the Gut spirochete group. Wood-roach clades are distributed throughout the entire Sec clade in positions that are not consistently basal to termite-derived groups as might be expected based on the taxonomic (sister group) relationship between the two insect types. A similar observation applies to the phylogeny of the 12 remaining

phylotypes, which mostly encode Cys FDH<sub>H</sub>. These cluster within the Gut spirochete group Cys clade, forming 4 monophyletic clades. Taken together, the location of wood-roach clades relative to termite clades and the branch lengths associated with each clade indicate an evolutionary radiation of FDH<sub>H</sub> phylotypes occurred in termites and wood-roach gut communities.

The pattern of Sec FDH<sub>H</sub> sequences grouping with each other to the exclusion of Cys sequences in the Gut spirochete group has held true in termites and the wood-roach but for one exception: the phylotype CpB10sec. This wood-roach sequence encodes a catalytic Sec but is phylogenetically “Cys-like.” We note that a SECIS element required for direction Sec insertion into a nascent polypeptide could not be identified with bSECIS (Zhang and Gladyshev, 2005) or mFOLD (Zuker, 2003). Thus the CpB10sec phylotype may represent a snapshot in time of paralog evolution.

**Figure 2.6.** Phylogeny of sequences within the Gut spirochete group recovered from the wood-roach *C. punctulatus* (black boxes) and termites. Monophyletic groups (roach clades C1-C10; *Z. nevadensis*, Z1-Z12; *R. hesperus*, R1-R7; *I. minor*, I1-I5) are indicated on the right. Tree was constructed with 563 aligned amino acids using a protein maximum likelihood algorithm (Phylip PROTML). The *Nasutitermes* higher termite gut metagenome sequence was added in by parsimony. Closed circles indicate nodes were supported by distance (Fitch), parsimony (Phylip PROPARS) and Phylip PROTML methods. Scale bar represents 0.1 amino acid changes per alignment position.



## Discussion

In this study, novel degenerate primers were designed to amplify genes for hydrogenase-linked formate dehydrogenase (FDH<sub>H</sub>) enzymes encoded by *fdhF* genes. These primers were applied to nucleic acids from bacterial isolates and insect gut microbial communities. To our knowledge, this study represents the first examination of FDH<sub>H</sub> diversity in any environment. The results show that (i) the presence of previously unknown *fdhF* genes in three enteric *Gammaproteobacteria*, an H<sub>2</sub> + CO<sub>2</sub> acetogenic treponeme, and an H<sub>2</sub> + CO<sub>2</sub> firmicute, (ii) genes for both Sec and Cys FDH<sub>H</sub> variants are broadly represented in the gut communities of three phylogenetically distant lower, wood-feeding termites and a wood-feeding roach, a member of the extant sister taxon to termites, (iii) genes for both Sec and Cys FDH<sub>H</sub> variants are transcribed by gut microbial communities, and (iv) nearly all gut sequences phylogenetically group with Sec and Cys FDH<sub>H</sub> in cultured acetogenic treponemes.

Previous direct investigation of CO<sub>2</sub>-reductive acetogenesis in wood-feeding insect gut microbial communities has demonstrated the unambiguous, H<sub>2</sub> dependent conversion of <sup>14</sup>CO<sub>2</sub> to <sup>14</sup>C-formate (Breznak and Switzer, 1986). While these results point to the importance of formate dehydrogenase for gut acetogenesis, the details underlying the production of formate from H<sub>2</sub> and CO<sub>2</sub> remain largely unknown. A study of the termite gut acetogenic spirochete *T. primitia* str. ZAS-2 (Matson *et al.*, 2010) indicated that it possesses two formate dehydrogenase enzymes that are likely hydrogenase-linked and encoded by *fdhF* homologs. The present survey of *fdhF* gene diversity in wood-feeding insect guts not only underscores observations in *T. primitia*, but suggests that many uncultured acetogenic

treponemes possess genes for  $\text{FDH}_\text{H}$ . This assertion is supported by the finding that all but two gut phylotypes grouped with  $\text{FDH}_\text{H}$  from acetogenic treponeme isolates to the exclusion of all other organisms based on phylogeny (Fig. 2.6) and the presence of a distinguishing sequence character (Appendix, Table 2.6). Analysis of a recently completed draft genome for *T. primitia* str. ZAS-2 (Genbank No. CP001843) confirms the absence of additional formate dehydrogenase gene homologs, demonstrating the relevance of *fdhF* for  $\text{CO}_2$ -reductive acetogenesis in these spirochetes. As recent studies have since reinforced the long-standing view that spirochetes are responsible for much of  $\text{CO}_2$ -reductive metabolism in termite guts (Salmassi and Leadbetter, 2003; Pester and Brune, 2006), we hypothesize that  $\text{FDH}_\text{H}$  enzymes may be important for acetogenesis in  $\text{H}_2$  rich termite hindgut environments.

$\text{FDH}_\text{H}$  was not an expected feature of the Wood-Ljungdahl pathway for several reasons. First,  $\text{FDH}_\text{H}$  in *E. coli* has only been shown to operate in the direction for formate oxidation (Zinoni *et al.*, 1986; Böhm *et al.*, 1990; Hakobyan *et al.*, 2005). Indeed, assays of formate dehydrogenase activity function in the oxidative direction (Ljungdahl and Andreesen, 1978), regardless of the reaction direction *in vivo*. Secondly, only one Wood-Ljungdahl pathway formate dehydrogenase has been biochemically characterized and it is not a hydrogenase-linked enzyme (Yamamoto *et al.*, 1983). The formate dehydrogenase in the classic *Firmicute* acetogen, *Moorella thermoacetica*, is a tungsten containing selenoprotein that uses electrons from the physiological electron donor NADPH to reduce  $\text{CO}_2$  to formate (Thauer, 1972; Yamamoto *et al.*, 1983). We note that this enzyme was purified from cells grown under glucose-driven acetogenic conditions (Yamamoto *et al.*, 1983) and that a gene predicted to encode a hydrogenase-linked formate dehydrogenase in *M. thermoacetica* has

been recently identified (Pierce *et al.*, 2008). We hypothesize that *M. thermoacetica* uses its  $\text{FDH}_H$ , rather than its NADPH-linked formate dehydrogenase, for  $\text{H}_2 + \text{CO}_2$  acetogenic metabolism. Our discovery of an  $\text{FDH}_H$  gene from *A. longum*, a termite gut  $\text{H}_2 + \text{CO}_2$  acetogenic *Firmicute*, and the identification of *fdhF* in the acetogen *C. carboxidovorans* is consistent with this proposal.

Surveys of *fdhF* in insect gut microbial communities indicated that the genes for both Sec and Cys variants of  $\text{FDH}_H$  are present in each examined wood-feeding species and that, in *Z. nevadensis*, both variants are transcribed. In addition, the numbers of unique Sec and Cys phylotypes recovered from each gut environment were not statistically different. One possible interpretation is that, like *T. primitia*, many other gut microbes possess genes for both Sec and Cys  $\text{FDH}_H$  variants and differentially transcribe them in response to fluctuations in selenium availability. Alternatively, the results could point to the existence of organisms that have specialized to using one or the other variant. In either case, the broad representation of both Sec and Cys variants in gut communities suggests the trace element selenium plays an important role in shaping the genomes of microbes inhabiting the guts of wood-feeding insects.

The phylogenetic separation of Gut spirochete group sequences into Sec and Cys sub-clades (Figs. 2.5, 2.6) has important implications for gene evolution in gut communities. The aforementioned tree topology suggests the duplication of an ancestral *fdhF* gene into Sec and Cys encoding forms occurred once, as the independent innovation of *fdhF*<sub>Sec</sub> and *fdhF*<sub>Cys</sub> in each examined insect lineage would result in sequences that cluster by insect of origin

before they cluster by the Sec/Cys character. This duplication appears to be followed by evolutionary radiations in several basal wood-feeding insect taxa. The presence of long branches near the base of each Sec and Cys clade suggests the duplication event may have occurred early during the evolution of lignocellulose-fermenting, insect gut microbial communities, perhaps in the wood-feeding progenitor to termites and *C. punctulatus*. The absence of *fdhF* in PCR assays with *P. americana* gut community DNA appears to support this hypothesis, but more extensive study in *Blattidae* is required as only one roach individual was used in this study. Successful radiation of dual *fdhF*<sub>Sec</sub> and *fdhF*<sub>Cys</sub> genes is not confined to wood-feeding insect gut communities. Based on the long branches and deep node separating Sec and Cys clades, the (convergent) invention of dual genes in the *Enterobacteriaceae* line of descent (Fig. 2.2) may also have been an early event during gut community evolution in mammals.

This study on hydrogenase-linked formate dehydrogenase enzyme diversity in wood-feeding insect gut microbial communities yields several insights into the physiological ecology of uncultured gut microbes. First, FDH<sub>H</sub> enzymes are predicted to play key roles in the metabolism of many uncultured acetogenic treponemes. Second, the results suggest that selenium availability has shaped the gene content of gut microbial communities in wood-feeding insects representing three different termite families and the sister lineage of termites. Third, it is likely that *fdhF*<sub>Sec</sub> and *fdhF*<sub>Cys</sub> variants have been maintained over long time scales in gut microbial communities, possibly since the divergence of termites from roaches over 100 mya (Grimaldi and Engel, 2005). Further studies are required to determine whether their presence and transcription in gut microbial communities is due to changes in selenium

levels in the insect host's diet, local changes in selenium concentration or redox state in the termite gut, or some other selective feature of lignocellulose-fermenting insect guts.

## **Acknowledgements**

This research was supported by NSF grant DEB-0321753 (JRL) and a NSF predoctoral fellowship (XZ). We thank J. Switzer-Blum for the microbial isolates *Citrobacter sp.* str. TSA-1 and J. Breznak for *S. grimesii* str. ZFX-1. We thank previous and current members of the Leadbetter lab for their helpful discussions and comments.

## Appendix

**Table 2.2.** Nucleotide accession numbers of sequences used for *fdhF* primer design.

**Table 2.3.** PCR primer combinations for *fdhF* amplification from pure culture and insect gut templates.

**Table 2.4.** FDH phylotype distribution in lower termite gut DNA, cDNA, and wood roach gut DNA.

**Table 2.5.** Summary of *fdhF* inventories generated from termite hindgut DNA, cDNA, and wood-roach hindgut DNA.

**Table 2.6.** Amino acid alignment in the area of a characteristic amino acid indel (bold) found only in Gut spirochete group FDH<sub>H</sub> sequences.

**Figure 2.7.** Mitochondrial cytochrome oxidase II phylogeny of insects.

**Figure 2.8.** Rarefaction curves each insect gut *fdhF* DNA or cDNA inventory.

**Table 2.2.** Nucleotide accession numbers of sequences used for *fdhF* primer design (see Figure 2.1).

Source	Accession Number <sup>1</sup>
<i>Aeromonas salmonicida</i> subsp. <i>salmonicida</i> A449	NC_009348.1:1906100-1908244
<i>Aggregatibacter aphrophilus</i> NJ8700	NC_012913.1:c1159571-1157412
<i>Citrobacter koseri</i> ATCC BAA-895 copy 1	NC_009792.1:3531364-3533511
<i>Citrobacter koseri</i> ATCC BAA-895 copy 2	NC_009792.1:1727418-1729565
<i>Citrobacter rodentium</i> ICC168 copy 1	NC_013716.1:c3662542-3660395
<i>Citrobacter rodentium</i> ICC168 copy 2	NC_013716.1:c3568359-3566212
<i>Citrobacter</i> sp. 30_2 copy 1	NZ_GG657366.1:c93031-90884
<i>Citrobacter</i> sp. 30_2 copy 2	NZ_GG657366.1:c1094197-1096347
<i>Citrobacter youngae</i> ATCC 29220 copy 1	NZ_ABWL01000021.1:c93031-90884
<i>Citrobacter youngae</i> ATCC 29220 copy 2	NZ_ABWL01000021.1:c24883-27030
<i>Clostridium bartlettii</i> DSM 16795	NZ_ABEZ02000007.1:c36324-34174
<i>Clostridium beijerinckii</i> NCIMB 8052	NC_009617.1:c4364248-4366389
<i>Clostridium bolteae</i> ATCC BAA-613	NZ_ABCC02000017.1:93731-95716
<i>Clostridium carboxidivorans</i> P7 copy 1	NZ_ACVI01000105.1:231-2378
<i>Clostridium carboxidivorans</i> P7 copy 2	NZ_ACVI01000010.1:36001-38157
<i>Clostridium difficile</i> 630	NC_009089.1:c3884230-3882086
<i>Cronobacter sakazakii</i> ATCC BAA-894	NC_009778.1:c1996280-1998430
<i>Cronobacter turicensis</i> copy 1	NC_013282.1:2002311-2004458
<i>Cronobacter turicensis</i> copy 2	NC_013282.1:1996635-1998845
<i>Dickeya dadantii</i> Ech586	NC_013592.1:2958853-2961003
<i>Dickeya dadantii</i> Ech703	NC_012880.1:c1450903-1453053
<i>Dickeya zeae</i> Ech1591	NC_012912.1:3084906-3087056
<i>Edwardsiella ictaluri</i> 93-146	NC_012779.1:3156478-3158622
<i>Edwardsiella tarda</i> EIB202	NC_013508.1:3053142-3055286
<i>Enterobacter cancerogenus</i> ATCC 35316	NZ_ABWM02000022.1:21042-23189
<i>Enterobacter</i> sp. 638 copy 1	NC_009436.1:c 329787-331934
<i>Enterobacter</i> sp. 638 copy 2	NC_009436.1:c1907448-1909598
<i>Escherichia coli</i> K-12 substr MG1655	NC_000913.2:c4295242..4297389
<i>Escherichia fergusonii</i> ATCC 35469	NC_011740.1:4397249..4399396
<i>Klebsiella pneumoniae</i> NTXH-K2044 copy 1	NC_012731.1:c358869-356722
<i>Klebsiella pneumoniae</i> NTXH-K2044 copy 2	NC_012731.1:3017444..3019594
<i>Pantoea</i> sp. At-9b	NZ_ACYJ01000001:122540..124690
<i>Pectobacterium atrosepticum</i> SCRI1043 copy 1	NC_004547.2:c1752061..1754157
<i>Pectobacterium atrosepticum</i> SCRI1043 copy 2	NC_004547.2:1420602..1422752
<i>Pectobacterium carotovorum</i> sbsp. <i>carotovorum</i> WPP14	NZ_ABVY01000027.1:c9266..11416
<i>Pectobacterium wasabiae</i> WPP163	NC_013421.1:c1930748..1932898
<i>Photobacterium profundum</i> 3TCK	NZ_AAPH01000003.1:97396-99486
<i>Proteus mirabilis</i> HI4320 copy 1	NC_010554.1:3909884-3912028
<i>Proteus mirabilis</i> HI4320 copy 2	NC_010554.1:c3265604..3267772
<i>Providencia alcalifaciens</i> DSM 30120	NZ_ABXW01000042.1:35044-37197

---

<i>Providencia rustigianii</i> DSM 4541	NZ_ABXV02000023.1:88004-90157
<i>Psychromonas</i> sp. CNPT3	NZ_AAPG01000013.1:c3595..5742
<i>Salmonella enterica</i> sbsp. <i>enterica</i> serovar Typhi CT18	NP_458584; NC_003198.1:4370484..4372631
<i>Salmonella typhimurium</i> LT2	NP_463150; NC_003197.1:c4525350..4527497
<i>Serratia proteamaculans</i> 568	NC_009832.1:c2657681..2659837
<i>Shigella</i> sp. D9	NZ_ACDL01000041.1:c37225..39372
<i>Treponema primitia</i> str. ZAS-2 copy 1 (Sec FDH)	FJ479768:50505..52697
<i>Treponema primitia</i> str. ZAS-2 copy 2 (Cys FDH)	FJ479768:30735..32933
<i>Vibrio angustum</i> S14	NZ_AAOJ01000001.1:c1074316..1076460
<i>Yersinia aldovae</i> ATCC 35236	NZ_ACCB01000002.1:136225..138372
<i>Yersinia bercovieri</i> ATCC 43970	NZ_AALC02000017.1:13658..15805
<i>Yersinia enterocolitica</i> subsp. <i>enterocolitica</i> 8081	NC_008800.1:3050211..3052358
<i>Yersinia frederiksenii</i> ATCC 33641 copy 1	NZ_AALE02000011.1:c133500..135647
<i>Yersinia frederiksenii</i> ATCC 33641 copy 2	NZ_AALE02000004.1:63404..65548
<i>Yersinia mollaretii</i> ATCC 43969 copy 1	NZ_AALD02000005.1:c25400..27571
<i>Yersinia mollaretii</i> ATCC 43969 copy 2	NZ_AALD02000036.1:52..2196
<i>Yersinia rohdei</i> ATCC 43380	NZ_ACCD01000002.1:c116227..118374
<i>Yersinia ruckeri</i> ATCC 29473	NZ_ACCC01000020.1:c42838..44961

---

<sup>1</sup> A 'c' before genome coordinates indicates complementary sequence

**Table 2.3.** PCR primer combinations for *fdhF* amplification from pure culture and insect gut templates. Sequenced amplicons were classified as *fdhF*<sub>Sec</sub> ('Sec') or *fdhF*<sub>Cys</sub> ('Cys') versions of *fdhF* based on whether their deduced amino acid translations encode a selenocysteine or cysteine, respectively, at the catalytic active site. All templates are DNA unless noted. Primer set 1: fdhF-F1, fdhF-F2, fdhF-F3, fdhF-R1, fdhF-R2. Primer set 2: universal primers EntfdhFunv- F1, TgfdhFunv-F1, and fdhFunv-R1.

Templates	Primer Combinations (μM)	Amplicon
<i>Treponema primitia</i> str. ZAS-2	fdhF-F1 (1.0), fdhF-R1 (1.0)	Sec
<i>T. primitia</i> str. ZAS-2	fdhF-F1 (1.0), fdhF-R2 (1.0)	Cys
<i>T. primitia</i> str. ZAS-1	fdhF-F1 (1.0), fdhF-R1 (1.0)	Sec
<i>T. primitia</i> str. ZAS-1	fdhF-F1 (1.0), fdhF-R2 (1.0)	Cys
<i>Buttiauxiella</i> sp. SN-1	fdhF-F3 (1.0), fdhF-R2 (1.0)	Sec
<i>Serratia grimesii</i> str. ZFX-1	fdhF-F2 (1.0), fdhF-R1 (1.0)	Cys
<i>Citrobacter</i> sp. TSA-1	Primer set 2 forward (0.5), reverse (1.0)	Sec
<i>Acetonema longum</i> str. APO-1	Primer set 2 forward (0.5), reverse (1.0)	Sec
<i>Pantoea stewartii</i> subsp. <i>stewartii</i>	Primer set 2 forward (0.5), reverse (1.0)	no product
<i>Zootermopsis nevadensis</i> collection ChiA1 gut DNA	Primer set 1 forward (0.3), reverse (0.3)	Sec, Cys
<i>Zootermopsis nevadensis</i> collection ChiA1 gut DNA	Primer set 2 forward (0.5), reverse (1.0)	Sec, Cys
<i>Zootermopsis nevadensis</i> collection ChiB gut cDNA	Primer set 2 forward (0.5), reverse (1.0)	Sec, Cys
<i>Reticulitermes hesperus</i> collection ChiA2 gut DNA	Primer set 2 forward (0.5), reverse (1.0)	Sec, Cys
<i>Incisitermes minor</i> isolate collection Pas1 gut DNA	Primer set 2 forward (0.5), reverse (1.0)	Sec, Cys
<i>Cryptocercus punctulatus</i> nymph gut DNA	Primer set 2 forward (1.0), reverse (1.0)	Sec, Cys

**Table 2.4.** FDH phylotype distribution in lower termite gut DNA, cDNA, and wood roach gut DNA. Phylotypes were identified using DOTUR at a cutoff of 97% protein similarity level (Jones-Thorton-Taylor corrected). The number of genotypes, inferred based on RFLP sorting, comprising each phylotype are listed. PS1: primers fdhF-F1, fdhF-F2, fdhF-F3, fdhF-R1, fdhF-R2. PS 2: universal primers EntfdhFunv-F1, TgfdhFunv-F1, and fdhFunv-R1.

<b>Clone Library (No. clones, primer set)</b>	<b>Phylotype (No. genotypes)</b>	<b>Abundance</b>
<i>Zootermopsis nevadensis</i> ChiA1 (84, PS 1)	Zn9cys (10)	42.9%
	Zn2cys (1)	22.6%
	Zn70sec (6)	17.9%
	Zn62sec (1)	13.1%
	Zn13cys (1)	1.2%
	Zn51sec (1)	1.2%
	Zn61sec (1)	1.2%
<i>Zootermopsis nevadensis</i> ChiA1 (86, PS 2)	ZnC1cys (2)	45.3%
	ZnD2sec (1)	23.3%
	ZnF7sec (3)	4.7%
	ZnH6cys (2)	3.5%
	ZnB3cys (1)	3.5%
	ZnB5sec (3)	3.5%
	ZnB8sec (1)	2.3%
	ZnC6sec (1)	2.3%
	ZnD3cys (1)	2.3%
	ZnA4cys (2)	2.3%
	ZnC8sec (2)	2.3%
	ZnB9cys (1)	1.2%
	ZnC11cys (1)	1.2%
	ZnE2cys (1)	1.2%
	ZnH8cys (1)	1.2%
<i>Zootermopsis nevadensis</i> ChiB (81, PS 2)	Zn5secRT (2)	29.6%
	Zn16secRT (2)	19.8%
	Zn2cysRT (3)	12.3%
	Zn25secRT (6)	11.1%
	Zn9cysRT (3)	6.2%
	Zn56secRT (2)	3.7%
	Zn55secRT (1)	2.5%
	Zn67cysRT (1)	2.5%
	Zn71cysRT (1)	2.5%
	Zn75cysRT (3)	2.5%
	Zn76secRT (1)	2.5%
	Zn36secRT (1)	1.2%
	Zn51secRT (1)	1.2%
	Zn61secRT (1)	1.2%
	Zn72secRT (1)	1.2%

---

*Reticulitermes hesperus* ChiA2 (89, PS 2)

Rh36cys (6)	30.3%
Rh2sec (5)	28.1%
Rh9sec (1)	10.1%
Rh15cys (3)	7.9%
Rh41sec (5)	6.7%
Rh24sec (2)	5.6%
Rh35sec (2)	2.2%
Rh53sec (2)	2.2%
Rh54cys (2)	2.2%
Rh47cys (1)	1.1%
Rh65cys (1)	1.1%
Rh71sec (1)	1.1%
Rh93cys (1)	1.1%

*Incisitermes minor* Pas1 (80, PS 2)

Im5cys (5)	18.8%
Im26sec (3)	17.5%
Im15sec (5)	16.3%
Im11cys (7)	13.8%
Im27sec (2)	10.0%
Im42cys (2)	6.3%
Im10sec (2)	5.0%
Im22sec (1)	5.0%
Im24cys (2)	2.5%
Im3sec (1)	2.5%
Im63sec (2)	2.5%

*Cryptocercus punctulatus nymph* (136, PS 2)

Cp16sec (11)	21.3%
Cp10sec (12)	17.6%
Cp3sec (4)	9.6%
CpF1cys (6)	8.8%
Cp14sec (4)	8.1%
Cp72cys (4)	4.4%
CpH1cys (2)	3.7%
CpD8sec (1)	2.9%
CpB3sec (3)	2.9%
Cp9cys (2)	2.9%
CpB10sec (1)	2.9%
CpC3sec (2)	2.9%
CpC1cys (1)	1.5%
CpD1cys (1)	1.5%
CpB2sec (1)	1.5%
CpE8cys (1)	1.5%
CpF9cys (1)	1.5%
CpF8cys (1)	1.5%
Cp28sec (1)	1.5%
Cp24sec (1)	1.5%
Cp34sec (1)	1.5%
Cp78sec (1)	1.5%
Cp82sec (1)	1.5%
Cp94sec (1)	1.5%

---

**Table 2.5.** Summary of *fdhF* inventories generated from termite hindgut DNA, cDNA, and wood-roach hindgut DNA.

Clone library templates	Sample Type	No. Clones Analyzed	No. of OTU <sup>1</sup>	Mean Chao1 (SD) <sup>2</sup>	95% LCI, HCI <sup>3</sup>	No. Sec, Cys OTU <sup>4</sup>
<i>Zootermopsis nevadensis</i> collection ChiA1	DNA	84	7	7.53 (1.88)	6.61, 17.23	4, 3
<i>Zootermopsis nevadensis</i> collection ChiA1	DNA	86	15	14.96 (2.78)	13.11, 27.81	6, 9
<i>Zootermopsis nevadensis</i> collection ChiB	cDNA	81	15	14.78 (2.52)	13.20, 26.79	10, 5
<i>Reticulitermes hesperus</i> collection ChiA2	DNA	89	13	13.66 (3.86)	11.49, 33.15	7, 6
<i>Incisitermes minor</i> isolate collection Pas1	DNA	80	11	10.92 (0.62)	10.69, 13.6	7, 4
<i>Cryptocercus punctulatus</i> nymph <sup>5</sup>	DNA	136	24	21.52 (2.97)	21.52, 37.03	15, 9

<sup>1</sup> Number of operational taxonomic units (OTU) determined using DOTUR (Schloss and Handelsman, 2005) based on > 3% amino acid distance between different phylotypes.

<sup>2</sup> Bias-corrected Chao1 diversity estimator calculated using EstimateS (Colwell, 2009) based on 100 randomizations, sampling without replacement; SD = standard deviation.

<sup>3</sup> Lower (LCI) and higher (HCI) 95% confidence interval limits for mean Chao1 as calculated by EstimateS.

<sup>4</sup> Number of unique Sec and Cys FDH<sub>H</sub> phylotypes.

<sup>5</sup> Sequences derived from PCR at different annealing temperatures (51°C and 57 °C) were combined for analyses.

**Table 2.6.** Amino acid alignment in the area of a characteristic amino acid indel (bold) found only in Gut spirochete group FDH<sub>H</sub> sequences. The alignment corresponds to amino acids 394-420 in the selenocysteine encoding FDH<sub>H</sub> of *T. primitia* str. ZAS-2. Sequences are listed in phylogenetic order (see Figure 2.6).

Sequence	Amino Acid Alignment
Zn16secRT	LSDQPGITLTLVPHHVL <b>HEKDP</b> AKQIHAYYIMGEDPGQSDPD
ZnB8sec	LSDQPGITLTLVPHHVL <b>HEKDP</b> AKQIHAYYIMGEDPGQSDPD
Zn72secRT	LSDQPGITLTLVPHHVL <b>HEKDP</b> AKQIHAYYIMGEDPGQSDPD
Zn51sec	LSDQPGITLTLVPHHVL <b>HEKDP</b> AKQIHAYYIMGEDPGQSDPD
Zn5secRT	LSDQPGITLTLVPHHVL <b>HEKDP</b> AKQIHAYYIMGEDPGQSDPD
Zn56secRT	LSDKPGITLTLVPHHVL <b>HEKDP</b> TKQIHAYYIMGEDPGQSDPD
ZnB5sec	LSDKPGITLTLVPHHVL <b>HEKDP</b> TKQIHAYYIMGEDPGQSDPD
Zn25secRT	LSDKPGITLTAVPHQVL <b>HEKDP</b> AKQIHAYYIMGEDPGQSDPD
ZnF7sec	LSDKPGITLTAVPHQVL <b>HEKDP</b> AKQIHAYYIMGEDPGQSDPD
ZnC6sec	LSDKAGITLTLVPHHVL <b>HEKDP</b> AKQIHAYYIMGEDPGQSDPD
Cp94sec	LSDQLGITLTTVPHHVL <b>HEKDP</b> KKRIHAYYIMGEGPGQSDPD
Cp14sec	LSDQPGITLTVPHQVL <b>HEKDP</b> AKQIHAYYIMGEDPGQSDPD
CpB3sec	LSDQPGITLTVPHHVL <b>HEKDP</b> AKQIHAYYIMGEDPGQSDPD
Cp34sec	LSDQPGITLTVPHHVL <b>HEKDP</b> AKQIHAYYIMGEDPGQSDPD
Cp82sec	LSDQPGITLTVPHHVL <b>HEKDP</b> AKQIHAYYIMGEDPGQSDPD
Cp10sec	LSDQPGITLTVPHHVL <b>HEKDP</b> AKQIHAYYIMGEDPGQSDPD
Rh71sec	LSPDVGITLTTVPHQVL <b>HETDP</b> KKKIHAYYIMGEDPAQSDPD
Rh9sec	LSPDVGITLTTVPHQVL <b>HETDP</b> KKQIHAYYIMGEDPAQSDPD
Rh24sec	LSADIGITLTTVPHQVL <b>HEKDP</b> KKQIHAYYIMGEDPGQSDPD
Zn36secRT	LSDQPGITLTTVPHQVL <b>HETDP</b> RKQIHAYYIMGEDPGQSDPD
Zn51secRT	LSPDLGITLTTVPHQVL <b>HEKDP</b> KKQIHAYYIMGEDPGQSDPD
Zn61sec	LSPDLGITLTTVPHQVL <b>HEKDP</b> KKQIHAYYIMGEDPGQSDSD
Cp16sec	LSDKLGITLTTVPHQVL <b>HETDP</b> TKQIHAYYIMGEDPGQSDPD
Cp3sec	LSDKAGITLTMVPHQVL <b>AEKDP</b> AKKIHAYYIMGEDPGQSDPD
CpC3sec	LSDKAGTTLTMVPHQVL <b>AETDP</b> AKKIHAYYIMGEDPGQSDPD
Cp24sec	LSDKPGITLTMVPHQVL <b>AETDP</b> AKKIHAYYIMGEDPGQSDPD
Cp78sec	LSDKPGITLTMVPHQVL <b>AETDP</b> AKKIHAYYIMGEDPGQSDPD
<i>Treponema primitia</i> str. ZAS-2 fdhFsec	LSDKPGITLTVPHHVL <b>HEKDP</b> TKQIHAYYIMGEDPVQSDPD
ZnC8sec	LSDKPGITLTVPHHVL <b>HEKDP</b> TKQIHAYYIMGEDPVQSDPD
Zn61secRT	LSDKPGITLTVPHHVL <b>HEKDP</b> TKQIHAYYIMGEDPVQSDPD
Zn70sec	LSDKPGITLTVPHHVL <b>HEKDP</b> TKQIHAYYIMGEDPVQSDPD
<i>Treponema primitia</i> str. ZAS-1 fdhFsec	LSDKPGITLTVPHHVL <b>HETDP</b> AKQIHAYYIMGEDPVQSDPD
CpD8sec	LSDQAGITLTVPHHVL <b>HEKDP</b> AKQIHAYYIMGEDPVQSDPD
Zn55secRT	LSDKAGITLTVPHQVL <b>HEKDP</b> KKQIHAYYIMGEDPVQSDPD
Rh41sec	LSDQPGITLTVPHHVL <b>HETDP</b> AKQIHAYYIMGEDPAQSDPD
Im22sec	LSGEPGITLTTVPQRVL <b>HEKDP</b> AKHIRAYYVMGEDPAQSDPD
Im26sec	LSDQPGITLTMVPHQVL <b>HEKDP</b> AKKIRGYIIMGEDPAQSDPD

---

Im27sec	LSDQPGITLTMVPHHVL <b>HEKDP</b> AKQIHAYYVLGEDPAQSDPD
Im63sec	LSDQPGITLTMVPHHVL <b>HEKDP</b> AKKIRAYYIMGEDPAQSDPD
Im10sec	LSDQAGITLTGVPHQVL <b>HETDP</b> AKKIRAYYIMGEDPAQSDPD
Rh2sec	LPDQNGITLTVVPHQVL <b>HEKDP</b> TKQIHGYYIMGEDPVQSDPD
Rh35sec	LPDQNGITLTVVPHQVL <b>HETDP</b> AKKIHGYYIMGEDPVQSDPD
Rh53sec	LPAENGITLTVVGRVL <b>HEKDP</b> SKQIHAYYIMGEDPVQSDPD
Zn62sec	LSNKIGIPLTQVPHYVL <b>HETE</b> -EKKIRAYYIFGEDPAQSDPD
ZnD2sec	LSNKIGIPLTQVPHYVL <b>HETE</b> -EKKIRAYYIFGEDPAQSDPD
Zn76secRT	LSNKIGIPITQVPRYVL <b>HEPE</b> -EKKIRAYYIFGEDPAQSDPD
Im3sec	LPEKVGIPLTQVPHYVL <b>HEPE</b> -ERKIRAYYVFGEDPAQSDPD
CpB2sec	LSDKVGITLTKVPHHVL <b>HEKG</b> -AKKIHAYYIMGEDPAQSDPD
Im15sec	LSDKVGCPIPTHVPHRVL <b>HEKDP</b> AKRIHAYYIFGEDPAQSDPD
ZnClcys	LDNKVGIQLTRIPFVI <b>HEKNP</b> ANRIHAYYITGEDPAQSDPD
ZnHcys	LDNKVGIQLTRIPFVI <b>HEKNP</b> ANRIHAYYITGEDPAQSDPD
CpE8cys	LDNKVGIQLTRIPFVI <b>HEQDP</b> AKRIHAYYITGEDPAQSDPD
Rh15cys	LDNKVGIQLTRIAEFTI <b>HQKDP</b> AKRIHAYYITGEDPAQSDPD
Rh54cys	LDNKVGIQLTRIAEFTI <b>HQADP</b> AKRIHAYYITGEDPAQSDPD
Zn13cys	LDNKVGVQLTRIPELVL <b>HEKDP</b> AKRIHAYYITGEDPAQSDPD
Rh36cys	LDNKVGIQLTRIPELVI <b>HEKDP</b> AKRIHAYYITGEDPAQSDPD
Rh47cys	LDNKVGIQLTRIPFVL <b>HEKDP</b> AKRIHAYYITGEDPAQSDPD
Rh65cys	LDNKVGIQLTRIPFVI <b>HEKDP</b> AKRIHAYYITGEDPAQSDPD
Zn71cysRT	LDNKVGLQLTRVPEFVL <b>QEKDP</b> AKQIHAYYITGEDPAQSDPD
ZnD3cys	LDNKVGLQLTRVPEFVL <b>HEKDP</b> AKCIHAYYITGEDPAQSDPD
Zn75cysRT	LDNKVGIQLTRVPEFVL <b>HEKDP</b> KKQLHAYYITGEDPAQSDPD
ZnA4cys	LDNKVGIQLTRVPEFVL <b>HEKDP</b> KKQLHAYYITGEDPAQSDPD
Cp72cys	LDNKVGIQLTRVPEFVI <b>HEKDP</b> AKRIHAYYITGEDPAQSDPD
CpF8cys	LDNKVGIQLTRVPEFVI <b>HEKDP</b> AKRIHAYYITGEDPAQSDPD
CpF9cys	LDNKVGIQLTRVPEFVI <b>HDKDP</b> AKRIHAYYITGEDPAQSDPD
CpClcys	LDNKVGIQLTRVPEFVV <b>HEKDP</b> AKRIHAYYITGEDPAQSDPD
CpH1cys	LDNKVGIQLTRVPEFVI <b>HEKDP</b> AKRIHAYYITGEDPAQSDPD
CpD1cys	LDNQVGLQLTRVPEFVL <b>HEKDP</b> KKRIHAYYITGEDPAQFDPD
CpF1cys	LDNKVGIQLTRVPEFVI <b>HEKDP</b> AKRIHAYYITGEDPAQSDPD
Im11cys	LDDKVGIQLTRVPEFVQ <b>HMSDP</b> AKRLHAYYITGEDPCQSDPD
Im42cys	LDDKAGIQLTRVPEFVQ <b>HESDP</b> AKRIHAYYITGEDPCQSDPD
<i>Treponema primitia</i> str. ZAS-2 fdhFcys	LSNKAGIQLTRVPEFVI <b>HEKDP</b> AKRIHAYYITGEDPAQSDPD
ZnH6cys	LSNKVGLQLTRVPEFVI <b>HEKDP</b> AKRIHAYYITGEDPAQSDPD
Zn9cys	LSNKAGIQLTRVPEFVI <b>HEKDP</b> AKRIHAYYITGEDPAQSDPD
Zn9cysRT	LSNKAGIQLTRVPEFVI <b>HEKDP</b> AKRIHAYYITGEDPAQSDPD
Zn67cysRT	LSNKAGIQLTRVPEFVI <b>HEKDP</b> AKRIHAYYITGEDPAQSDPD
ZnB9cys	LSNKAGIQLTRVPEFVI <b>HEKDP</b> AKRIHAYYITGEDPAQSDPD
<i>Treponema primitia</i> str. ZAS-1 fdhFcys	LSNKAGIQLTRVPEFVI <b>HEKDP</b> AKRIHAYYITGEDPAQSDPD
Zn2cys	LSNKVGIQLTRVPEFVI <b>HEKDP</b> AKRIHAYYITGEDPAQSDPD
Zn2cysRT	LSNKVGIQLTRVPEFVI <b>HEKDP</b> AKRIHAYYITGEDPAQSDPD

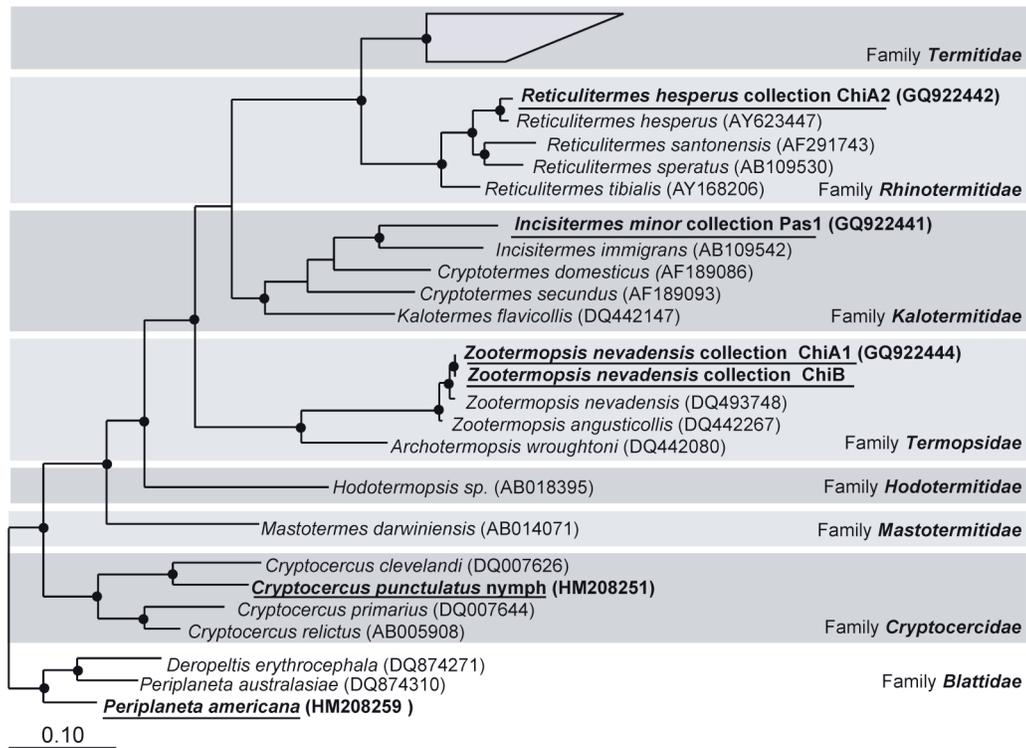
---

---

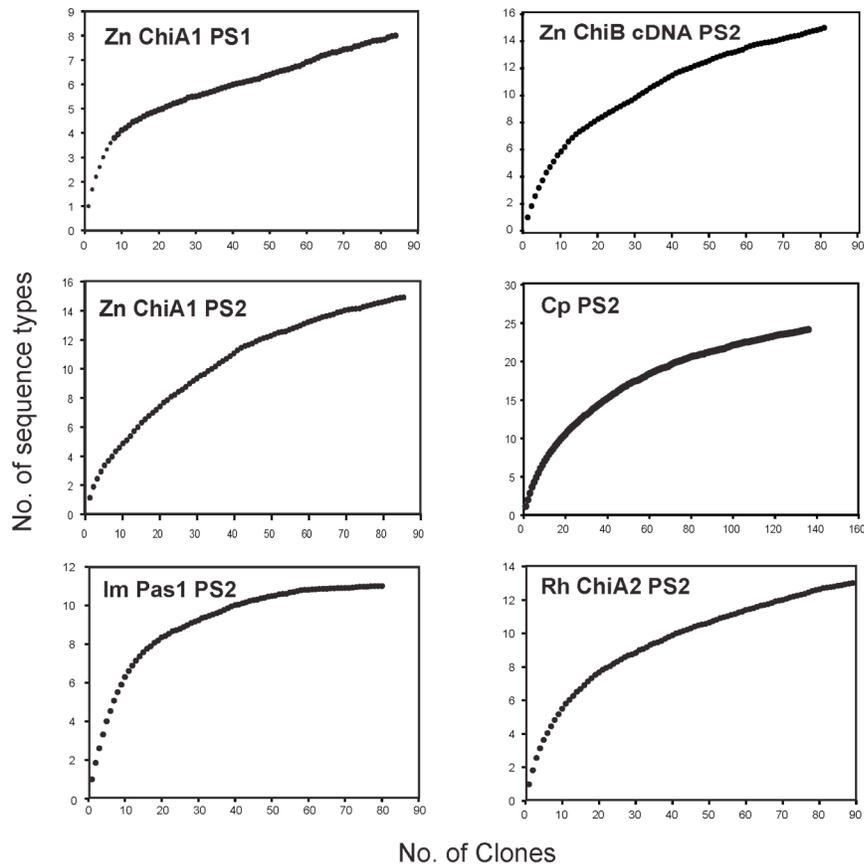
ZnB3cys	LSNKVGIQLTRVPEFV <b>HEKDP</b> AKRIHAYYITGEDPAQSDPD
Im24cys	LSDKVGLALTRVPERVL <b>HEEDP</b> AKRIHAYYIFGEDPGQSDPD
Im5cys	LSDKVGLALTRVPERVL <b>HEEDP</b> AKRIHAYYIFGEDPGQSDPD
Cp9cys	LPPEVGLQLTRVPEKVI <b>HEKDP</b> AKRIHAYYIFGEDPAQSDPD
Rh93cys	LPDQPLQLTRVPERVV <b>HEKDP</b> AKQIHAYYIFGEDPAQSDPD
ZnC11cys	LSPAVGLHVTRVPEFVL <b>DPPEE</b> AKRIHAYYVYGEDLAHSDPN
ZnH8cys	LSPTVGLHVTRVPEFVL <b>KEPDP</b> AKQIHAYYVYGEDPAHSDPN
CpB10sec	LSPNAGLHVTRVPEHVLE <b>PPSPE</b> KAIHGYYVYGEDPAHSDPN
<i>Clostridium bartlettii</i> DSM 16795	LPSPVGLKLTVEVPHAVLE----EHKIKAYYIFGEDPVQSDPD
<i>Clostridium difficile</i> 630	LSPNNGYSLTQVPNLVLK----EKKLKAYYIFGEDPVQSDPD
<i>Clostridium beijerinckii</i> NCIMB 8052	LSDKNGYFLTQVPELVLK----EDKIKAYYIFGEDPVQSDPN
<i>Clostridium carboxidivorans</i> P7 copy 2	LSDKVGYHLTEVPKLVLK----ENKLKAYYIMGEDTVQSDPN
<i>Clostridium carboxidivorans</i> P7 copy 1	LPNKVGYHLTEVPHLVLK----EDKIKAYYIMGEDPVQSDPD
<i>Acetonea longum</i> APO-1	LPAKPGYHLTEVPHLAR-----EGKIKAYYIFGEDPVQSDPD
<i>Citrobacter koseri</i> ATCC BAA-895	LPAHTGYRISELPHRAA-----HGEVRAAYIMGEDPLQTDAE
<i>Escherichia coli</i> str K-12	LPAHTGYRISELPHRAA-----HGEVRAAYIMGEDPLQTDAE
<i>Buttiauxiella</i> SN1	LPAHTGYRISELPHRVA-----HGEVYAAYIMGEDPLQTDAE
<i>Proteus mirabilis</i> HI4320	MPEEVGYALSEVPHNID-----HGLIKAHYVMGEDPLQTEPD
Cp28sec	NSREKGYPLSELPHNAI-----SGKVKAFYVMGEDPMQTEPD
<i>Yersinia frederiksenii</i> ATCC 33641	LPAHVGYSITDVPHKVA-----EGKLKAYYVFGEDPIQTEPD
ZnE2cys	FPEKVGHLHLTEVPHAVH-----EGKLKAFYIMGEDPLQTEPD
<i>Yersinia frederiksenii</i> ATCC 33641	LSGKIGYSLTDVPHKVK-----EGKIKANYVMGEDPLQTEPD
<i>Citrobacter koseri</i> ATCC BAA-895	MDDKVGTRITEVPHLAM-----EGKIKAYYIMGEDPLQTEAD
<i>Proteus mirabilis</i> HI4320	LDPQVGYRITEVPHLAI-----EGKVKAYYIMGEDPLQTEAD

---

**Figure 2.7.** Mitochondrial cytochrome oxidase II phylogeny of insects representing major termite (*Mastotermitidae*, *Hodotermitidae*, *Termopsidae*, *Kalotermitidae*, *Rhinotermitidae*, *Termitidae*), wood-feeding roach (*Cryptocercidae*), and omnivorous roach (*Blattidae*) families. *Serritermitidae*, a rare group of lower termites from Brazil, is usually classified as a seventh termite family (Krishna, 1970; Grimaldi and Engel, 2005). Insects examined in this study are underlined. Families in which the wood-feeding ability has been well-established are highlighted by shaded boxes. Members of the first 5 termite families are classified as “lower” termites; those within the *Termitidae* are “higher” termites. 11 cytochrome oxidase sequences (*Amitermes dentatus* acc. no. DQ442065, *Amitermes evuncifer* DQ442066, *Cornitermes pugnax* DQ442106, *Cornitermes walkeri* AB005577, *Labiatermes labralis* DQ442149, *Microcerotermes newmani* DQ442166, *Microcerotermes parvus* DQ442167, *Nasutitermes corniger* AB037327, *Nasutitermes ephratae* AB037328, *Nasutitermes* sp. Warnecke-2007 EU236539, *Nasutitermes nigriceps* AB037329) comprise the grouped clade *Termitidae*. The tree was calculated based on 393 aligned nucleotides using the maximum likelihood algorithm AxML. Filled circles indicate nodes supported by three different tree construction methods (Fitch distance, Phylip DNA parsimony, and AxML). The scale bar represents 0.1 nucleotide changes per alignment position.



**Figure 2.8.** Rarefaction curves calculated using EstimateS for each insect gut *fdhF* DNA or cDNA inventory. Sequences were first binned into operational taxonomic units at a cutoff of 97% amino acid similarity (Jones-Thorton-Taylor corrected 3% amino acid difference) using DOTUR. Inventory templates (Zn, *Zootermopsis nevadensis*; Rh, *Reticulitermes hesperus*; Im, *Incisitermes minor*; Cp, *Cryptocercus punctulatus*) and primer sets (PS1, primer set 1; PS2, primer set 2) are designated in the upper left corner. Primer set definitions are listed in Table 2.3.



## References

- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389-3402.
- Axley, M.J., Böck, A., and Stadtman, T.C. (1991) Catalytic properties of an *Escherichia coli* formate dehydrogenase mutant in which sulfur replaces selenium. *Proc Natl Acad Sci U S A* **88**: 8450-8454.
- Berghöfer, Y., Agha-Amiri, K., and Klein, A. (1994) Selenium is involved in the negative regulation of the expression of selenium-free [NiFe] hydrogenases in *Methanococcus voltae*. *Mol Gen Genet* **242**: 369-373.
- Berry, M.J., Maia, L., Kieffer, J.D., Harney, J.W., and Larsen, P.R. (1992) Substitution of cysteine for selenocysteine in type I iodothyronine deiodinase reduces the catalytic efficiency of the protein but enhances its translation. *Endocrinology* **131**: 1848-1852.
- Böck, A. (2000) Biosynthesis of selenoproteins: an overview. *BioFactors* **11**: 77.
- Böhm, R., Sauter, M., and Böck, A. (1990) Nucleotide sequence and expression of an operon in *Escherichia coli* coding for formate hydrogenylase components. *Mol Microbiol* **4**: 231-243.
- Boyington, J.C., Gladyshev, V., Khangulov, S.V., Stadtman, T., and Sun, P.D. (1997) Crystal structure of formate dehydrogenase H: catalysis Involving Mo, molybdopterin, selenocysteine, and an Fe<sub>4</sub>S<sub>4</sub> cluster. *Science* **275**: 1305-1308.
- Brauman, A., Kane, M.D., Labat, M., and Breznak, J.A. (1992) Genesis of acetate and methane by gut bacteria of nutritionally diverse termites. *Science* **257**: 1384-1387.
- Breznak, J.A. (1982) Intestinal microbiota of termites and other xylophagous insects. *Annu Rev Microbiol* **36**: 323-343.
- Breznak, J.A., and Switzer, J.M. (1986) Acetate synthesis from H<sub>2</sub> plus CO<sub>2</sub> by termite gut microbes. *Applied and Environment Microbiology* **52**: 623-630.
- Breznak, J.A., and Brune, A. (1994) Role of microorganisms in the digestion of lignocellulose by termites. *Ann Rev Entomol* **39**: 453-487.
- Breznak, J.A., and Warnecke, F. (2008) *Spirochaeta cellobiosiphila* sp. nov., a facultatively anaerobic, marine spirochaete. *Int J Syst Evol Microbiol* **58**: 2762-2768.
- Chao, A. (1987) Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* **43**: 783-791.
- Colwell, R.K. (2009). EstimateS: statistical estimation of species richness and shared species from samples. version 8.2.0. URL <http://viceroy.eeb.uconn.edu/EstimateS>

- Drake, H., Küsel, K., and Matthies, C. (2002) Ecological consequences of the phylogenetic and physiological diversities of acetogens. *Antonie van Leeuwenhoek* **81**: 203-213.
- Drake, H., Gossner, A., and Daniel, S. (2007) Old acetogens, new light. *Ann N Y Acad Sci* **1125**: 100-128.
- Drake, H.L. (1994) Introduction to acetogenesis. In *Acetogenesis*. Drake, H.L. (ed). New York: Chapman and Hall, pp. 3-60.
- Drake, H.L., Küsel, K., and Matthies, C. (2006) Acetogenic prokaryotes. In *The Prokaryotes*. Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.H., and Stackebrandt, E. (eds): Springer, pp. 354-420.
- Ebert, A., and 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). *Appl Environ Microbiol* **63**: 4039-4046.
- Felsenstein, J. (1989) PHYLIP - Phylogeny inference package (version 3.2). *Cladistics* **5**: 164 - 166.
- Gladyshev, V.N., Khangulov, S.V., Milton, J.A., and Stadtman, T.C. (1994) Coordination of selenium to molybdenum in formate dehydrogenase H from *Escherichia coli*. *Proceedings of the National Academy of Sciences U S A* **91**: 7708-7711.
- Graber, J.R., and Breznak, J.A. (2004) Physiology and nutrition of *Treponema primitia*, an H<sub>2</sub>/CO<sub>2</sub>-acetogenic spirochete from termite hindguts. *Appl Environ Microbiol* **70**: 1307-1314.
- Graber, J.R., and Breznak, J.A. (2005) Folate cross-feeding supports symbiotic homoacetogenic spirochetes. *Appl Environ Microbiol* **71**: 1883-1889.
- Graber, J.R., Leadbetter, J.R., and Breznak, J.A. (2004) Description of *Treponema azotonutricium* sp. nov. and *Treponema primitia* sp. nov., the first spirochetes isolated from termite guts. *Appl Environ Microbiol* **70**: 1315-1320.
- Grimaldi, D., and Engel, M.S. (2005) *Evolution of the insects*. New York, NY: Cambridge University Press.
- Gromer, S., Johansson, L., Bauer, H., Arscott, L.D., Rauch, S., Ballou, D.P. *et al.* (2003) Active sites of thioredoxin reductases: why selenoproteins? *Proc Natl Acad Sci USA* **100**: 12618 - 12623.
- Hakobyan, M., Sargsyan, H., and Bagramyan, K. (2005) Proton translocation coupled to formate oxidation in anaerobically grown fermenting *Escherichia coli*. *Biophys Chem* **115**: 55-61.

- Honigberg, B.M. (1970) Protozoa associated with termites and their role in digestion. In *Biology of Termites*. Krishna, K., and Weesner, F.M. (eds). New York: Academic Press, pp. 1-36.
- Inoue, T., Kitade, O., Yoshimura, T., and Yamaoka, I. (2000) Symbiotic association with protists. In *Termites: Evolution, Sociality, Symbioses*. Abe, T., Bignell, D.E., and Higashi, M. (eds). Dordrecht, The Netherlands: Kluwer Academic Publishers, pp. 275-288.
- Jones, J.B., and Stadtman, T.C. (1981) Selenium-dependent and selenium-independent formate dehydrogenases of *Methanococcus vannielii*. Separation of the two forms and characterization of the purified selenium-independent form. *J Biol Chem* **256**: 656-663.
- Jormakka, M., Byrne, B., and Iwata, S. (2003) Formate dehydrogenase-a versatile enzyme in changing environments. *Curr Opin Struct Biol* **13**: 418-423.
- Jormakka, M., Tornroth, S., Byrne, B., and Iwata, S. (2002) Molecular basis of proton motive force generation: structure of formate dehydrogenase-N. *Science* **295**: 1863-1868.
- Kambhampati, S., and Eggleton, P. (2000) Taxonomy and phylogeny of termites. In *Termites: Evolution, Sociality, Symbioses, Ecology*. Abe, T., Bignell, D.E., and Higashi, M. (eds). Dordrecht, The Netherlands: Kluwer Academic Publishers, pp. 1-24.
- Kane, M.D., and Breznak, J.A. (1991) *Acetonema longum* gen. nov. sp. nov., an H<sub>2</sub>/CO<sub>2</sub> acetogenic bacterium from the termite, *Pterotermes occidentis*. *Arch Microbiol* **156**: 91-98.
- Kim, H.Y., and Gladyshev, V.N. (2005) Different catalytic mechanisms in mammalian selenocysteine- and cysteine-containing methionine-R-sulfoxide reductases. *PLoS Biol* **3**: e375.
- Krishna, K. (1970) Taxonomy, phylogeny, and distribution of termites. In *Biology of Termites*. Krishna, K., and Weesner, F.M. (eds). New York: Academic Press, pp. 127-152.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H. *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947-2948.
- Leadbetter, J.R., Schmidt, T.M., Graber, J.R., and Breznak, J.A. (1999) Acetogenesis from H<sub>2</sub> plus CO<sub>2</sub> by spirochetes from termite guts. *Science* **283**: 686-689.
- Lee, S.R., Bar-Noy, S., Kwon, J., Levine, R.L., Stadtman, T.C., and Rhee, S.G. (2000) Mammalian thioredoxin reductase: Oxidation of the C-terminal cysteine/selenocysteine active site forms a thioselenide, and replacement of selenium with sulfur markedly reduces catalytic activity. *Proc Natl Acad Sci USA* **97**: 2521-2526.
- Lilburn, T.G., Kim, K.S., Ostrom, N.E., Byzek, K.R., Leadbetter, J.R., and Breznak, J.A. (2001) Nitrogen fixation by symbiotic and free-living spirochetes. *Science* **292**: 2495-2498.
- Liou, J.S.-C., Balkwill, D.L., Drake, G.R., and Tanner, R.S. (2005) *Clostridium carboxidivorans* sp. nov., a solvent-producing clostridium isolated from an agricultural

settling lagoon, and reclassification of the acetogen *Clostridium scatologenes* strain SL1 as *Clostridium drakei* sp. nov. *Int J Syst Evol Microbiol* **55**: 2085-2091.

Ljungdahl, L.G. (1986) The autotrophic pathway of acetate synthesis in acetogenic bacteria. *Annu Rev Microbiol* **40**: 415-450.

Ljungdahl, L.G., and Andreesen, J.R. (1978) Formate dehydrogenase, a selenium--tungsten enzyme from *Clostridium thermoaceticum*. *Methods Enzymol* **53**: 360-372.

Lo, N., Tokuda, G., Watanabe, H., Rose, H., Slaytor, M., Maekawa, K. *et al.* (2000) Evidence from multiple gene sequences indicates that termites evolved from wood-feeding cockroaches *Curr Biol* **10**: 801-804

Ludwig, W., Strunk, O., Westram, R., Richter, L., Meier, H., Yadhukumar *et al.* (2004) ARB: a software environment for sequence data. *Nucl Acids Res* **32**: 1363-1371.

Matson, E.G., Ottesen, E., and Leadbetter, J.R. (2007) Extracting DNA from the gut microbes of the termite *Zootermopsis nevadensis*. *J Vis Exp* **4**: 195.

Matson, E.G., Zhang, X., and Leadbetter, J.R. (2010) Selenium controls expression of paralogous formate dehydrogenases in the termite gut acetogen *Treponema primitia*. *Environ Microbiol* **Accepted**.

Odelson, D.A., and Breznak, J.A. (1983) Volatile fatty acid production by the hindgut microbiota of xylophagous termites. *Appl Environ Microbiol* **45**: 1602-1613.

Park, Y.C., Maekawa, K., Matsumoto, T., Santoni, R., and Choe, J.C. (2004) Molecular phylogeny and biogeography of the Korean woodroaches *Cryptocercus* spp. *Molecular Phylogenetics and Evolution* **30**: 450-464.

Pecher, A., Zinoni, F., and Böck, A. (1985) The seleno-polypeptide of formic dehydrogenase (formate hydrogen-lyase linked) from *Escherichia coli*: genetic analysis. *Arch Microbiol* **141**: 359-363.

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

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

Pierce, E., Xie, G., Barabote, R., Saunders, E., Han, C., Detter, J. *et al.* (2008) The complete genome sequence of *Moorella thermoacetica* (f. *Clostridium thermoaceticum*). *Environ Microbiol* **10**: 2550-2573.

Raaijmakers, H., Macieira, S., Dias, J.M., Teixeira, S., Bursakov, S., Huber, R. *et al.* (2002) Gene sequence and the 1.8 Å crystal structure of the tungsten-containing formate dehydrogenase from *Desulfovibrio gigas*. *Structure* **10**: 1261-1272.

- Romão, M. (2009) Molybdenum and tungsten enzymes: a crystallographic and mechanistic overview. *Dalton Trans*, : 4053 - 4068.
- Salmassi, T.M., and Leadbetter, J.R. (2003) Analysis of genes of tetrahydrofolate-dependent metabolism from cultivated spirochaetes and the gut community of the termite *Zootermopsis angusticollis*. *Microbiology* **149**: 2529–2537.
- Sawers, G. (1994) The hydrogenases and formate dehydrogenases of *Escherichia coli*. *Antonie van Leeuwenhoek* **66**: 57-88.
- Schloss, P.D., and Handelsman, J. (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* **71**: 1501-1506.
- Schultz, J.E., and Breznak, J.A. (1978) Heterotrophic bacteria present in hindguts of wood-eating termites [*Reticulitermes flavipes* (Kollar)]. *Appl Environ Microbiol* **35**: 930-936.
- Stamatakis, A.P., Ludwig, T., and Meier, H. (2004) The AxML program family for maximum likelihood-based phylogenetic tree inference. *Concurrency and Computation: Practice and Experience* **16**: 975-988.
- Thauer, R.K. (1972) CO<sub>2</sub>-reduction to formate by NADPH. The initial step in the total synthesis of acetate from CO<sub>2</sub> in *Clostridium thermoaceticum*. *FEBS Lett* **27**: 111-115.
- Valente, F.M., Almeida, C.C., Pacheco, I., Carita, J., Saraiva, L.M., and Pereira, I.A. (2006) Selenium is involved in regulation of periplasmic hydrogenase gene expression in *Desulfovibrio vulgaris* Hildenborough. *J Bacteriol* **188**: 3228-3235.
- Vorholt, J.A., and Thauer, R.K. (2002) Molybdenum and tungsten enzymes in C1 metabolism. *Met Ions Biol Syst* **39**: 571-619.
- Vorholt, J.A., Vaupel, M., and Thauer, R.K. (1997) A selenium-dependent and a selenium-independent formylmethanofuran dehydrogenase and their transcriptional regulation in the hyperthermophilic *Methanopyrus kandleri*. *Mol Microbiol* **23**: 1033-1042.
- Warnecke, F., Luginbühl, P., Ivanova, N., Ghassemian, M., Richardson, T., Stege, J. *et al.* (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**: 560-565.
- Wood, H.G., and Ljungdahl, L.G. (1991) Autotrophic character of the acetogenic bacteria. In *Variations in Autotrophic Life*. Shively, J.M., and Barton, L.L. (eds). New York: Academic Press, pp. 201-250.
- Yamamoto, I., Saiki, T., Liu, S.M., and Ljungdahl, L.G. (1983) Purification and properties of NADP-dependent formate dehydrogenase from *Clostridium thermoaceticum*, a tungsten-selenium-iron protein. *J Biol Chem* **258**: 1826-1832.

Zhang, Y., and Gladyshev, V.N. (2005) An algorithm for identification of bacterial selenocysteine insertion sequence elements and selenoprotein genes. *Bioinformatics* **21**: 2580-2589.

Zinoni, F., Birkmann, A., Stadtman, T.C., and Böck, A. (1986) Nucleotide sequence and expression of the selenocysteine-containing polypeptide of formate dehydrogenase (formate-hydrogen-lyase-linked) from *Escherichia coli*. *Proceedings of the National Academy of Sciences U S A* **83**: 4650-4654.

Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucl Acids Res* **31**: 3406 - 3415.

# **Formate dehydrogenase gene phylogeny in higher termites suggests gut microbial communities have undergone an evolutionary bottleneck, convergent evolution, and invasion**

## **Abstract**

The majority of termites and termite species on the planet belong to the phylogenetically 'higher' termite family *Termitidae*. Higher termites thrive on diverse lignocellulosic substrates with the aid of symbiotic gut microbiota.  $H_2$  consuming  $CO_2$  reductive acetogenic bacteria are an important group of symbionts that produce a significant fraction of the acetate used by their insect host as its primary carbon and energy source. A recent metagenomic analysis of the hindgut paunch bacterial community of a wood-feeding higher termite suggested spirochetes are the dominant acetogens in higher termites, as they appear to be in phylogenetically lower termites. However, a certain genetic feature of acetogenesis in higher termites was not resolved. Genes for hydrogenase-linked formate dehydrogenase ( $FDH_H$ ), an enzyme implicated in  $H_2$  turnover and  $CO_2$  fixing capacities of a termite gut acetogenic spirochete isolate and many uncultured lower termite gut acetogens, were notably depleted with respect to abundance and diversity relative to other acetogenesis genes in the metagenome and the gut communities of lower termites. Here, we use  $FDH_H$  primers to determine whether higher termite gut communities are as poor in  $FDH_H$  genes as previous data suggest. We

report that each and every  $FDH_H$  gene inventory generated from the whole gut communities of 8 species of taxonomically and nutritionally diverse higher termites (subfamilies *Nasutitermitinae* and *Termitinae*) was considerably more diverse than the metagenomic data set (4-15 phlotypes versus 1 phlotype), indicating the near absence of  $FDH_H$  genes in the metagenomic data set may result from artifacts of sampling or methodology. Phylogenetic analysis of higher termite  $FDH_H$  sequences also supports the concept that spirochetes dominate acetogenesis in lignocellulose-feeding higher termites. More significantly, we present evidence that suggests that acetogenic spirochete populations have undergone extinctions and radiations associated with an evolutionary bottleneck, convergent evolutions, and possibly even invasion during higher termite evolution. We posit that the extinction of flagellates and any associated bacteria – implied by the absence of flagellates in all higher termites – as the likely genetic bottleneck underlying such phylogenetic patterns.

## Introduction

All phylogenetically “higher” termites belong to the family, *Termitidae* (22, 26), within the arthropod order *Isoptera*. This single family encompasses the majority of termite individuals on earth and also comprises ~84% of all 2,900 extant termite species described to date (18, 52). The numerical abundance of higher termites establishes *Termitidae* as important members of many tropical and subtropical terrestrial ecosystems (3, 6). Most higher termites live in tropical ecosystems, wherein several termites have been credited for as much as 50% of plant biomass turnover (18) and the maintenance of soil fertility (6). The ecological success of the *Termitidae* has been correlated with their ability to subsist – with the aid of symbiotic gut microbiota – on recalcitrant substrates other than wood-derived lignocellulose (2, 37). Higher termites, engaging in obligate nutritional mutualisms, are able to eat dry grass, dung, decayed roots, lichen, leaf litter, fungus, and humus-rich soil in addition to wood, the predominant food source for phylogenetically “lower” (less derived) termites (3, 7).

Investigations on the nature of termite-microbe nutritional mutualisms indicate lignocellulose degradation by gut microbes is stepwise and results in the production of substantial levels of acetate, the main carbon and energy source of the insect host (11, 39, 43). Polysaccharides are first hydrolyzed from wood and fermented to acetate, H<sub>2</sub>, and CO<sub>2</sub>. CO<sub>2</sub> reductive bacteria, using the Wood-Ljungdahl pathway for acetogenesis, then consume the great majority of fermentation-derived H<sub>2</sub> and CO<sub>2</sub> (i.e., 82–100% in lower termites) and produce additional acetate for the insect host (11, 39, 43). Acetate generated from CO<sub>2</sub> reductive acetogenesis may account for up to 30% of gut acetate (11,

39). The remaining H<sub>2</sub> from fermentation does not benefit the host, but is instead consumed by methanogenic *Archaea* and emitted as methane.

Studies of lower termite gut microbiota have attributed fermentation and acetogenesis to cellulolytic flagellate protozoa and acetogenic spirochetes, respectively (10, 25, 27, 42). The microbes responsible for such processes in higher termites are relatively unstudied, but the noticeable lack of flagellate protozoa in *all* higher termites described thus far (25) implies bacteria play a greater role in lignocellulose digestion within higher termites. The increased complexity of gut structure in higher termites is also quite noticeable. Whereas all key steps of lignocellulose degradation occur in the single hindgut paunch of lower termites, higher termite hindguts are composed of a series of chambers, each potentially characterized by its own pH (4, 5, 13, 49) and microbial community (46, 47, 51, 53).

Investigations aimed at elucidating the processes involved in digestion of non-woody substrates have also been undertaken, but interpretations have been challenged by the complex nature of food substrates like soil. Nevertheless, several important observations have been made. Radiotracer studies comparing carbon and reductant flows in higher termites with different feeding habits revealed rates of acetogenesis and methanogenesis could vary by an order of magnitude (8, 9, 11). In these experiments, CO<sub>2</sub> reduction to acetate was the dominant terminal electron accepting process in grass- and wood-feeding termites, but methanogenesis outcompeted acetogenesis for H<sub>2</sub> in fungus- and soil-feeding termites (9, 50). Efforts aimed at understanding the organisms responsible for such differences have been largely focused on ribosome-based identifications (1, 8, 35,

41, 47, 51). However, such methods can not reliably identify acetogenic bacteria since acetogens are paraphyletic (15), thus information on functional genes encoding acetogenesis enzymes in higher termites is also required.

Such information was recently provided by a metagenomic analysis of the gut bacterial community inhabiting the largest gut compartment (P3) of a wood-feeding *Nasutitermes* higher termite (54). Phylogenetic analysis revealed numerous gene variants (14–37) for all Wood-Ljungdahl pathway enzymes but formate dehydrogenase (FDH), for which only two gene variants were identified. The near absence of FDH genes was striking in light of the absolute necessity of FDH for acetogenesis from  $H_2 + CO_2$ , a process firmly established in wood-feeding higher termite guts (9, 11). However, the phylogeny of one gene variant was consistent with that of other acetogenesis genes (54). This FDH gene affiliated with hydrogenase-linked FDH (FDH<sub>H</sub>) sequences identified in the termite gut acetogenic spirochete, *Treponema primitia*, and the gut communities of lower termites and a wood-feeding roach (Chapter 2), in support of the prediction that spirochetes dominate acetogenesis in higher termites (54). The function and origin of second gene were not as clear.

Taken together, the findings suggest four hypotheses: (i) FDH genes are absent from wood-feeding higher termite gut communities; (ii) FDH genes are located elsewhere in the gut tract and, thus, were not sampled for metagenomic analysis – this implies acetogens within the hindgut paunch rely on an outside supply of formate (i.e., formate transfer between gut chambers) (45); (iii) FDH genes in the *Nasutitermes* metagenome

may not be recognizable by bioinformatics methods; or (iv) metagenome results may be inaccurate with respect to FDH due to cloning and other methodological artifacts. Here, we explore these hypotheses by surveying FDH<sub>H</sub> gene (*fdhF*) diversity in the whole gut microbial communities of 8 species of taxonomically diverse higher termites (*Nasutitermitinae*, *Termitinae*) which represent different nesting strategies (arboreal, subterranean), habitats (tropical, desert), feeding habits (wood, leaf litter, roots/soil, dry grass/soil), and levels of soil exposure. In particular, we compare and contrast *fdhF* diversity between higher and lower termites, different species of higher termites, and termites with different lifestyles to ascertain whether FDH<sub>H</sub> genes present in lower termites are absent from *Nasutitermes* as metagenomics suggests and explore the evolution of hydrogenase-linked FDH enzymes within *Termitidae*, the most ecologically successful lineage of termites on the planet.

## Materials and Methods

### Insect collection and identification

Several termite species were collected in Costa Rica. *Nasutitermes* sp. Cost003 was arboreal and collected from its nest on a guava tree (*Psidium guajaba*) located in the forest preserve of the National Biodiversity Institute of Costa Rica (INBio), near the city of Guápiles. *Rhynchotermes* sp. Cost004 was collected after amongst leaf litter near the root zone of an unidentified *Bromeliad* sp. within the same INBio forest. *Amitermes* sp. Cost010 was collected from decayed sugar cane roots encrusted with soil at a sugar cane plantation in Grecia, Costa Rica. *Nasutitermes corniger* Cost007 was collected from its

nest carton located on an unidentified species of palm tree, which was growing in sandy soil within the forest/beach transition zone in Cahuita National Park (CNP), Costa Rica. *Microcerotermes* sp. Cost 006 and *Microcerotermes* sp. Cost008 were collected from a nest in a palm tree and a nest at the base of a palm tree, respectively, within CNP. Both trees were growing in sandy soil. *Coptotermes* sp. Cost 009 (lower termite, family *Rhinotermitidae*) was collected near sulfidic smelling soil in the forest/beach transition zone near the Kelly Creek Ranger Station (CNP).

Termites were also collected from Joshua Tree National Park, CA. *Amitermes* sp. JT2 and *Gnathamitermes* sp. JT5 were collected from subterranean nests; *Reticulitermes tibialis* JT1 (lower termite, family *Rhinotermitidae*) was collected from a decayed log found in a dry stream bed.

### **DNA extraction**

For each termite species, the entire hindguts of 20 worker termites were extracted within 48 hours of collection, pooled into 500  $\mu\text{l}$  1X Tris-EDTA buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8), and stored at  $-20^{\circ}\text{C}$  until DNA extraction. Whole gut community DNA was obtained using the method described by Matson *et al.* (31).

### ***fdhF* amplification and cloning**

PCR reactions were assembled as previously described in Chapter 2 (1  $\mu\text{M}$ , each universal primer), except polymerase (0.07 – 0.14  $\text{U} \cdot \mu\text{l}^{-1}$ ) and gut DNA template concentrations (0.05 – 1  $\text{ng} \cdot \mu\text{l}^{-1}$ ) were adjusted so that reactions would yield similar

amounts of PCR product. Thermocycling conditions for PCR on a Mastercycler Model 5331 thermocycler (Eppendorf, Westbury, NY) were: 2 min at 94°C, 25 cycles of (denaturation at 94°C for 30 sec, annealing at 51°C, 53.6°C or 55°C for 1 min, extension at 68°C for 2 min 30 sec), followed by 10 min of final extension at 68°C. Details of PCR reaction composition and amplification can be found in Table 3.4 (Appendix 3). Amplification of templates at an annealing temperature of 51°C (used to generate lower termite inventories in Chapter 2) yielded multiple sized products upon electrophoresis with 1.5% w/v agarose (Invitrogen). The correct-sized bands were excised and gel purified with a QIAquick Gel Extraction Kit (QIAGEN, Valencia, CA). To ensure product specificity, PCR was performed at higher annealing temperatures (53.6 °C for Cost008, Cost010; 55°C for Cost003, Cost004). This second set of reactions yielded a single product band upon electrophoresis. All PCR products were cloned using a TOPO-TA cloning kit (Invitrogen, Carlsbad, CA).

Clones (30-107 per termite species) were screened for the presence of the correct sized insert by PCR and gel electrophoresis. PCR reactions (10 µL) contained T3 (1 µM) and T7 (1 µM) primers, 1X FAILSAFE Premix D (EPICENTRE Biotechnologies, Madison, WI), 0.05 U · µl<sup>-1</sup> Taq polymerase (New England Biolabs, Beverly, MA) and 1 µL of cells lysed in 1X TE as template. Thermocycling conditions were 2 min at 95°C, 30 cycles of (95°C for 30 sec, 55°C for 1 min, 72°C for 2 min 30 sec), followed by 10 min at 72°C.

**RFLP analysis, sequencing, diversity assessment**

Most inventories were subject to RFLP typing, wherein correct-sized products generated by screening PCRs were digested with the restriction enzyme *RsaI* (New England Biolabs) and electrophoresed on a 2.5% (w/v) agarose gel (Invitrogen). Plasmids from clones with unique RFLP patterns were purified using a QIAprep Spin Miniprep Kit (QIAGEN). For a few inventories, plasmids from clones having the correct-sized products were purified for sequencing without RFLP typing. Plasmids were sequenced with T3 and T7 primers at Laragen, Inc. (Los Angeles, CA) using an Applied Biosystems Incorporated ABI3730 automated sequencer. Lasergene (DNASTAR, Inc., Madison, WI) software was used to assemble and edit sequences. Sequences were grouped into operational taxonomic units at a 97% protein similarity level based on distance calculations (Phylip Distance Matrix using a JTT correction) and DOTUR (44). The program EstimateS v8.2.0 (14) was used to assess *fdhF* inventory diversity.

**COII amplification for termite identification**

A fragment of the mitochondrial cytochrome oxidase subunit II (COII) gene in Costa Rican termites was amplified from DNA containing both insect and gut community material using primers A-tLEU and B-tLYS at concentrations and thermocycling conditions described by Miura *et al.* (33, 34). For each species of Joshua tree termite, COII gene fragments were amplified using the supernatant of a mixture containing an individual termite head crushed in 1X TE as template. Primers and PCR conditions were identical to those employed for Costa Rican termite COII. PCR products were purified

using a QIAquick PCR purification kit (QIAGEN), sequenced, and analyzed to verify the species identity of termite specimens.

### **Primer design and PCR for a major clade of lower termite and wood roach Cys FDH<sub>H</sub> alleles**

Degenerate primers (Cys499F1b, 1045R) for a major clade of selenium independent (Cys) FDH<sub>H</sub> alleles present in lower termites and the wood roach *C. punctulatus* were designed manually using all sequences recovered from these insects (Chapter 2). Forward primer Cys499F1b (5'– ATG TCS CTK TCS ATI CCG GAA A –3') specificity is as follows: 38.9% of the sequences are perfectly matched, 22.2% have 1 mismatch, 27.8% have 2 mismatches, and 8.3% have 3 mismatches. No mismatches are in located in the terminal 3' position. The reverse primer 1045R (5'– CIC CCA TRT CGC AGG YIC CCT G –3') was designed based on 154 sequences from higher termites, lower termites and *C. punctulatus*. The primer targets both Sec and Cys *fdhF* variants; 60.3% of the sequences have 0 primer mismatches, 32.4% have 1, 5.8% have 2, and 1.3% have 3 mismatches. All sequences are perfectly matched at the terminal 3' position. PCR reactions contained 0.4 ng · μl<sup>-1</sup> of DNA template, 200 nM of Cys499F1b, 200 nM 1045R, 1X FAILSAFE Premix D (EPICENTRE), and 0.05 U · μl<sup>-1</sup> Taq polymerase (New England Biolabs). Thermocycling conditions were 2 min at 95°C, 30 cycles of (95°C for 30 sec, 60°C for 30 sec, 72°C for 45 sec), followed by 10 min at 72°C.

### **Primer design and PCR for a novel group of FDH<sub>H</sub> alleles identified in subterranean and litter feeding termites.**

'*Amitermes-Gnathamitermes-Rhynchotermes*' clade FDH<sub>H</sub> sequences were amplified using a nested PCR approach in which the amplicon from the first PCR reaction, generated with universal *fdhF* primers (TgfdhF-unvF1, EntfdhF-unvF1, and fdhF-unvR1), was used as the template for the second PCR reaction, containing clade specific primers (193F, 1045R). Forward primer 193F (5'- AGG CTT ACC AAG CCG CCT ATC AGA - 3') targets 55.6% of the sequences in the clade with 4 or fewer mismatches, none of them at the terminal 3' end. PCR amplification of all *fdhF* types was achieved using the PCR reaction compositions and thermocycling conditions (51°C annealing temperature) previously specified for inventories. Clade specific PCR reactions contained 1 µl of diluted product from the first reaction (1:1000 in water), 250 nM 193F, 250 nM 1045R, 1X FAILSAFE Premix D (EPICENTRE), and 0.07 U · µl<sup>-1</sup> of EXPAND High Fidelity polymerase (Roche). Thermocycling conditions were 2 min at 95°C, 25 cycles of (95°C for 15 sec, 60°C for 30 sec, 72°C for 1 min), followed by 10 min at 72°C.

### **Phylogenetic and Principle Component Analysis**

Phylogenetic analyses of protein and nucleotide sequences were performed with ARB version 09.08.29 (29). COII DNA phylogeny was generated with the AxML method (48). FDH protein phylogenies were calculated with the Phylip protein maximum likelihood (PROTML) algorithm (20). Details of tree construction can be found in figure legends. The same filter and alignments were employed when additional tree algorithms

(Fitch distance, Phylip protein parsimony) were used to infer node robustness (20). All phylogenetic inference models were run assuming a uniform rate of change for each nucleotide or amino acid position. Principal component analysis of FDH<sub>H</sub> phylogeny and environment data was performed using the phylogenetic analysis software Unifrac (28).

## Results

### Termite classification

Our collection of six species of Costa Rican higher termites and two species of Californian higher termites enabled comparisons of *fdhF* diversity in higher termites with different phylogenies, habitats, and lifestyles (Table 3.1). Termites were identified based on morphological characteristics, feeding behavior and diet (when observed), and their mitochondrial cytochrome oxidase 2 (COII) gene sequence (Figure 3.1). Together, the termites examined in this study represent two subfamilies (*Nasutitermitinae*, *Termitinae*) within the higher termite family *Termitidae*, generally recognized as comprising four subfamilies (26).

**Table 3.1.** Characteristics of insects examined in this study.

Insect	Family (Subfamily) <sup>1</sup>	Nest type/ Collection Site <sup>2</sup>	Habitat <sup>3</sup>	Pro- bable Food <sup>4</sup>	Soil Expo- sure <sup>5</sup>
<i>Nasutitermes</i> sp. Cost003	<i>Termitidae</i> ( <i>Nasutitermitinae</i> )	Arboreal, Forest (CR)	Premontane-wet rainforest transition	wood	low
<i>Nasutitermes corniger</i> Cost007	<i>Termitidae</i> ( <i>Nasutitermitinae</i> )	Arboreal, Forest-beach transition (CR)	Lowland moist forest	palm	low
<i>Rhynchotermes</i> sp. Cost004	<i>Termitidae</i> ( <i>Nasutitermitinae</i> )	Arboreal, Forest (CR)	Premontane-wet rainforest transition	leaf- litter	med
<i>Microcerotermes</i> sp. Cost006	<i>Termitidae</i> ( <i>Termitinae</i> )	Arboreal, Forest-beach transition (CR)	Lowland moist forest	palm	low
<i>Microcerotermes</i> sp. Cost008	<i>Termitidae</i> ( <i>Termitinae</i> )	Arboreal, Forest-beach transition (CR)	Lowland moist forest	palm	low
<i>Amitermes</i> sp. Cost010	<i>Termitidae</i> ( <i>Termitinae</i> ) <sup>6</sup>	Subterranean, root zone (CR)	Premontane wet forest	roots/ soil	high
<i>Amitermes</i> sp. JT2	<i>Termitidae</i> ( <i>Termitinae</i> ) <sup>6</sup>	Subterranean galleries, desert (JT)	Warm temperate desert	dry grass/ soil	high
<i>Gnathamitermes</i> sp. JT5	<i>Termitidae</i> ( <i>Termitinae</i> ) <sup>6</sup>	Subterranean galleries, desert (JT)	Warm temperate desert	dry grass/ <i>Yucca</i> / soil	high

<sup>1</sup> Termite family classifications were based on Kambhampati and Eggleton (26) and Grimaldi and Engel (22).

<sup>2</sup> Nest type (Arboreal versus subterranean) and collection location, CR = Costa Rica, JT= Joshua Tree, CA.

<sup>3</sup> Ecosystem terminology is based on the Holdridge life zone classification of land areas, which relies on climate data (24). Life zone categories for collection sites are based on maps in Enquist *et al.* (19) and Lugo *et al.* (30).

<sup>4</sup> Possible food source based on vegetation near collection location, insect trails, and/or laboratory feeding studies.

<sup>5</sup> Predicted level of soil exposure based on nest location (subterranean or above ground), food substrate, and foraging style.

<sup>6</sup> It is unclear whether *Amitermes* sp. affiliate within the subfamily *Termitinae* or rather constitute their own subfamily (26).

Estimates of global termite abundance indicate *Nasutitermitinae* and *Termitinae* are the two most numerically abundant and species rich subfamilies of the *Termitidae* (17). In most cases, we could not establish termite identity beyond the genus level due to the patchy distribution of COII gene sequences in NCBI databases. Genus names for *Rhynchotermes* sp. Cost004 and *Gnathamitermes* sp. JT5 specimens were assigned based solely on morphology since COII sequences only allowed definitive phylogenetic placement at the subfamily level. However, COII analysis indicates the 8 species of termites are phylogenetically distinct and represent a diversity of *Termitidae* lineages.

Other than phylogeny, the termites could be differentiated based on geography, nesting strategies, habitats (19, 24, 30), diet, and soil exposure levels (Table 3.1). Termites collected in Costa Rica showed greater variation with respect to each parameter than those collected in the California desert. While there are certainly other environmental factors that may influence gut microbial community structure and function, we consider insect phylogeny, geography, habitat, diet, and soil exposure the most obvious set of possible guiding parameters for interpreting gene inventory data.

### ***fdhF* alleles are present in the guts communities of every higher termite**

Our examination of *fdhF* diversity in 8 species of higher termite yielded *fdhF* genes from every higher termite species (Table 3.2), including *Nasutitermes* sp. Cost003 which is phylogenetically identical to the *Nasutitermes* sampled for metagenomic analysis and was collected within 100 m of the latter insect sample. Multiple *fdhF* genotypes (8–59) were recovered from each higher termite (Table 3.2). In particular, analyses revealed that 37

*fdhF* genotypes are encoded by the whole gut community in *Nasutitermes* sp. Cost003; this is nearly 20-fold greater than the number of FDH genotypes recovered from metagenomic analysis. Genotype diversity is likely much greater for inventory sequences that were subject to RFLP typing before sequencing.

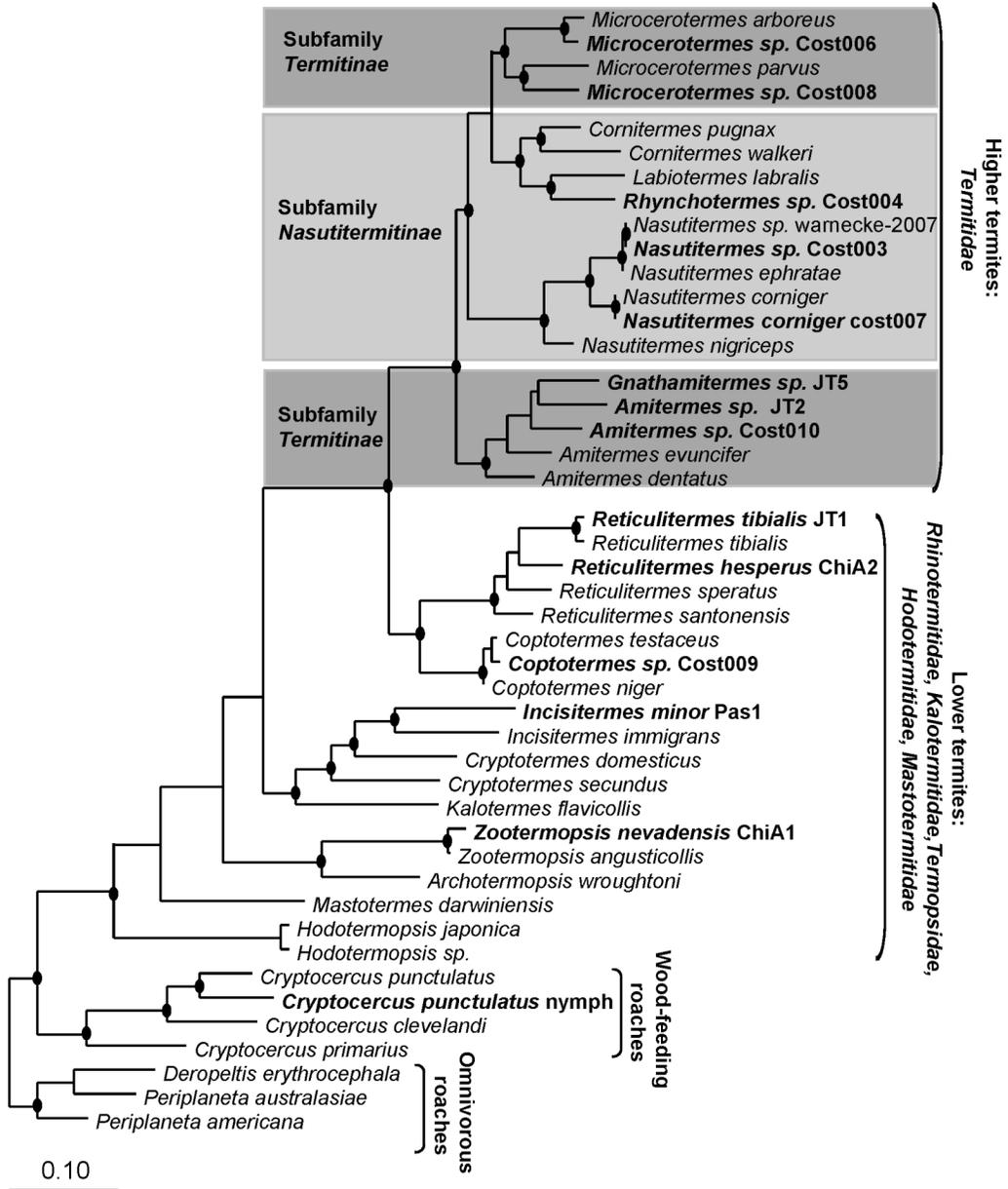
FDH<sub>H</sub> diversity is still greater than that observed in the metagenome when the deduced amino acid translations of genotypes are sorted into phylotypes (operational taxonomic units defined as 97% protein similarity). Each higher termite species encodes 4–15 phylotypes. Inventories from subterranean grass/soil-feeding (Cost010, JT2, JT5) and arboreal leaf litter-feeding termites (Cost004) contain 8–15 phylotypes. Wood-feeding termite inventories (Cost003, Cost007, Cost006, Cost008) contain noticeably fewer phylotypes, 4–8. Chao1 estimates of phylotype abundance indicate our sequencing efforts recovered the majority of diversity present in each termite. This allows meaningful comparisons of phylotype abundances.

### **Phylotype abundance related to lifestyle and insect phylogeny**

FDH<sub>H</sub> phylotype abundance appears to be more strongly related to termite lifestyle (e.g., diet similarities and soil exposure) than phylogeny within the higher termite lineage. This is evidenced by the grouping of Cost004 with Cost010, JT2, and JT5, rather than with other *Nasutitermitinae*, in support of an association between lifestyle and phylotype abundance in higher termites. Geography, nest type, and habitat are not as clearly associated with phylotype abundance in higher termites.

Insect phylogeny may be related to phylotype abundance at higher taxonomic scales for wood-feeding termites. Inventories from the lower wood-feeding termites *Zootermopsis nevadensis*, *Reticulitermes hesperus*, and *Incisitermes minor* comprise 11–15 phylotypes (Chapter 2). In contrast, the abundances are, on average, only half that in wood-feeding higher termites.

**Figure 3.1.** Mitochondrial cytochrome oxidase II (COII) phylogeny of termites and related roaches. Family names and other descriptions are located on the right side of the tree. Only two of four subfamilies (*Macrotermitinae*, *Apicotermitinae*, *Nasutitermitinae*, and *Termitinae*) in the higher termite family *Termitidae* are shown (26). Subfamily *Termitinae* is paraphyletic (26). The gut communities of insect species highlighted in bold have been examined for *fdhF* using inventory and/or PCR screening techniques. Tree was constructed with 393 aligned nucleotides using the maximum likelihood phylogenetic algorithm PHYML. Filled circles at nodes indicate support from PHYML, parsimony (Phylip DNAPARS), and Fitch distance methods. Scale bar corresponds to 0.1 nucleotide changes per alignment position.



**Table 3.2.** FDH inventories constructed in this study.

Species	Inven- tory <sup>1</sup>	No. clones	No. geno- types	No. OTU <sup>2</sup>	Mean Chao1 (SD) <sup>3</sup>	95% LCI, HCI Chao1 <sup>4</sup>	No. geno- types per species <sup>5</sup>	No. OTU per species <sup>5</sup>
<i>Nasutitermes sp.</i> Cost003	3L1	87	20	4	3.42 (0.12)	3.42, 3.42	37	6
<i>Nasutitermes sp.</i> Cost003	3L2	17	17 <sup>6</sup>	4	3.87 (1.33)	3.28, 10.54		
<i>Nasutitermes corniger</i> Cost007	7L1	30	19	8	7.3 (1.1)	6.8, 12.7	19	8
<i>Rhynchotermes sp.</i> Cost004	4L1	85	37	14	13.4 (1.8)	12.4, 22.4	59	15
<i>Rhynchotermes sp.</i> Cost004	4L2	22	22 <sup>6</sup>	8	8.3 (2.9)	6.7, 23.0		
<i>Microcerotermes sp.</i> Cost006	6L1	74	8	6	5.3 (0.9)	5.0, 9.8	8	6
<i>Microcerotermes sp.</i> Cost008	8L1	84	10	4	4.0 (0.01)	4.0, 4.0	10	4
<i>Amitermes sp.</i> Cost010	10L1	78	28	8	7.4 (1.0)	7.0, 12.4	51	12
<i>Amitermes sp.</i> Cost010	10L2	23	23 <sup>6</sup>	9	8.0 (1.5)	7.3, 15.8		
<i>Amitermes sp.</i> JT2	Jt2L1	101	18	8	7.4 (0.9)	7.1, 11.4	18	8
<i>Gnathamitermes sp.</i> JT5	Jt5L1	84	30	10	9.8 (0.5)	9.7, 11.8	30	10

<sup>1</sup> Two libraries were constructed for each of the following templates: Cost003, Cost004, and Cost010.

These differ most significantly in PCR annealing temperature (details in Table 3.4). PCR was performed at 55°C for libraries 3L1 and 4L1, 53.6 °C for 10L1, and 51°C for 3L2, 4L2, and 10L2.

<sup>2</sup> Number of operational taxonomic units (OTUs) defined at 97% amino acid similarity; calculated using Phylip Distance Matrix (JTT correction) and DOTUR.

<sup>3</sup> Mean of the diversity estimator Chao1 (SD, standard deviation) calculated using EstimateS.

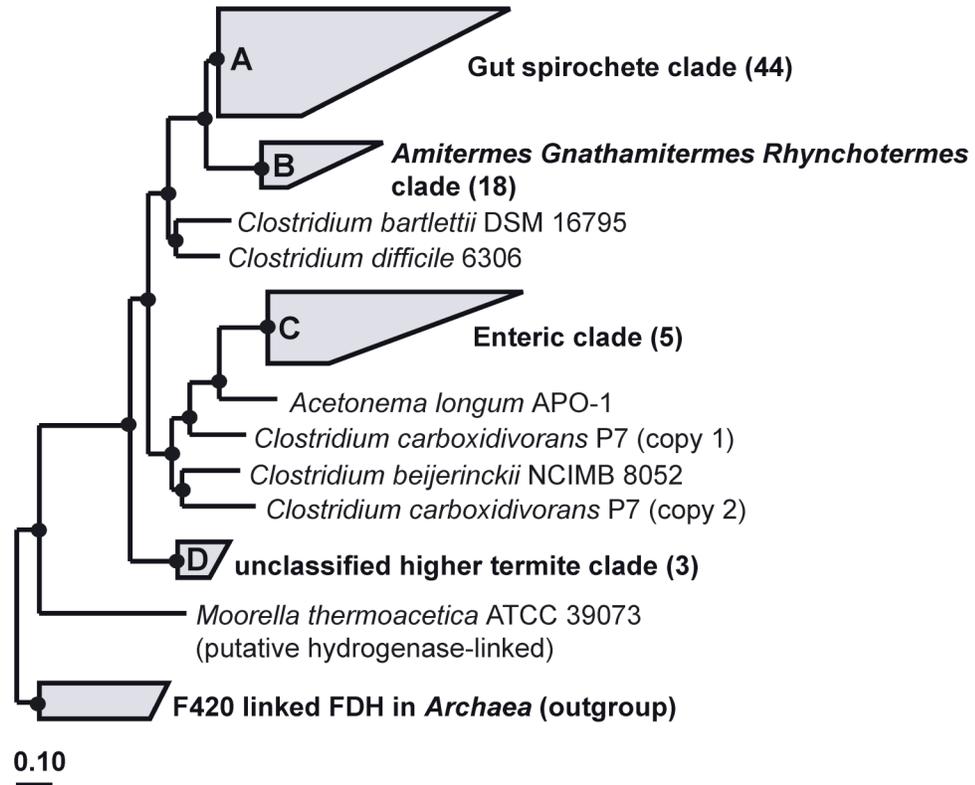
<sup>4</sup> Lower (LCI) and higher (HCI) 95% confidence interval for mean Chao1.

<sup>5</sup> Number of unique genotypes/OTUs when sequences from L1 and L2 libraries are combined for cost003, cost 004, and cost 010. For other templates, this column is equivalent to column 4. The distribution of each OTU can be found in Table 3.5 (Appendix).

<sup>6</sup> All clones that were picked were sequenced, rather than being presorted by RFLP typing prior to sequencing. All sequences were unique at the DNA level.

**Higher termite sequences affiliate with four major FDH<sub>H</sub> clades**

Higher termite FDH<sub>H</sub> sequences phylogenetically cluster into four major clades (Figure 3.2, clades A-D) within the FDH<sub>H</sub> family of enzymes, composed of sequences from enteric *γ-Proteobacteria*, *Spirochaetes*, *Firmicutes*, and uncultured organisms from lower termite and wood-roach hindguts. The relative abundances of the different sequence types in each inventory are listed in Table 3.3.



**Figure 3.2.** Higher termite, lower termite, wood-roach, and pure culture sequences form four major FDH<sub>H</sub> clades (A, B, C, D). The numbers of sequences within grouped clades are indicated in parentheses. Tree was constructed with 542 aligned amino acids with the maximum likelihood phylogenetic algorithm Phylip PROTML. A metagenomic FDH<sub>H</sub> sequence fragment (tgut2b\_BHZN47861\_b2) from the gut of *Nasutitermes* sp. Warnecke-2007 (54) was added in by parsimony and falls within the Gut spirochete clade (clade A). Filled circles indicate nodes supported by PROTML and parsimony (Phylip PROPARS, 100 bootstraps) methods. The tree was outgrouped with F420-linked FDH from methanogenic *Archaea*. Scale bar indicates 0.1 units of amino acid change per alignment position.

**Table 3.3.** Distribution of clones in each major FDH<sub>H</sub> clade.

Library	A	B	C	D
	Gut Spirochete	<i>Amit.-Gnath.-Rhyncho.</i>	Enteric Proteobacteria	Unclassified
<i>Nasutitermes sp.</i> Cost003 <sup>1</sup>	99	0	0	1
<i>Nasutitermes corniger</i> Cost007	86	0	7	7
<i>Rhynchotermes sp.</i> Cost004 <sup>1</sup>	47	51	2	0
<i>Microcerotermes sp.</i> Cost006	96	0	4	0
<i>Microcerotermes sp.</i> Cost008	100	0	0	0
<i>Amitermes sp.</i> Cost010 <sup>1</sup>	85	13	0	2
<i>Amitermes sp.</i> JT2	92	8	0	0
<i>Gnathamitermes sp.</i> JT5	74	8	18	0

<sup>1</sup> Libraries L1 and L2 were combined for abundance calculation.

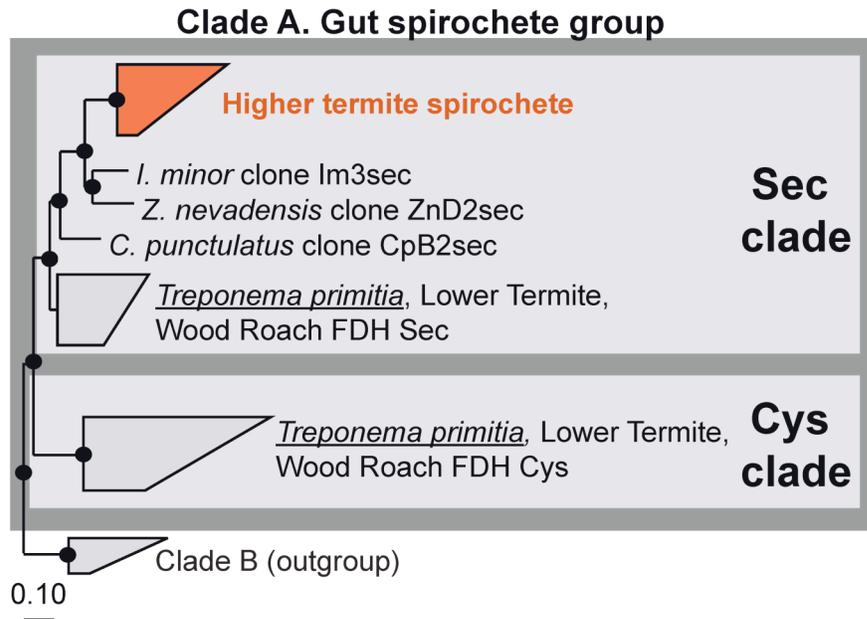
### Sweeping loss of ‘Cys clade’ alleles from higher termite gut communities

Previously, in Chapter 2 we reported that *fdhF* genes of phylogenetically lower termites (*Zootermopsis nevadensis*, *Reticulitermes hesperus*, *Incisitermes minor*) and a wood-roach (*Cryptocercus punctulatus*, the extant sister taxon of termites), could be broadly categorized into two major phylogenetic clades, which we refer to here as ‘Sec’ and ‘Cys clades’ (Figure 3.3). The Sec clade is comprised mainly of selenium-dependent FDH<sub>H</sub> enzymes, which encode selenocysteine (Sec, a non-canonical amino acid) at the enzyme active site. In contrast, most sequences in the Cys clade encode selenium-independent FDH<sub>H</sub> enzymes, which contain a cysteine (Cys), instead of selenocysteine, at the active site. Phylotype abundances for Sec clade and Cys clade FDH<sub>H</sub> variants were roughly equivalent in the guts of each of these evolutionarily primitive wood-feeding insects.

Phylogenetic analysis of higher termite sequences revealed a striking absence of Cys clade sequences from every higher termite (Figure 3.3). We therefore hypothesized that Cys clade alleles, previously identified in evolutionarily primitive wood-feeding insects,

were lost from the FDH<sub>H</sub> gene pool of higher termite gut communities. To test this hypothesis and ascertain whether Cys clade genes may have been present in higher termites but were not recovered due to inventory artifacts, we designed Cys clade specific primers (Cys499F1b, 1045R). We used these primers to screen the gut DNA of higher termites, 3 species of Southern California lower termites representing 3 termite families, and *C. punctulatus* for Cys clade *fdhF* genes. We did not detect product in any higher termite species after 30 cycles of PCR amplification (Figure 3.7, Appendix). In contrast, all amplifications from lower termites and roach yielded robust products.

We then hypothesized that the absence of Cys clade alleles in higher termites may be related to insect habitat. To explore the relationship between habitat and the presence of Cys clade genes, we performed PCR screens of two lower termite species collected in the same habitats as certain higher termites (Costa Rican lower termite *Coptotermes* sp. Cost009 collected near Cost006 and Cost008; desert-adapted lower termite *R. tibialis* sp. JT1 collected near JT1 and JT5). PCR amplicons were observed for each lower termite sample (Figure 3.7, Appendix). Dilution-to-extinction PCRs suggest that Cys clade alleles are at least 1000-fold more abundant in lower termites than higher termites (calculations in the legend of Figure 3.7, Appendix). Taken together, the results of targeted PCR assays are consistent with inventory findings and the hypothesis that sweeping gene loss has occurred in the FDH<sub>H</sub> gene pool of higher termites.



**Figure 3.3.** Sec and Cys clades within the “Gut spirochete clade” (Clade A, Figure 3.2). Higher termite sequences, marked in red, form the “higher termite spirochete group”. Tree was constructed using the methods described in the legend of Figure 3.2. Filled circles indicate nodes were supported by PROTML and parsimony methods of analyses. Scale bar indicates 0.1 amino acid changes per alignment position.

### **Higher termite FDH<sub>H</sub> sequences form a single clade within the ‘Sec clade’ of the Gut spirochete group**

The vast majority of higher termite sequences cluster into one phylogenetic group (44 phylotypes, Figure 3.3) within the Sec clade of the Gut spirochete group (Clade A, Figure 3.2). The latter encompasses hydrogenase-linked FDHs from acetogenic spirochetes *Treponema primitia* str. ZAS-1 and ZAS-2, lower termites, and *C. punctulatus*. We infer higher termite sequences belong to uncultured acetogenic spirochetes based on phylogeny – *T. primitia* is the nearest pure culture relative – and the presence of a diagnostic amino acid character shared by every sequence in the Gut spirochete group, but absent from sequences outside the group.

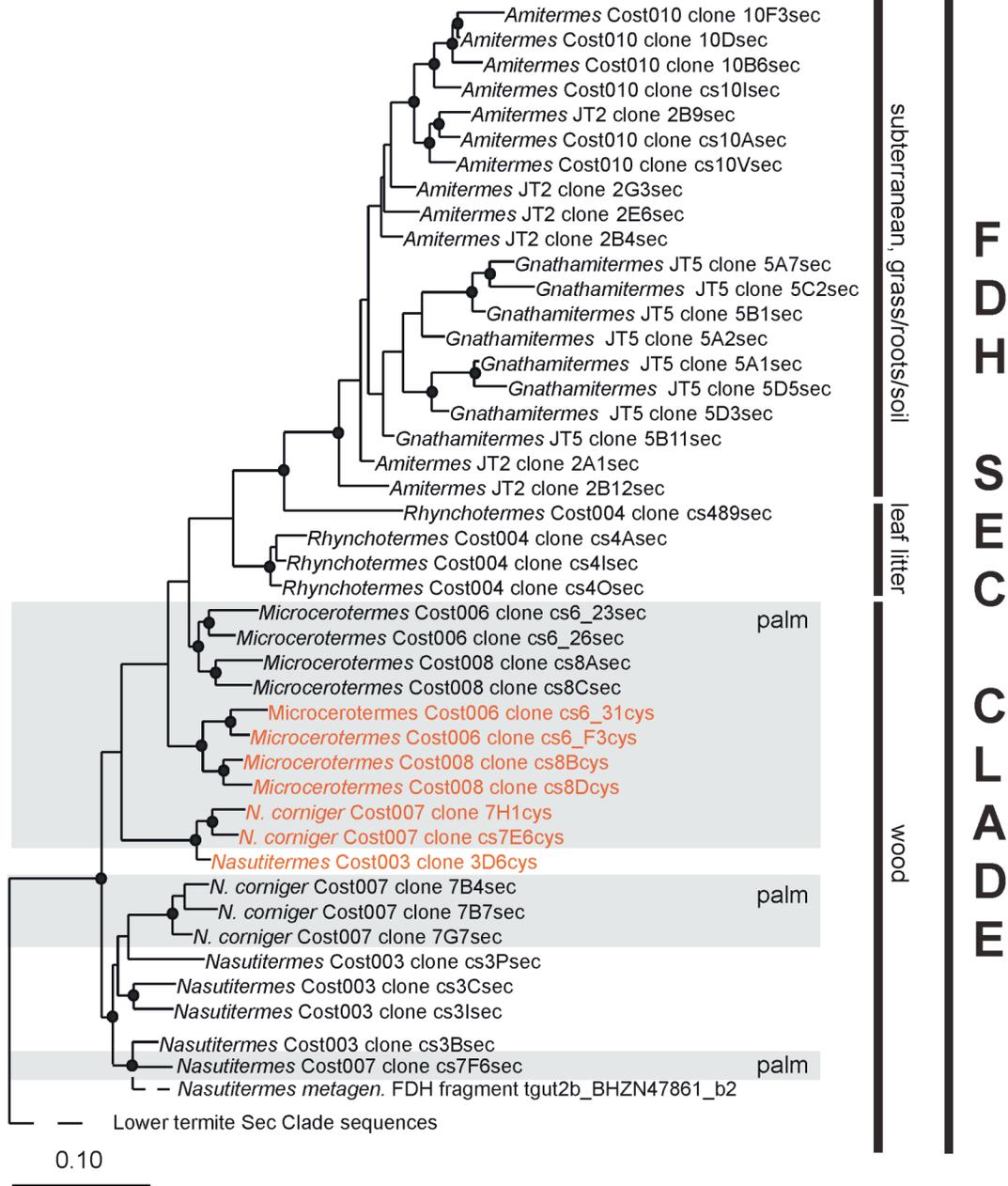
Besides being the largest, the “higher termite spirochete” clade is also the most diverse group with regard to termite species representation (Table 3.3). The relative abundances of higher termite spirochete clade sequences in the inventories (74-100%) indicate spirochete-like FDH<sub>H</sub> types dominate FDH<sub>H</sub> diversity in all higher termite species but the litter-feeding termite Cost004, in which they are the second most abundant FDH<sub>H</sub> type (47%).

The broad distribution of higher termite spirochete clade sequences among termites from different subfamilies, coupled with the finding that they are the *only* spirochete-like FDH<sub>H</sub> types in the Gut spirochete group, indicates the higher termite spirochete clade represents an important evolutionary radiation within the FDH<sub>H</sub> gene pool of acetogenic spirochetes. We hypothesize this radiation is associated with the loss of most Sec clade

and all Cys clade alleles previously identified in the gut communities of evolutionarily primitive wood-feeding insects (Figure 3.3).

Figure 3.4 shows a detailed phylogeny of higher termite spirochete FDH<sub>H</sub> alleles (red clade in Figure 3.3). It appears phylogeny tracks the level of soil exposure: sequences from subterranean *Amitermes* and *Gnathamitermes* spp. are more derived than sequences from leaf litter-feeding *Rhynchotermes*, which are more derived than sequences from wood-feeding termites.

**Figure 3.4.** Phylogeny of higher termite spirochete FDH<sub>H</sub> sequences within the Sec clade of the Gut spirochete group (red colored clade in Figure 3.3). FDH<sub>H</sub> sequences predicted to encode cysteine in the position of selenocysteine are highlighted in red. Grey box highlights sequences recovered from palm-feeding termites collected at a beach in Costa Rica. Clone names containing 'sec' correspond to selenocysteine encoding sequences; those with 'cys' correspond to cysteine sequences. Tree was constructed with 601 aligned amino acids using PROTML. The branching position of a *Nasutitermes* metaganomic FDH<sub>H</sub> fragment (added in by parsimony using 250 amino acids) is indicated with a dashed line; phylogenetic distance represented by this dashed line is not comparable to any other sequence. Filled circles indicate nodes were supported by PROTML, parsimony (Phylip PROPARS), and distance (Fitch) methods of tree construction. Scale bar corresponds 0.1 units of amino acid change per alignment position.



### **Multiple reinventions of selenium-independent FDH<sub>H</sub> alleles in arboreal termites**

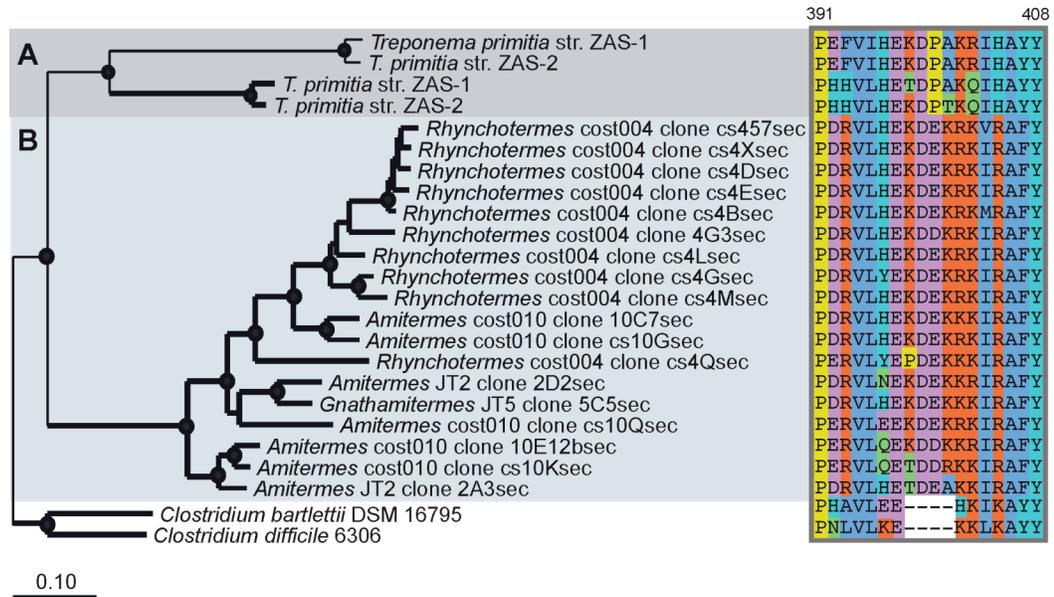
We analyzed spirochete-like FDH<sub>H</sub> alleles in arboreal higher termites (palm-feeding beach termites Cost007, Cost006, Cost008 and rainforest dwelling Cost 003) and identified several selenium-independent (cysteine encoding) FDH<sub>H</sub> alleles (highlighted in red, Figure 3.4). These appear to have been “reinvented” from selenocysteine-encoding FDH<sub>H</sub> alleles within the Sec clade of the Gut spirochete group (Figure 3.3, Figure 3.4), as they are nested within the higher termite spirochete clade, which is comprised primarily of selenocysteine-encoding FDH<sub>H</sub> sequences. This topology represents the first example of convergent evolution within the gut spirochete FDH<sub>H</sub> lineage. Moreover, the clustering of Cost006 and Cost008 cysteine-encoding FDH<sub>H</sub> sequences with each other to the exclusion of cysteine-encoding FDH<sub>H</sub> in Cost007 and Cost 003 point to two instances of convergent evolution, one in the *Microcerotermes* FDH<sub>H</sub> lineage and one in the *Nasutitermes* lineage. This suggests that the convergent evolution of cysteine-encoding FDH<sub>H</sub> in termites may have been driven by the sweeping loss of all Cys clade genes from the FDH<sub>H</sub> gene pool in higher termites followed by a major perturbation that decreased selenium availability in the gut community.

### **Sequences from subterranean and litter-feeding termites form a novel FDH<sub>H</sub> clade**

The guts of subterranean and litter-feeding termites harbored novel FDH<sub>H</sub> alleles, not identified in any other termite. The sequences phylogenetically group together into one clade (34 phylotypes), which we designate as the *Amitermes-Gnathamitermes-Rhynchotermes* group (Figure 3.5; Clade B, Figure 3.2). *Amitermes-Gnathamitermes-Rhynchotermes* group phylotypes represent 51% of Cost004, 13% of Cost010, 8% of JT2,

and 8% of JT5 inventories (Table 3.3; Figure 3.5, left panel). We were not able to infer the identity of the uncultured organisms encoding sequences in this clade, as the clade has no pure culture representatives and falls outside the Gut spirochete group. However, the presence of an amino acid indel (Figure 3.5, right panel) characteristic of the Gut spirochete group may indicate a spirochetal origin. If this origin is confirmed, these FDH<sub>H</sub> types would function in the direction of CO<sub>2</sub> fixation within the context Wood-Ljungdahl pathway. No additional information could be extracted from the genomic context of FDH<sub>H</sub> in *Clostridium difficile*, which falls basal to the *Amitermes-Gnathamitermes-Rhynchotermes* clade. However, we note that *C. difficile* possesses several Wood-Ljungdahl pathway genes.

Based on our inventory findings, we hypothesized that *Amitermes-Gnathamitermes-Rhynchotermes* FDH<sub>H</sub> sequences represent a group of gut symbionts present only in subterranean and litter-feeding termites (Table 3.3). To identify *Amitermes-Gnathamitermes-Rhynchotermes* FDH<sub>H</sub> sequences in other termites, we designed clade-specific primers (193F, 1045R) and screened lower and higher termite gut DNA using nested PCR methods (Figure 3.9, Appendix). Robust amplicons were detected in every subterranean and litter-feeding termite, but in no other termite species. This result implies *Amitermes-Gnathamitermes-Rhynchotermes* clade alleles are only present in subterranean and litter-feeding termites gut communities.



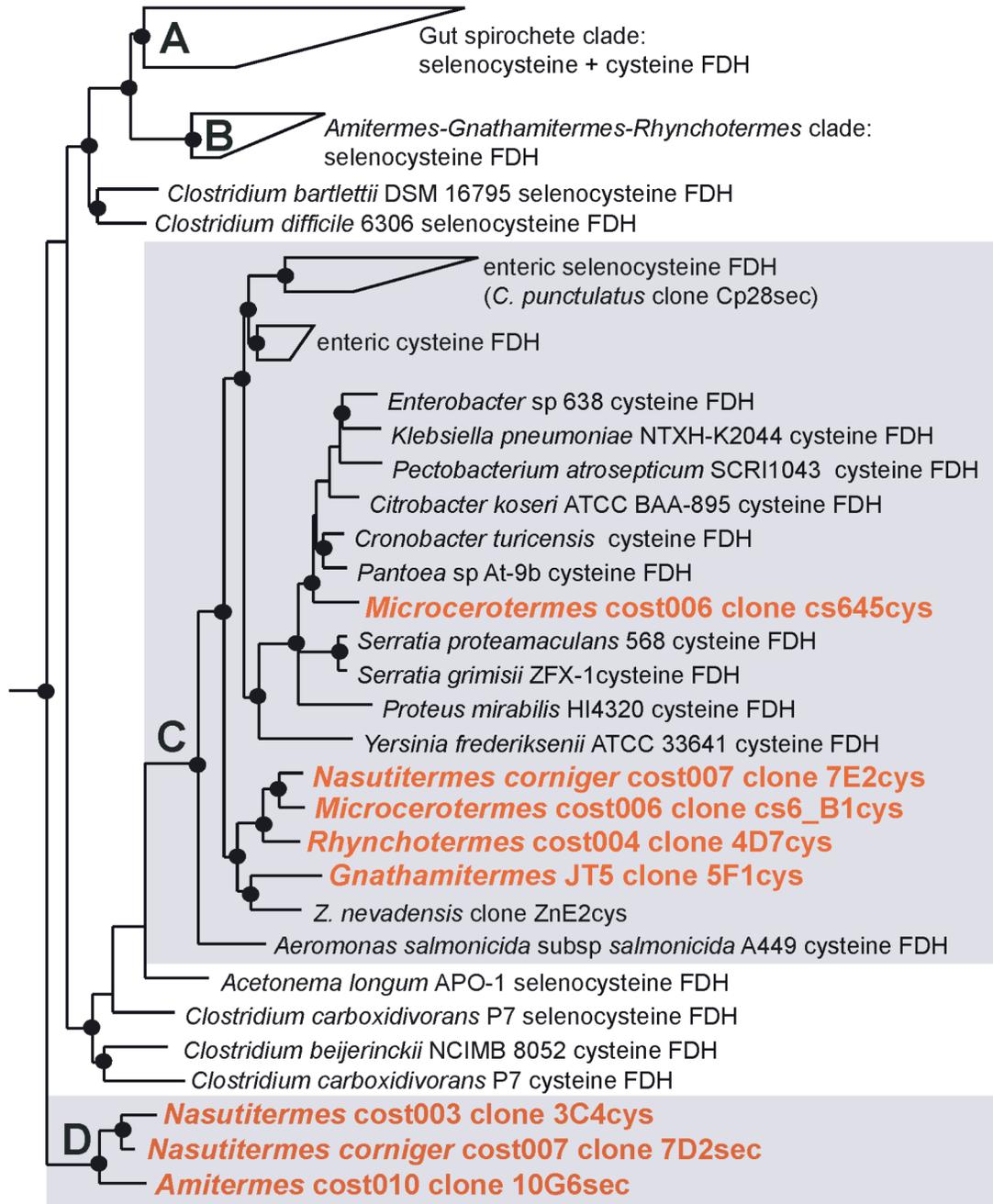
**Figure 3.5.** *Amitermes-Gnathamitermes-Rhynchotermes* FDH<sub>H</sub> clade (Clade B, Figure 3.2) detailed phylogeny (left panel, grey box B) and amino acid character analysis (right panel). Dark grey box A highlights the Gut spirochete group (Clade A, Figure 3.2), which is represented by *T. primitia*. Clone names containing ‘sec’ and branches in bold correspond to selenocysteine encoding FDH<sub>H</sub>; those with ‘cys’ correspond to cysteine FDH<sub>H</sub>. Tree was constructed with 595 aligned amino acids using the maximum likelihood algorithm Phylip PROTML. Filled circles indicate nodes were supported by PROTML, parsimony, and distance methods of analyses. Scale bar indicates 0.1 amino acid changes per alignment position. Numbers above the alignment refer to amino acid positions in the Sec FDH<sub>H</sub> of *T. primitia* str. ZAS-2.

**Enteric *Proteobacteria* and unclassified FDH<sub>H</sub> alleles**

Clade C in Figure 3.2 and Figure 3.6 (5 phylotypes) clusters within a clade of enteric *Proteobacteria* defined by the FDH<sub>H</sub> from *Aeromonas salmonicida* and likely represents uncultured enteric bacteria which operate FDH<sub>H</sub> in the oxidative direction during sugar fermentation (21). Enteric-like phylotypes account for 18% of JT5 clones but less than 7% in other termites (Table 3.3).

Clade D is a novel FDH<sub>H</sub> clade (Figure 3.6), consisting of rare sequence types found in Cost 010, Cost003, and Cost007. Its basal position relative to all other FDH<sub>H</sub> types make a prediction of function and 16S rRNA organism identity impossible. We therefore designate it as “unclassified.”

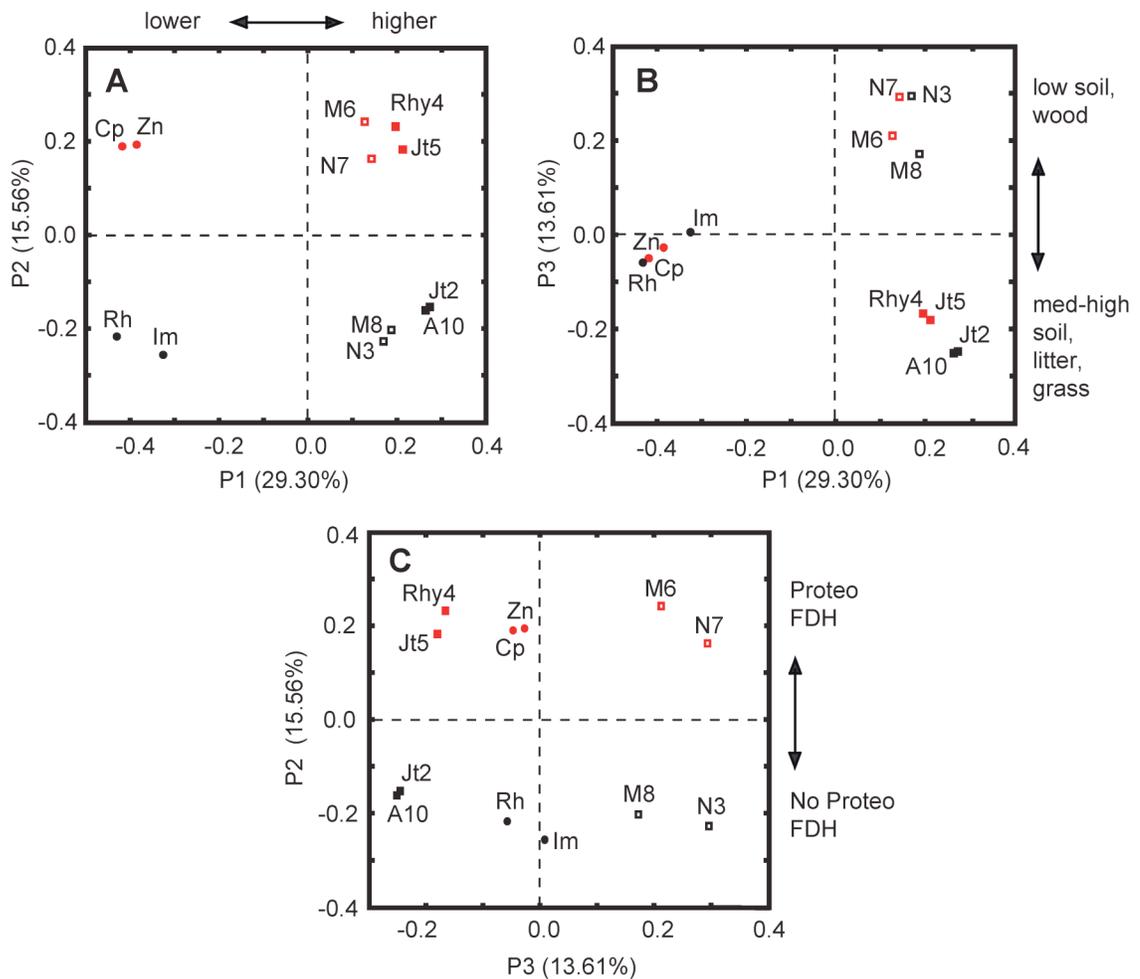
**Figure 3.6.** Detailed phylogenies of enteric *Proteobacteria* (Clade C, Figure 3.2) and unclassified FDH<sub>H</sub> sequences (Clade D, Figure 3.2). Sequences from higher termites are highlighted in red. Clone names containing ‘sec’ correspond to selenocysteine encoding FDHs; those with ‘cys’ correspond to cysteine FDHs. Tree was constructed using methods and setting specified in the legend of Figure 3.2. Filled circles indicate nodes were supported by PROTML and parsimony methods of analyses. Scale bar indicates 0.1 amino acid changes per alignment position.



**Principal component analysis reveals data cluster by higher order insect taxonomy, presence of enteric-like FDH<sub>H</sub>, and degree of soil exposure.**

Both phylotype abundance and FDH<sub>H</sub> phylogeny suggest that (i) insect phylogeny at higher taxonomic levels and (ii) degree of soil exposure may be important factors associated with FDH<sub>H</sub> phylogeny, and indirectly, the community structure of FDH<sub>H</sub>-bearing organisms. We performed a principal component analysis (Figure 3.7) using the phylogeny statistics software Unifrac (28) to explore these relationships. The first principal component (Figure 3.7, panels A and B) accounts for 29.30% of total variance and clearly separates lower termites from higher termites (x-axis, Figure 3.7, panel A). It does appear to differentiate the termites within these groups at finer phylogenetic scales (i.e., family, subfamily). The first principal component can also be viewed as tracking the presence (lower termite) or absence (higher termite) of flagellate protozoa. The second and third principal components account for similar levels of variance (15.56%, 13.61%). Principal component 2 (Figure 3.7, panel C) clusters inventories containing enteric fermentative FDH<sub>H</sub> types together (Figure 3.2, Figure 3.6, Table 3.3), whereas principal component 3 (Figure 3.7, panel B) is associated with the degree of soil exposure and diet (e.g., wood versus grass, roots, litter). This analysis suggests that geography, nest type, and habitat are not strongly associated with FDH<sub>H</sub> phylogeny.

**Figure 3.7.** Unifrac principal component analysis of FDH<sub>H</sub> phylogeny in termites and related insects. Principal components P1 – P3 (accounting for 58.4% of variance) are plotted against each other (panels A, B, C). The tree depicted in Figure 3.2 was analyzed with termite species set as the environment variable and 100 permutations. Wood-roach and lower termites: Cp, *C. punctulatus*; Zn, *Z. nevadensis*; Rh, *R. hesperus*; Im, *I. minor*. Higher termites: N3, *Nasutitermes* sp. Cost003; N7, *Nasutitermes corniger* Cost007; M6, *Microcerotermes* sp. Cost006; M8, *Microcerotermes* sp. Cost008; Rhy4, *Rhynchotermes* sp. Cost004; A10, *Amitermes* sp. Cost010; Jt2, *Amitermes* sp. JT2; Jt5, *Gnathamitermes* sp. JT5. *Proteobacteria*-like FDH<sub>H</sub> types: Cp, Zn, N7, M6, Rhy4, Jt5. Low soil exposure: Cp, Zn, Rh, Im, N3, N7, M6, M8. Medium soil exposure: Rhy4. High soil exposure: A10, Jt2, Jt5. Circles (Square) denote lower (higher) termites. Red color (black) denotes presence (absence) of enteric *Proteobacteria*. Filled (unfilled) symbols denote med-high soil (low) exposure.



## Discussion

In this study, we hypothesized that hydrogenase-linked formate dehydrogenase (FDH<sub>H</sub>) genes are absent from higher termite gut communities based on the anomalously poor recovery of genes for formate dehydrogenase from the gut metagenome of a wood-feeding higher termite (54). To investigate this unresolved feature of the metagenomic study, we constructed and analyzed FDH<sub>H</sub> gene inventories from 8 species of higher termite. The results indicate that FDH<sub>H</sub> genes are a common feature in the symbiotic gut communities of taxonomically and geographically diverse higher termites whose lifestyles vary with respect to nesting strategy, diet, and soil exposure.

We suggest that compartment specific sampling efforts and/or methodological artifacts may be the cause of low FDH recovery in the metagenome. Both of these possibilities have merit. With respect to compartment specific sampling, studies of methanogenesis in the soil-feeding termite *Cubitermes* demonstrate cross-epithelial H<sub>2</sub> transfer between gut compartments and imply formate transfer between gut compartments is also possible (50). This suggests FDH genes may be present outside the gut chamber sampled for metagenomic analysis. If this is true, acetogens inhabiting the largest gut compartment would be heterotrophs, as they would rely on outside sources of formate for acetate production. Gut compartment specific inventories of FDH should clarify the issue. With respect to methodology, one study of spirochete acetogenesis genes suggests that spirochete DNA is difficult to clone (32). Thus DNA toxicity issues may also contribute to low metagenome FDH gene recovery.

In addition to showing that  $FDH_H$  genes are present and diverse in higher termites, our inventory analysis revealed that  $FDH_H$  sequences affiliated with a clade of selenium-independent  $FDH_H$  alleles, widely distributed in lower termites and a wood roach, are absent in all higher termites. We confirmed the absence of these ‘Cys clade’ sequences in inventories by performing clade specific PCR amplifications of gut DNA from diverse lignocellulose-feeding insects. Phylogenetic analysis also revealed a novel clade of  $FDH_H$  comprised of sequences from subterranean and litter feeding termites. We provided additional support for the absence of *Amitermes-Gnathamitermes-Rhynchotermes* clade sequences in wood-feeding arboreal higher termites, lower termites, and a wood-roach, with clade-specific PCR amplifications. Taken together, these phylogenetic patterns have important implications for the ecology and evolutionary biology of uncultured acetogenic spirochetes and other termite gut bacteria. We discuss these implications in the following sections.

### **$FDH_H$ phylogeny and diet**

Spirochete-like  $FDH_H$  sequences are by far the most abundant  $FDH_H$  type in phylogenetically diverse higher termites whose diets consist primarily of lignocellulose (i.e., those eating wood or dried grass, Cost003, Cost007, Cost006, Cost008, JT2, JT5). This result, along with similar observations made in lower wood-feeding termites, suggests lignocellulose-degrading gut communities harbor a stable niche for  $H_2$ -utilizing acetogenic spirochetes.

Spirochete-like FDH<sub>H</sub> sequences are the second most abundant type in leaf litter-feeding Cost004, who presumably consumes higher levels of tannins due to the elevated tannin levels in leaves relative to other plant parts (23). The most abundant FDH<sub>H</sub> type in Cost004 belongs to the *Amitermes-Gnathamitermes-Rhynchotermes* clade, which contains sequences from all soil-exposed termites (subterranean and litter-feeding). We identified these alleles at lower levels in subterranean termites Cost010, JT2, and JT5, whose diets consist of monocots (sugarcane root, grass) that are low in tannin (23). None were recovered from termites (Cost003, Cost007, Cost006, Cost008) feeding on woods, which tend to have the lowest tannin levels of all plant parts (23).

Rates of acetogenesis have not been measured for litter-feeding termites, so it is unclear whether the shift in FDH<sub>H</sub> diversity is associated with a less productive acetogenic treponeme population. If acetogenesis rates in litter-feeding termites are comparable to those in wood-feeding termites, the *Amitermes-Gnathamitermes-Rhynchotermes* group of FDH<sub>H</sub> types may represent a novel group of uncultured acetogens, which have greater tolerance to phenolic compounds like tannin. Alternatively, they may belong to a group of tannin tolerant fermenting bacteria that utilize residual leaf sugars. In any case, the phylogenetic isolation of the *Amitermes-Gnathamitermes-Rhynchotermes* clade from other major groups suggests that a niche which was previously small or absent in wood-feeding termites gained importance in termites that feed on decaying plants and have substantial contact with soil. In the latter case, the presence of *Amitermes-Gnathamitermes-Rhynchotermes* clade FDH<sub>H</sub> alleles would signal the influx of new FDH<sub>H</sub> gene stock into the gut community. This could occur by lateral gene transfer from

an organism passing through the gut to an established gut symbiont or the acquisition of a new symbiont tolerant of phenolic compounds from the surrounding soil environment.

### **FDH<sub>H</sub> phylogeny and acetogenic spirochete evolution**

Initial glimpses into the evolutionary histories of host-symbiont and symbiont-symbiont relationships within termite gut microbial communities have been provided by 16S rRNA surveys of bacterial diversity [Ohkuma *et al.* (40), Eggleton (16), and references therein]. These studies suggest the relationships are highly complex, showing signs of coevolution, symbiont loss, and acquisition (36) at varying taxonomic scales. In particular, the community structure of spirochetes does not track host phylogeny at family or subfamily levels, but shows signs of extinction, evolutionary radiations, and multiple instances of symbiont acquisition [reviewed by (40)]. However, the species richness of gut microbiota (and their insect hosts) may prove prohibitive to gaining a comprehensive understanding of bacterial evolution based on 16S rRNA. More importantly, our ability to infer the factors and impacts associated with evolutionary patterns is ultimately limited by our meager knowledge of the various roles different symbiont populations play in diverse nutritional mutualisms.

In light of these concerns, we took a focused approach and used a functional gene (*fdhF*), used in fermentation and acetogenesis, to identify evolutionary patterns for metabolically similar organisms within the gut communities of termites belonging to different lineages and characterized by different habitats and lifestyles. We note that a functional gene approach has its own drawbacks, namely the decoupling of an organism from its genes

when horizontal gene transfer is at play. Nevertheless, we believe *fdhF* inventory data highlights intriguing patterns of diversity that shed light on the complex evolutionary history of termites and their gut symbionts.

Phylogenetic patterns within the Gut spirochete clade (Figure 3.2) imply acetogenic spirochetes in higher termite gut communities have experienced events and challenges not faced or reflected in the lower termite and wood-roaches. Several Sec and Cys clade FDH<sub>H</sub> alleles of likely spirochete origin were previously identified in lower termites and *C. punctulatus*. Our analysis of FDH<sub>H</sub> in 8 higher termite species yielded zero sequences that grouped within the Cys clade, comprised of sequences from extant primitive termites and *C. punctulatus*. Instead, all spirochete-like FDHs in higher termites affiliated with a single FDH<sub>H</sub> lineage located within the Sec clade of the Gut spirochete group. This topology is strong evidence that dramatic restructuring within the FDH gene pool occurred during evolution of higher termite subfamilies *Nasutitermitinae* and *Termitinae*. If we consider the absence of flagellate protozoa in extant higher termites (12), the results suggest that a sweeping loss of FDH<sub>H</sub> genes may have accompanied the extinction of flagellates, as any FDH<sub>H</sub>-bearing acetogenic spirochetes physically associated with flagellates or dependent on flagellate metabolites would also go extinct. Alternatively, sweeping genes loss may have resulted from a “molting bottleneck” (i.e., incomplete gut community transfer during the re-inoculation of freshly molted termites by their nest-mates).

Our hypothesis that the FDH<sub>H</sub> gene pool has undergone sweeping gene loss associated with an evolutionary bottleneck is consistent with (i) the total absence of Cys clade FDH<sub>H</sub> genes in every higher termite examined herein and (ii) the independent (re)invention of cysteine-encoding FDH<sub>H</sub> from Sec clade FDH<sub>H</sub> gene stock by acetogenic spirochetes in two different termite subfamilies (*Nasutitermitinae*, *Termitinae*). We posit that the presence of ‘lower termite-type’ Cys clade FDH<sub>H</sub> genes at any (biologically) significant abundance in higher termite gut communities should preclude convergent evolution, as the organisms bearing such FDH<sub>H</sub> variants would proliferate under environmental selection (e.g., low selenium conditions) and out-compete organisms that have only the Sec clade alleles. Alternatively, convergent evolution would not be required if Cys clade alleles were laterally transferred from a population less fit in higher termite guts for reasons unrelated to selenium. In any case, gene inventories showed no signs that either of the preceding two scenarios occurred, leaving sweeping gene loss (i.e., genetic extinction of lower termite type Cys alleles) as the most reasonable conclusion. We also note that phylogenetic patterns consistent with an evolutionary radiation of a “founding” Sec FDH<sub>H</sub> allele within a surviving population of acetogenic spirochetes serves as additional support for sweeping gene loss due to an evolutionary bottleneck.

The selective forces behind the convergent evolution of cysteine-encoding FDH<sub>H</sub> variants in higher termites are unclear. We postulate that dietary selenium (Se) may play a role, as the majority of reinvented cysteine FDH<sub>H</sub> alleles were identified in termites collected from a beach area (Cost006, Cost007, and Cost008), which may be regularly submerged in low Se seawater. Nriagu *et al.* (38) estimate total Se concentrations in ocean surface

mixed layers are 4-orders of magnitude lower than in surface soils. A reasonable assumption is that this low Se seawater flushes out Se from beach soil, reducing Se levels in plants, and consequently the diet of termites. This hypothesis is consistent with the finding that gene transcription of the Cys FDH<sub>H</sub> allele in the acetogenic treponeme, *T. primitia*, is controlled by media Se concentration (32).

Even if dietary Se were the driver for convergent evolution of genes for cysteine-encoding FDH<sub>H</sub> in Cost006, Cost007, and Cost008, the larger question of why selenocysteine FDH<sub>H</sub> genes are favored to the apparent exclusion of all cysteine FDH<sub>H</sub> genes in higher termites remains unanswered. Was it a shift in gut structure, from a single hindgut paunch to a gut tract characterized by multiple chambers, which relaxed or removed the selective pressure of Se limitation on the gut community? If so, what led to the invention of a multi-chamber gut? These and many other important questions remain unanswered, but need to be explored given the abundance and species-richness of higher termites.

## Appendix

**Figure 3.8.** Targeted PCR assays on termite and roach gut DNA using lower termite spirochete group Cys clade *fdhF* specific primers.

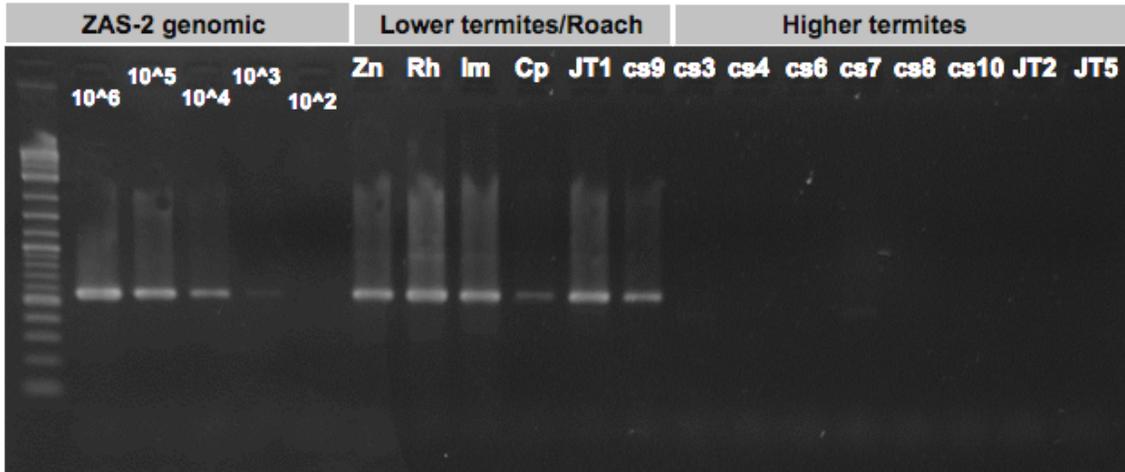
**Figure 3.9.** Targeted PCR assays using universal *fdhF* primers followed by *Amitermes-Gnathamitermes-Rhychotermes* clade specific primers on gut templates.

**Table 3.4.** PCR conditions for clone library construction.

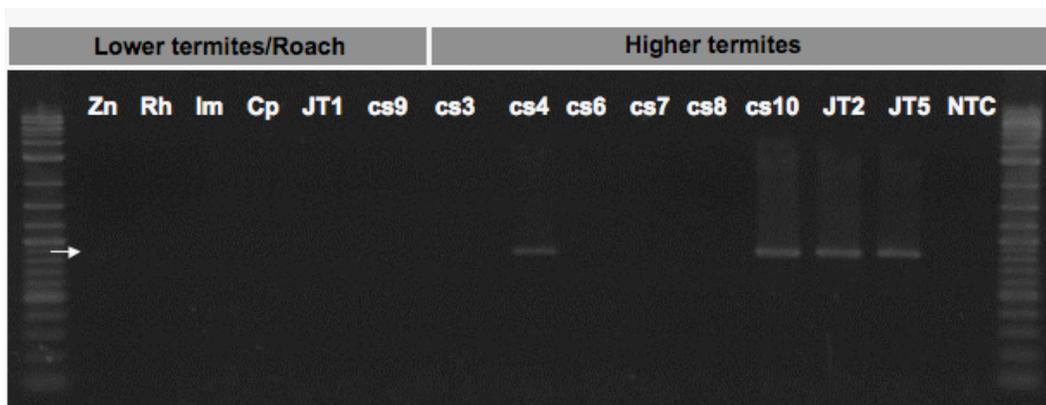
**Table 3.5.** Phylotype distribution in each library.

**Table 3.6.** Sequences used in phylogenetic analysis.

**Figure 3.8.** Targeted PCR assays on termite and roach gut DNA using Cys clade specific *fdhF* primers (Cys499F1b, 1045R), which yield a ca. 600 bp product. Templates are: ZAS-2, *T. primitia* str. ZAS-2 genomic DNA; Zn, *Z. nevadensis*; Rh, *R. hesperus*; Im, *I. minor*; Cp, *C. punctulatus*; JT1, *R. tibialis*; cs9, *Coptotermes* sp. Cost009; cs3, *Nasutitermes* sp. Cost003; cs4, *Rhynchotermes* sp. Cost004; cs6, *Microcerotermes* sp. Cost006; cs7, *Nasutitermes corniger* Cost007; cs8, *Microcerotermes* sp. Cost008; cs10, *Amitermes* sp. Cost010; JT2, *Amitermes* sp. JT2; JT5, *Gnathamitermes* sp. JT5. Numbers in ZAS-2 genomic lanes refer to the number of genome copies per reaction. Copy numbers ( $10^6$  copies/gut) in the lower termite *Z. nevadensis* were estimated from band strength in dilution-to-extinction PCR of *T. primitia* ZAS-2 DNA (assuming a yield of 1  $\mu\text{g}$  total DNA/gut typically observed in QIAGEN DNA extractions, 10% derived from prokaryotes, and  $10^4$  copies/ng gut DNA in *Z. nevadensis*). As Cys bands were not present in higher termites, the detection limit (100 copies/ng gut DNA) was used to estimate a maximum abundance of  $10^3$  copies/gut for lower termite Cys clade FDH genes in higher termites (assuming a yield of 0.25  $\mu\text{g}$  total DNA/gut, 100% derived from prokaryotes).



**Figure 3.9.** Products from nested PCR reactions using (i) universal *fdhF* primers followed by (ii) *Amitermes-Gnathamitermes-Rhychotermes* clade specific primers on gut templates. Template designations can be found in the legend of Figure 3.8. (Note, slight band in Zn lane.)



**Table 3.4.** PCR conditions for clone library construction. Shaded grey rows highlight templates for which multiple libraries were created. Thermocycling conditions for each PCR reaction were 94°C for 2 min, 25 cycles of (94°C for 30 s, annealing for 1 min, 68°C for 2 min 30 s), then 68°C for 10 min.

Source of gut DNA	Library	Polymerase (U/ $\mu$ l)	Template (ng/ $\mu$ l)	Annealing Temp °C
<i>Nasutitermes sp.</i> isolate Cost003	3L1	0.035	0.25	55
<i>Nasutitermes sp.</i> isolate Cost003	3L2	0.14	0.25	51
<i>Nasutitermes corniger</i> isolate Cost007	7L1	0.14	0.25	51
<i>Rhynchotermes sp.</i> isolate Cost004	4L1	0.035	0.25	55
<i>Rhynchotermes sp.</i> isolate Cost004	4L2	0.14	0.25	51
<i>Microcerotermes sp.</i> isolate Cost006	6L1	0.14	0.25	51
<i>Microcerotermes sp.</i> isolate Cost008	8L1	0.035	1	53.6
<i>Amitermes sp.</i> isolate Cost010	10L1	0.035	0.5	53.6
<i>Amitermes sp.</i> isolate Cost010	10L2	0.14	0.25	51
<i>Amitermes sp.</i> isolate JT2	Jt2L1	0.07	0.05	51
<i>Gnathamitermes sp.</i> isolate JT5	Jt5L1	0.07	0.05	51

**Table 3.5.** Phylotype distribution in each library.

Source of gut template	Phylotype	Abundance (%) in Library		
<i>Nasutitermes</i> sp. Cost003		<b><u>3L1</u></b>	<b><u>3L2</u></b>	
	cs3Csec	73.6	70.6	
	cs3Bsec	20.7	17.6	
	cs3Isec	4.6	0.0	
	cs3Psec	1.1	0.0	
	3C4cys	0.0	5.9	
	3D6cys	0.0	5.9	
	total clones	87	17	
<i>Rhynchotermes</i> sp. Cost004		<b><u>4L1</u></b>	<b><u>4L2</u></b>	
	cs4Asec	27.1	9.1	
	cs4Isec	12.9	0.0	
	cs4Esec	11.8	9.1	
	cs4Osec	8.2	22.7	
	cs4Bsec	8.2	0.0	
	cs4Gsec	7.1	31.8	
	cs4Dsec	7.1	4.5	
	cs4Msec	4.7	4.5	
	cs4Lsec	4.7	0.0	
	cs457sec	3.5	0.0	
	cs4Xsec	2.4	0.0	
	cs489sec	1.2	4.5	
	cs4Qsec	1.2	0.0	
	4D7cys	0.0	9.1	
	4G3sec	0.0	4.5	
	total clones	85	22	
	<i>Microcerotermes</i> sp. Cost006		<b><u>6L1</u></b>	
		cs6_23sec	47.3	
		cs6_31cys	32.4	
cs6_26sec		14.9		
cs6_B1cys		2.7		
cs6_F3cys		1.4		
cs6_45cys		1.4		
total clones		74		
<i>Nasutitermes</i> corniger Cost007		<b><u>7L1</u></b>		
	cs7F6sec	43.3		
	cs7E6cys	16.7		

---

	7B7sec	13.3	
	7D2sec	6.7	
	7E2cys	6.7	
	7H1cys	6.7	
	7B4sec	3.3	
	7G7sec	3.3	
	total clones	30	
<i>Microcerotermes</i> sp. Cost008			
		<b><u>8L1</u></b>	
	cs8Csec	42.9	
	cs8Asec	39.3	
	cs8Bcys	10.7	
	cs8Dcys	7.1	
	total clones	84	
<i>Amitermes</i> sp. Cost010			
		<b><u>10L1</u></b>	<b><u>10L2</u></b>
	cs10Dsec	62.8	30.4
	cs10Asec	21.8	21.7
	cs10Gsec	7.7	8.7
	cs10Isec	2.6	4.3
	cs10Ksec	2.6	0.0
	cs10Qsec	1.3	0.0
	cs10Vsec	1.3	0.0
	10B6sec	0	13.0
	10C7sec	0	4.3
	10G6sec	0	8.7
	10E12bsec	0	4.3
	10F3sec	0	4.3
	total clones	78	23
<i>Amitermes</i> sp. JT2			
		<b><u>It2L1</u></b>	
	2A1sec	68.3	
	2D2sec	6.9	
	2B4sec	5.9	
	2G3sec	8.9	
	2B9sec	6.9	
	2E6sec	1.0	
	2B12sec	1.0	
	2A3sec	1.0	
	total clones	101	
<i>Gnathamitermes</i> sp. JT5			
		<b><u>It5L1</u></b>	
	5F1cys	17.9	

---

---

5A1sec	23.8
5C2sec	20.2
5C5sec	8.3
5B11sec	11.9
5D3sec	4.8
5D5sec	4.8
5A7sec	3.6
5A2sec	2.4
5B1sec	2.4
total clones	84

---

**Table 3.6.** Sequences used in phylogenetic analyses. COII, cytochrome oxidase. FDH-H, hydrogenase-linked formate dehydrogenase. FDH-NAD, NAD-linked formate dehydrogenase. FDH-F420, F420-linked formate dehydrogenase.

Source	Gene	Accession
<i>Amitermes dentatus</i>	COII	DQ442065
<i>Amitermes evuncifer</i>	COII	DQ442066
<i>Archotermopsis wroughtoni</i>	COII	DQ442080
<i>Coptotermes niger</i>	COII	DQ442104
<i>Coptotermes testaceus</i>	COII	DQ442102
<i>Cornitermes pugnax</i>	COII	DQ442106
<i>Cornitermes walkeri</i>	COII	AB005577
<i>Cryptotermes domesticus</i>	COII	AF189086
<i>Cryptotermes secundus</i>	COII	AF189093
<i>Cryptocercus clevelandi</i>	COII	DQ007626
<i>Cryptocercus primarius</i>	COII	DQ007644
<i>Cryptocercus punctulatus</i>	COII	AB005462
<i>Deropeltis erythrocephala</i>	COII	DQ874271
<i>Hodotermopsis japonica</i>	COII	AB018391
<i>Hodotermopsis sp.</i>	COII	AB018395
<i>Incisitermes minor</i> isolate Pas1	COII	GQ922441
<i>Incisitermes immigrans</i>	COII	AB109542
<i>Kalotermes flavicollis</i>	COII	DQ442147
<i>Labiotermes labralis</i>	COII	DQ442149
<i>Mastotermes darwiniensis</i>	COII	AB014071
<i>Microcerotermes arboreus</i>	COII	DQ442164
<i>Microcerotermes parvus</i>	COII	DQ442167
<i>Nasutitermes corniger</i>	COII	AB037327
<i>Nasutitermes ephratae</i>	COII	AB037328
<i>Nasutitermes sp.</i> warnecke-2007	COII	EU236539
<i>Nasutitermes nigriceps</i>	COII	AB037329

---

<i>Periplaneta americana</i>	COII	M83971
<i>Periplaneta australasiae</i>	COII	DQ874310
<i>Reticulitermes hesperus</i> isolate ChiA2	COII	GQ922442
<i>Reticulitermes santonensis</i>	COII	AF291743
<i>Reticulitermes speratus</i>	COII	AB109530
<i>Reticulitermes tibialis</i>	COII	AY168206
<i>Zootermopsis nevadensis</i> isolate ChiA1	COII	GQ922444
<i>Zootermopsis angusticollis</i>	COII	DQ442267
<i>Cryptocercus punctulatus</i> nymph	COII	HM208251
<i>Nasutitermes</i> sp. Cost003	COII	HM208252
<i>Rhynchotermes</i> sp. Cost004	COII	HM208253
<i>Microcerotermes</i> sp. Cost008	COII	HM208254
<i>Amitermes</i> sp. Cost010	COII	HM208255
<i>Microcerotermes</i> sp. Cost006	COII	HM208256
<i>Nasutitermes corniger</i> Cost007	COII	HM208257
<i>Coptotermes</i> sp. Cost009	COII	HM208258
<i>Reticulitermes tibialis</i> JT1	COII	HM208248
<i>Gnathamitermes</i> sp. JT5	COII	HM208249
<i>Amitermes</i> sp. JT2	COII	HM208250
<i>Aeromonas salmonicida</i> subsp. <i>salmonicida</i> A449	FDH-H	YP_001141645
<i>Aggregatibacter aphrophilus</i> NJ8700	FDH-H	YP_003007599, YP_003007598
<i>Acetonema longum</i> APO-1	FDH-H	GQ922445
<i>Buttiauxiella</i> SN1	FDH-H	GQ922446
<i>Citrobacter koseri</i> ATCC BAA-895 (copy 2)	FDH-H	YP_001453385
<i>Citrobacter koseri</i> ATCC BAA-895 (copy 1)	FDH-H	YP_001455313, YP_001455315
<i>Citrobacter</i> TSA-1	FDH-H	GQ922447
<i>Clostridium bartlettii</i> DSM 16795	FDH-H	ZP_02210704
<i>Clostridium beijerinckii</i> NCIMB 8052	FDH-H	YP_001310874
<i>Clostridium carboxidivorans</i> P7 (copy 1)	FDH-H	ZP_05394379, ZP_05394380

---

---

<i>Clostridium carboxidivorans</i> P7 (copy 2)	FDH-H	ZP_05390901
<i>Clostridium difficile</i> 630	FDH-H	YP_001089834
<i>C. punctulatus nymph</i> gut clone Cp10sec	FDH-H	GU563433
<i>C. punctulatus nymph</i> gut clone Cp14sec	FDH-H	GU563436
<i>C. punctulatus nymph</i> gut clone Cp16sec	FDH-H	GU563432
<i>C. punctulatus nymph</i> gut clone Cp24sec	FDH-H	GU563451
<i>C. punctulatus nymph</i> gut clone Cp28sec	FDH-H	GU563450
<i>C. punctulatus nymph</i> gut clone Cp34sec	FDH-H	GU563452
<i>C. punctulatus nymph</i> gut clone Cp3sec	FDH-H	GU563434
<i>C. punctulatus nymph</i> gut clone Cp72cys	FDH-H	GU563437
<i>C. punctulatus nymph</i> gut clone Cp78sec	FDH-H	GU563453
<i>C. punctulatus nymph</i> gut clone Cp82sec	FDH-H	GU563454
<i>C. punctulatus nymph</i> gut clone Cp94sec	FDH-H	GU563455
<i>C. punctulatus nymph</i> gut clone Cp9cys	FDH-H	GU563441
<i>C. punctulatus nymph</i> gut clone CpB10sec	FDH-H	GU563442
<i>C. punctulatus nymph</i> gut clone CpB2sec	FDH-H	GU563446
<i>C. punctulatus nymph</i> gut clone CpB3sec	FDH-H	GU563440
<i>C. punctulatus nymph</i> gut clone CpC1cys	FDH-H	GU563444
<i>C. punctulatus nymph</i> gut clone CpC3sec	FDH-H	GU563443
<i>C. punctulatus nymph</i> gut clone CpD1cys	FDH-H	GU563445
<i>C. punctulatus nymph</i> gut clone CpD8sec	FDH-H	GU563439
<i>C. punctulatus nymph</i> gut clone CpE8cys	FDH-H	GU563447
<i>C. punctulatus nymph</i> gut clone CpF1cys	FDH-H	GU563435
<i>C. punctulatus nymph</i> gut clone CpF8cys	FDH-H	GU563449
<i>C. punctulatus nymph</i> gut clone CpF9cys	FDH-H	GU563448
<i>C. punctulatus nymph</i> gut clone CpH1cys	FDH-H	GU563438
<i>Cronobacter turicensis</i> (copy 2)	FDH-H	YP_003210268
<i>Cronobacter turicensis</i> (copy 1)	FDH-H	YP_003210272, YP_003210273
<i>Dickeya dadantii</i> Ech703	FDH-H	YP_002986892

---

---

<i>Edwardsiella ictaluri</i> 93-146	FDH-H	YP_002934652, YP_002934653
<i>Enterobacter</i> sp. 638 (copy 1)	FDH-H	YP_001175022, YP_001175021
<i>Escherichia coli</i> str. K-12 substr MG1655	FDH-H	NP_418503
<i>I. minor</i> Pas1 gut clone Im10sec	FDH-H	GQ922349
<i>I. minor</i> Pas1 gut clone Im11cys	FDH-H	GQ922364
<i>I. minor</i> Pas1 gut clone Im15sec	FDH-H	GQ922351
<i>I. minor</i> Pas1 gut clone Im22sec	FDH-H	GQ922353
<i>I. minor</i> Pas1 gut clone Im24cys	FDH-H	GQ922369
<i>I. minor</i> Pas1 gut clone Im26sec	FDH-H	GQ922354
<i>I. minor</i> Pas1 gut clone Im27sec	FDH-H	GQ922355
<i>I. minor</i> Pas1 gut clone Im3sec	FDH-H	GQ922356
<i>I. minor</i> Pas1 gut clone Im42cys	FDH-H	GQ922371
<i>I. minor</i> Pas1 gut clone Im5cys	FDH-H	GQ922373
<i>I. minor</i> Pas1 gut clone Im63sec	FDH-H	GQ922361
<i>Klebsiella pneumoniae</i> NTXH-K2044 (copy 1)	FDH-H	YP_002917305
<i>Klebsiella pneumoniae</i> NTXH-K2044 (copy 2)	FDH-H	YP_002919873
<i>Methanocaldococcus jannaschii</i> DSM 2661	FDH-F420	P61159
<i>Moorella thermoacetica</i> ATCC 39073	FDH-F420	YP_431025
<i>Methanococcus maripaludis</i> S2	FDH-F420	CAF29694
<i>Methanococcus vannielii</i> SB	FDH-F420	ABR54514
<i>Pantoea</i> sp. At-9b	FDH-H	ZP_05726796
<i>Pectobacterium atrosepticum</i> SCRI1043 (copy 2)	FDH-H	CAG74160
<i>Proteus mirabilis</i> HI4320 (copy 2)	FDH-H	YP_002152680
<i>Proteus mirabilis</i> HI4320 (copy 1)	FDH-H	YP_002153253
<i>R. hesperus</i> ChiA2 gut clone Rh15cys	FDH-H	GQ922398
<i>R. hesperus</i> ChiA2 gut clone Rh24sec	FDH-H	GQ922383
<i>R. hesperus</i> ChiA2 gut clone Rh2sec	FDH-H	GQ922381
<i>R. hesperus</i> ChiA2 gut clone Rh35sec	FDH-H	GQ922385
<i>R. hesperus</i> ChiA2 gut clone Rh36cys	FDH-H	GQ922410

---

---

<i>R. hesperus</i> ChiA2 gut clone Rh41sec	FDH-H	GQ922386
<i>R. hesperus</i> ChiA2 gut clone Rh47cys	FDH-H	GQ922402
<i>R. hesperus</i> ChiA2 gut clone Rh53sec	FDH-H	GQ922389
<i>R. hesperus</i> ChiA2 gut clone Rh54cys	FDH-H	GQ922404
<i>R. hesperus</i> ChiA2 gut clone Rh65cys	FDH-H	GQ922406
<i>R. hesperus</i> ChiA2 gut clone Rh71sec	FDH-H	GQ922391
<i>R. hesperus</i> ChiA2 gut clone Rh93cys	FDH-H	GQ922409
<i>R. hesperus</i> ChiA2 gut clone Rh9sec	FDH-H	GQ922397
<i>Salmonella typhimurium</i> LT2	FDH-H	NP_463150
<i>Serratia proteamaculans</i> 568	FDH-H	YP_001478653
<i>Serratia grimesii</i> ZFX-1	FDH-H	GQ922448
<i>Shigella</i> sp. D9	FDH-H	ZP_05433594, ZP_054335931
<i>Yersinia frederiksenii</i> ATCC 33641 (copy 1)	FDH-H	ZP_04632644
<i>Yersinia frederiksenii</i> ATCC 33641 (copy 2)	FDH-H	ZP_04631307
<i>Treponema primitia</i> str. ZAS-1 (copy 2)	FDH-H	GQ922450
<i>Treponema primitia</i> str. ZAS-1 (copy 1)	FDH-H	GQ922449
<i>Treponema primitia</i> str. ZAS-2 (copy 2)	FDH-H	FJ479767
<i>Treponema primitia</i> str. ZAS-2 (copy 1)	FDH-H	FJ479767
<i>Nasutitermes</i> sp. metagenome contig tgut2b_BHZN47861_b2	FDH-H	IMG Gene object ID: 2004163507
<i>Z. nevadensis</i> ChiA1 gut clone Zn13cys	FDH-H	GQ922430
<i>Z. nevadensis</i> ChiA1 gut clone Zn2cys	FDH-H	GQ922431
<i>Z. nevadensis</i> ChiA1 gut clone Zn51sec	FDH-H	GQ922423
<i>Z. nevadensis</i> ChiA1 gut clone Zn61sec	FDH-H	GQ922426
<i>Z. nevadensis</i> ChiA1 gut clone Zn70sec	FDH-H	GQ922428
<i>Z. nevadensis</i> ChiA1 gut clone Zn9cys	FDH-H	GQ922435
<i>Z. nevadensis</i> ChiA1 gut clone ZnA4cys	FDH-H	GU563456
<i>Z. nevadensis</i> ChiA1 gut clone ZnB3cys	FDH-H	GU563459
<i>Z. nevadensis</i> ChiA1 gut clone ZnB5sec	FDH-H	GU563460
<i>Z. nevadensis</i> ChiA1 gut clone ZnB8sec	FDH-H	GU563461

---

---

<i>Z. nevadensis</i> ChiA1 gut clone ZnB9cys	FDH-H	GU563462
<i>Z. nevadensis</i> ChiA1 gut clone ZnC11cys	FDH-H	GU563466
<i>Z. nevadensis</i> ChiA1 gut clone ZnC1cys	FDH-H	GU563463
<i>Z. nevadensis</i> ChiA1 gut clone ZnC6sec	FDH-H	GU563464
<i>Z. nevadensis</i> ChiA1 gut clone ZnC8sec	FDH-H	GU563465
<i>Z. nevadensis</i> ChiA1 gut clone ZnD2sec	FDH-H	GU563467
<i>Z. nevadensis</i> ChiA1 gut clone ZnD3cys	FDH-H	GU563468
<i>Z. nevadensis</i> ChiA1 gut clone ZnE2cys	FDH-H	GU563469
<i>Z. nevadensis</i> ChiA1 gut clone ZnF7sec	FDH-H	GU563458
<i>Z. nevadensis</i> ChiA1 gut clone ZnH6cys	FDH-H	GU563457
<i>Z. nevadensis</i> ChiA1 gut clone ZnH8cys	FDH-H	GU563470
<i>Z. nevadensis</i> ChiA1 gut clone ZnHcys	FDH-H	GQ922420
<i>Amitermes</i> sp. Cost010 gut clone cs10Dsec	FDH-H	HM208218
<i>Amitermes</i> sp. Cost010 gut clone cs10Isec	FDH-H	HM208221
<i>Amitermes</i> sp. Cost010 gut clone 10B6sec	FDH-H	HM208225
<i>Amitermes</i> sp. Cost010 gut clone 10C7sec	FDH-H	HM208226
<i>Amitermes</i> sp. Cost010 gut clone 10E12bsec	FDH-H	HM208228
<i>Amitermes</i> sp. Cost010 gut clone 10F3sec	FDH-H	HM208229
<i>Amitermes</i> sp. Cost010 gut clone 10G6sec	FDH-H	HM208227
<i>Amitermes</i> sp. Cost010 gut clone cs10Asec	FDH-H	HM208219
<i>Amitermes</i> sp. Cost010 gut clone cs10Gsec	FDH-H	HM208220
<i>Amitermes</i> sp. Cost010 gut clone cs10Ksec	FDH-H	HM208222
<i>Amitermes</i> sp. Cost010 gut clone cs10Qsec	FDH-H	HM208223
<i>Amitermes</i> sp. Cost010 gut clone cs10Vsec	FDH-H	HM208224
<i>Amitermes</i> sp. JT2 gut clone 2A1sec	FDH-H	HM208230
<i>Amitermes</i> sp. JT2 gut clone 2A3sec	FDH-H	HM208237
<i>Amitermes</i> sp. JT2 gut clone 2B12sec	FDH-H	HM208236
<i>Amitermes</i> sp. JT2 gut clone 2B4sec	FDH-H	HM208232
<i>Amitermes</i> sp. JT2 gut clone 2B9sec	FDH-H	HM208234

---

---

<i>Amitermes</i> sp. JT2 gut clone 2D2sec	FDH-H	HM208231
<i>Amitermes</i> sp. JT2 gut clone 2E6sec	FDH-H	HM208235
<i>Amitermes</i> sp. JT2 gut clone 2G3sec	FDH-H	HM208233
<i>Gnathamitermes</i> sp. JT5 gut clone 5A1sec	FDH-H	HM208239
<i>Gnathamitermes</i> sp. JT5 gut clone 5A2sec	FDH-H	HM208246
<i>Gnathamitermes</i> sp. JT5 gut clone 5A7sec	FDH-H	HM208245
<i>Gnathamitermes</i> sp. JT5 gut clone 5B11sec	FDH-H	HM208242
<i>Gnathamitermes</i> sp. JT5 gut clone 5B1sec	FDH-H	HM208247
<i>Gnathamitermes</i> sp. JT5 gut clone 5C2sec	FDH-H	HM208240
<i>Gnathamitermes</i> sp. JT5 gut clone 5C5sec	FDH-H	HM208241
<i>Gnathamitermes</i> sp. JT5 gut clone 5D3sec	FDH-H	HM208243
<i>Gnathamitermes</i> sp. JT5 gut clone 5D5sec	FDH-H	HM208244
<i>Gnathamitermes</i> sp. JT5 gut clone 5F1cys	FDH-H	HM208238
<i>Microcerotermes</i> sp. Cost006 gut clone cs6_23sec	FDH-H	HM208200
<i>Microcerotermes</i> sp. Cost006 gut clone cs6_26sec	FDH-H	HM208202
<i>Microcerotermes</i> sp. Cost006 gut clone cs6_31cys	FDH-H	HM208201
<i>Microcerotermes</i> sp. Cost006 gut clone cs6_45cys	FDH-H	HM208205
<i>Microcerotermes</i> sp. Cost006 gut clone cs6_B1cys	FDH-H	HM208203
<i>Microcerotermes</i> sp. Cost006 gut clone cs6_F3cys	FDH-H	HM208204
<i>Microcerotermes</i> sp. Cost008 gut clone cs8Asec	FDH-H	HM208215
<i>Microcerotermes</i> sp. Cost008 gut clone cs8Bcys	FDH-H	HM208216
<i>Microcerotermes</i> sp. Cost008 gut clone cs8Csec	FDH-H	HM208214
<i>Microcerotermes</i> sp. Cost008 gut clone cs8Dcys	FDH-H	HM208217
<i>Nasutitermes</i> sp. Cost003 gut clone 3D6cys	FDH-H	HM208184
<i>Nasutitermes</i> sp. Cost003 gut clone cs3Bsec	FDH-H	HM208180
<i>Nasutitermes</i> sp. Cost003 gut clone cs3Csec	FDH-H	HM208179
<i>Nasutitermes</i> sp. Cost003 gut clone cs3Isec	FDH-H	HM208181
<i>Nasutitermes</i> sp. Cost003 gut clone cs3Psec	FDH-H	HM208182
<i>Nasutitermes</i> sp. Cost003 gut clone 3C4cys	FDH-H	HM208183

---

---

<i>Nasutitermes corniger</i> Cost007 gut clone 7B4sec	FDH-H	HM208212
<i>Nasutitermes corniger</i> Cost007 gut clone 7B7sec	FDH-H	HM208208
<i>Nasutitermes corniger</i> Cost007 gut clone 7D2sec	FDH-H	HM208209
<i>Nasutitermes corniger</i> Cost007 gut clone 7E2cys	FDH-H	HM208210
<i>Nasutitermes corniger</i> Cost007 gut clone 7G7sec	FDH-H	HM208213
<i>Nasutitermes corniger</i> Cost007 gut clone 7H1cys	FDH-H	HM208211
<i>Nasutitermes corniger</i> Cost007 gut clone cs7E6cys	FDH-H	HM208207
<i>Nasutitermes corniger</i> Cost007 gut clone cs7F6sec	FDH-H	HM208206
<i>Rhynchotermes</i> sp. Cost004 gut clone 4D7cys	FDH-H	HM208198
<i>Rhynchotermes</i> sp. Cost004 gut clone 4G3sec	FDH-H	HM208199
<i>Rhynchotermes</i> sp. Cost004 gut clone cs457sec	FDH-H	HM208194
<i>Rhynchotermes</i> sp. Cost004 gut clone cs489sec	FDH-H	HM208196
<i>Rhynchotermes</i> sp. Cost004 gut clone cs4Asec	FDH-H	HM208185
<i>Rhynchotermes</i> sp. Cost004 gut clone cs4Bsec	FDH-H	HM208189
<i>Rhynchotermes</i> sp. Cost004 gut clone cs4Dsec	FDH-H	HM208191
<i>Rhynchotermes</i> sp. Cost004 gut clone cs4Esec	FDH-H	HM208187
<i>Rhynchotermes</i> sp. Cost004 gut clone cs4Gsec	FDH-H	HM208190
<i>Rhynchotermes</i> sp. Cost004 gut clone cs4Isec	FDH-H	HM208186
<i>Rhynchotermes</i> sp. Cost004 gut clone cs4Lsec	FDH-H	HM208193
<i>Rhynchotermes</i> sp. Cost004 gut clone cs4Msec	FDH-H	HM208192
<i>Rhynchotermes</i> sp. Cost004 gut clone cs4Osec	FDH-H	HM208188
<i>Rhynchotermes</i> sp. Cost004 gut clone cs4Qsec	FDH-H	HM208197
<i>Rhynchotermes</i> sp. Cost004 gut clone cs4Xsec	FDH-H	HM208195

---

## References

1. **Bauer, S., A. Tholen, J. Overmann, and A. Brune.** 2000. Characterization of abundance and diversity of lactic acid bacteria in the hindgut of wood- and soil-feeding termites by molecular and culture-dependent techniques. *Arch Microbiol* **173**:126-37.
2. **Bignell, D. E.** 2000. Introduction to symbiosis, p. 189–208. *In* T. Abe, D. E. Bignell, and M. Higashi (ed.), *Termites: Evolution, Sociality, Symbioses, Ecology*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
3. **Bignell, D. E.** 2006. Termites as soil engineers and soil processors, p. 183-220. *In* H. König and A. Varma (ed.), *Intestinal microorganisms of termites and other invertebrates*. Springer, Heidelberg, Germany.
4. **Bignell, D. E., and J. M. Anderson.** 1980. Determination of pH and oxygen status in the guts of lower and higher termites. *J Insect Physiol* **26**:183-188.
5. **Bignell, D. E., and P. Eggleton.** 1995. On the elevated intestinal pH of higher termites (*Isoptera, Termitidae*). *Insectes Sociaux* **42**:57-69.
6. **Bignell, D. E., and P. Eggleton.** 2000. Termites in Ecosystems, p. 363-387. *In* T. Abe, D. E. Bignell, and M. Higashi (ed.), *Termites: Evolution, Sociality, Symbioses, Ecology*. Springer, Dordrecht, The Netherlands.
7. **Brauman, A., D. E. Bignell, and I. Tayasu.** 2000. Soil-feeding termites: biology, microbial associations and digestive mechanisms, p. 233-259. *In* T. Abe, D. E. Bignell, and M. Higashi (ed.), *Termites: Evolution, Sociality, Symbioses, Ecology*. Kluwer Academic Publishers,, Dordrecht, The Netherlands.
8. **Brauman, A., J. Dore, P. Eggleton, D. Bignell, J. A. Breznak, and M. D. Kane.** 2001. Molecular phylogenetic profiling of prokaryotic communities in guts of termites with different feeding habits. *FEMS Microbiol Ecol* **35**:27-36.
9. **Brauman, A., M. D. Kane, M. Labat, and J. A. Breznak.** 1992. Genesis of acetate and methane by gut bacteria of nutritionally diverse termites. *Science* **257**:1384-1387.
10. **Breznak, J. A.** 2000. Ecology of prokaryotic microbes in the guts of wood-and litter-feeding termites, p. 209-231. *In* T. Abe, D. E. Bignell, and M. Higashi (ed.), *Termites: Evolution, Sociality, Symbiosis, Ecology* Kluwer Academic Publishers Dordrecht, The Netherlands.
11. **Breznak, J. A., and J. M. Switzer.** 1986. Acetate synthesis from H<sub>2</sub> plus CO<sub>2</sub> by termite gut microbes. *Appl Environ Microbiol* **52**:623–630.
12. **Brugerolle, G., and R. Radek.** 2006. Symbiotic Protozoa of Termites, p. 244-269. *In* H. König and A. Varma (ed.), *Intestinal microorganisms of termites and other invertebrates*. Springer, Heidelberg, Germany.
13. **Brune, A., and M. Kühl.** 1996. pH profiles of the extremely alkaline hindguts of soil-feeding termites (*Isoptera: Termitidae*) determined with microelectrodes. *J Insect Physiol* **42**:1121–1127.
14. **Colwell, R. K.** 2009, posting date. EstimateS: statistical estimation of species richness and shared species from samples. version 8.2.0. [Online.]
15. **Drake, H. L., K. Küsel, and C. Matthies.** 2006. Acetogenic prokaryotes, p. 354-420. *In* M. Dworkin, S. Falkow, E. Rosenber, K. H. Schleifer, and E. Stackebrandt (ed.), *The Prokaryotes*, 3 ed. Springer.

16. **Eggleton, P.** 2006. The Termite Gut Habitat: Its Evolution and Co-Evolution, p. 373-404. *In* H. König and A. Varma (ed.), *Intestinal microorganisms of termites and other invertebrates*. Springer, Heidelberg, Germany.
17. **Eggleton, P., D. E. Bignell, W. A. Sands, N. A. Mawdsley, J. H. Lawton, T. G. Wood, and N. C. Bignell.** 1996. The diversity, abundance and biomass of termites under differing levels of disturbance in the Mbalmayo Forest Reserve, southern Cameroon. *Philosophical Transactions - Royal Society Biological sciences* **351**:51-68.
18. **Engel, M. S., D. A. Grimaldi, and K. Krishna.** 2009. Termites (*Isoptera*): Their Phylogeny, Classification, and Rise to Ecological Dominance. *American Museum Novitates* **3650**:1-27.
19. **Enquist, C. A. F.** 2002. Predicted regional impacts of climate change on the geographical distribution and diversity of tropical forests in Costa Rica. *Journal of Biogeography* **29**:519-534.
20. **Felsenstein, J.** 1989. PHYLIP - Phylogeny inference package (version 3.2). *Cladistics* **5**:164 - 166.
21. **Ferry, J. G.** 1990. Formate Dehydrogenase: Microbiology, Biochemistry and Genetics, p. 117-141. *In* G. A. Codd, L. Dijkhuizen, and F. R. Tabita (ed.), *Autotrophic Microbiology and One Carbon Metabolism*. Kluwer Academic Publishers, Dordrecht.
22. **Grimaldi, D., and M. S. Engel.** 2005. *Evolution of the insects*. Cambridge University Press, New York, NY.
23. **Hernes, P. J., and J. I. Hedges.** 2004. Tannin signatures of barks, needles, leaves, cones, and wood at the molecular level. *Geochim Cosmochim Acta* **68**:1293-1307.
24. **Holdridge, L. R., W. C. Grenke, W. H. Hatheway, and J. A. Tosi.** 1971. *Forest environments in tropical life zones. A pilot study*. Pergamon Press, New York.
25. **Inoue, T., O. Kitade, T. Yoshimura, and I. Yamaoka.** 2000. Symbiotic association with protists, p. 275-288. *In* T. Abe, D. E. Bignell, and M. Higashi (ed.), *Termites: Evolution, Sociality, Symbioses*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
26. **Kambhampati, S., and P. Eggleton.** 2000. Taxonomy and phylogeny of termites, p. 1-24. *In* T. Abe, D. E. Bignell, and M. Higashi (ed.), *Termites: Evolution, Sociality, Symbioses, Ecology*. Kluwer Academic Publishers, Dordrecht, The Netherlands.
27. **Leadbetter, J. R., T. M. Schmidt, J. R. Graber, and J. A. Breznak.** 1999. Acetogenesis from H<sub>2</sub> plus CO<sub>2</sub> by spirochetes from termite guts. *Science* **283**:686-689.
28. **Lozupone, C., and R. Knight.** 2005. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* **71**:8228-8235.
29. **Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Forster, I. Brettske, S. Gerber, A. W. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. König, T. Liss, R. Lussmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N. Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode, and K.-H. Schleifer.**

2004. ARB: a software environment for sequence data. *Nucl Acids Res* **32**:1363-1371.
30. **Lugo, A. E., S. L. Brown, R. Dodson, T. S. Smith, and H. H. Shugart.** 1999. The Holdridge life zones of the conterminous United States in relation to ecosystem mapping. *Journal of Biogeography* **26**:1025-1038.
  31. **Matson, E. G., E. Ottesen, and J. R. Leadbetter.** 2007. Extracting DNA from the gut microbes of the termite *Zootermopsis nevadensis*. *J Vis Exp* **4**:195.
  32. **Matson, E. G., X. Zhang, and J. R. Leadbetter.** 2010. Selenium controls expression of paralogous formate dehydrogenases in the termite gut acetogen *Treponema primitia*. *Environ Microbiol* *Accepted*.
  33. **Miura, T., K. Maekawa, O. Kitade, T. Abe, and T. Matsumoto.** 1998. Phylogenetic relationships among subfamilies in higher termites (*Isoptera: Termitidae*) based on mitochondrial COII gene sequences. *Annals of the Entomological Society of America* **91**:515-521.
  34. **Miura, T., Y. Roisin, and T. Matsumoto.** 2000. Molecular phylogeny and biogeography of the Nasute termite genus *Nasutitermes* (*Isoptera: Termitidae*) in the Pacific Tropics. *Molecular Phylogenetics and Evolution* **17**:1-10.
  35. **Miyata, R., N. Noda, H. Tamaki, K. Kinjyo, H. Aoyagi, H. Uchiyama, and H. Tanaka.** 2007. Influence of feed components on symbiotic bacterial community structure in the gut of the wood-feeding higher termite *Nasutitermes takasagoensis*. *Biosci Biotechnol Biochem* **71**:1244-1251.
  36. **Noda, S., M. Ohkuma, A. Yamada, Y. Hongoh, and T. Kudo.** 2003. Phylogenetic position and in situ identification of ectosymbiotic spirochetes on protists in the termite gut. *Appl Environ Microbiol* **69**:625-633.
  37. **Noirot, C.** 1992. From wood- to humus-feeding: an important trend in termite evolution. , p. 107-119. *In* J. Billen (ed.), *Biology and Evolution of Social Insects*. University Press, Leuven, Belgium.
  38. **Nriagu, J. O.** 1989. Global cycling of selenium, p. 327-340. *In* M. Ichnat (ed.), *Occurrence and Distribution of Selenium*. CRC Press, Boca Baton, Florida.
  39. **Odelson, D. A., and J. A. Breznak.** 1983. Volatile fatty acid production by the hindgut microbiota of xylophagous termites. *Appl Environ Microbiol* **45**:1602-1613.
  40. **Ohkuma, M., Y. Hongoh, and T. Kudo.** 2006. Diversity and Molecular Analyses of Yet-Uncultivated Microorganisms. *In* H. König and A. Varma (ed.), *Intestinal microorganisms of termites and other invertebrates*. Springer, Heidelberg, Germany.
  41. **Paster, B. J., F. E. Dewhirst, S. M. Cooke, V. Fussing, L. K. Poulsen, and J. A. Breznak.** 1996. Phylogeny of not-yet-cultured spirochetes from termite guts. *Appl Environ Microbiol* **62**:347-52.
  42. **Pester, M., and A. Brune.** 2006. Expression profiles of fhs (FTHFS) genes support the hypothesis that spirochaetes dominate reductive acetogenesis in the hindgut of lower termites. *Environ Microbiol* **8**:1261-1270.
  43. **Pester, M., and A. Brune.** 2007. Hydrogen is the central free intermediate during lignocellulose degradation by termite gut symbionts. *ISME J* **1**:551-65.

44. **Schloss, P. D., and J. Handelsman.** 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* **71**:1501-1506.
45. **Schmitt-Wagner, D., and A. Brune.** 1999. Hydrogen profiles and localization of methanogenic activities in the highly compartmentalized hindgut of soil-feeding higher termites (*Cubitermes* spp.). *Appl Environ Microbiol* **65**:4490-4496.
46. **Schmitt-Wagner, D., M. W. Friedrich, B. Wagner, and A. Brune.** 2003. Axial dynamics, stability, and interspecies similarity of bacterial community structure in the highly compartmentalized gut of soil-feeding termites (*Cubitermes* spp.). *Appl Environ Microbiol* **69**:6018–6024
47. **Schmitt-Wagner, D., M. W. Friedrich, B. Wagner, and A. Brune.** 2003. Phylogenetic diversity, abundance, and axial distribution of bacteria in the intestinal tract of two soil-feeding termites (*Cubitermes* spp.). *Appl Environ Microbiol* **69**:6007-17.
48. **Stamatakis, A. P., T. Ludwig, and H. Meier.** 2004. The AxML program family for maximum likelihood-based phylogenetic tree inference. *Concurrency and Computation: Practice and Experience* **16**:975-988.
49. **Strassert, J. F., M. S. Desai, A. Brune, and R. Radek.** 2009. The true diversity of Devescovinid flagellates in the termite *Incisitermes marginipennis*. *Protist*.
50. **Tholen, A., and A. Brune.** 1999. Localization and *in situ* activities of homoacetogenic bacteria in the highly compartmentalized hindgut of soil-feeding higher termites (*Cubitermes* spp.). *Appl Environ Microbiol* **65**:4497-505.
51. **Thongaram, T., Y. Hongoh, S. Kosono, M. Ohkuma, S. Trakulnaleamsai, N. Noparatnaraporn, and T. Kudo.** 2005. Comparison of bacterial communities in the alkaline gut segment among various species of higher termites. *Extremophiles* **9**:229-238.
52. **Thorne, B. L., D. Grimaldi, and K. Krishna.** 2000. Early fossil history of termites, p. 77-93. *In* T. Abe, D. E. Bignell, and M. Higashi (ed.), *Termites: Evolution, Sociality, Symbioses, Ecology*. Springer, Dordrecht, The Netherlands.
53. **Tokuda, G. I. Y., and H. Noda.** 2000. Localization of symbiotic clostridia in the mixed segment of the termite *Nasutitermes takasagoensis* (Shiraki). *Appl Environ Microbiol* **66**:2199–2207.
54. **Warnecke, F., P. Luginbühl, N. Ivanova, M. Ghassemian, T. Richardson, J. Stege, M. Cayouette, A. Mchardy, G. Djordjevic, N. Aboushadi, R. Sorek, S. Tringe, M. Podar, H. Martin, V. Kunin, D. Dalevi, J. Madejska, E. Kirton, D. Platt, E. Szeto, A. Salamov, K. Barry, N. Mikhailova, N. Kyrpides, E. Matson, E. Ottesen, X. Zhang, M. Hernández, C. Murillo, L. Acosta, I. Rigoutsos, G. Tamayo, B. Green, C. Chang, E. Rubin, E. Mathur, D. Robertson, P. Hugenholtz, and J. Leadbetter.** 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**:560-565.

# RNA-Seq and microfluidic digital PCR identification of transcriptionally active spirochetes in termite gut microbial communities

## Abstract

CO<sub>2</sub>-reductive acetogenesis in termite hindguts is a bacterial process with significant impact on the nutrition of wood-feeding termites. Acetogenic spirochetes have been identified as key mediators of acetogenesis. Here, we use high-throughput, short transcript sequencing (RNA-Seq) and microfluidic, multiplex digital PCR to identify uncultured termite gut spirochetes transcribing genes for hydrogenase-linked formate dehydrogenase (FDH<sub>H</sub>) enzymes, which are required for acetogenic metabolism in the spirochete, *Treponema primitia*. To assess FDH<sub>H</sub> gene (*fdhF*) transcription within the gut community of a wood-feeding termite, we sequenced ca. 28,000,000 short transcript reads of gut microbial community RNA using Illumina Solexa technology. RNA-Seq results indicate that *fdhF* transcription in the gut is dominated by two *fdhF* genotypes: ZnD2sec and Zn2cys. This finding was independently corroborated with cDNA inventory and qRT-PCR transcription measurements. We, therefore, propose that RNA-Seq mapping of microbial community transcripts is specific and quantitative. Following transcriptional assessments, we performed microfluidic, multiplex digital PCR on single termite gut bacterial cells to discover the identity of uncultured bacteria encoding *fdhF*

genotypes ZnD2sec and Zn2cys. We identified the specific 16S rRNA gene ribotype of the bacterium encoding Zn2cys *fdhF* and report that the bacterium is a spirochete. Phylogenetic analysis reveals that this uncultured spirochete, like *T. primitia*, possesses genes for acetogenic metabolism – formyl-tetra-hydrofolate synthetase and both selenocysteine and cysteine variants of formate dehydrogenase. Microfluidic results also imply a spirochetal origin for ZnD2sec *fdhF*, but further gene pair associations are required for verification. Taken together, the results (i) show novel transcriptomic and single cell approaches can be successfully combined to study active microbes in natural microbial communities, (ii) underscore the continued relevance of leveraging investigations of uncultured bacteria with the results from pure culture studies, and (iii) imply that termite gut acetogenesis is largely mediated by spirochetes which represent only a small portion of total acetogenic spirochete diversity.

## Introduction

The structure and function of natural microbial communities are primary targets of study for microbial ecologists. Molecular profiling using 16S rRNA gene inventory (14, 27, 40) and metagenomic (2, 12, 40) techniques have proved incredibly useful for elucidating community structure, particularly in environments that have yielded few culturable microbes. Similar methods have been utilized to outline a suite of potential functions encoded in community DNA (2, 40, 41). These efforts have led to surveys of actual community function at the level of transcription [e.g., (9, 16, 21, 31, 39, 46)]. Quantitative reverse transcriptase PCR (qRT-PCR), microarray, cDNA inventory, and mRNA-based terminal-restriction fragment length polymorphism (T-RFLP) techniques have been commonly employed to monitor community transcription (7, 28, 29, 38). However, all these environmental transcriptomic methods suffer drawbacks related to primer/probe binding specificity and/or PCR biases (1, 8, 20, 30, 43). Furthermore, the extent of primer/probe cross-binding can not be easily assessed, as data sets yield accurate information on either transcript abundance or transcript sequence.

The recent advent of high-throughput Illumina/Solexa Genome Analyzer sequencing technology (26) has enabled researchers to obtain massive amounts of short DNA sequences (37-75 base pairs) from their sample quickly, with no primer or cloning bias. Unlike previous transcriptomic methods, this technology and similar high-throughput sequencing methods (e.g., Roche 454 GS20 pyrosequencing) yield *both* transcript sequence (i.e., verification) and transcript abundance in a single data set. However, the difficulty of sequence fragment mapping makes data interpretation a major challenge.

Previous knowledge of gene sequence, which can serve as a “scaffold” for fragment mapping, is usually required. Recent studies have utilized Solexa technology to deep sequence transcriptomes (RNA-Seq) for eukaryotes (13, 23, 44) or defined cultures of prokaryotes (33, 45), but studies of natural microbial community transcription have so far only utilized 454 pyrosequencing technology (9, 42).

Here, we demonstrate that high-throughput sequencing of transcripts via Illumina-Solexa RNA-Seq can be leveraged by traditional DNA and cDNA library data (used as scaffolding for fragment assembly and interpretation) to rapidly assess environmental functional gene transcription in microbial communities. This approach differs from mRNA-T-RLFP, as the entire length of scaffold sequence is informative and, more importantly, the sequence fragment serves as both signal (abundance of particular fragment) and verification (sequences can be mapped to library scaffolds). While this approach is still scaffold-limited, we expect that a combination of RNA-Seq and inventory data can serve as a tool for microbial ecologists interested in assessing transcription in environments with high allelic diversity.

With some knowledge of community structure and function in hand, microbial ecologists then face the challenge of linking community members (structure) with the respective activities they carry out (function). This is straight-forward when pure culture isolates representing different functional groups are available, but in the majority of cases, researchers find themselves confronted with a diversity of 16S rRNA and functional gene sequences from uncultured organisms which can not be related to one another based on

phylogenetic inference. Ottesen *et al.* (26) have recently shown that microfluidic multiplex digital PCR assays on single cells can resolve such relationships in natural microbial communities.

In this study, we combine RNA-Seq and single cell techniques to investigate functionally important uncultured bacteria in the symbiotic microbial community of a wood-feeding termite. All phylogenetically “lower” wood-feeding termites harbor a species-rich hindgut community of symbiotic protozoa and bacteria that efficiently degrades lignocellulose into acetate, the major carbon and energy source of their insect host (3, 4, 6). CO<sub>2</sub>-reducing acetogens play an important role in this nutritional mutualism: these bacteria consume the majority of the lignocellulose fermentation byproducts H<sub>2</sub> and CO<sub>2</sub>, generating up to a third of gut acetate (5, 6). Inventory surveys of key acetogenesis genes (28, 35) and pure culture studies (17) imply that spirochetes of the bacterial phylum *Treponema* are responsible for acetogenesis in wood-feeding termites. Moreover, phylotype abundance for the functional gene encoding the hydrogenase-linked acetogenesis enzyme formate dehydrogenase (FDH<sub>H</sub>, *fdhF*) in the wood-feeding termite *Zootermopsis nevadensis* suggest the acetogenic spirochete population comprises as many as 7–15 different types of Treponemes (Chapter 2). This estimate is roughly consistent with phylotype abundance (3–11) observed for another key acetogenesis enzyme (formyl-tetrahydrofolate synthetase, FTHFS) in *Zootermopsis* and other wood-feeding termites (28, 35).

Sequence and phylogenetic analyses revealed *fdhF* phlotypes in *Z. nevadensis* could be classified into two clades: one comprised of sequences encoding selenocysteine (Sec) at the FDH<sub>H</sub> active site (*fdhF*<sub>Sec</sub>) and the other of sequences that encode cysteine (Cys) at the homologous position (*fdhF*<sub>Cys</sub>). Studies with the pure culture acetogenic spirochete *Treponema primitia* str. ZAS-2 indicated *fdhF*<sub>Sec</sub> and *fdhF*<sub>Cys</sub> could be present in the same organism and that *both* are transcriptionally controlled by the trace element selenium (22). It is unknown whether various *fdhF*<sub>Sec</sub> and *fdhF*<sub>Cys</sub> variants in *Z. nevadensis* belong to the same spirochete or are differentially transcribed. Here, we use novel sequencing and single-cell techniques (26) to (i) assess transcription of hydrogenase-linked FDH genes within the species-rich symbiotic gut microbial community of *Z. nevadensis* and (ii) determine the 16S rRNA sequence identity of uncultured termite gut bacteria encoding transcriptionally active FDH<sub>H</sub> genes.

## Materials and Methods

### Termite collection

Worker specimens of the dampwood termite *Zootermopsis nevadensis* were collected in the San Gabriel Mountains of California. Some were maintained in plastic boxes at 95% humidity in foil-covered glass aquaria in the laboratory. The entire gut tracts of ~5 worker termites were preserved in 50 – 200 µl of RNA stabilization buffer (RNA Protect Bacteria Reagent, QIAGEN, Valencia, CA) at -80°C until nucleic acid extraction for RNA-Seq and inventory experiments.

**Termite gut nucleic acid extraction**

100  $\mu$ l of TE buffer (1 mM Tris-HCl, 0.1 mM EDTA, pH 8.0) was added to an ice-thawed tube containing worker guts. Guts were then homogenized (3 x 30 sec) by bead beating with sterile zirconia/silica beads (0.1 mm) using a MiniBeadbeater-8 (BioSpec Products, Inc., Bartlesville, OK). Lysozyme (Sigma, St. Louis, MO) was added to the homogenate (1 mg); this mixture was incubated at room temperature for 15 min. DNA and total RNA were extracted from 150  $\mu$ l aliquots of gut homogenate using a DNeasy Tissue Kit (QIAGEN) and RNeasy Kit (QIAGEN), respectively. Purification details for total RNA can be found in Chapter 2. Total RNA was used for Illumina RNA-Seq and cDNA library experiments.

**RNA-Seq: Processing and sequencing**

Samples were prepared using the Illumina protocol for RNA-Seq sample preparation V2 (<https://icom.illumina.com>). Briefly, total RNA (at least 5  $\mu$ g) was fragmented using an Ambion RNA fragmentation kit and then converted to single-strand cDNA using an Invitrogen SuperScript II kit (Invitrogen, Carlsbad, CA). Second Strand Buffer (500 mM Tris-HCl, pH 7.8, 50 mM MgCl<sub>2</sub>, 10mM DTT), dNTP (0.3 mM), RNaseH (2 U  $\cdot$   $\mu$ l<sup>-1</sup>, Invitrogen) and DNA polymerase I (Invitrogen) were then added to the first-strand reaction to synthesize second strand cDNA (16°C, 2.5 hours). Fragmented second strand cDNA samples were sequenced as 37-mers using the standard Solexa (Illumina) protocol and pipeline at Caltech's Sequencing Core Facility (Pasadena, CA).

### **RNA-Seq Data Analysis**

Illumina raw data, obtained using GERALD (a software package within the Illumina pipeline), was aligned to a FASTA file containing FDH gene sequences (Table 4.4, Appendix) with the Maq short read aligning program (18). Samples were analyzed for perfect matches only. Signal intensities were visualized graphically by converting Maq aligned reads into a .BAR file using the Cisgenome software (15) and viewed on the Cisgenome browser and on the IGB genomic browser (<http://www.affymetrix.com>).

### **cDNA inventories**

Separate cDNA libraries for *fdhF<sub>Sec</sub>* and *fdhF<sub>Cys</sub>* gene variants were generated from gut cDNA. A forward primer for *fdhF<sub>Sec</sub>* (Sec427F, Table 4.1) that targets the selenocysteine FDH<sub>H</sub> active site was designed manually. Sec427F was used with 1045R (Chapter 3, Table 4.1) to amplify *fdhF<sub>Sec</sub>* from gut cDNA. The *fdhF<sub>Cys</sub>* cDNA library was constructed with primers Cys499F1b and 1045R (Chapter 3, Table 4.1). PCR reactions contained 200 nM forward primer (Sec427F or Cys499F1b), 200 nM 1045R, 1X FAILSAFE Premix D (EPICENTRE, Madison, WI), 0.07 U ·  $\mu\text{l}^{-1}$  of EXPAND High Fidelity polymerase (Roche Applied Science, Indianapolis, IN), and 0.5 ng ·  $\mu\text{l}^{-1}$  gut cDNA. Thermocycling conditions on a Mastercycler Model 5331 thermocycler (Eppendorf, Westbury, NY) were 2 min at 95°C, 30 cycles of (95°C for 30 sec, 60°C for 1 min, 72°C for 1 min), followed by 10 min at 72°C. Amplicon size was checked on 1.5% agarose gels (Invitrogen) and the products were TOPO-TA cloned (Invitrogen). Plasmids were extracted (QIAprep Spin Miniprep Kit, QIAGEN) from 48 randomly chosen clones and sequenced (Laragen Inc., Los Angeles, CA).

### Quantitative PCR

Quantitative RT-PCR for select FDH genotypes (ZnD2sec, ZnB5sec, *T. primitia fdhF<sub>Sec</sub>*) was performed on termite gut cDNA and DNA. Quantitative PCR primers for these genotypes were: ZnD2sec (ZnO-1636F, 5'– ACT ATG ACC GGC AAT TGT CGC CTG TT –3'; ZnO-1729R, 5'– TCA GAC CCA TAT CAC GGC AAA GTT –3'), ZnB5sec (ZnB5-1636F, 5'– ACG ATG ACG GGC AAC TGC CGG ATG TT –3'; ZnB5-1729R, 5'– TAT GCC GAG AGC ATT GGC ATC TT –3'), and *T. primitia fdhF<sub>Sec</sub>* (ZAS-1636F, 5'– ACC ATG ACC GGT AAC TGC CGG ACC CT –3'; ZAS-1729R, 5'– TTA TAC CGA GCT TTT CCG CAT CCC –3'). Primers were designed with Primer3 software (34) and amplify the same region in *fdhF* genes to avoid primer site biases. Standard curves (10-fold dilutions ranging from 10<sup>9</sup> – 10<sup>6</sup> copies/reaction) were generated from TOPO-TA plasmid templates containing the relevant inserts. QPCR reactions (20 µl) contained iQ SYBR Green Supermix (Bio-Rad laboratories, Irvine, CA), 500 nM forward primer, 500 nM reverse primer, 5 ng cDNA or 10 ng DNA. All reactions were run in duplicate. Thermocycling conditions on a Bio-Rad DNAEngine thermocycler (Chromo4 real time detector) were: 3 min at 95°C, followed by 44 cycles of 95°C for 15 sec, and 60°C for 30 sec.

### Microfluidic multiplex digital PCR

For each microfluidic chip experiment, the entire gut tract of one worker termite was extracted and suspended in 250 µl Synthetic Gut Fluid solution (25) containing 0.5 µg · mL<sup>-1</sup> Dnase-free RNase (Roche Applied Science). Cells were released from the gut tract by aspirating the sample 3 – 5 times with a sterile 200 µl pipet tip. Large particles were

allowed to sediment for ~5 sec. Cell dilutions ( $10^{-5}$  –  $2.5 \times 10^{-6}$  range) were added (1:20 v/v) to PCR reactions.

FDH<sub>H</sub> genes in spirochetes were surveyed using multiplex digital PCR. PCR reactions (20  $\mu$ l) contained iQ Multiple Powermix (Bio-Rad Laboratories, Discontinued Cat. No. 170-8848), 0.1% Tween-20, and 75 – 175 nM ROX standard. Final reaction concentrations of primer and probes (Table 4.1) were 100 – 400 nM. Specific concentrations for each chip experiment are described in Table 4.5 (Appendix). 16S rRNA primers and a general bacterial 16S rRNA probe (1389Prb) were designed by Ottesen *et al.* (26). A new 16S rRNA probe specific for spirochetes (1409RaPrb) was designed based on 1409Ra, a spirochete-specific primer (26). Functional gene primers (Cys499F1b, 1045R) for formate dehydrogenase genes have been described in Chapter 3. Sec427F and Cys538F primers were designed to target *fdhF*<sub>Sec</sub> and *fdhF*<sub>Cys</sub> gene variants, respectively. Sec427F targets all *fdhF*<sub>Sec</sub> genes, whereas Cys538F has a more limited target range for *fdhF*<sub>Cys</sub>. PCR reactions were loaded on microfluidic chips (Biomark 12.765 Digital Array series) purchased from Fluidigm Corporation (San Francisco, CA). Microfluidic chip thermocycling conditions were: 2 min at 95°C, 45 cycles of (95°C for 15 sec, 60°C for 1 min, 72°C for 1 min), followed by 10 min at 72°C.

Samples were retrieved based on amplification of spirochete DNA, accomplished using spirochete specific-primers and a general bacterial 16S rRNA probe, or general bacterial 16S rRNA primers and a spirochete-specific 16S rRNA probe (Table 4.2). Fluorescence above background for amplification-positive wells was typically detected  $\leq$  cycle 35.

Total bacterial concentration in panels sampled for retrieval was inferred from the total number of positive 16S rRNA gene amplifications observed in a separate panel loaded with template at the same dilution and general bacterial 16S rRNA primers/probes. Only panels that corresponded to template dilutions resulting in < 250 all bacteria hits (~1/3 of all chambers) were sampled for retrieval. Samples were manually retrieved into 10 µl TE from chip chambers using a dissecting microscope and 30 gauge needles (Becton, Dickinson, and Company, Franklin Lakes, NJ) as described by Ottensen *et al.* (26).

Chip samples were screened for 16S rRNA and *fdhF* gene products via simplex PCR with microfluidic chip primers on a Mastercycler Model 5331 thermocycler (Eppendorf, Westbury, NY) and agarose gel electrophoresis (1.5%, Invitrogen). PCR reactions (50 µl) contained iQ Multiple Powermix (Bio-Rad Laboratories), 200 – 300 nM of each primer, and 2.5 µl of template. Benchtop thermocycling conditions were 2 min at 95°C, 30 or 35 cycles of (95°C for 15 sec, 60°C for 1 min, 72°C for 1 min), followed by 10 min at 72°C. Products from samples that yielded both 16S rRNA and *fdhF* amplicons were PCR purified (QIAquick PCR purification, QIAGEN). 16S rRNA PCR products were cloned in TOPO-TA vectors (TOPO-TA cloning kit, Invitrogen) for low-yield PCR purifications; plasmids from 8 randomly chosen clones were purified (QIAprep Spin Miniprep, QIAGEN). 16S rRNA PCR products and plasmids were sequenced with the internal primers 533F and 1100R (26); *fdhF* products were sequenced with microfluidic chip primers. All sequencing reactions were performed at Laragen, Inc. (Los Angeles, CA).

**Sequence Analysis**

Sequences were assembled and edited using DNA-Star Lasagene software (Madison, WI). The software DOTUR was used to group sequences into operational taxonomic units (OTU) based on 8% Phylip DNA distance between OTUs, a cutoff which corresponds to the definition (3% amino acid distance) used to distinguish protein phylotypes in Chapters 2 and 3 (36). Phylogenetic trees were constructed using algorithms implemented within the ARB software environment (19). Tree construction details can be found in figure legends. The accession numbers of sequences used for phylogenetic analysis appear in Table 4.4 (Appendix).

**Table 4.1.** Primer and probes used in this study.

Primer	Sequence	Target <sup>1</sup>	Experiments <sup>2</sup>	Reference
357F	5' – CTC CTA CGG GAG GCA GCA G – 3'	Gen Bac 16S rRNA	chip 1-5	(26)
1409Ra	5' – GGG TAC CTC CAA CTC GGA TGG TG – 3'	Spirochete 16S rRNA	chip 1, 2	(26)
1492RL2D	5' – TAC GGY TAC CTT GTT ACG ACT T – 3'	Gen Bac 16S rRNA	chip 1-5	(26)
1389Prb	5' – HEX-CTT GTA CAC ACC GCC CGT C-3BHQ1 – 3'	Gen Bac 16S rRNA (probe)	chip 1-5	(26)
1409RaPrb	5' – HEX-CGG GTA CCT CCA ACT CGG ATG GTG-3BHQ1 – 3'	Spirochete 16S rRNA (probe)	chip 3-5	this study
Sec427F	5' – CGI ATA TGA CAC GCT CCT TCT GTA GC – 3'	<i>fdhF</i> <sub>Sec</sub>	chip 1-5, <i>fdhF</i> <sub>Sec</sub> lib.	this study
Cys538F	5' – TAY AAY GCG GCG GCI TCC CAC – 3'	<i>fdhF</i> <sub>Cys</sub>	chip 1, 2	this study
Cys499F1b	5' – ATG TCS CTK TCS ATI CCG GAA A – 3'	<i>fdhF</i> <sub>Cys</sub>	chip 3-5, Cys lib.	Chap. 3
1045R	5' – CIC CCA TRT CGC AGG YIC CCT G – 3'	<i>fdhF</i> <sub>Sec</sub> + <i>fdhF</i> <sub>Cys</sub>	chip 1-5, <i>fdhF</i> <sub>Sec</sub> , <i>fdhF</i> <sub>Cys</sub> lib.	Chap. 3

<sup>1</sup> Gen Bac, general bacterial.

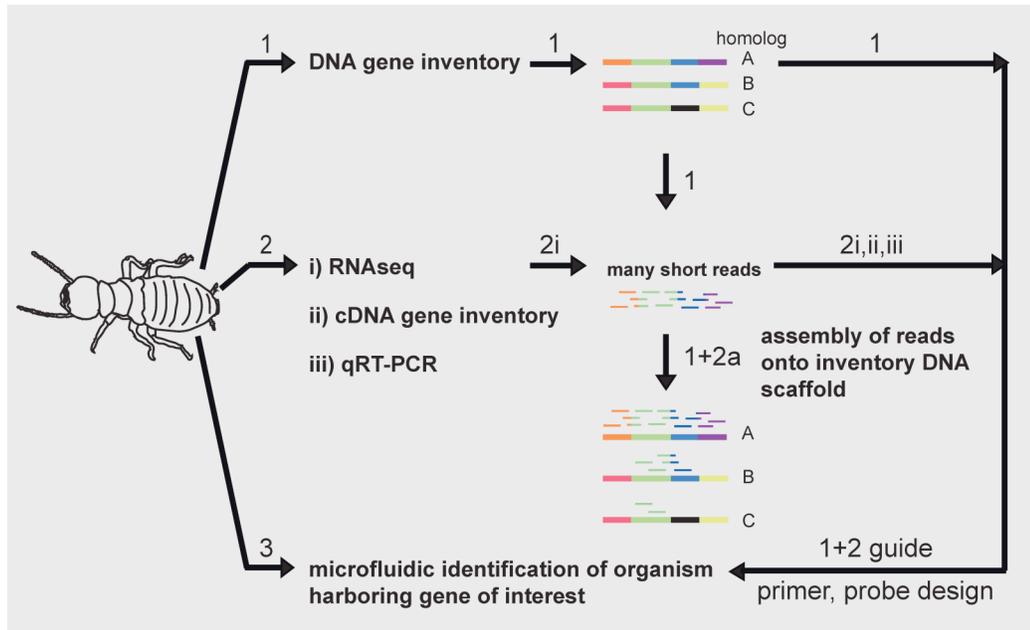
<sup>2</sup> chip, microfluidic chip experiment (Table 4.5, Appendix); lib. = cDNA library

## Results

In this study, we employ an approach that combines gene inventory, Illumina RNA-Seq, and microfluidic digital PCR techniques to assess transcription of a key acetogenesis gene (*fdhF*) in the gut community of a wood-feeding termite and identify bacteria encoding transcriptionally active *fdhF* genotypes. Figure 4.1 outlines the components of this approach. Briefly, we first mapped Illumina transcript reads of gut community RNA to gene inventory and pure culture sequence data to identify highly transcribed *fdhF* genotypes (arrows leading to 1+2a in Figure 4.1). We then corroborated the results using two independent methods. Finally, we performed microfluidics to discover the identity of organisms encoding transcribed *fdhF* genotypes (arrows leading to 3 in Figure 4.1).

### **RNA-Seq and other transcriptional assessments reveal two *fdhF* phylotypes dominate gut community *fdhF* transcription**

Total RNA was extracted from two collections of worker termites and sequenced by Illumina Solexa; one set was processed immediately after field collection, the other was maintained in the laboratory. RNA-Seq runs yielded 13,913,270 total 37-base pair reads (37-mers) for lab maintained termites and 14,043,698 reads for field-collected termites. Accounting for ribosomal RNA (~ 90% total) (24) and protozoa RNA [~ 90% of gut volume, (3)], we estimate bacterial functional gene transcripts only represent ~300,000 of total reads. We combined RNA-Seq reads from two Illumina runs into one large dataset (~28,000,000 reads) to increase bacterial functional gene read density.



**Figure 4.1.** Schematic of gene inventory, RNA-Seq, microfluidic PCR work-flow.  $FDH_H$  gene inventories and NCBI database sequences serve as scaffolds for RNA-Seq read mapping and data analysis. RNA-Seq based identifications of candidate genotypes belonging to transcriptionally important organisms can be corroborated using independent transcriptomic methods (cDNA gene inventory, qRT-PCR). Microfluidic, multiplex digital PCR on single cells can then be employed to obtain more genetic information on these important organisms.

To identify *fdhF* genotypes transcribed in the gut community, we analyzed the combined 28,000,000-read dataset using three nucleotide scaffold data sets, which contained: (i) 44 *fdhF* genotypes representing 23 phylotypes (8% DNA distance between OTUs) from *Z. nevadensis*; (ii) 92 *fdhF* genotypes representing 71 phylotypes from *Z. nevadensis*, two other phylogenetically “lower” termites, and a wood-feeding roach; (iii) 224 FDH genotypes representing 167 phylotypes for hydrogenase (*fdhF*), *Archaeal* F420, NAD(P)H, and respiratory chain-linked FDH enzymes from insects and the NCBI public database (Table 4.4, Appendix). Only reads that were perfectly matched to scaffold data set sequences were counted as hits. Reads were considered “unique” when they could only be mapped to genotypes within one FDH phylotype.

A total of 69 unique reads mapped onto the *Z. nevadensis* data set; these reads were distributed amongst 10 phylotypes (Table 4.2). Nearly half of all hits (30 reads, 43.4%) mapped to a single *Treponeme*-like *fdhF*<sub>sec</sub> phylotype (ZnD2sec). Almost all hits (27 out of 30) were distributed at unique positions along the entire length of the scaffold sequence, consistent with our inference that ZnD2sec is highly transcribed within the gut community. ZnD2sec also represented the majority of hits when laboratory maintained and field collected RNA-Seq reads were considered separately. The remaining 39 hits mapped to ZnHcys, Zn2cys, *fdhF*<sub>sec</sub> in *T. primitia* str. ZAS-2, and other *Treponeme*-like phylotypes (Table 4.2). All ZnHcys hits were derived from the field termite RNA-Seq dataset. Hits for other phylotypes were approximately evenly split between lab and field termites. The number of unique hits (69 reads) did not increase when reads were mapped to the 92-genotype data set, which contained sequences from four insect species.

Mapping onto the largest data set (224 genotypes representing 167 phylotypes for functionally diverse FDH enzymes) yielded only 1 more unique hit. This read mapped to an *Escherichia coli* hydrogenase-linked FDH<sub>H</sub> gene (Table 4.2). Although the total number of hits (70 reads) is small based on the abundance of ribosome and eukaryote transcripts, our results indicate RNA-Seq reads can be mapped to specific genotypes and phylotypes within an inventory containing several different homologs of a functional gene. We expect that increased sequencing, combined with effective rRNA depletion methods, will yield a more finely-resolved assessment of transcription.

To verify RNA-Seq results, we constructed separate cDNA libraries for *fdhF<sub>Sec</sub>* and *fdhF<sub>Cys</sub>* genes and performed SYBR-green qPCR assays using genotype specific primers (Table 4.3). In particular, we sought to determine whether ZnD2sec transcription was dominant relative to other *fdhF* sequences as RNA-Seq results indicated. Analysis of the *fdhF<sub>Sec</sub>* cDNA inventory from lab-maintained termite guts indicated the ZnD2sec phylotype accounts for 67% of all clone sequences. Comparison of ZnD2sec transcription with that of ZnB5sec and *T. primitia* Sec *fdhF<sub>Sec</sub>* using SYBR green qPCR assays yielded further confirmation of cDNA and RNA-Seq transcriptional patterns. Transcription of ZnD2cys was highest, followed by Zn70sec transcription; ZnB5sec transcription was not consistently detected. This order is consistent with the order of transcriptional abundance observed in RNA-Seq and cDNA library data. Both RNA-Seq and *fdhF<sub>Cys</sub>* cDNA libraries also identified the Zn2cys phylotype of *fdhF<sub>Cys</sub>* variants as relatively transcriptionally active. The absence of ZnHcys sequences from the cDNA dataset can be explained by samples differences, as all RNA-Seq reads mapping to this phylotype were

from field collected termite gut cDNA sample, which we did not analyze using cDNA inventory techniques. While we believe more RNA-Seq reads and qPCR assays are needed for an accurate picture of *fdhF* transcription in the gut, at least two independent methods indicate *ZnD2sec* and *Zn2cys* transcripts are relatively abundant in the *fdhF* transcript pool.

**Table 4.2.** *Z. nevadensis* gut community FDH gene transcription: RNA-Seq, *fdhF<sub>Sec</sub>* and *fdhF<sub>Cys</sub>* cDNA libraries, qRT-PCR.

Phylotype (FDH category) <sup>1</sup>	RNA-Seq Unique hits (% total) <sup>2</sup>	<i>fdhF<sub>Sec</sub></i> cDNA library (%) <sup>3</sup>	<i>fdhF<sub>Cys</sub></i> cDNA library (%) <sup>3</sup>	qPCR (copies/ng gut cDNA) <sup>4</sup>
Gut clone ZnD2sec ( <i>fdhF<sub>Sec</sub></i> )	30 (42.8)	67	—	131
Gut clone ZnHcys ( <i>fdhF<sub>Cys</sub></i> )	11 (15.7)	—	—	—
Gut clone Zn2cys ( <i>fdhF<sub>Cys</sub></i> )	10 (14.3)	—	54	—
<i>T. primitia</i> ( <i>fdhF<sub>Sec</sub></i> )	7 (10.0)	—	—	44
<i>T. primitia</i> ( <i>fdhF<sub>Cys</sub></i> )	3 (4.3)	—	2	—
Gut clone ZnB5sec ( <i>fdhF<sub>Sec</sub></i> )	3 (4.3)	2	—	NCD
Gut clone Zn61sec ( <i>fdhF<sub>Sec</sub></i> )	2 (2.9)	—	—	—
Gut clone ZnF7sec ( <i>fdhF<sub>Sec</sub></i> )	1 (1.4)	—	—	—
Gut clone ZnB8sec ( <i>fdhF<sub>Sec</sub></i> )	1 (1.4)	—	—	—
Gut clone Zn72secRT ( <i>fdhF<sub>Sec</sub></i> )	1 (1.4)	—	—	—
<i>Escherichia coli</i> ( <i>fdhF<sub>Sec</sub></i> )	1 (1.4)	—	—	—

<sup>1</sup> RNA-Seq reads were mapped to a dataset containing genes for hydrogenase-, NADPH-, F420-, respiratory chain-linked FDH enzymes ('FDH category'). All FDH reads mapped to hydrogenase-linked FDH genes (*fdhF*). Selenocysteine *fdhF* variants are denoted as *fdhF<sub>Sec</sub>*; cysteine variants are denoted as *fdhF<sub>Cys</sub>*. *Zn* gut clone phylotypes recovered from *Z. nevadensis* inventories are likely encoded by uncultured acetogenic spirochetes, as they phylogenetically group with *T. primitia* sequences (see Chapter 2).

<sup>2</sup> Reads were drawn from the combined 28 million read RNA-Seq dataset. Only reads that perfectly matched scaffold sequences within the same phylotype were considered "unique hits."

<sup>3</sup> Percentage of clones from *fdhF<sub>Sec</sub>* or *fdhF<sub>Cys</sub>* inventories constructed from laboratory maintained termite gut cDNA.

<sup>4</sup> Copies/ng lab maintained termite gut cDNA. NCD = not consistently detected.  $1\sigma < 2$  copies/ng.

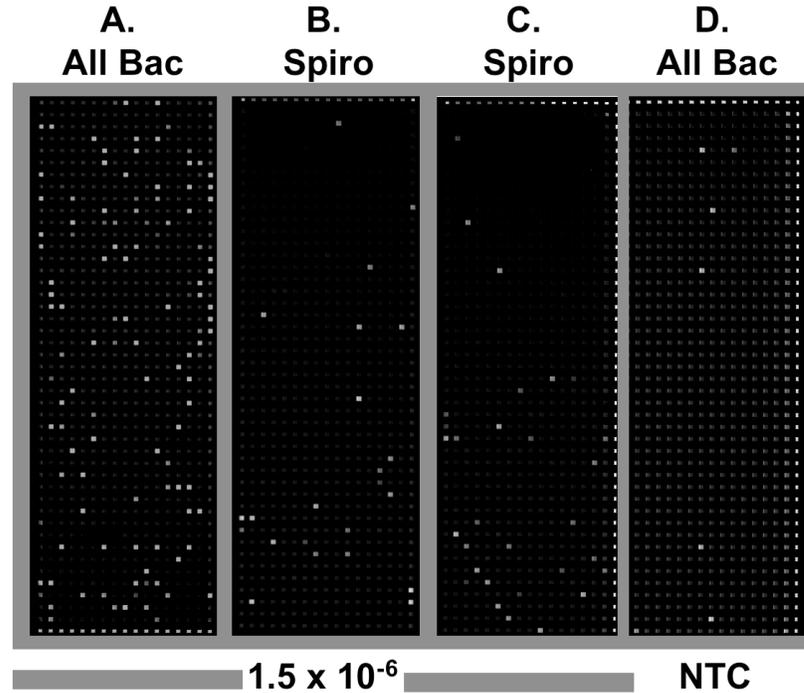
**Microfluidic digital PCR identification of two important *fdhF*-bearing spirochetes**

Previously in Chapter 2, we hypothesized that *fdhF<sub>Sec</sub>* and *fdhF<sub>Cys</sub>* phylotypes ZnD2sec and Zn2cys belong to spirochetes and that each of these spirochetes harbors both *fdhF<sub>Sec</sub>* and *fdhF<sub>Cys</sub>* gene variants. This was based on the phylogenetic clustering of ZnD2sec and Zn2cys with the dual *fdhF<sub>Sec</sub>* and *fdhF<sub>Cys</sub>* genes in *T. primitia* str. ZAS-1 and ZAS-2 [Chapter 2, (22)]. However, closely related treponemes like *T. azotonutricium* str. ZAS-9 are not acetogenic (10), nor encode FDH genes of any type (unpublished closed genome). Here, we performed microfluidic, multiplex digital PCR with 16S rRNA and *fdhF* primers on single termite gut bacterial cells to determine (i) whether uncultured spirochetes in *Z. nevadensis* guts encode *fdhF* and (ii) whether these spirochetes possess dual *fdhF<sub>Sec</sub>* and *fdhF<sub>Cys</sub>* as observed in *T. primitia*.

We utilized a broad approach, which differs from that of Ottesen *et al.* (26), as the design of functional gene probes with broad target ranges for *fdhF* was highly problematic. Microfluidic chip PCR reactions contained 16S rRNA primer and probes sets targeting spirochetes, and *fdhF* primers targeting Sec and/or Cys gene variants, but no functional gene probe. Most microfluidic PCR reactions were constructed to amplify 16S rRNA and *fdhF<sub>Sec</sub>* genes (i.e., duplex PCR); a few reactions targeted 16S rRNA, *fdhF<sub>Sec</sub>*, and *fdhF<sub>Cys</sub>* genes (i.e., triplex PCR). Samples were retrieved based on 16S rRNA probe fluorescence for spirochete 16S rRNA ribotypes rather than probe fluorescence for the functional gene; retrieved samples were then screened off-chip in simplex PCR reactions for the presence of 16S rRNA and *fdhF* gene products.

Spirochetes 16S rRNA genes were initially targeted with spirochete specific primers (357F, 1409Ra) and a general bacterial 16S rRNA probe (1389Prb) (microfluidic chip experiments 1, 2, Appendix 4, Table 4.5). However, the presence of non-spirochete 16S rRNA sequences within the same chamber as the target sequence could not be ruled out, since these sequences would not be amplified by 16S rRNA spirochete primers. Therefore, chip experiments 3 – 5 (Appendix 4, Table 4.5) were run with general bacterial 16S rRNA primers (357F, 1492RL2D), a spirochete specific 16S rRNA probe (1409RaPrb), and un-probed *fdhF* primers. Despite the increased sampling and screening steps associated this approach, we can, nevertheless, identify organisms encoding vastly different *fdhF* types (including those that are transcribed) as well as detect multiple 16S rRNA ribotypes in a sample to verify single cell amplification.

Microfluidic chip panels loaded with  $\sim 1 - 2 \times 10^{-6}$  dilutions of *Z. nevadensis* gut contents were sampled for retrieval. Panel A of Figure 4.2 shows end-point amplification from a typical gut dilution that yields  $< 150$  positive amplifications when general 16S rRNA bacteria gene primers are used. Assuming the distribution of cells on-chip follows a Poisson distribution, we estimate  $\sim 2.6\%$  of chambers contain more than a single cell. Replicate panels B and C (Figure 4.2) show the same gut dilution run with general bacterial 16S rRNA gene primers and a spirochete 16S rRNA gene probe. Well-separated amplification positive wells in spirochete specific panels were sampled for retrieval. Spirochetes accounted for  $12.5 \pm 6.5\%$  ( $1\sigma$ ) of all bacteria amplified on chip, consistent with previous observations in *Zootermopsis* (26). No template controls for PCR reactions targeting all bacteria (Panel D, Figure 4.2) typically yielded  $< 15$  positive amplifications.



**Figure 4.2.** Microfluidic digital PCR using an all-bacterial 16S rRNA gene primers-probe sets (panel A, D) or a spirochete specific 16S rRNA primer-probe set (panel B, C) with *fdhF* primers. Panels A, B, C contain *Z. nevadensis* gut contents diluted by  $1.5 \times 10^6$ . Panel D shows the no template control panel. Samples are retrieved from panels like B and C, which contain gut dilutions resulting in  $> 70\%$  chambers empty of any bacteria (Panel A).

On average,  $9.5 \pm 0.6\%$  of spirochetes retrieved from triplex (16S rRNA – *fdhF*<sub>Sec</sub> – *fdhF*<sub>Cys</sub>) microfluidic chip PCR reactions run with Sec427F and Cys538F primers (Table 4.5, Appendix, experiments 1 and 2) were positive for *fdhF* amplification upon simplex screening. Screens of duplex (16S rRNA – *fdhF*<sub>Sec</sub>) chip reactions indicated  $10.1 \pm 5.9\%$  of spirochetes had a gene for *fdhF*<sub>Sec</sub> (Table 4.5, Appendix, experiments 3a, 4, 5). Microfluidic chip triplex PCR with Sec427F and Cys519Fb primers yielded a much higher *fdhF* amplification rate from spirochete samples (40%) but this is likely due to the low sampling effort (only 5 wells were sampled in Experiment 3b).

We estimate 10% of all spirochetes and ~ 1% of all bacteria in *Z. nevadensis* carry *fdhF* genes. The latter result is consistent with Ottesen *et al.* (26), which estimates that 1% of all bacteria harbor the acetogenesis marker gene FTHFS, and with our previous findings that the majority of *fdhF* recovered from *Z. nevadensis* guts phylogenetically group with acetogenic spirochete sequences (Chapter 2). However, we note that *fdhF* genes were not recovered from 90% of spirochete on-chip retrievals. The effect of primer efficiency needs to be determined before a better estimate can be made of the true percentage of spirochetes harboring *fdhF*.

**Table 4.3.** Microfluidic chip retrieval of *fdhF* from spirochetes.

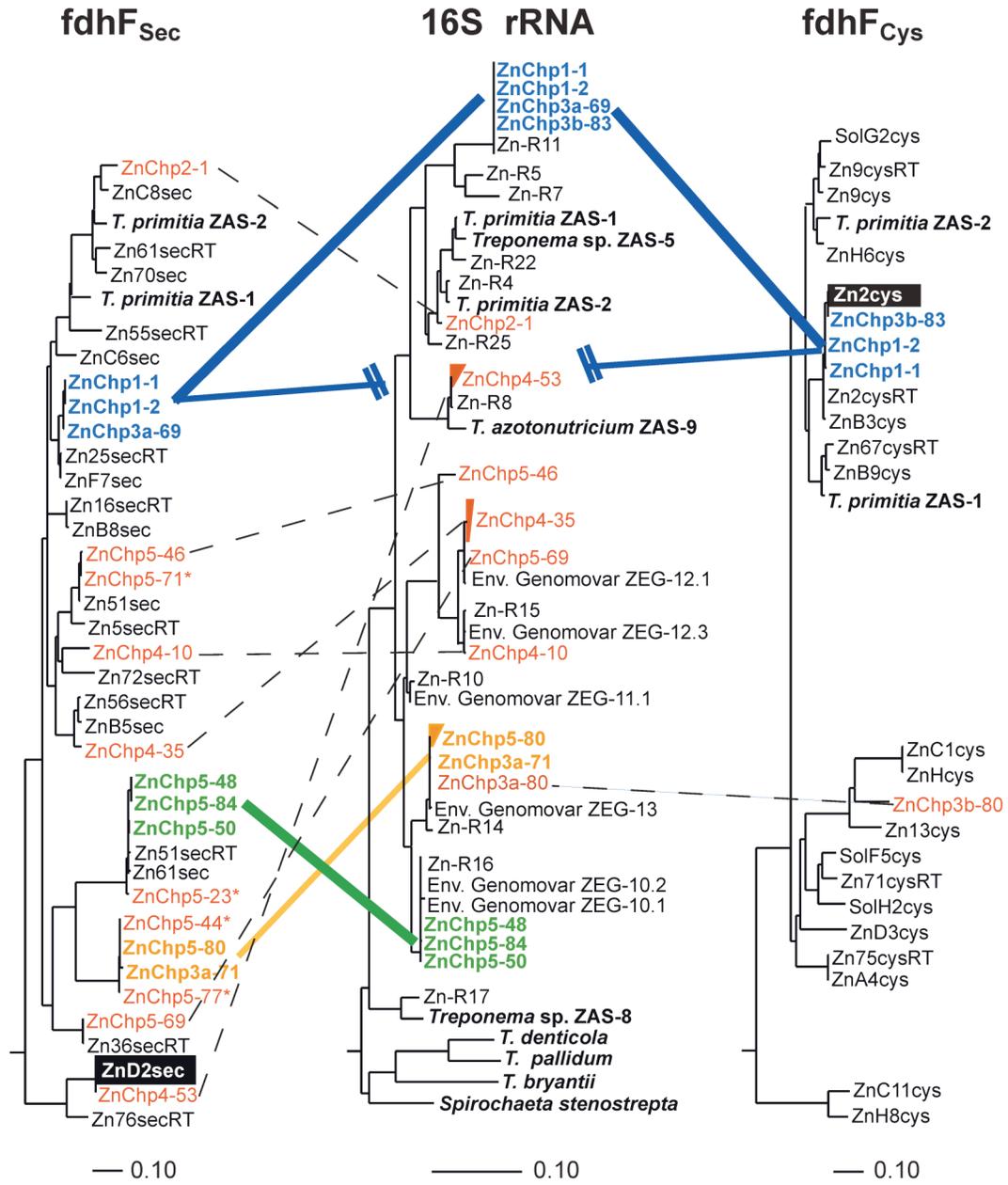
Chip Experiment	Targets	Spirochete with <i>fdhF</i> (%)
1	16S rRNA – <i>fdhF</i> <sub>Sec</sub> – <i>fdhF</i> <sub>Cys</sub>	9.1
2	16S rRNA – <i>fdhF</i> <sub>Sec</sub> – <i>fdhF</i> <sub>Cys</sub>	10.0
3a	16S rRNA – <i>fdhF</i> <sub>Sec</sub>	16.7
3b	16S rRNA – <i>fdhF</i> <sub>Sec</sub> – <i>fdhF</i> <sub>Cys</sub>	40.0
4	16S rRNA – <i>fdhF</i> <sub>Sec</sub>	5.3
5	16S rRNA – <i>fdhF</i> <sub>Sec</sub>	8.3

Figure 4.3 shows microfluidic chip sequence phylogeny. The results support the hypothesis that *fdhF* are encoded by uncultured spirochetes (Figure 4.3, gene pairs marked in orange, blue, green). The results also indicate at least one uncultured spirochete [Zn-R11 ribotype, (25)] possesses both *fdhF*<sub>Sec</sub> and *fdhF*<sub>Cys</sub>, like *T. primitia* (ZnF7sec and Zn2cys phylotypes). We believe the 16S rRNA – *fdhF* gene pairs in Figure 4.3 are colocalized to single bacterial cells as mixed templates were not apparent in 16S

rRNA sequence traces. Multiple identifications of the same gene pair serve as further evidence of single cell amplification.

We found that the most commonly retrieved gene pairs were associated with the Zn-R11 16S rRNA ribotype, suggesting it is an abundant member of the gut spirochete population. RNA-Seq and cDNA analyses, which indicate this organism is responsible for a significant proportion of the *fdhF* transcript pool, are consistent with our inference that the Zn-R11 ribotype represents an important acetogen in the gut community. Moreover, Ottesen *et al.* (25) have previously co-localized FTHFS and Clp protease (ClpX) genes to the Zn-R11 16S rRNA ribotype (FTHFS, Zn-F8; ClpX, Zn-X3). Our results extend the genetic inventory of this uncultured spirochete (*Zootermopsis* environmental genomovar, ZEG 16) to include two more functional genes associated with acetogenesis. More experiments are needed to confirm the 16S rRNA identity (ZnR8 ribotype) of the organism encoding ZnD2cys (the most highly transcribed *fdhF* within the gut community) as well as other single 16S – *fdhF* colocalizations (Figure 4.2, gene pairs connected by dashed lines).

**Figure 4.3.** 16S rRNA (middle panel) and *fdhF* phylogeny (left panel, *fdhF*<sub>Sec</sub>; right panel, *fdhF*<sub>Cys</sub>) of microfluidic chip sequences. Chip samples are labeled “ZnChp(Chip number)-sample” and highlighted in color (orange, red, green, blue). Pure culture sequences are highlighted in bold. Environmental genomovars are uncultured spirochetes from *Z. nevadensis* that encode the canonical acetogenesis marker gene FTHFS (26). *fdhF* sequences outlined by black boxes were highly transcribed in RNA-Seq and cDNA datasets. Lines connecting sequences highlight 16S rRNA - *fdhF* colocalizations (duplex gene pairs for all but ZnChp1-1, ZnChp1-2 samples, which contained 16S rRNA, *fdhF*<sub>Sec</sub> and *fdhF*<sub>Cys</sub> gene products). Line thickness corresponds to the number of repeated colocalizations and indicates our confidence in the observed associations. Dotted lines denote only one instance of colocalization. Blue line with hatch marks connects *fdhF*<sub>Sec</sub> and *fdhF*<sub>Cys</sub>. Grouped clades are composed of chip 16S rRNA sequences that were re-amplified and cloned into plasmids prior to sequencing; all other sequences were from PCR products. 16S rRNA tree was constructed using the neighbor joining algorithm implemented in ARB (19) based on 705 SINA (SILVA Incremental aligner) aligned nucleotides (32). ZnChp5-84, ZnChp2-1, and ZnChp4-10 sequences were added in by parsimony using 600 aligned nucleotides. A PhyML-maximum likelihood (11) *fdhF* tree was constructed using 1818 aligned nucleotides from *fdhF*<sub>Sec</sub> and *fdhF*<sub>Cys</sub> genes. Chip *fdhF* sequences were added in by parsimony using 380 aligned nucleotides using ARB. The tree was then split into *fdhF*<sub>Sec</sub> and *fdhF*<sub>Cys</sub> clades for ease of viewing. Scale bars denote 0.1 base pair changes per alignment position.



## Discussion

In this study, we used gene inventories as guides for transcriptome and single cell experiments on uncultured bacteria from termite hindgut communities. In particular, we interpreted microbial community RNA-Seq data based on the results of previous gene inventory and pure culture studies. We then corroborated RNA-Seq results with traditional transcription assays. Lastly, we employed transcript and gene inventory information as a guide for microfluidic experiments, which aimed at identifying uncultured organisms possessing target genes of interest.

The analysis of *fdhF* transcription within the symbiotic gut microbial community of a wood-feeding termite using new RNA-Seq, cDNA inventory, and qRT-PCR techniques revealed that two *fdhF* phylotypes (ZnD2sec and Zn2cys) account for a significant proportion of all *fdhF* gene transcripts. These results indicate that RNA-Seq reads can be mapped to specific *fdhF* genotypes/phylotypes to obtain a snapshot of transcription in a species-rich community. However, it is obvious that RNA-Seq read depth for bacterial functional genes needs to be increased. Read depth can be enhanced with mRNA enrichment techniques, which can increase mRNA sample content by 5-50% (37). Although less significant, total *fdhF* read density may also be improved by extending functional gene library coverage. We did not construct a scaffold library with DNA from termites used for RNA-Seq, thus any genotypes (and transcripts) unique to these samples would not be detected.

Following transcriptome analyses, we performed microfluidic multiplex digital PCR on single termite gut bacterial cells to learn more about the uncultured bacteria encoding highly transcribed *fdhF* genotypes. Microfluidic chip experiments indicated that these *Treponeme*-like *fdhF* genotypes are encoded by uncultured spirochetes (Zn2cys encoded by spirochete with the Zn-R11 ribotype; ZnD2sec encoded by spirochete with the Zn-R8 ribotype). These results not only provide additional support to the concept that spirochetes dominate acetogenesis in termite guts, they also suggest that the bulk of acetogenesis in the gut may be due to relatively few spirochete species.

We also identified an uncultured spirochete that possesses a repertoire of acetogenesis genes similar to *T. primitia*. The spirochete defined by the Zn-R11 ribotype encodes the highly transcribed Zn2Cys *fdhF* and possesses genes for FTHFS and a second *fdhF* allele (ZnF7sec) for selenocysteine FDH<sub>H</sub>. This finding underscores the relevance of *T. primitia* to understanding carbon and energy flows mediated by uncultured acetogenic bacteria in the gut community. Additionally, the genomic context provided by multiplex digital PCR enables environmental transcription studies of this specific uncultured organisms (Zn-R11), wherein the organism's ClpX gene (Zn-X3) can be used as a quantitative internal transcription standard rather than total RNA content, which does not indicate whether transcriptional changes are due to variations in organism abundance or transcriptional upregulation.

The results presented herein provide a framework for future studies of transcriptionally active spirochetes. We employed a degenerate primer approach to identify spirochetes

that highly transcribe *fdhF*. However, a targeted primer and probe strategy may prove more time efficient. After 16S rRNA identification, other techniques like fluorescent in-situ hybridisation (FISH) can be used to study the environment niches uncultured organisms occupy. Targeted approaches are especially appropriate for confirming the 16S rRNA ribotype (Zn-R8) associated with the ZnD2sec *fdhF* phylotype. Phylogenetic analysis (Chapter 3) indicates the ZnD2sec phylotype is basal to a lineage of spirochete-like *fdhF* genes that have likely persisted through a sweeping gene loss within gut communities during termite evolution to subsequently diversify in the guts of ecologically successful phylogenetically “higher” termites. The fact that ZnD2sec is highly transcribed underscores the importance of obtaining more genetic information on the uncultured bacterium bearing this apparently successful functional gene allele. We envision that 16S rRNA identification will enable future studies (e.g., FISH, single cell whole genome amplification) that enhance our understanding of this organism’s ecological role in termite gut communities.

## Acknowledgements

Much of this work would not have been possible without Adam Rosenthal. He performed the RNA-Seq and qRT-PCR portions of the study and should be considered an equal contributor to the work presented herein.

## Appendix

**Table 4.4.** RNA-Seq scaffold data set.

**Table 4.5.** Microfluidic chip experiment details.

**Table 4.4.** RNA-Seq scaffold data set. FDH nucleotide sequences are categorized into Sec and Cys enzyme variants and major FDH types (FDH-H, hydrogen-linked; FDH-NADH, NADH-linked; FDH-N and FDH-O, respiratory chain linked).

Sequence Source	Variant	FDH type	Nucleotide Accession <sup>1</sup>
<i>Cryptocercus puntulatus</i> gut clone Cp10sec	Sec	FDH-H	GU563433
<i>Cryptocercus puntulatus</i> gut clone Cp14sec	Sec	FDH-H	GU563436
<i>Cryptocercus puntulatus</i> gut clone Cp16sec	Sec	FDH-H	GU563432
<i>Cryptocercus puntulatus</i> gut clone Cp24sec	Sec	FDH-H	GU563451
<i>Cryptocercus puntulatus</i> gut clone Cp28sec	Sec	FDH-H	GU563450
<i>Cryptocercus puntulatus</i> gut clone Cp34sec	Sec	FDH-H	GU563452
<i>Cryptocercus puntulatus</i> gut clone Cp3sec	Sec	FDH-H	GU563434
<i>Cryptocercus puntulatus</i> gut clone Cp72cys	Cys	FDH-H	GU563437
<i>Cryptocercus puntulatus</i> gut clone Cp78sec	Sec	FDH-H	GU563453
<i>Cryptocercus puntulatus</i> gut clone Cp82sec	Sec	FDH-H	GU563454
<i>Cryptocercus puntulatus</i> gut clone Cp94sec	Sec	FDH-H	GU563455
<i>Cryptocercus puntulatus</i> gut clone Cp9cys	Cys	FDH-H	GU563441
<i>Cryptocercus puntulatus</i> gut clone CpB10sec	Sec	FDH-H	GU563442
<i>Cryptocercus puntulatus</i> gut clone CpB2sec	Sec	FDH-H	GU563446
<i>Cryptocercus puntulatus</i> gut clone CpB3sec	Sec	FDH-H	GU563440
<i>Cryptocercus puntulatus</i> gut clone CpC1cys	Cys	FDH-H	GU563444
<i>Cryptocercus puntulatus</i> gut clone CpC3sec	Sec	FDH-H	GU563443
<i>Cryptocercus puntulatus</i> gut clone CpD1cys	Cys	FDH-H	GU563445
<i>Cryptocercus puntulatus</i> gut clone CpD8sec	Sec	FDH-H	GU563439
<i>Cryptocercus puntulatus</i> gut clone CpE8cys	Cys	FDH-H	GU563447
<i>Cryptocercus puntulatus</i> gut clone CpF1cys	Cys	FDH-H	GU563435
<i>Cryptocercus puntulatus</i> gut clone CpF8cys	Cys	FDH-H	GU563449
<i>Cryptocercus puntulatus</i> gut clone CpF9cys	Cys	FDH-H	GU563448
<i>Cryptocercus puntulatus</i> gut clone CpH1cys	Cys	FDH-H	GU563438

---

<i>Incisitermes minor</i> gut clone Im10sec	Sec	FDH-H	GQ922349
<i>Incisitermes minor</i> gut clone Im11cys	Cys	FDH-H	GQ922364
<i>Incisitermes minor</i> gut clone Im15sec	Sec	FDH-H	GQ922351
<i>Incisitermes minor</i> gut clone Im22sec	Sec	FDH-H	GQ922353
<i>Incisitermes minor</i> gut clone Im24cys	Cys	FDH-H	GQ922369
<i>Incisitermes minor</i> gut clone Im26sec	Sec	FDH-H	GQ922354
<i>Incisitermes minor</i> gut clone Im27sec	Sec	FDH-H	GQ922355
<i>Incisitermes minor</i> gut clone Im3sec	Sec	FDH-H	GQ922356
<i>Incisitermes minor</i> gut clone Im42cys	Cys	FDH-H	GQ922371
<i>Incisitermes minor</i> gut clone Im5cys	Cys	FDH-H	GQ922373
<i>Incisitermes minor</i> gut clone Im63sec	Sec	FDH-H	GQ922361
<i>Reticulitermes hesperus</i> gut clone Rh15cys	Cys	FDH-H	GQ922398
<i>Reticulitermes hesperus</i> gut clone Rh24sec	Sec	FDH-H	GQ922383
<i>Reticulitermes hesperus</i> gut clone Rh2sec	Sec	FDH-H	GQ922381
<i>Reticulitermes hesperus</i> gut clone Rh35sec	Sec	FDH-H	GQ922385
<i>Reticulitermes hesperus</i> gut clone Rh36cys	Cys	FDH-H	GQ922410
<i>Reticulitermes hesperus</i> gut clone Rh41sec	Sec	FDH-H	GQ922386
<i>Reticulitermes hesperus</i> gut clone Rh47cys	Cys	FDH-H	GQ922402
<i>Reticulitermes hesperus</i> gut clone Rh53sec	Sec	FDH-H	GQ922389
<i>Reticulitermes hesperus</i> gut clone Rh54cys	Cys	FDH-H	GQ922404
<i>Reticulitermes hesperus</i> gut clone Rh65cys	Cys	FDH-H	GQ922406
<i>Reticulitermes hesperus</i> gut clone Rh71sec	Sec	FDH-H	GQ922391
<i>Reticulitermes hesperus</i> gut clone Rh93cys	Cys	FDH-H	GQ922409
<i>Reticulitermes hesperus</i> gut clone Rh9sec	Sec	FDH-H	GQ922397
<i>Treponema primitia</i> str. ZAS-1	Cys	FDH-H	GQ922450
<i>Treponema primitia</i> str. ZAS-1	Sec	FDH-H	GQ922449
<i>Treponema primitia</i> str. ZAS-2	Cys	FDH-H	FJ479767
<i>Treponema primitia</i> str. ZAS-2	Sec	FDH-H	FJ479767
<i>Zootermopsis nevadensis</i> gut clone Zn13cys	Cys	FDH-H	GQ922430

---

---

<i>Zootermopsis nevadensis</i> gut clone Zn16secRT	Sec	FDH-H	GU563476
<i>Zootermopsis nevadensis</i> gut clone Zn25secRT	Sec	FDH-H	GU563475
<i>Zootermopsis nevadensis</i> gut clone Zn2cys	Cys	FDH-H	GQ922431
<i>Zootermopsis nevadensis</i> gut clone Zn2cysRT	Cys	FDH-H	GU563472
<i>Zootermopsis nevadensis</i> gut clone Zn36secRT	Sec	FDH-H	GU563477
<i>Zootermopsis nevadensis</i> gut clone Zn51sec	Sec	FDH-H	GQ922423
<i>Zootermopsis nevadensis</i> gut clone Zn51secRT	Sec	FDH-H	GU563478
<i>Zootermopsis nevadensis</i> gut clone Zn55secRT	Sec	FDH-H	GU563479
<i>Zootermopsis nevadensis</i> gut clone Zn56secRT	Sec	FDH-H	GU563473
<i>Zootermopsis nevadensis</i> gut clone Zn5secRT	Sec	FDH-H	GU563471
<i>Zootermopsis nevadensis</i> gut clone Zn61sec	Sec	FDH-H	GQ922426
<i>Zootermopsis nevadensis</i> gut clone Zn61secRT	Sec	FDH-H	GU563480
<i>Zootermopsis nevadensis</i> gut clone Zn62sec	Sec	FDH-H	GQ922427
<i>Zootermopsis nevadensis</i> gut clone Zn67cysRT	Cys	FDH-H	GU563482
<i>Zootermopsis nevadensis</i> gut clone Zn70sec	Sec	FDH-H	GQ922428
<i>Zootermopsis nevadensis</i> gut clone Zn71cysRT	Cys	FDH-H	GU563483
<i>Zootermopsis nevadensis</i> gut clone Zn72secRT	Sec	FDH-H	GU563484
<i>Zootermopsis nevadensis</i> gut clone Zn75cysRT	Cys	FDH-H	GU563481
<i>Zootermopsis nevadensis</i> gut clone Zn76secRT	Sec	FDH-H	GU563485
<i>Zootermopsis nevadensis</i> gut clone Zn9cys	Cys	FDH-H	GQ922435
<i>Zootermopsis nevadensis</i> gut clone Zn9cysRT	Cys	FDH-H	GU563474
<i>Zootermopsis nevadensis</i> gut clone ZnA4cys	Cys	FDH-H	GU563456
<i>Zootermopsis nevadensis</i> gut clone ZnB3cys	Cys	FDH-H	GU563459
<i>Zootermopsis nevadensis</i> gut clone ZnB5sec	Sec	FDH-H	GU563460
<i>Zootermopsis nevadensis</i> gut clone ZnB8sec	Sec	FDH-H	GU563461
<i>Zootermopsis nevadensis</i> gut clone ZnB9cys	Cys	FDH-H	GU563462
<i>Zootermopsis nevadensis</i> gut clone ZnC11cys	Cys	FDH-H	GU563466
<i>Zootermopsis nevadensis</i> gut clone ZnC1cys	Cys	FDH-H	GU563463

---

---

<i>Zootermopsis nevadensis</i> gut clone ZnC6sec	Cys	FDH-H	GU563464
<i>Zootermopsis nevadensis</i> gut clone ZnC8sec	Sec	FDH-H	GU563465
<i>Zootermopsis nevadensis</i> gut clone ZnD2sec	Sec	FDH-H	GU563467
<i>Zootermopsis nevadensis</i> gut clone ZnD3cys	Cys	FDH-H	GU563468
<i>Zootermopsis nevadensis</i> gut clone ZnE2cys	Cys	FDH-H	GU563469
<i>Zootermopsis nevadensis</i> gut clone ZnF7sec	Sec	FDH-H	GU563458
<i>Zootermopsis nevadensis</i> gut clone ZnH6cys	Cys	FDH-H	GU563457
<i>Zootermopsis nevadensis</i> gut clone ZnH8cys	Cys	FDH-H	GU563470
<i>Zootermopsis nevadensis</i> gut clone ZnHcys	Cys	FDH-H	GQ922420
<i>Zootermopsis nevadensis</i> gut clone ZnJcys	Cys	FDH-H	GQ922417
<i>Zootermopsis nevadensis</i> gut clone ZnKcys	Cys	FDH-H	GQ922418
<i>Zootermopsis nevadensis</i> gut clone ZnLsec	Sec	FDH-H	GQ922412
<i>Zootermopsis nevadensis</i> gut clone ZnMsec	Sec	FDH-H	GQ922413
<i>Zootermopsis nevadensis</i> gut clone ZnOsec	Sec	FDH-H	GQ922415
<i>Zootermopsis nevadensis</i> gut clone ZnPcys	Cys	FDH-H	GQ922419
<i>Aeromonas salmonicida</i> subsp. <i>salmonicida</i> A449	Sec	FDH-H	NC_009348.1:1906100-1908244
<i>Aggregatibacter aphrophilus</i> NJ8700	Sec	FDH-H	NC_012913.1: c1159571-1157412
<i>Acetoneama longum</i> APO-1	Sec	FDH-H	GQ922445
<i>Buttiauxiella</i> SN1	Sec	FDH-H	GQ922446
<i>Carboxydotherrmus hydrogenoformans</i> Z-2901	Sec	FDH-NAD	NC_007503.1:646163-648844
<i>Carboxydotherrmus hydrogenoformans</i> Z-2901	Sec	FDH-O	NC_007503.1:702113-705121
<i>Citrobacter koseri</i> ATCC BAA-895	Cys	FDH-H	NC_009792.1:1727418-1729565
<i>Citrobacter koseri</i> ATCC BAA-895	Sec	FDH-H	NC_009792.1:3531364-3533511
<i>Citrobacter rodentium</i> ICC168 fdhFsec	Sec	FDH-H	NC_013716.1:c3662542-3660395
<i>Citrobacter rodentium</i> ICC168 fdhFsec	Sec	FDH-H	NC_013716.1:c3568359-3566212
<i>Citrobacter</i> str. TSA-1	Sec	FDH-H	GQ922447
<i>Citrobacter</i> sp. 30_2	Cys	FDH-H	NZ_GG657366.1:c1094197-1096347
<i>Citrobacter</i> sp. 30_2	Sec	FDH-H	NZ_GG657366.1:c93031-90884

---

---

<i>Citrobacter</i> sp. 30_2	Sec	FDH-N	NZ_GG657366.1:c1468196-1465035
<i>Citrobacter</i> sp. 30_2	Sec	FDH-O	NZ_GG657366.1:c37521-34471
<i>Citrobacter youngae</i> ATCC 29220	Sec	FDH-H	NZ_ABWL01000021.1::c93031-90884
<i>Citrobacter youngae</i> ATCC 29220	Cys	FDH-H	NZ_ABWL01000021.1:c24883-27030
<i>Citrobacter youngae</i> ATCC 29220	Sec	FDH-O	NZ_ABWL01000021.1:c43554-40504
<i>Clostridium bartlettii</i> DSM 16795	Cys	FDH-NAD	NZ_ABEZ02000007.1:22423-25119
<i>Clostridium bartlettii</i> DSM 16795	Sec	FDH-H	NZ_ABEZ02000007.1:c36324-34174
<i>Clostridium beijerinckii</i> NCIMB 8052	Cys	FDH-H	NC_009617.1:c4364248-4366389
<i>Clostridium bolteae</i> ATCC BAA-613	Cys	FDH-H	NZ_ABCC02000017.1:93731-95716
<i>Clostridium carboxidivorans</i> P7	Sec	FDH-H	NZ_ACVI01000105.1:231-2378
<i>Clostridium carboxidivorans</i> P7	Cys	FDH-H	NZ_ACVI01000010.1:36001-38157
<i>Clostridium difficile</i> 630	Sec	FDH-H	NC_009089.1:c3884230-3882086
<i>Cronobacter turicensis</i>	Cys	FDH-H	NC_013282.1:1996635-1998845
<i>Cronobacter turicensis</i>	Sec	FDH-H	NC_013282.1:2002311-2004458
<i>Cronobacter turicensis</i>	Cys	FDH-NAD	NC_013282.1:c1009687-1006715
<i>Desulfotobacterium hafniense</i> DCB-2	Sec	FDH-NAD	NC_011830.1:1504497-1507178
<i>Dickeya dadantii</i> Ech586	Cys	FDH-N	NC_013592.1:c3063358-3066408
<i>Dickeya dadantii</i> Ech586	Cys	FDH-H	NC_013592.1:2958853-2961003
<i>Dickeya dadantii</i> Ech703	Cys	FDH-H	NC_012880.1:c1450903-1453053
<i>Dickeya dadantii</i> Ech703	Cys	FDH-N	NC_012880.1:c2955857-2958907
<i>Dickeya dadantii</i> Ech703	Cys	FDH-O	NC_012880.1:c1523376-1526423
<i>Dickeya zaeae</i> Ech1591	Cys	FDH-H	NC_012912.1:3084906-3087056
<i>Desulfatibacillum alkenivorans</i> AK-01	Sec	FDH-NAD	NC_011768.1:5447766-5450528
<i>Desulfobacterium autotrophicum</i> HRM2	Cys	FDH-NAD	NC_012108.1:1930486-1933251
<i>Desulfotomaculum acetoxidans</i> 5575	Sec	FDH-NAD	NC_013216.1:c3713225-3715906
<i>Escherichia coli</i> O157:H7 str. FRIK2000	Sec	FDH-H	NZ_ACXO01000060.1:c38313-36585
<i>Escherichia coli</i> O157:H7 str. FRIK966	Sec	FDH-H	NZ_ACXN01000050.1:79269-81416
<i>Escherichia coli</i> 83972	Sec	FDH-H	NZ_ACGN01000114.1:89871-92018
<i>Escherichia coli</i> APEC O1	Sec	FDH-H	NC_008563.1:c4646031-4643884

---

---

<i>Escherichia coli</i> O157:H7 str. EC4024	Sec	FDH-H	NZ_ABJT01000004.1:c104404-106551
<i>Escherichia coli</i> O157:H7 str. TW14588	Sec	FDH-H	NZ_ABKY02000001.1:1646350-1648497
<i>Escherichia</i> sp. 4_1_40B	Sec	FDH-H	NZ_ACDM01000067.1:c85542-83814
<i>Escherichia coli</i> BL21(DE3)	Sec	FDH-H	NC_012947.1:4135920-4138067
<i>Escherichia coli</i> SE11	Sec	FDH-H	NC_011415.1:c4568500-4570647
<i>Escherichia coli</i> UMN026	Sec	FDH-H	NC_011751.1:c4792216-4790069
<i>Edwardsiella ictaluri</i> 93-146	Sec	FDH-H	NC_012779.1:3156478-3158622
<i>Edwardsiella tarda</i> EIB202	Sec	FDH-H	NC_013508.1:3053142-3055286
<i>Eggerthella lenta</i> VPI 0255	Cys	FDH-H	NC_013204.1:c3320160..3322586
<i>Cronobacter (Enterobacter) sakazakii</i> ATCC BAA-894	Cys	FDH-NAD	NC_009778.1:2900970-2903942
<i>Cronobacter (Enterobacter) sakazakii</i> ATCC BAA-894	Cys	FDH-H	NC_009778.1:c1996280-1998430
<i>Enterobacter</i> sp. 638	Sec	FDH-H	NC_009436.1:c 329787-331934
<i>Enterobacter</i> sp. 638	Cys	FDH-H	NC_009436.1:c1907448-1909598
<i>Enterobacter cancerogenus</i> ATCC 35316	Sec	FDH-H	NZ_ABWM02000022.1:21042-23189
<i>Enterococcus faecalis</i> V583	Cys	FDH-NAD	NC_004668.1:1367291-1370011
<i>Escherichia fergusonii</i> ATCC 35469	Sec	FDH-H	NC_011740.1:4397249-4399396
<i>Escherichia fergusonii</i> ATCC 35469	Sec	FDH-N	NC_011740.1:1525306..1528353
<i>Escherichia fergusonii</i> ATCC 35469	Sec	FDH-O	NC_011740.1:3984322..3987372
<i>Escherichia coli</i> str. K-12 substr. MG1655	Sec	FDH-N	NC_000913.2:1545425..1548472
<i>Escherichia coli</i> str. K-12 substr. MG1655	Sec	FDH-O	NC_000913.2:c4080795..4083845
<i>Escherichia coli</i> str. K-12 substr. MG1655	Sec	FDH-H	NC_000913.2:c4295242..4297389
<i>Eubacterium acidaminophilum</i>	Sec	FDH-NAD	AJ312124.1:11347..14028
<i>Eubacterium acidaminophilum</i>	Sec	FDH-NAD	AJ312125.1:2250..4943
<i>Heliobacterium modesticaldum</i> Icel: NC_010337	Cys	FDH-NAD	NC_010337.2:1747735..1750623
<i>Klebsiella pneumoniae</i> subsp. <i>pneumoniae</i> MGH 78578	Cys	FDH-H	NC_009648.1:2290424..2292574
<i>Klebsiella pneumoniae</i> subsp. <i>pneumoniae</i> MGH 78578	Sec	FDH-H	NC_009648.1:c4907710-4905563
<i>Klebsiella pneumoniae</i> 342	Cys	FDH-H	NC_011283.1:c2310716-2308566
<i>Klebsiella pneumoniae</i> 342	Sec	FDH-H	NC_011283.1:5239144-5241291
<i>Klebsiella pneumoniae</i> NTUH-K2044	Sec	FDH-O	NC_012731.1:c46019..49069

---

<i>Klebsiella pneumoniae</i> NTUH-K2044	Sec	FDH-H	NC_012731.1:c358869-356722
<i>Klebsiella pneumoniae</i> NTUH-K2044	Sec	FDH-N	NC_012731.1:c2794353..2797400
<i>Klebsiella pneumoniae</i> NTUH-K2044	Cys	FDH-H	NC_012731.1:3017444..3019594
<i>Klebsiella pneumoniae</i> 342	Sec	FDH-N	NC_011283.1:2546701..2549748
<i>Klebsiella pneumoniae</i> 342	Sec	FDH-O	NC_011283.1:5557641..5560691
<i>Mannheimia succiniciproducens</i> MBEL55E	Cys	FDH-NAD	NC_006300.1:684085..686892
<i>Moorella thermoacetica</i> ATCC 39073	Sec	FDH-NAD	NC_007644.1:c2432486..2435188
<i>Moorella thermoacetica</i> ATCC 39073	Sec	FDH-H	NC_007644.1:c2292497..2294737
<i>Methanococcus maripaludis</i> S2	Sec	FDH-F420	BX950229.1:145038..147068
<i>Methanococcus vannielii</i> SB	Sec	FDH-F420	CP000742.1:c663600..665624
<i>Natranaerobius thermophilus</i> JW/NM-WN-LF	Sec	FDH-NAD	NC_010718.1:115206..117887
<i>Oxalobacter formigenes</i> HOxBLS	Cys	FDH-H	NZ_GG658151.1:2458842..2460998
<i>Pantoea</i> sp. At-9b	Cys	FDH-H	NZ_ACYJ01000001.1:122540..124690
<i>Pantoea</i> sp. At-9b	Sec	FDH-O	NZ_ACYJ01000014.1:c128676..131723
<i>Pectobacterium carotovorum</i> subsp. carotovorum WPP14	Cys	FDH-H	NZ_ABVY01000027.1:c9266..11416
<i>Pectobacterium carotovorum</i> subsp. brasiliensis PBR1692	Cys	FDH-H	NZ_ABVX01000086.1:c2739..4889
<i>Pectobacterium atrosepticum</i> SCRI1043	Cys	FDH-H	NC_004547.2:1420602..1422752
<i>Pectobacterium atrosepticum</i> SCRI1043	Cys	FDH-H	NC_004547.2:c1752061..1754157
<i>Pectobacterium atrosepticum</i> SCRI1043	Cys	FDH-H	BX950851.1:1752061..175415
<i>Pectobacterium wasabiae</i> WPP163	Cys	FDH-H	NC_013421.1:c1930748..1932898
<i>Photobacterium profundum</i> 3TCK	Sec	FDH-H	NZ_AAPH01000003.1:97396-99486
<i>Pelobacter propionicus</i> DSM 2379	Cys	FDH-H	NZ_AAJH01000001.1:11892..14606
<i>Proteus mirabilis</i> ATCC 29906	Cys	FDH-H	NZ_ACLE01000010.1:50054..52222
<i>Proteus mirabilis</i> ATCC 29906	Sec	FDH-H	NZ_ACLE01000010.1:30536-32701
<i>Providencia alcalifaciens</i> DSM 30120	Sec	FDH-H	NZ_ABXW01000042.1:35044-37197
<i>Providencia alcalifaciens</i> DSM 30120	Sec	FDH-NAD	NZ_ABXW01000042.1:c37197-35044
<i>Providencia alcalifaciens</i> DSM 30120	Sec	FDH-O	NZ_ABXW01000042.1:c129523-126476
<i>Providencia alcalifaciens</i> DSM 30120	Sec	FDH-N	NZ_ABXW01000042.1:235693-238740

---

<i>Proteus mirabilis</i> HI4320	Cys	FDH-H	NC_010554.1:c3265604..3267772
<i>Proteus mirabilis</i> HI4320	Sec	FDH-H	NC_010554.1:3909884-3912028
<i>Providencia rettgeri</i> DSM 1131	Sec	FDH-N	NZ_ACCI02000030:c33183-30136
<i>Providencia rustigianii</i> DSM 4541	Sec	FDH-H	NZ_ABXV02000023.1:88004-90157
<i>Providencia rustigianii</i> DSM 4541	Sec	FDH-N	NZ_ABXV02000023.1:70811-73858
<i>Psychromonas</i> sp. CNPT3 fdhFsec	Sec	FDH-H	NZ_AAPG01000013.1:c5742-3595
<i>Ruminococcus</i> sp. 5_1_39B_FAA	Cys	FDH-NAD	NZ_GG696049.1:c238140..240848
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Typhi str. CT18	Sec	FDH-H	NC_003198.1:4370484..4372631
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Typhimurium str. LT2	Sec	FDH-H	AE006468.1:c4525350..4527497
<i>Salmonella enterica</i> subsp. <i>enterica</i> serovar Typhi str. CT18	Sec	FDH-O	NC_003198.1:3697528..3700578
<i>Salmonella typhimurium</i> LT2	Sec	FDH-H	NC_003197.1:c4525350..4527497
<i>Salmonella typhimurium</i> LT2	Sec	FDH-N	NC_003197.1:c1650442..1653489
<i>Salmonella typhimurium</i> LT2	Sec	FDH-O	NC_003197.1:c4244758..4247808
<i>Serratia proteamaculans</i> 568	Cys	FDH-H	NC_009832.1:c2657681..2659837
<i>Serratia proteamaculans</i> 568	Sec	FDH-N	NC_009832.1:87013..90060
<i>Serratia grimesii</i> ZFX-1	Cys	FDH-H	GQ922448
<i>Shigella flexneri</i> 2a str. 301	Sec	FDH-O	NC_004337.1:c4098182..4101232
<i>Shigella</i> sp. D9	Sec	FDH-H	NZ_ACDL01000041:c39372-37225
<i>Shigella sonnei</i> Ss046	Sec	FDH-O	NC_007384.1:c4296262..4299312
<i>Shigella sonnei</i> Ss046	Sec	FDH-N	NC_007384.1:c1741118..1744165
<i>Vibrio angustum</i> S14	Sec	FDH-H	NZ_AAOJ01000001.1:c1074316..1076460
<i>Yersinia aldovae</i> ATCC 35236	Cys	FDH-H	NZ_ACCB01000002.1:136225..138372
<i>Yersinia aldovae</i> ATCC 35236	Sec	FDH-O	NZ_ACCB01000003.1:36348..39395
<i>Yersinia bercovieri</i> ATCC 43970	Cys	FDH-H	NZ_AALC02000017.1:13658..15805
<i>Yersinia bercovieri</i> ATCC 43970	Sec	FDH-O	NZ_AALC02000005.1:103163..106210
<i>Yersinia enterocolitica</i> subsp. <i>enterocolitica</i> 8081	Cys	FDH-H	NC_008800.1:3050211..3052358
<i>Yersinia enterocolitica</i> subsp. <i>enterocolitica</i> 8081	Sec	FDH-O	NC_008800.1:c4525888..4528935
<i>Yersinia frederiksenii</i> ATCC 33641	Cys	FDH-H	NZ_AALE02000011.1:c133500..135647
<i>Yersinia frederiksenii</i> ATCC 33641	Cys	FDH-H	NZ_AALE02000004.1:63404..6554

---

<i>Yersinia frederiksenii</i> ATCC 33641	Sec	FDH-O	NZ_AAAL02000005.1:c136955-133908
<i>Yersinia intermedia</i> ATCC 29909	Cys	FDH-H	NZ_AAALF02000015.1:c38542..40698
<i>Yersinia intermedia</i> ATCC 29909	Sec	FDH-N	NZ_AAALF02000012.1:109282-112284
<i>Yersinia kristensenii</i> ATCC 33638	Cys	FDH-H	NZ_ACCA01000001.1:c40178..42325
<i>Yersinia kristensenii</i> ATCC 33638	Cys	FDH-H	NZ_ACCA01000002.1:c40178..42325
<i>Yersinia kristensenii</i> ATCC 33638	Sec	FDH-O	NZ_ACCA01000015.1:8904-11951
<i>Yersinia mollaretii</i> ATCC 43969	Cys	FDH-H	NZ_AALD02000036.1:52..2196
<i>Yersinia mollaretii</i> ATCC 43969	Cys	FDH-H	NZ_AALD02000005.1:c25400..27571
<i>Yersinia mollaretii</i> ATCC 43969	Sec	FDH-O	NZ_AALD02000033.1:c13893..16940
<i>Yersinia pestis</i> KIM	Cys	FDH-H	NC_004088.1:678737..680884
<i>Yersinia pseudotuberculosis</i> IP 32953	Cys	FDH-H	NC_006155.1:474164..476311
<i>Yersinia pseudotuberculosis</i> IP 32953	Cys	FDH-H	NC_009708.1:c4151279..4153426
<i>Yersinia rohdei</i> ATCC 43380	Cys	FDH-H	NZ_ACCD01000002.1:c116227..118374
<i>Yersinia rohdei</i> ATCC 43380	Sec	FDH-N	NZ_ACCD01000004.1:c74607-71605
<i>Yersinia ruckeri</i> ATCC 29473	Cys	FDH-H	NZ_ACCC01000020.1:c42838..4496
<i>Yersinia ruckeri</i> ATCC 29473	Sec	FDH-N	NZ_ACCC01000005.1:93044-96546

<sup>1</sup> 'c' in front of genome coordinates indicates complement sequence

**Table 4.5.** Microfluidic chip experiment details.

Chip Exp.	Fluidigm Chip No.	Primers and Probes ( $\mu$ M)	Targets	Spirochete Specificity via
1	1151-005-038	357F (200 nM), 1409Ra (200 nM), 1389Prb (300 nM), Sec427F (200 nM), Cys538F (200 nM), 1045R (200 nM)	Spirochete 16S rRNA, <i>fdhF</i> <sub>Sec</sub> , <i>fdhF</i> <sub>Cys</sub>	Spirochete specific primers, All Bacteria Probe
2	1151-026-033	357F (200 nM), 1409Ra (200 nM), 1389Prb (300 nM), Sec427F (200 nM), Cys538F (200 nM), 1045R (200 nM)	Spirochete 16S rRNA, <i>fdhF</i> <sub>Sec</sub> , <i>fdhF</i> <sub>Cys</sub>	Spirochete specific primers, All Bacteria Probe
3a	1151-067-035	357F (200 nM), 1492RL2D (200 nM), 1409RaPrb (300 nM), Sec427F (200 nM), 1045R (175 nM)	Spirochete 16S rRNA, <i>fdhF</i> <sub>Sec</sub> ,	All Bacteria primers, Spirochete specific Probe
3b	1151-067-035	357F (200 nM), 1492RL2D (200 nM), 1409RaPrb (300 nM), Sec427F (200 nM), Cys499F1b (125 nM), 1045R (175 nM)	Spirochete 16S rRNA, <i>fdhF</i> <sub>Sec</sub> , <i>fdhF</i> <sub>Cys</sub>	All Bacteria primers, Spirochete specific Probe
4	1151-067-038	357F (200 nM), 1492RL2D (200 nM), 1409RaPrb (300 nM), Sec427F (200 nM), 1045R (200 nM)	Spirochete 16S rRNA, <i>fdhF</i> <sub>Sec</sub> ,	All Bacteria primers, Spirochete specific Probe
5	1151-067-041	357F (200 nM), 1492RL2D (200 nM), 1409RaPrb (300 nM), Sec427F (200 nM), 1045R (200 nM)	Spirochete 16S rRNA, <i>fdhF</i> <sub>Sec</sub> ,	All Bacteria primers, Spirochete specific Probe

## References

1. **Acinas, S. G., R. Sarma-Rupavtarm, V. Klepac-Ceraj, and M. F. Polz.** 2005. PCR-Induced sequence artifacts and bias: Insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl Environ Microbiol* **71**:8966-8969.
2. **Allen, E. E., and J. F. Banfield.** 2005. Community genomics in microbial ecology and evolution. *Nature Rev Microbiol* **3**:489-498.
3. **Breznak, J. A.** 2000. Ecology of prokaryotic microbes in the guts of wood-and litter-feeding termites, p. 209-231. *In* T. Abe, D. E. Bignell, and M. Higashi (ed.), *Termites: Evolution, Sociality, Symbiosis, Ecology* Kluwer Academic Publishers Dordrecht, The Netherlands.
4. **Breznak, J. A., and A. Brune.** 1994. Role of microorganisms in the digestion of lignocellulose by termites. *Ann Rev Entomol* **39**:453-487.
5. **Breznak, J. A., and J. M. Switzer.** 1986. Acetate synthesis from H<sub>2</sub> plus CO<sub>2</sub> by termite gut microbes. *Appl Environ Microbiol* **52**:623-630.
6. **Brune, A.** 2006. Symbiotic associations between termites and prokaryotes, p. 439-474. *In* M. Dworkin, S. Falkow, E. Rosenber, K. H. Schleifer, and E. Stackebrandt (ed.), *The Prokaryotes*, 3 ed, vol. 1. Springer, New York.
7. **Dennis, P., E. A. Edwards, S. N. Liss, and R. Fulthorpe.** 2003. Monitoring gene expression in mixed microbial communities by using DNA microarrays. *Appl Environ Microbiol* **69**:769-778.
8. **Egert, M., and M. W. Friedrich.** 2003. Formation of pseudo-terminal restriction fragments, a PCR-related bias affecting terminal restriction fragment length polymorphism analysis of microbial community structure. *Appl Environ Microbiol* **69**:2555-2562.
9. **Frias-Lopez, J., Y. Shi, G. W. Tyson, M. L. Coleman, S. C. Schuster, S. W. Chisholm, and E. F. DeLong.** 2008. Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci USA* **105**:3805-3810.
10. **Graber, J. R., J. R. Leadbetter, and J. A. Breznak.** 2004. Description of *Treponema azotonutricium* sp. nov. and *Treponema primitia* sp. nov., the first spirochetes isolated from termite guts. *Appl Environ Microbiol* **70**:1315-1320.
11. **Guindon, S., F. Lethiec, P. Duroux, and O. Gascuel.** 2005. PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. *Nucl Acids Res* **33**:W557-559.
12. **Handelsman, J.** 2004. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* **68**:669-685.
13. **Hillier, L. W., V. Reinke, P. Green, M. Hirst, M. A. Marra, and R. H. Waterston.** 2009. Massively parallel sequencing of the polyadenylated transcriptome of *C. elegans*. *Genome Res* **19**:657-666.
14. **Hugenholtz, P., B. M. Goebel, and N. Pace.** 1998. Impact of culture-independent studies on the emerging phylogenetic view of Bacterial diversity. *J Bacteriol* **180**: 4765-4774.

15. **Ji, H., H. Jiang, W. Ma, D. S. Johnson, R. M. Myers, and W. H. Wong.** 2008. An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nat Biotechnol* **26**:1293-300.
16. **Kolb, S., C. Knief, S. Stubner, and R. Conrad.** 2003. Quantitative detection of methanotrophs in soil by novel pmoA-targeted real-time PCR assays. *Appl Environ Microbiol* **69**:2423–2429.
17. **Leadbetter, J. R., T. M. Schmidt, J. R. Graber, and J. A. Breznak.** 1999. Acetogenesis from H<sub>2</sub> plus CO<sub>2</sub> by spirochetes from termite guts. *Science* **283**:686-689.
18. **Li, H., J. Ruan, and R. Durbin.** 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* **18**:1851-1858.
19. **Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Forster, I. Brettske, S. Gerber, A. W. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. König, T. Liss, R. Lussmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N. Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode, and K.-H. Schleifer.** 2004. ARB: a software environment for sequence data. *Nucl Acids Res* **32**:1363-1371.
20. **Lueders, T., and M. W. Friedrich.** 2003. Evaluation of PCR amplification bias by terminal restriction fragment length polymorphism analysis of small-subunit rRNA and mcrA genes by using defined template mixtures of methanogenic pure Cultures and soil DNA extracts. *Appl Environ Microbiol* **69**:320-326.
21. **Luo, Y. H., L. Steinberg, S. Suda, S. Kumazawa, and A. Mitsui.** 1991. Extremely low D/H ratios of photoproduced hydrogen by cyanobacteria. *Plant Cell Physiol* **32**:897-900.
22. **Matson, E. G., X. Zhang, and J. R. Leadbetter.** 2010. Selenium controls expression of paralogous formate dehydrogenases in the termite gut acetogen *Treponema primitia*. *Environ Microbiol* *Accepted*.
23. **Nagalakshmi, U., Z. Wang, K. Waern, C. Shou, D. Raha, M. Gerstein, and M. Snyder.** 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**:1344-1349.
24. **Neidhardt, F. C., and R. Curtiss.** 1996. *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2 ed. ASM press, Washington D.C.
25. **Ottesen, E. A.** 2008. The biology and community structure of CO<sub>2</sub>-reducing acetogens in the termite hindgut. Ph.D. dissertation. California Institute of Technology, Pasadena.
26. **Ottesen, E. A., J. W. Hong, S. R. Quake, and J. R. Leadbetter.** 2006. Microfluidic digital PCR enables multigene analysis of individual environmental bacteria. *Science* **314**:1464-1467.
27. **Pace, N.** 1997. A molecular view of microbial diversity and the biosphere. *Science* **276**:734-740.
28. **Pester, M., and A. Brune.** 2006. Expression profiles of fhs (FTHFS) genes support the hypothesis that spirochaetes dominate reductive acetogenesis in the hindgut of lower termites. *Environ Microbiol* **8**:1261-1270.

29. **Pester, M., M. W. Friedrich, B. Schink, and A. Brune.** 2004. *pmoA*-based analysis of methanotrophs in a littoral lake sediment reveals a diverse and stable community in a dynamic environment. *Appl Environ Microbiol* **70**:3138-42.
30. **Polz, M. F., and C. M. Cavanaugh.** 1998. Bias in template-to-product ratios in multitemplate PCR. *Appl Environ Microbiol* **64**:3724-3730.
31. **Poretzky, R. S., N. Bano, A. Buchan, G. LeClerc, J. Kleikemper, M. Pickering, W. M. Pate, M. A. Moran, and J. T. Hollibaugh.** 2005. Analysis of microbial gene transcripts in environmental samples. *Appl Environ Microbiol* **71**:4121-4126.
32. **Pruesse, E., C. Quast, K. Knittel, B. M. Fuchs, W. Ludwig, J. Peplies, and F. O. Glockner.** 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucl Acids Res* **35**:7188-7196.
33. **Rosenthal, A. Z., E. Matson, A. Eldar, and J. R. Leadbetter.** Deep-transcript sequencing reveals multifaceted interactions between two termite gut spirochetes in co-culture. *Unpublished*.
34. **Rozen, S., and H. J. Skaletsky.** 2000. Primer3 on the WWW for general users and for biologist programmers. , p. 365-386. *In* S. Krawetz and S. Misener (ed.), *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ.
35. **Salmassi, T. M., and J. R. Leadbetter.** 2003. Analysis of genes of tetrahydrofolate-dependent metabolism from cultivated spirochaetes and the gut community of the termite *Zootermopsis angusticollis*. *Microbiology* **149**:2529-2537.
36. **Schloss, P. D., and J. Handelsman.** 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* **71**:1501-1506.
37. **Sorek, R., and P. Cossart.** 2010. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nat Rev Genet* **11**:9-16.
38. **Steinberg, L. M., and J. M. Regan.** 2009. *mcrA*-targeted real-time quantitative PCR method to examine methanogen communities. *Appl Environ Microbiol* **75**:4435-4442.
39. **Todaka, N., T. Inoue, K. Saita, M. Ohkuma, C. A. Nalepa, M. Lenz, T. Kudo, and S. Moriya.** 2010. Phylogenetic analysis of cellulolytic enzyme genes from representative lineages of termites and a related cockroach. *PLoS One* **5**:e8636.
40. **Tringe, S. G., and E. M. Rubin.** 2005. Metagenomics: DNA sequencing of environmental samples. *Nature Rev Gen* **6**:805-814.
41. **Tringe, S. G., C. von Mering, A. Kobayashi, A. A. Salamov, K. Chen, H. W. Chang, M. Podar, J. M. Short, E. J. Mathur, J. C. Detter, P. Bork, P. Hugenholtz, and E. M. Rubin.** 2005. Comparative metagenomics of microbial communities. *Science* **308**:554-557.
42. **Urich, T., A. Lanzen, J. Qi, D. H. Huson, C. Schleper, and S. C. Schuster.** 2008. Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS One* **3**:e2527.
43. **Wang, Z., M. Gerstein, and M. Snyder.** 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**:57-63.

44. **Wilhelm, B. T., S. Marguerat, S. Watt, F. Schubert, V. Wood, I. Goodhead, C. J. Penkett, J. Rogers, and J. Bahler.** 2008. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* **453**:1239-1243.
45. **Yoder-Himes, D. R., P. S. G. Chain, Y. Zhu, O. Wurtzel, E. M. Rubin, J. M. Tiedje, and R. Sorek.** 2009. Mapping the *Burkholderia cenocepacia* niche response via high-throughput sequencing. *Proc Natl Acad Sci USA* **106**:3976-3981.
46. **Zhang, L., T. Hurek, and B. Reinhold-Hurek.** 2007. A nifH-based oligonucleotide microarray for functional diagnostics of nitrogen-fixing microorganisms. *Microb Ecol* **53**:456-470.

# Large D/H variations in bacterial lipids reflect central metabolic pathways

*Originally presented in:*

Xinning Zhang, Aimee L. Gillespie, and Alex L. Sessions. 2009. Large D/H variations in bacterial lipids reflect central metabolic pathways. *Proceedings of the National Academy of Sciences* 106:12580-12586.

## **Abstract**

Large hydrogen-isotopic (D/H) fractionations between lipids and growth water have been observed in most organisms studied to date. These fractionations are generally attributed to isotope effects in the biosynthesis of lipids, and are frequently assumed to be approximately constant for the purpose of reconstructing climactic variables. Here, we report D/H fractionations between lipids and water in 4 cultured members of the phylum *Proteobacteria*, and show that they can vary by up to 500‰ in a single organism. The variation cannot be attributed to lipid biosynthesis as there is no significant change in these pathways between cultures, nor can it be attributed to changing substrate D/H ratios. More importantly, lipid/water D/H fractionations vary systematically with metabolism: chemoautotrophic growth (approximately -200 to -400‰), photoautotrophic growth (-150 to -250‰), heterotrophic growth on sugars (0 to -150‰), and heterotrophic growth on TCA-cycle precursors and intermediates (-50 to +200‰) all yield different fractionations. We hypothesize that the D/H ratios of lipids are controlled largely by

those of NADPH used for biosynthesis, rather than by isotope effects within the lipid biosynthetic pathway itself. Our results suggest that different central metabolic pathways yield NADPH — and indirectly lipids — with characteristic isotopic compositions. If so, lipid  $\delta D$  values could become an important biogeochemical tool for linking lipids to energy metabolism, and would yield information that is highly complementary to that provided by  $^{13}C$  about pathways of carbon fixation.

## Introduction

The hydrogen-isotopic composition ( $^2\text{H}/^1\text{H}$  or D/H ratio, commonly expressed as a  $\delta\text{D}$  value) of lipids is being explored by scientists with diverse interests, including the origins of natural products (1, 2), biogeochemical cycles (3), petroleum systems (4), and paleoclimate (5–7). Because the D/H ratios of lipids are generally conserved over  $\sim 10^6$ -year time scales (8), they are a potentially useful tracer of biogeochemical pathways and processes in the environment. Most research to date has focused on higher plants, in which environmental water is the sole source of external hydrogen and consequently provides primary control over the D/H ratio of biosynthesized lipids (9). Although  $\delta\text{D}$  values for plant lipids and environmental water are generally well correlated, they are also substantially offset from each other. The biochemical basis for this lipid/water fractionation is not well understood. It is generally assumed to arise from a combination of isotope effects during photosynthesis and the biosynthesis of lipids (9–12), and is often treated as approximately constant to reconstruct isotopic compositions of environmental water as a paleoclimate proxy.

There is, however, mounting evidence that the net D/H fractionation between lipids and water can vary by up to 150‰ in plants, even in the same organism (12–16). Modest fractionations associated with fatty acid elongation and desaturation have been documented (2, 14, 15) but are unlikely to account for all of the observed variability. Recent surveys of lipids in marine environments have hinted at even greater variability in isotopic compositions. Jones *et al.* (17) measured fatty acids extracted from coastal marine particulate organic matter (POM) and found  $\delta\text{D}$  values ranging from -73 to -

237‰. Measurements of lipids from marine sediments have extended this range from -32 to -348‰ for lipids with *n*-alkyl skeletons and -148 to -469‰ for those with isoprenoid skeletons (18). Because the lipids measured by both studies likely derive from marine organisms inhabiting seawater of essentially constant  $\delta D$  value ( $\sim 0$ ‰), such differences cannot be due to varying environmental water. Rather, they must relate to more fundamental differences in metabolism.

Culture studies, although limited in number, support the occurrence of highly variable D/H fractionations. Hydrocarbons produced by the green alga *Botryococcus braunii* were depleted in D relative to growth water by 197 to 358‰ (16). Fatty acids from the aerobic methanotroph *Methylococcus capsulatus* (19) were depleted by 20 to 70‰, whereas those in the sulfate reducing chemoautotroph *Desulfobacterium autotrophicum* were depleted by 190 to 360‰ (20). The most strongly fractionating organism reported to date is an  $H_2 + CO_2$  using acetogen, *Sporomusa* sp. DSM 58, which produced fatty acids with depletions in D of nearly 400‰ (21).

Some of this reported variability can be ascribed to systematic differences between lipids with *n*-alkyl versus isoprenoid skeletons. Isoprenoid lipids are typically D depleted relative to *n*-alkyl lipids by 100‰ or more, a pattern now widely confirmed in both culture (13, 16, 19) and environmental samples (12, 18). However, significant variability within single classes of lipids (e.g., fatty acids) cannot be explained because the chemical mechanisms of lipid biosynthesis are strongly conserved across most bacterial and eukaryotic phyla (22–25). Thus, many important questions linger. Do D/H fractionations associated with different metabolic lifestyles (i.e., photoautotrophy,

chemoautotrophy, or heterotrophy) systematically differ? Do they differ in bacteria versus eukaryotes? Perhaps most fundamentally, how and why do biosynthetic processes fractionate hydrogen isotopes at the molecular level and lead to lipids with such diverse isotopic compositions? These questions lie at the heart of our ability to use and interpret lipid  $\delta D$  values from all types of environmental samples. To explore such issues, we measured lipid D/H fractionations in 4 metabolically versatile bacteria grown under photoautotrophic, photoheterotrophic, chemoautotrophic, and heterotrophic conditions on a range of carbon sources metabolized by different pathways of central metabolism.

## Results

### Cultures and Fatty Acids.

Four species of bacteria, chosen to provide a sampling of metabolic diversity, were grown in batch culture on varying substrates (Table 5.1, see Appendix Supplementary Methods for details). *Cupriavidus oxalaticus* str. OX1 and *C. necator* str. H16 are facultative chemoautotrophic  $\beta$ -*Proteobacteria* commonly found in soil and freshwater environments (26), and were grown as aerobic heterotrophs and chemoautotrophs. The model organism *Escherichia coli* K-12 str. MG1655, an obligate heterotrophic  $\alpha$ -*Proteobacterium*, was grown aerobically. The purple non-sulfur anoxygenic phototroph *Rhodospseudomonas palustris* str. TIE-1 is an  $\alpha$ -*Proteobacterium* and was grown under anaerobic photoautotrophic, anaerobic photoheterotrophic, and aerobic heterotrophic conditions. Organic substrates were chosen based on their catabolic relationship to the different pathways of central metabolism. They include those that feed into glycolysis

(glucose, fructose, gluconate, pyruvate), the tricarboxylic acid (TCA) cycle (acetate, succinate), and chemoautotrophic<sup>\*</sup> metabolism [formate, oxalate (27–29)].

Most cultures were harvested during exponential growth (Appendix, Fig. 5.4). Fatty acids were solvent-extracted, derivatized as methyl esters, and quantified by gas chromatography/mass spectrometry (GC/MS; Appendix, Table 5.2). The most abundant fatty acids in *C. oxalaticus* and *C. necator* were palmitic (16:0), palmitoleic (16:1), and oleic (18:1) acids. An additional fatty acid, cyclopropylheptadecanoic acid (cyc-17), was abundant in *E. coli*. *R. palustris* produced significant amounts of 18:1, 18:0 (stearic acid), and 16:0 fatty acids. Relative abundances of fatty acids varied by < 35% between cultures of each bacterial species, with no systematic relationship between growth substrate and fatty acid abundance (Appendix, Table 5.2). Values of  $\delta D$  for individual fatty acids varied widely between cultures (-362 to +331‰), but typically by less than < 30‰ between different fatty acids from the same culture (Appendix, Table 5.3). For simplicity we report and discuss the  $\delta D$  values for palmitic acid as representative of each culture, both because it was present in every organism and because it was generally the most abundant fatty acid.

---

\* Growth on oxalate is classically regarded as heterotrophic, not chemoautotrophic, metabolism. However, conservation of energy during growth is similar to that on formate in that 1-carbon reactions form the basis for generation of reducing power, analogous to “true” chemoautotrophy (e.g., growth on  $H_2 + CO_2$ ) (29, 30). Hence we refer here to growth on formate and oxalate as chemoautotrophic for the purpose of describing hydrogen, rather than carbon, metabolism.

**Table 5.1.** Summary of culture experiments.

Organism-Substrate	$\delta D_s$ (‰)*	$\delta D_w$ , † (‰)	Growth Rate ‡ (h <sup>-1</sup> )	Cultures §
<i>C. oxalaticus</i>				
oxalate	-	-68.6 to +218.3	0.29	Co1 - I,II,III,IV
oxalate	-	-68.6 to +218.3	0.28	Co2 - I,II,III,IV
formate	972	-68.6 to +218.3	0.33	Co3 - I,II,III,IV
acetate	-76	-68.6 to +218.3	0.50	Co4 - I,II,III,IV
succinate	-97	-64.3 to +214.1	0.60	Co5 - I,II,III,IV
succinate	-97	-41.1 to +214.1	NA	Co6 - II,III,IV
<i>C. necator</i>				
formate	972	-68.3	0.17	Cn1 - I
fructose	-22	-65.5	0.34	Cn2 - I
gluconate	NA	-68.1	0.36	Cn3 - I
pyruvate	-12	-64.4	0.61	Cn4 - I
acetate	-76	-68.5	0.35	Cn5 - I
succinate	-97	-68.6	0.48	Cn6 - I
succinate	-97	-68.6	NA	Cn7 - I
<i>E. coli</i>				
glucose	-60	-61.9	0.64	Ec1 - I
gluconate	NA	-62.2	0.57	Ec2 - I
pyruvate	-12	-68.1	0.37	Ec3 - I
acetate	-76	-62.4	0.31	Ec4 - I
glucose	-60	-60.0 to +314.0	0.66	Ec5 - I,II,III,IV
LB	NA	-60.0 to +152.0	NA	Ec6 - I,II,III
<i>R. palustris</i>				
acetate	NA	-53.6	0.1	Rp1 - I
acetate, light	NA	-53.6	0.068	Rp2 - I
CO <sub>2</sub> , light	-	-53.6	0.015	Rp3 - I

\*  $\delta D$  of non-exchangeable C-bound H in the growth substrate (see Appendix Supplementary Methods for calculation details). Uncertainties are likely < 20‰. - indicates a substrate with no H, NA, not available.

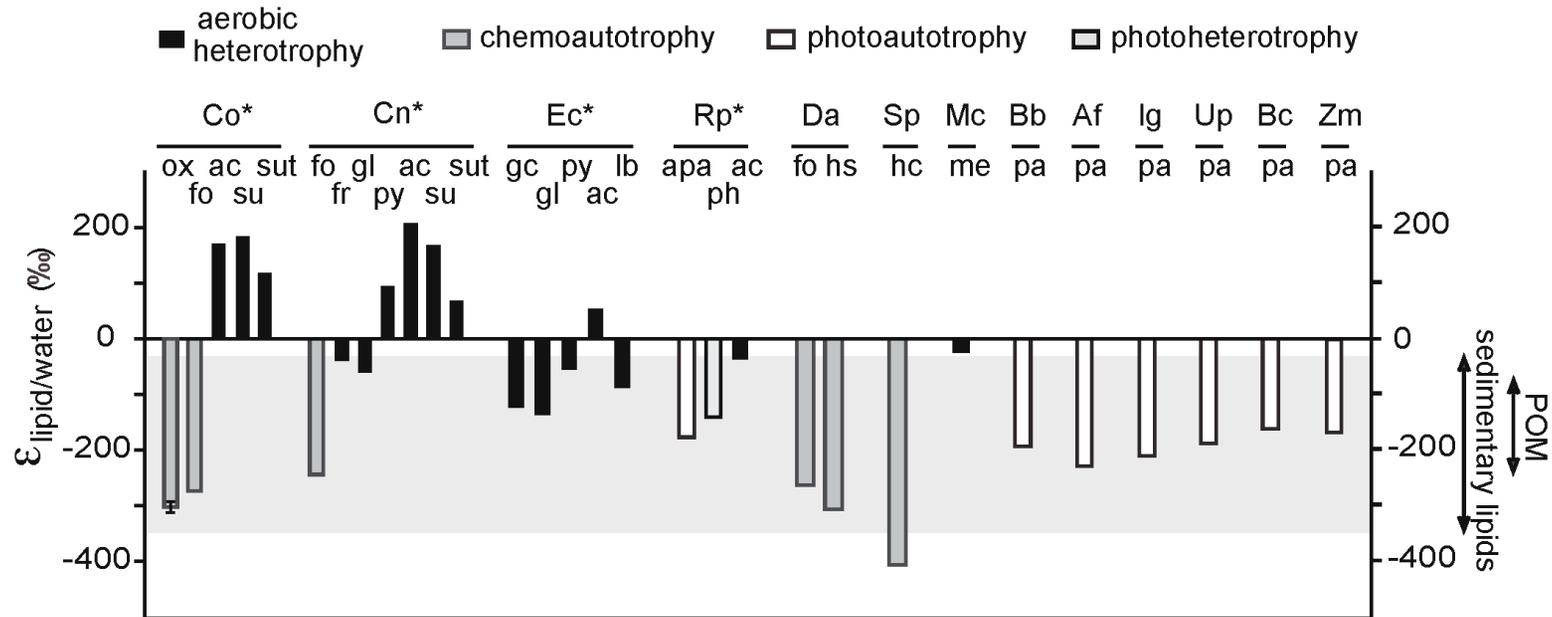
†  $\delta D$  of culture medium before inoculation. Average analytical uncertainty ( $1\sigma$ ) is 0.7‰. Range refers to the span of values covered by 4 replicate cultures, each differing by <100‰.

‡  $1\sigma$  was  $\leq 0.04$  for  $n \geq 2$  cultures in experiments Co1-Co5.

§ Multiple numbers indicate parallel cultures grown in medium with different  $\delta D_w$  values. Cultures Co6-II,III,IV and Cn7-I were harvested in stationary phase, all others were harvested in exponential phase. OD values at harvest are in Appendix (Fig. 5.2).

### **Growth on Different Substrates Leads to Varying Fractionation Between Lipids and Water.**

Fig. 5.1 summarizes current and previous culture data and shows that the net D/H fractionations between lipids and culture water vary in all analyzed strains. The 2 *Cupriavidus* strains, which have the most metabolically versatile carbon metabolisms, exhibited the largest variability (up to 500‰). This is the largest range of isotopic fractionations yet recorded for any individual organism, and includes instances of both D enrichment and D depletion relative to water. The observed range of fractionations is substantially larger than has been previously observed in environmental samples, and — if expressed in nature — would have the potential to explain all such environmental D/H variability. At the same time, lipid/water fractionation during anoxygenic photoautotrophic growth of *R. palustris* was within the range commonly observed for plants. For heterotrophic cultures, the  $\delta D_s$  values of most supplied organic substrates ( $\delta D_s$ ) differ by  $< 100\text{‰}$  (Table 5.1) and cannot explain the range of observed lipid  $\delta D$  values. For example, growth of *E. coli* on glucose ( $\delta D_s = -60\text{‰}$ ) led to D depletion of lipids relative to both water and substrate, whereas growth on acetate ( $\delta D_s = -76\text{‰}$ ) led to D enrichment.



**Fig. 5.1.** Summary of D/H fractionations between fatty acids and water observed in culture experiments, native specimens, and marine organic matter fractions. Plotted fractionations are based on the average  $\delta\text{D}$  value of palmitic acid, or the fatty acid/alkane of nearest chain length, in the culture with water  $\delta\text{D}$  closest to 0‰. Bacterial cultures from this study (\*) are *C. oxalaticus* (Co), *C. necator* (Cn), *E. coli* (Ec), and *R. palustris* (Rp). Organisms from other studies (13, 16, 19–21, 61) are cultured bacteria *D. autotrophicum* (Da), *Sporumusa* sp. (Sp), and *M. capsulatus* (Mc), cultured phytoplankton *B. braunii* (Bb), *Alexandrium fundyense* (Af), *Isochrysis galbana* (Ig), and natural specimens of brown alga *Undaria pinnatifida* (Up), red alga *Binghamia californica* (Bc), and seagrass *Zostera marina* (Zm). The gray box covers the range of fractionations observed in marine POM and sediments (17, 18). Growth substrates are oxalate (ox), formate (fo), fructose (fr), glucose (gc), gluconate (gl), pyruvate (py), acetate (ac), succinate (su, sut), LB (lb),  $\text{H}_2 + \text{CO}_2$  (hc),  $\text{H}_2 + \text{SO}_4^{2-}$  (hs), methane (me), acetate + light (ph),  $\text{CO}_2 + \text{light}$  (pa),  $\text{S}_2\text{O}_3^{2-} + \text{CO}_2 + \text{light}$  (apa). Error bars for culture Co\*-ox are the standard deviation ( $\pm 1\sigma$ ) for 4 sets of biological replicates; other cultures were not replicated. Typical analytical uncertainties are  $< 3.5\text{‰}$  for all cultures.

Our data show a strong correspondence between the pathways of substrate metabolism and lipid  $\delta D$  values. Growth on formate or oxalate, catabolized through 1-carbon reactions (29, 30), yielded lipids depleted in D by 200 to 300‰ relative to water (Fig. 5.1, Appendix Table 5.3). Heterotrophic growth on sugars and photoautotrophic growth on  $CO_2$  produced lipids depleted by 50 to 190‰ relative to water. Growth on a direct precursor (acetate) and intermediate (succinate) of the TCA cycle yielded lipids that were generally D-enriched relative to water (-50 to +200‰), with growth phase modulating the level of enrichment (compare “su” vs. “sut” in Fig. 5.1). These patterns are most strongly exhibited in the 2 *Cupriavidus* strains, but are also present in *E. coli* and *R. palustris*.

### **Manipulation of Growth Water $\delta D$ : Fractionation Factor Curves.**

Additional information on the biochemical causes of these variable fractionations can be deduced from experiments in which the isotopic composition of culture water is experimentally manipulated (19). Conceptually, the net fractionations between lipids and each external H source (i.e., water and organic substrate) can be treated as distinct, yielding the isotopic mass balance

$$R_l = X_w \alpha_{l/w} R_w + (1 - X_w) \alpha_{l/s} R_s \quad [1]$$

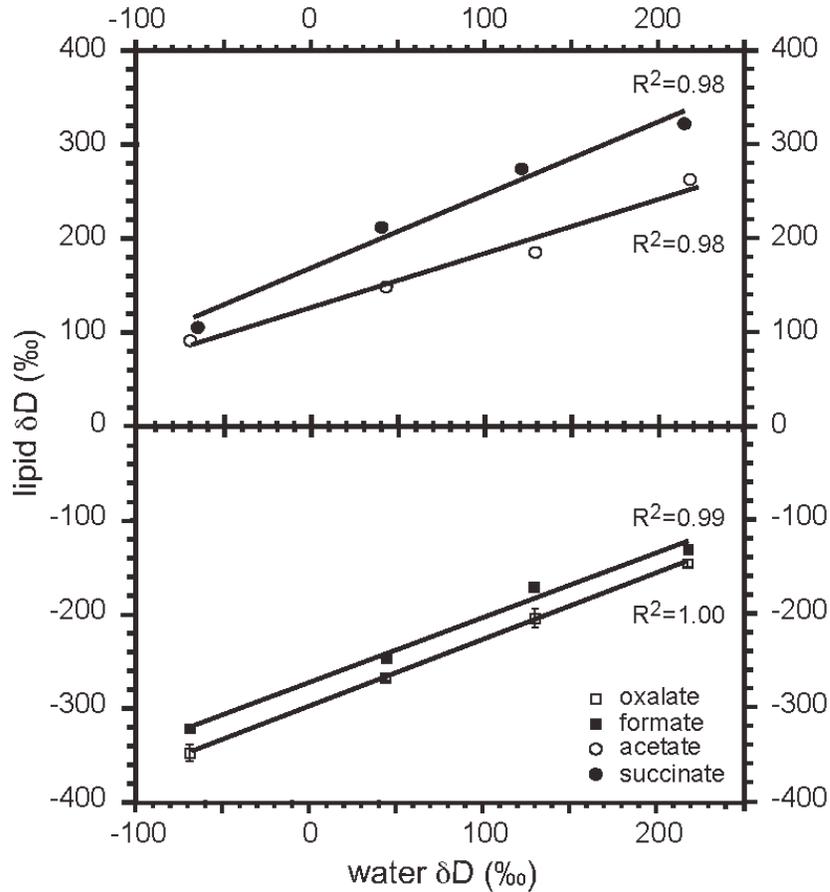
where  $R_l$ ,  $R_w$ , and  $R_s$  denote the D/H ratios of lipids, water, and substrates, respectively (31).  $X_w$  is the mole fraction of lipid H derived from external water, whereas  $\alpha_{l/w}$  and  $\alpha_{l/s}$  represent the net isotopic fractionations associated with uptake and utilization of water and substrate hydrogen, respectively. Eq. 1 represents the overall isotopic relationship

between lipids and external sources of H, and does not imply that the sources must be directly involved in lipid biosynthesis (e.g., glucose may contribute H to lipids by way of metabolic intermediates even though it does not directly participate in the biosynthetic reactions). Measurements of  $R_w$ ,  $R_s$ , and  $R_l$  for parallel cultures in which only 1 parameter,  $R_w$ , is varied experimentally form the basis for regression of  $R_l$  on  $R_w$ . This yields a unique slope and intercept that can be used to constrain the relevant fractionations.

To this end, *C. oxalaticus* and *E. coli* were grown on glucose, acetate, succinate, formate, oxalate, and Lysogeny broth (LB) using waters with  $\delta D$  values ranging from -68 to +314‰. Strong linear relationships ( $R^2 \geq 0.98$ ) between  $\delta D$  values of fatty acids and water were obtained for all such experiments (Fig. 2), implying that the uncertainty in  $\delta D$  values for each culture is minimal (probably  $< 20\%$ ). This is consistent with several “true” biologic replicates grown in identical waters, for which  $\delta D$  values differed by  $< 11\%$  ( $1\sigma$ ).

For cultures where  $X_w$  is known,  $\alpha_{l/w}$  and  $\alpha_{l/s}$  can be calculated directly from the slope and intercept of the regression. Oxalate carries no H at physiological pH, so  $X_w = 1$  for *C. oxalaticus* growing on oxalate. The corresponding value of  $\alpha_{l/w}$  ranged from 0.64 to 0.73 for different fatty acids (Appendix, Table 5.4). Growth on formate yielded nearly identical results, even though formate is a potential source of H. We infer that H on formate exchanges with water to preclude the transmission of substrate H to fatty acids, in accord with our understanding of formate metabolism (29) and the recent results of

Campbell et al. (20).

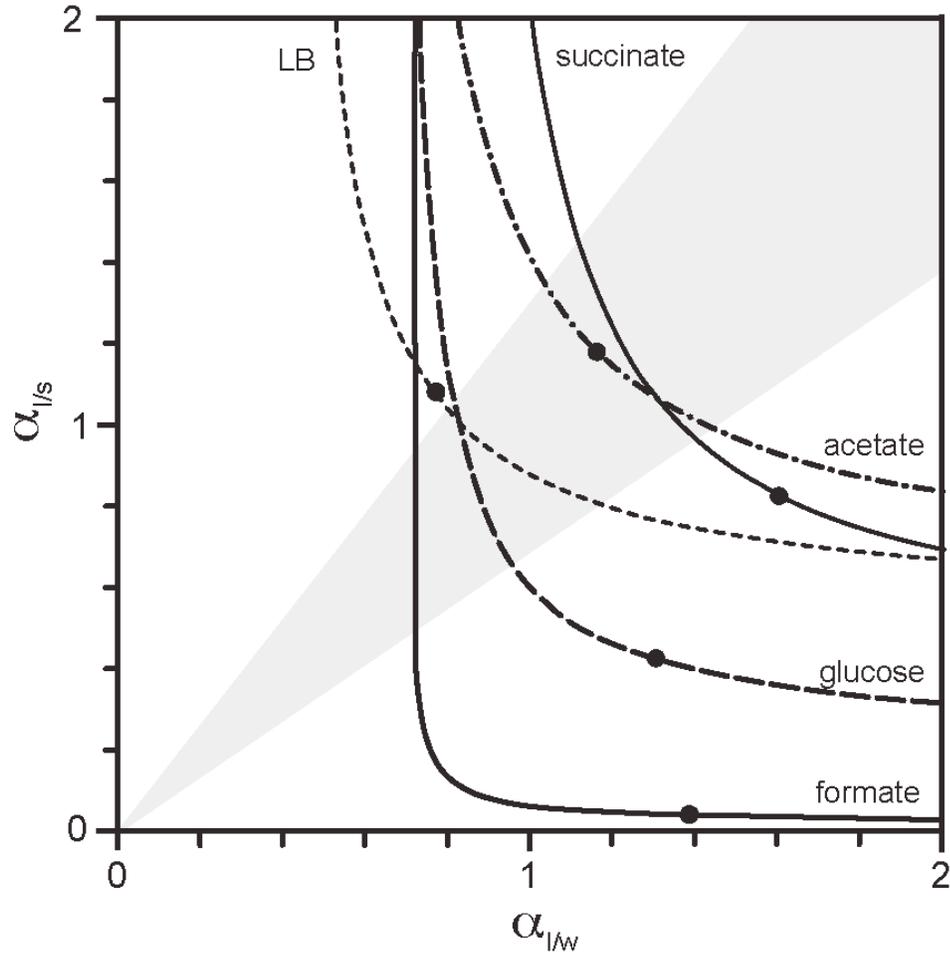


**Fig. 5.2.** Regressions of  $\delta D$  values for palmitic acid versus water for *C. oxalaticus* grown on oxalate, formate, acetate, and succinate. Error bars represent  $1\sigma$  uncertainty for biological replicates. Data for other lipids are Fig. 5.4 and Table 5.4 (Appendix). Each regression provides constraints on fractionations that can be described most succinctly as a single fractionation curve (see Fig. 5.3).

If  $X_w$  is unknown, as is the case for most heterotrophic growth conditions, then a unique solution for  $X_w$ ,  $\alpha_{l/w}$ , and  $\alpha_{l/s}$  is not possible (31). However, the results of each regression can be depicted as a curve relating  $\alpha_{l/s}$  to  $\alpha_{l/w}$ , with each point on the curve representing a different possible combination of values for  $X_w$ ,  $\alpha_{l/w}$ , and  $\alpha_{l/s}$  (Fig. 5.3). These

relationships, which we term “fractionation curves,” are unique for each set of culture conditions and provide a useful means for comparison even while the true values of  $X_w$ ,  $\alpha_{l/w}$ , and  $\alpha_{l/s}$  remain unknown. A basic understanding of the behavior of such curves aids in their interpretation: When comparing 2 fractionation curves, representing 2 different culture conditions, an increase in the true value of  $\alpha_{l/w}$  between conditions will shift the curves horizontally to the right; an increase in the true value of  $\alpha_{l/s}$  will shift them vertically upward; and an increase in  $X_w$  will shift them diagonally downward and to the right (Appendix, Fig. 5.6).

Two key inferences can be drawn from the fractionation curves in Fig. 5.3. First, within a plausible range of values for the 2 fractionation factors (0.5 to 2.0), equivalent to  $\pm$  1,000‰ and 2-fold larger than any net fractionations yet measured for biosynthetic processes, many fractionation curves do not intersect. This is possible only if both  $\alpha_{l/w}$  and  $\alpha_{l/s}$  vary between conditions (e.g., compare *C. oxalaticus* on formate vs. *E. coli* on glucose vs. *C. oxalaticus* on succinate). Changes in fractionation associated with the assimilation of substrate,  $\alpha_{l/s}$ , cannot by itself explain our results, and it is clear that the concept of a nearly constant lipid/water fractionation must be discarded. Because  $\alpha_{l/w}$  changes even in a single organism, whereas pathways of lipid biosynthesis are not known to change significantly with growth on different substrates, we infer that the magnitude of  $\alpha_{l/w}$  is not set primarily by lipid biosynthetic reactions.



**Fig. 5.3.** Fractionation factor curves for palmitic acid in *C. oxalaticus* grown on formate, acetate, and succinate, and in *E. coli* grown on glucose and LB. Each curve represents the set of all possible combinations of  $\alpha_{l/s}$  and  $\alpha_{l/w}$  satisfying the constraints imposed by parallel cultures with differing  $R_w$  (i.e., 1 linear regression in Fig. 5.2). Filled circles indicate values corresponding to  $X_w = 0.5$ . Gray shaded area defines up to 600% variation between  $\alpha_{l/s}$  and  $\alpha_{l/w}$ . Increases in the true value of  $\alpha_{l/s}$ ,  $\alpha_{l/w}$ , and  $X_w$  shift curves up, to the right, or down a diagonal as detailed in Fig. 5.6 (Appendix).

Second, fractionation curves for *E. coli* growth on glucose and LB are offset along a diagonal, and thus are likely related by similar fractionations ( $\alpha$  values) with a decrease in  $X_w$  from glucose to LB. This is consistent with the assimilation of preformed cell constituents by cells grown on the complex LB medium, leading to smaller  $X_w$ . Curves for growth on acetate and succinate also appear related by similar fractionations with

a decrease in  $X_w$  from succinate to acetate, and are consistent with the role of acetate as a direct precursor for lipid biosynthesis.

### **Similarity of Lipid-Substrate and Lipid-Water Fractionations.**

Although it is convenient to treat the net lipid/water and lipid/substrate fractionations as independent, this distinction is probably artificial. Approximately 3/4 of fatty acid H derives from central metabolites (acetate or NADPH; see *Discussion*). H derived from water and growth substrates is comingled in these and virtually all other metabolites because many common classes of reactions (isomerization, hydrolysis, rearrangement, exchange) lead to significant scrambling of C-bound H (32). At this point the 2 “external” sources of H should be affected by many of the same reactions and thus the same isotope effects. The possibility of markedly different values for  $\alpha_{l/s}$  and  $\alpha_{l/w}$  is therefore difficult to envision. Indeed, having widely different fractionations for water and substrate H would practically require that these two H pools remain metabolically distinct, but they do not. Although a strong covariance of  $\alpha_{l/w}$  and  $\alpha_{l/s}$  does reduce the utility of our isotopic labeling approach, there is also a substantial benefit in that the inverse problem — inferring biochemical processes from measured lipid  $\delta D$  values — is made much easier. The strong correspondence of metabolic pathways and lipid  $\delta D$  values implied by Fig. 5.1 would be highly improbable if  $\alpha_{l/s}$  and  $\alpha_{l/w}$  both varied widely and independently.

To extract further insight from Fig. 5.3, we therefore assume that  $\alpha_{l/s}$  and  $\alpha_{l/w}$  for any single culture differ from each other by  $< 600\%$  (i.e., within the gray region in Fig. 5.3).

This is an arbitrary limit, chosen to encompass the  $\sim 500\text{‰}$  range of fractionations we observe, but more conservative limits would yield similar conclusions. Probable limits for  $X_w$ ,  $\alpha_{l/w}$ , and  $\alpha_{l/s}$  can then be calculated: formate ( $> 95\%$ , 0.73, NA), glucose (68–80%, 0.80–0.95, 0.67–1.05), LB (33–47%, 0.80–1.15, 0.82–1.05), succinate (58–70%, 1.13–1.38, 0.98–1.48), and acetate (40–56%, 1.03–1.43, 1.0–1.35). The fact that both  $\alpha_{l/w}$  and  $\alpha_{l/s}$  vary from D depletion ( $\alpha < 1$ ) to D enrichment ( $\alpha > 1$ ) strongly supports our contention that variability in fatty acid  $\delta D$  values cannot be explained solely by modulation of a single fractionating step in their biosynthesis.

Four conclusions arise from this analysis. First, more H is transmitted from water to fatty acids when growing on sugars than on acetate or succinate. This may reflect greater exchange of H associated with sugar isomerization reactions, and/or the fact that acetate feeds directly into fatty acid biosynthesis. Second, no H is transmitted from “chemoautotrophic” substrates to lipids, including formate, oxalate, and  $H_2$ . Third, the fractionations associated with different metabolic pathways are distinct but partially overlapping, in the order chemoautotrophy  $<$  photoautotrophy  $<$  heterotrophic growth on sugars  $<$  growth on TCA-cycle substrates. This pattern of changing fractionations mirrors the observed shifts in fatty acid  $\delta D$  values, and forms the basis for a plausible mechanistic link between lipid  $\delta D$  and metabolism. Fourth, the net lipid/water fractionation for a fatty acid is characteristic of a particular metabolism despite a wide possible range of  $\delta D$  values for potential organic substrates, because relatively little substrate H is incorporated into fatty acids. This means useful metabolic information can be extracted from the  $\delta D$  values of environmental lipids without the necessity of knowing

or measuring  $\delta D$  values of precursor substrates.

## Discussion

Our results demonstrate that 4 metabolically diverse *Proteobacteria* produce fatty acids with  $\delta D$  values that vary systematically with the utilization of key metabolic pathways. The pattern is supported by previous culture studies of bacteria (19–21) and plants (11, 33) and is consistent with field observations. For example, fatty acids with odd carbon-numbered chains are substantially D-enriched (by up to 100‰) relative to those with even-numbered chains in marine POM (17). The former are attributed to heterotrophic bacteria, whereas the latter are likely the products of photosynthetic algae. Similar D enrichments of many bacterial fatty acids and all hopanols relative to their algal counterparts (i.e., even-numbered fatty acids and sterols) have also been observed in marine sediments (18). A 16:1 fatty acid with a  $\delta D$  value as low as -348‰ was detected in those sediments and is apparently generated in situ within the zone of maximal sulfate reduction. Its occurrence and isotopic composition are consistent with origins from chemoautotrophic sulfate reducing bacteria (18). Studies of higher plants have shown several cases of D enrichment resulting from increased reliance on stored carbohydrates for growth and maintenance, and are consistent with a shift from photosynthesis to glycolysis as the primary metabolism (32, 34).

These similarities lead us to propose that the systematic variations reported here are a general feature resulting from the commonality of central metabolic pathways present in

most microbes. Most of the same pathways are also present in higher life forms, although the resulting isotopic patterns will likely be complicated by compartmentalization and transport of lipids, metabolites, and water. Because we have sampled only a minute fraction of extant biota, further culture-based evidence is required to address these issues. In the mean time, confidence that we have observed a general phenomenon can be improved by understanding the mechanistic link between metabolism and D/H fractionation in lipids. Although we do not yet have sufficient data to prove such a link, we suggest that fractionations accompanying the reduction of  $\text{NADP}^+$  provide a plausible — and perhaps unavoidable — mechanism. The basis for this hypothesis is summarized next.

#### **Sources of D/H Variability in Fatty Acids.**

Isotopic labeling studies of fatty acid biosynthesis *in vitro* provide a rough accounting of the H sources for fatty acids (1, 35, 36). They indicate that the most important cellular H source is NAD(P)H, providing ~ 50% of fatty acid H (Appendix, Fig. 5.7). In most cases this comes solely from NADPH, but in some organisms such as *E. coli* it derives equally from both NADPH and NADH (22–24, 37). The methyl group of acetyl-CoA (25%) and water (25%) are of lesser importance. Four possible sources of isotopic variability in fatty acids can then be considered: (i) fractionations associated with substrate uptake and utilization; (ii) the isotopic composition of cellular water, acetate, and NADPH used for biosynthesis; (iii) variations in fractionation associated with reactions of the fatty acid biosynthetic pathway itself, including transfer of H from water and/or NADPH to fatty acids; and (iv) fractionations downstream from fatty acid biosynthesis such as

desaturation and cellular transport.

Analysis of fractionation curves (see *Results*) removes *i* as a possibility. The lack of significant changes in fatty acid abundance and structure between culture conditions with very different lipid  $\delta D$  values rules out *iv*. Option *iii* is worth considering in some detail because multiple enzyme variants exist for 1 of the 2 reductive steps in fatty acid biosynthesis.  $\beta$ -ketoacyl ACP reductase transfers H from NADPH to odd-numbered carbon positions on the nascent fatty acid (step 4 in Appendix, Fig. 5.7), whereas enoyl ACP reductase transfers H from either NADPH or NADH to even and odd positions (step 6 in Appendix, Fig. 5.7) (22, 24). Only one type of  $\beta$ -ketoacyl ACP reductase (FabG) is known, but multiple variants (FabI, FabK, FabL) of enoyl reductase exist, sometimes in the same bacterium (23–25). Thus, differences in fractionations between flavin-free pyridine nucleotide-dependent enzymes like FabG and FabI versus flavoproteins like FabK might provide a mechanism for lipid D/H variability. The former catalyze direct hydride (H<sup>-</sup>) transfer from NAD(P)H to fatty acids (24, 38,39), whereas in the latter H<sup>-</sup> is transferred via the flavin ring, which is susceptible to isotopic exchange with water (36, 40, 41). Differing fractionations between the enzyme types are therefore plausible. However, *E. coli* has only 1 of each reductase enzyme (i.e., a single FabG and FabI) yet lipid/water fractionations still vary substantially (23, 37, 42, 43). *C. necator* has several putative  $\beta$ -ketoacyl ACP reductases and FabI-type enoyl ACP reductases (26), but fractionations in *C. necator* are similar to those of *E. coli*. Thus, current data are inconsistent with fatty acid biosynthetic enzymes causing large variability in fatty acid  $\delta D$  values.

This leads us to consider option *ii*, the H-isotopic composition of water, acetate, and NAD(P)H, as sources for variability in fatty acid  $\delta D$  values. Variations in the D/H ratio of intracellular water have been observed in rapidly-growing *E. coli* and were attributed to the accumulation of water derived from the oxidation of organic substrates (44). Given that substrate  $\delta D$  values did not approach -200‰ or -300‰ in our study, and no systematic relationship between growth rate and fatty acid  $\delta D$  value was observed, the influence of varying intracellular water  $\delta D$  can be ruled out. We eliminate acetate as a source for isotopic variability on the following grounds. When *C. oxalaticus* grows on oxalate, it synthesizes acetate by converting oxalate to 3-phosphoglycerate (Box 3 in Appendix, Fig. 5.8. and Fig. 5.11) (29). In contrast, 3-phosphoglycerate is generated by CO<sub>2</sub> fixation in the Calvin cycle when *C. oxalaticus* grows on formate (Box 7 in Appendix, Fig. 5.8) (29). Thus, 2 very different modes of acetate synthesis (growth on oxalate vs. formate) yield similar fractionations, whereas synthesis of acetate by the same pathway of carbon fixation in different organisms (e.g., *C. oxalaticus* grown on formate and all higher plants) yields very different fractionations.

By process of elimination then, we arrive at the inference that the isotopic composition of NAD(P)H is likely responsible for observed variations in lipid/water fractionation. This conclusion is consistent both with the role of NAD(P)H as the major source of H in fatty acids, and with the ability of flavin-free reductases to transmit isotopic signals from NAD(P)H to fatty acids via hydride transfer reactions (38, 39). But why should the H isotopic composition of NAD(P)H vary so greatly?

**Isotopic Composition of NADPH.**

Biosynthetic reactions commonly require NADPH rather than NADH (27, 28, 45), thus we consider here only the cellular sources of NADPH for simplicity. Similar arguments apply to sources of NADH. In heterotrophic metabolism, the main sources of NADPH are the oxidative reactions catalyzed by glucose-6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase in the pentose phosphate pathway, isocitrate dehydrogenase and malic enzyme in the TCA cycle, and the NADH-NADPH converting transhydrogenase (Appendix, Fig. 5.8 to 5.13) (28, 45). These oxidation reactions typically involve direct  $H^+$  transfer from substrate to  $NADP^+$  (38, 39, 46, 47). Thus, the isotopic composition of each reduced NADPH will depend on that of the reaction substrate plus any isotope effects associated with the  $H^+$  transfer, and the total pool of NADPH will reflect the relative contributions of different pathways of energy metabolism. In oxygenic photoautotrophs,  $NADP^+$  is reduced via the oxidation of water by ferredoxin-NADP oxidoreductase (48), providing yet another source of NADPH.

Isotope fractionations associated with many NADPH generating reactions have been studied *in vitro*. Although their magnitudes vary greatly (up to 3,500‰, see Appendix, Fig. 5.8) (49–52), it is not possible to confidently predict *in vivo* fractionations, and thus the isotopic composition of generated NADPH, from these data. In part this is because kinetic isotope effects will not be fully expressed as isotopic fractionations in committed pathways where the reactant is completely consumed (53). Nor can data from cultures be used, because even organisms grown on a single substrate generate NADPH via multiple

pathways (54). Nevertheless, it is reasonable to expect that the known variability in enzymatic isotope effects will manifest itself as varying  $\delta D$  values for NADPH generated by the respective pathways.

The enrichment of D in lipids from cultures grown on acetate and succinate is particularly interesting, because all of the relevant NADPH-generating reactions have normal isotope effects that should result in depletion of D. We consider 2 possible explanations here. First, D enrichments may arise during  $NADP^+$  reduction in the TCA cycle. Both malic enzyme and isocitrate dehydrogenase generate NADPH via  $H^+$  transfer from an  $OH-C-H$  position (C-2 in isocitrate and malate). The expression of (normal) isotope effects in these reactions should be limited, because there is only one H available for abstraction. However, in both cases the substrate  $OH-C-H$  group is generated by the upstream removal of H from a corresponding methylene ( $H-C-H$ ) position in the precursor molecule (succinate and citrate in Box 4 of Appendix, Fig. 5.8 and Fig. 5.12). These upstream reactions are catalyzed by succinate dehydrogenase and aconitase. Both have been shown to express significant normal isotope effects (50, 52, 55), and are present in *C. necator* and *E. coli* (26, 43). For example, an average kinetic fractionation of 4,400‰ was measured for the removal of pro-*R* H from methylene groups in succinate by flavoprotein succinate dehydrogenase (55). The scale of this isotope effect is consistent with that expected for other flavoproteins, which have been proposed to break C—H bonds by H-tunneling mechanisms (56). These reactions should leave the remaining  $OH-C-H$  position very strongly enriched in D, a signal that can be transferred to NADPH. A second possible route to D enrichment of NADPH is through the action of

transhydrogenases. These balance overproduction of NADPH due to high TCA cycle flux by converting it to NADH (54). Kinetic fractionations for transhydrogenases are typically large (800 to 3,500‰) (51), and should also leave the remaining NADPH strongly D-enriched.

The preceding discussion indicates that (i) NAD(P)H is the source for ~ 50% of fatty acid hydrogen, (ii) the D/H ratios of NAD(P)H generated in different metabolic pathways probably vary over a large range, and (iii) those isotopic signals can be transmitted to fatty acids via hydride transfer reactions. Given these constraints, we should ask whether a link between metabolism and fatty acid  $\delta D$  values can in any way be avoided? Practically the only possibilities are if the relative fluxes of different NADPH-generating pathways remain constant, or if isotopic exchange homogenizes the NADPH pool with water. The former can be dismissed because metabolic flux studies in *E. coli* conclusively show that NADPH sources vary significantly during growth on different substrates (54, 57, 58).

The possibility of hydrogen exchange warrants further consideration because the relevant H position in NAD(P)H is moderately acidic. Experiments with D-labeled NADPH added to purified fatty acid biosynthetic enzymes *in vitro* showed complete conservation of the label in resultant fatty acids, whereas addition to crude cell extracts showed loss of the label (35). The latter result was attributed to isotope exchange via flavoproteins unrelated to lipid biosynthesis in the extract. If isotopic exchange of NADPH also occurs *in vivo*, it could serve to partially or entirely mute the fractionations accompanying NADPH

production. As a concrete example, the slower growth rate of *R. palustris* might explain the smaller fractionations exhibited by this organism growing on acetate relative to *C. oxalaticus*. The extent to which this process is relevant in environmental samples is currently unknown, and must depend on turnover times for NADPH in growing cells. Although very little is known about the turnover of NADPH specifically, we note that the time scales for in vitro experiments (~ 5 h) are quite long compared with typical turnover times for many common metabolic intermediates.

## Conclusions

Existing culture and field data indicate that the D/H ratios of lipids vary substantially with growth conditions, and are systematically related to pathways of central metabolism. Organisms growing on heterotrophic substrates exhibit lipid/water fractionations ranging between approximately -150 to +200‰, photoautotrophic growth yields moderate D depletions (-150 to -250‰), whereas chemoautotrophic growth yields very strong D depletions of -200 to -400‰. We suggest that fractionations in the various pathways that reduce  $\text{NADP}^+$  are the likely source of these variations. However, regardless of mechanism, such patterns hold enormous potential as biogeochemical tracers if they are shown to be widespread. This is particularly so given that the information provided by lipid  $\delta\text{D}$  values would be highly complementary to that encoded by molecular structure and C and N stable isotopes. Whereas  $^{13}\text{C}$  largely records carbon fixation pathways in autotrophs,  $^2\text{H}$  will respond to pathways of energy conservation. In heterotrophs,  $^{13}\text{C}$  and  $^{15}\text{N}$  generally reflect the history of substrate transfers through successive trophic levels, whereas  $^2\text{H}$  could provide a snapshot of the metabolic pathways used for energy

generation by individual organisms. The ability to connect lipids with energy metabolism could find numerous applications, from assessing the in situ metabolic lifestyle of facultative heterotrophs, to identifying modern and ancient communities based on chemoautotrophy, to apportioning the relative contributions of primary production and heterotrophic recycling to sedimentary organic matter.

## Materials and Methods

For details, see Supplementary Methods in the Appendix.

### Culture Strains and Growth.

*Cupriavidus oxalaticus* str. OX1, *Cupriavidus necator* str. H16, *Escherichia coli* K-12 str. MG1655, and *Rhodopseudomonas palustris* str. TIE-1 were grown in batch culture on a variety of substrates (Table 5.1). The  $\delta D$  of culture water was manipulated for *C. oxalaticus* and *E. coli* cultures by volumetrically diluting 99.9% purity  $D_2O$  with distilled deionized water. Substrate  $\delta D$  was not manipulated. Defined carbon sources were provided at 15 mM (except for 22.2 mM glucose in Ec5) for minimal media cultures of *Cupriavidus* and *E. coli*. Undefined carbon source LB was used for Ec6 cultures. *R. palustris* was cultivated in minimal medium with 20 mM thiosulfate for photoautotrophy and 20 mM acetate for both photoheterotrophy and aerobic heterotrophy. All media were 0.2- $\mu m$  filter sterilized and inoculated with single colonies from rich media plates. Culture purity was checked by microscopy, colony morphology on plates, and — for *Cupriavidus* — the ability to grow on oxalate. Optical density (OD) at 600 nm (Cary 50 Bio, UV-Vis Spec) was used in conjunction with growth curve data (Appendix, Fig. 5.4) to harvest cultures at a specific growth phase, generally mid-log phase. To harvest,  $\sim 0.4$

L was centrifuged for 20 min at 4,500 x g, yielding cell pellets ranging from 0.2 to 0.7 g of wet mass. These were stored at -20°C before extraction.

### **Lipid Extraction and Quantification.**

Frozen cell pellets were lyophilized, then ~ 20 mg of biomass was simultaneously transesterified and extracted in hexane/ methanol/ acetyl chloride at 100°C for 10 min (59). The extract was concentrated under N<sub>2</sub> at room temperature. Fatty acid methyl esters (FAMES) were analyzed by gas chromatography/mass spectrometry (GC/MS) on a Thermo-Scientific Trace/ DSQ with a ZB-5ms column and PTV injector operated in splitless mode. Peaks were identified by comparison of mass spectra and retention times to authentic standards and library data. Relative abundances were calculated based on peak areas from the total ion chromatogram without further calibration. They are thus only semiquantitative, but still serve to demonstrate that fatty acid compositions did not change appreciably with growth substrate.

### **Isotopic Analyses.**

The  $\delta D$  values of the most abundant FAMES were measured by GC/pyrolysis/isotope-ratio mass spectrometry (IRMS) on a Thermo-Scientific Delta<sup>+</sup> XP. Chromatographic conditions were identical as for GC/MS analyses, and peaks were identified by retention order and relative height. Data are reported in the conventional  $\delta D$  notation versus the VSMOW standard, and are corrected for the addition of methyl H in the derivative. The root-mean-square (RMS) error of all external standards analyzed with these samples was 2.9‰. Typical precision (10) for replicate analyses of analytes was 3.4‰. The  $\delta D$  values

of culture media, subsampled (1 mL) before inoculation, were measured on a Los Gatos Research DLT-100 liquid water isotope analyzer. Samples were calibrated against 3 working standards with  $\delta D$  values ranging from -59 to +290‰. These were in turn calibrated against the VSMOW, GISP, and SLAP international standards (60). Average precision was 0.7‰ (10).  $\delta D$  values of nonexchangeable H in selected organic substrates were analyzed by Dr. A. Schimmelmann (Indiana University, Bloomington) by double-equilibration following the description in the Appendix (Supplementary Methods).

## **Acknowledgements**

We thank A. Schimmelmann for substrate D/H analyses, M. Eek, and L. Zhang for assistance with lipid D/H analyses and J. Leadbetter and D. Newman and their respective research groups for providing insightful discussion and assistance with microbial cultures. This work was supported by National Science Foundation Grant EAR-0645502 (to A.L.S.) and a predoctoral fellowship (to X.Z.).

## References

1. Robins RJ, et al. (2003) Measurement of  $^2\text{H}$  distribution in natural products by quantitative  $^2\text{H}$  NMR: An approach to understanding metabolism and enzyme mechanism? *Phytochem Rev* 2:87.
2. Billault I, Guiet S, Mabon F, Robins RJ (2001) Natural deuterium distribution in long-chain fatty acids is nonstatistical: A site-specific study by quantitative  $^2\text{H}$  NMR spectroscopy. *ChemBioChem* 2:425–431.
3. Krull E, Sachse D, Mügler I, Thiele A, Gleixner G (2006) Compound-specific  $\delta^{13}\text{C}$  and  $\delta^2\text{H}$  analyses of plant and soil organic matter: A preliminary assessment of the effects of vegetation change on ecosystem hydrology. *Soil Biol Biochem* 38:3211–3221.
4. Pond KL, Huang Y, Wang Y, Kulpa CF (2002) Hydrogen isotopic composition of individual *n*-alkanes as an intrinsic tracer for bioremediation and source identification of petroleum contamination. *Environ Sci Technol* 36:724–728.
5. Sauer PE, Eglinton TI, Hayes JM, Schimmelmann A, Sessions AL (2001) Compound specific D/H ratios of lipid biomarkers from sediments as a proxy for environmental and climatic conditions. *Geochim Cosmochim Acta* 65:213–222.
6. Huang Y, Shuman B, Wang Y, Webb T (2002) Hydrogen isotope ratios of palmitic acid in lacustrine sediments record late Quaternary climate variations. *Geology* 30:1103–1106.
7. Sachse D, Radke J, Gleixner G (2004) Hydrogen isotope ratios of recent lacustrine sedimentary *n*-alkanes record modern climate variability. *Geochim Cosmochim Acta* 68:4877–4889.
8. Sessions AL, Sylva SP, Summons RE, Hayes JM (2004) Isotopic exchange of carbon-bound hydrogen over geologic timescales. *Geochim Cosmochim Acta* 68:1545–1559.
9. Sternberg L (1988) D/H ratios of environmental water recorded by D/H ratios of plant lipids. *Nature* 333:59–61.
10. Luo YH, Steinberg L, Suda S, Kumazawa S, Mitsui A (1991) Extremely low D/H ratios of photoproduct hydrogen by cyanobacteria. *Plant Cell Physiol* 32:897–900.
11. Yakir D, Deniro MJ (1990) Oxygen and hydrogen isotope fractionation during cellulose metabolism in *Lemna gibba* L1. *Plant Physiol* 93:325–332.
12. Chikaraishi Y, Naraoka H, Poulson SR (2004) Hydrogen and carbon isotopic fractionations of lipid biosynthesis among terrestrial (C3, C4 and CAM) and aquatic plants. *Phytochemistry* 65:1369–1381.

13. Sessions AL, Burgoyne TW, Schimmelmann A, Hayes JM (1999) Fractionation of hydrogen isotopes in lipid biosynthesis. *Org Geochem* 30:1193–1200.
14. Chikaraishi Y, Naraoka H, Poulson SR (2004) Carbon and hydrogen isotopic fractionation during lipid biosynthesis in a higher plant (*Cryptomeria japonica*). *Phytochemistry* 65:323–330.
15. Chikaraishi Y, Suzuki Y, Naraoka H (2004) Hydrogen isotopic fractionations during desaturation and elongation associated with polyunsaturated fatty acid biosynthesis in marine macroalgae. *Phytochemistry* 65:2293–2300.
16. Zhang Z, Sachs JP (2007) Hydrogen isotope fractionation in freshwater algae: I. Variations among lipids and species. *Org Geochem* 38:582–608.
17. Jones A, Sessions AL, Campbell B, Li C, Valentine D (2008) D/H ratios of fatty acids from marine particulate organic matter in the California Borderland Basins. *Org Geochem* 39:485–500.
18. Li C, Sessions AL, Kinnaman FS, Valentine DL (2009) Hydrogen-isotopic variability in lipids from Santa Barbara Basin sediments. *Geochim Cosmochim Acta* 73: 4803-4823.
19. Sessions AL, Jahnke LL, Schimmelmann A, Hayes JM (2002) Hydrogen isotope fractionation in lipids of the methane-oxidizing bacterium *Methylococcus capsulatus*. *Geochim Cosmochim Acta* 66:3955–3969.
20. Campbell BJ, Li C, Sessions AL, Valentine DL (2009) Hydrogen isotopic fractionation in lipid biosynthesis by H<sub>2</sub>-consuming *Desulfobacterium autotrophicum*. *Geochim Cosmochim Acta* 73:2744–2757.
21. Valentine DL, Sessions AL, Tyler SC, Chidthaisong A (2004) Hydrogen isotope fractionation during H<sub>2</sub>/CO<sub>2</sub> acetogenesis: Hydrogen utilization efficiency and the origin of lipid-bound hydrogen. *Geobiology* 2:179–188.
22. Marrakchi H, Zhang Y, Rock CO (2002) Mechanistic diversity and regulation of Type II fatty acid synthesis. *Biochem Soc Trans* 30:1050–1055.
23. Campbell JW, Cronan JE (2001) Bacterial fatty acid biosynthesis: Targets for antibacterial drug discovery. *Annu Rev Microbiol* 55:305–332.
24. White SW, Zheng J, Zhang Y, Rock CO (2005) The structural biology of Type II fatty acid biosynthesis. *Annu Rev Biochem* 74:791–831.
25. Rock CO, Jackowski S (2002) Forty years of bacterial fatty acid synthesis. *Biochem Biophys Res Commun* 292:1155–1166.
26. Pohlmann A, et al. (2006) Genome sequence of the bioplastic-producing “Knallgas”

- bacterium *Ralstonia eutropha* H16. *Nat Biotechnol* 24:1257–1262.
27. Gottschalk G (1986) *Bacterial Metabolism* (Springer, New York).
28. White D (2000) *The physiology and biochemistry of prokaryotes* (Oxford Univ Press, New York).
29. Quayle JR (1961) Metabolism of C1 compounds in autotrophic and heterotrophic microorganisms. *Annu Rev Microbiol* 15:119–152.
30. Friedrich CG, Bowien B, Friedrich B (1979) Formate and oxalate metabolism in *Alcaligenes eutrophus*. *J Gen Microbiol* 115:185–192.
31. Sessions AL, Hayes JM (2005) Calculation of hydrogen isotopic fractionations in biogeochemical systems. *Geochim Cosmochim Acta* 69:593–597.
32. Yakir D (1992) Variations in the natural abundance of oxygen-18 and deuterium in plant carbohydrates. *Plant Cell Environ* 15:1005–1020.
33. Luo YH, Sternberg L (1991) Deuterium heterogeneity in starch and cellulose nitrate of CAM and C3 plants. *Phytochemistry* 30:1095–1098.
34. Sessions AL (2006) Seasonal changes in D/H fractionation accompanying lipid biosynthesis in *Spartina alterniflora*. *Geochim Cosmochim Acta* 70:2153–2162.
35. Saito K, Kawaguchi A, Okuda S, Seyama Y, Yamakawa T (1980) Incorporation of hydrogen atoms from deuterated water and stereospecifically deuterium-labeled nicotinamide nucleotides into fatty acids with the *Escherichia coli* fatty acid synthetase system. *Biochim Biophys Acta* 618:202–213.
36. Schmidt HL, Werner RA, Eisenreich W (2003) Systematics of  $^2\text{H}$  patterns in natural compounds and its importance for the elucidation of biosynthetic pathways. *Phytochem Rev* 2:61–85.
37. Heath RJ, Rock CO (1995) Enoyl-acyl carrier protein reductase (fabI) plays a determinant role in completing cycles of fatty acid elongation in *Escherichia coli*. *J Biol Chem* 270:26538–26542.
38. Popjak G (1970) in *The Enzymes*, ed Boyer PD (Academic, New York), pp 115–215.
39. McMurry JE, Begley TP (2005) *The Organic Chemistry of Biological Pathways* (Roberts and Company, Englewood, NJ).
40. Simon H, Kraus A (1976) in *Isotopes in Organic Chemistry*, eds Buncl E, Lee CC (Elsevier Science, Amsterdam), pp 153–229.

41. Ghisla S, Massey V (1989) Mechanisms of flavoprotein catalyzed reactions. *Eur J Biochem* 181:1–17.
42. Bergler H, Fuchsbichler S, Högenauer G, Turnowsky F (1996) The enoyl-[acyl-carrierprotein] reductase (FabI) of *Escherichia coli*, which catalyzes a key regulatory step in fatty acid biosynthesis, accepts NADH and NADPH as cofactors and is inhibited by palmitoyl-CoA. *Eur J Biochem* 242:689–694.
43. Riley M, et al. (2006) *Escherichia coli* K-12: A cooperatively developed annotation snapshot-2005. *Nucleic Acids Res* 34:1–9.
44. Kreuzer-Martin HW, Lott MJ, Ehleringer JR, Hegg EL (2006) Metabolic processes account for the majority of the intracellular water in log-phase *Escherichia coli* cells as revealed by hydrogen isotopes. *Biochemistry* 45:13622–13630.
45. Ingraham JL, Maaløe O, Neidhardt FC (1983) *Growth of the Bacterial Cell* (Sinauer, Sunderland, MA).
46. Edens WA, Urbauer JL, Cleland WW (1997) Determination of the chemical mechanism of malic enzyme by isotope effects. *Biochemistry* 36:1141–1147.
47. Jackson J (2003) Proton translocation by transhydrogenase. *FEBS Lett* 545:18–24.
48. Shin M (2004) How is ferredoxin-NADP reductase involved in the NADP photoreduction of chloroplasts? *Photosynth Res* 80:307–313.
49. O’Leary MH (1989) Multiple isotope effects on enzyme-catalyzed reactions. *Annu Rev Biochem* 59:377–401.
50. Thomson JF, Nance SL, Bush KJ, Szczepanik PA (1966) Isotope and solvent effects of deuterium on aconitase. *Arch Biochem Biophys* 117:65–74.
51. Jackson JB, Peake SJ, White SA (1999) Structure and mechanism of proton-translocating transhydrogenase. *FEBS Lett* 464:1–8.
52. Lenz H, et al. (1971) Stereochemistry of si-citrate synthase and ATP-citrate-lyase reactions. *Eur J Biochem* 24:207–215.
53. Hayes JM (2001) Fractionation of carbon and hydrogen isotopes in biosynthetic processes. *Rev Mineral Geochem* 43:225–277.
54. Sauer U, Canonaco F, Heri S, Perrenoud A, Fischer E (2004) The soluble and membrane-bound transhydrogenases UdhA and PntAB have divergent functions in NADPH metabolism of *Escherichia coli*. *J Biol Chem* 279:6613–6619.
55. Retey J, et al. (1970) Stereochemical studies of the exchange and abstraction of

succinate hydrogen on succinate dehydrogenase. *Eur J Biochem* 14:232–242.

56. Nesheim JC, Lipscomb JD (1996) Large kinetic isotope effects in methane oxidation catalyzed by methane monooxygenase: Evidence for C—H bond cleavage in a reaction cycle intermediate. *Biochem* 35:10240–10247.

57. Zhao J, Baba T, Mori H, Shimizu K (2004) Effect of *zwf* gene knockout on the metabolism of *Escherichia coli* grown on glucose or acetate. *Metab Eng* 6:164–174.

58. Zhao J, Shimizu K (2003) Metabolic flux analysis of *Escherichia coli* K12 grown on <sup>13</sup>C-labeled acetate and glucose using GC-MS and powerful flux calculation method. *J Biotechnol* 101:101–117.

59. Rodríguez-Ruiz J, Belarbi E, Sánchez JLG, Alonso DL (1998) Rapid simultaneous lipid extraction and transesterification for fatty acid analyses. *Biotechnol Tech* 12:689–691.

60. Coplen TB (1988) Normalization of oxygen and hydrogen isotope data. *Chem Geol* 72:293–297.

61. Chikaraishi Y (2003) Compound-specific  $\delta D$ – $\delta^{13}C$  analyses of *n*-alkanes extracted from terrestrial and aquatic plants. *Phytochemistry* 63:361–371.

## Appendix

### Supplementary Methods.

**Table 5.2.** Relative abundances of fatty acids in bacterial cultures.

**Table 5.3.** Measured  $\delta D$  values of fatty acids and culture media water.

**Table 5.4.** Coefficients for regression of  $R_l$  on  $R_w$  and their standard errors (SE). Intercepts and their standard errors are  $\times 10^6$ .

**Fig. 5.4.** Representative growth curves for *C. oxalaticus*, *C. necator*, *E. coli*, and *R. palustris* cultures on selected substrates.

**Fig. 5.5.** Relationship between fatty acid and water  $\delta D$  values for *C. oxalaticus* and *E. coli* cultures.

**Fig. 5.6.** Fractionation curves for hypothetical sets of cultures that differ only in a single parameter ( $\alpha_{l/s}$ ,  $\alpha_{l/w}$ , or  $X_w$ ).

**Fig. 5.7.** Fatty acid biosynthetic pathway highlighting cellular sources of H.

**Fig. 5.8.** Schematic summary of major central metabolic pathways highlighting the most important sources of NADPH reduction.

**Fig. 5.9.** Panel detail of pathway 1 (Fig. 5.8): Embden–Meyerhoff–Parnas Pathway (glycolysis).

**Fig. 5.10.** Panel detail of pathway 2 (Fig. 5.8): Oxidative pentose phosphate pathway and Entner–Doudoroff pathway (sugar degradation, glycolysis).

**Fig. 5.11.** Panel detail of pathway 3 (Fig. 5.8): Oxalate degradation and assimilation via the glycerate pathway.

**Fig. 5.12.** Panel detail of pathway 4 (Fig. 5.8): Tricarboxylic acid cycle and glyoxylate shunt.

**Fig. 5.13.** Panel details of pathways 5 and 6 (Fig. 5.8): Transhydrogenase and Light reactions of photosynthesis.

## Supplementary Methods.

### Culture Media and Growth Conditions.

*Cupriavidus oxalaticus* str. OX1 (DSM 1105T) and *Cupriavidus necator* str. H16 (previously known as *Ralstonia eutropha*, DSM 428) were cultivated in minimal medium described by Dijkhuizen and Harder (1). We used 15 mM rather than 20 mM phosphate buffering (pH 7.2) for pH control. *Escherichia coli* K-12 str. MG1655 was cultivated in M9 minimal media (2). Glucose (22.2 mM) was replaced with other carbon sources at 15 mM for the appropriate experiments (see Table 5.1). All minimal media cultures were amended with EDTA-chelated trace elements formulated according to Flagan et al. (3). *Rhodopseudomonas palustris* str. TIE-1 was cultivated in freshwater minimal medium (4) according to Rashby et al. (5) containing 20 mM bicarbonate buffer and 20 mM thiosulfate for photoautotrophic growth, 20 mM N-Tris(hydroxymethyl)methyl-2-aminoethanesulfonic acid buffer and 20 mM acetate for photoheterotrophic growth, and 20 mM acetate for aerobic heterotrophic growth. Growth substrates were potassium oxalate monohydrate (Sigma), sodium formate (Mallinckrodt), anhydrous D-glucose (Mallinckrodt), anhydrous sodium acetate (Sigma), sodium succinate hexahydrate ( $\geq 99\%$ , Sigma), sodium pyruvate ( $\geq 99\%$ , Sigma), D-(–)-fructose ( $\geq 99\%$ , Sigma), and sodium D-gluconate (97%, Sigma). Aerobic cultures (0.5 L) of *C. oxalaticus*, *C. necator*, and *E. coli* were grown shaking at 200 or 250 rpm in 1-L combusted Pyrex flasks. *Cupriavidus* was cultivated at 30 °C, *E. coli* at 37 °C. Phototrophic cultures of *R. palustris* were incubated in 2-L flasks with 1 L of N<sub>2</sub> headspace and 2,000-lux illumination at room temperature. Aerobic heterotrophic cultures (1 L) were grown at 30 °C, shaking at 250 rpm in the dark.

**Isotopic Analyses.**

The  $\delta\text{D}$  values of the most abundant FAMES were measured on a ThermoScientific GC coupled to a Delta<sup>+</sup>XP isotope-ratio mass spectrometer (IRMS) via the GC/TC pyrolysis interface. Chromatographic conditions were identical as for GC/MS analysis, and peaks were identified by retention order and relative height. The  $\text{H}_3$ -factor was calibrated daily using multiple peaks of  $\text{H}_2$  reference gas at varying intensity, and was stable at  $\sim 4.7$ – $4.8$  ppm/mV. Data are reported in the conventional  $\delta\text{D}$  notation versus the VSMOW standard, and are corrected for the addition of methyl H in the derivative. The  $\delta\text{D}$  value of added methyl H used for this correction was determined by analyzing the dimethyl derivative of phthalic acid for which the  $\delta\text{D}$  value of ring H is known. Replicate analyses were performed for all samples except Co1-Co4, and an external standard containing either 16 *n*-alkanes (*R. palustris* and *E. coli* data) or 8 FAMES (*Cupriavidus* data) of known  $\delta\text{D}$  value was analyzed every 5th injection. The root-mean-square (RMS) error of all external standards analyzed with these samples was 2.9‰. Typical precision ( $1\sigma$ ) based on multiple analyses of analytes was 3.4‰. The  $\delta\text{D}$  of culture media was measured on a Los Gatos Research DLT-100 liquid water isotope analyzer. This instrument measures by absorption spectroscopy, and has been evaluated in detail by Lis et al. (6). Six sequential aliquots (0.8  $\mu\text{L}$  of each) of each sample were injected, with the first 3 discarded, to minimize memory effects.

Substrate  $\delta\text{D}$  values were measured by equilibrating selected aliquots with at least 2 waters of differing D/H ratio (as steam) to control for the presence of exchangeable H before conversion to  $\text{H}_2$  by sequential combustion/reduction (7) and analysis by dual-inlet

IRMS. The isotope ratio of nonexchangeable H was then calculated by mass balance from the 2 exchanged samples (8). Because of significant uncertainties associated with this correction, uncertainties in reported values may be as high as  $\pm 20\%$ . Gluconate was also analyzed by this method, but did not yield reliable results for unknown reasons. The extreme D enrichment of formate indicated by this method is similar to that obtained for a separate formate sample, from a different supplier, measured by a different lab using a different analytical method (9) and is considered reliable.

1. Dijkhuizen L, Harder W (1975) Substrate inhibition in *Pseudomonas oxalaticus* OX1: A kinetic study of growth inhibition by oxalate and formate using extended cultures. *Antonie van Leeuwenhoek* 41:135–146.
2. Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Plainview, NY).
3. Flagan S, Ching WK, Leadbetter JR (2003) *Arthrobacter* strain VAI-A utilizes acylhomoserine lactone inactivation products and stimulates quorum signal biodegradation by *Variovorax paradoxus*. *Appl Environ Microbiol* 69:909–916.
4. Ehrenreich A, Widdel F (1994) Anaerobic oxidation of ferrous iron by purple bacteria, a new type of phototrophic metabolism. *Appl Environ Microbiol* 60:4517–4526.
5. Rashby SE, Sessions AL, Summons RE, Newman DK (2007) Biosynthesis of 2-methylbacteriohopanepolyols by an anoxygenic phototroph. *Proc Natl Acad Sci USA* 104:15099.
6. Lis G, Wassenaar LI, Hendry MJ (2008) High-precision laser spectroscopy D/H and  $^{18}\text{O}/^{16}\text{O}$  measurements of microliter natural water samples. *Anal Chem* 80:287–293.
7. Schimmelmann A, DeNiro M (1993) Preparation of organic and water hydrogen for stable isotope analysis: Effects due to reaction vessels and zinc reagent. *Anal Chem* 65:789–792.
8. Schimmelmann A (1991) Determination of the concentration and stable isotopic composition of nonexchangeable hydrogen in organic matter. *Anal Chem* 63:2456–2459.
9. Campbell BJ, Li C, Sessions AL, Valentine DL (2009) Hydrogen isotopic fractionation

in lipid biosynthesis by H<sub>2</sub>-consuming *Desulfobacterium autotrophicum*. *Geochim Cosmochim Acta* 73:2744–2757.

**Table 5.2.** Relative abundances of fatty acids in bacterial cultures.

Culture	Fatty Acid*											
	12:0	14:1	14:0	15:0	16:1	16:0	cyc17	17:0	18:1	18:0	cyc19	19:0
Co1-I	-	-	-	-	0.39	0.38	-	-	0.22	-	-	-
Co1-II	-	-	-	-	0.35	0.39	-	-	0.25	0.01	-	-
Co1-III	-	-	-	-	0.36	0.36	0.01	-	0.27	0.01	-	-
Co1-IV	-	-	-	-	0.37	0.35	0.01	-	0.27	0.01	-	-
Co2-I	-	-	-	-	0.38	0.34	-	-	0.26	0.01	-	-
Co2-II	-	-	0.02	-	0.4	0.3	0.01	-	0.26	0.01	-	-
Co2-III	-	-	-	-	0.39	0.31	-	-	0.29	0	-	-
Co2-IV	-	-	-	-	0.37	0.33	-	-	0.29	0.01	-	-
Co3-I	-	-	-	-	0.4	0.34	-	-	0.25	0.01	-	-
Co3-II	-	-	-	-	0.4	0.34	-	-	0.25	0.01	-	-
Co3-III	-	-	-	-	0.39	0.36	-	-	0.24	0.01	-	-
Co3-IV	-	-	-	-	0.4	0.35	-	-	0.24	0.01	-	-
Co4-I	-	-	-	-	0.39	0.38	-	-	0.22	0.01	-	-
Co4-II	-	-	-	-	0.39	0.41	-	-	0.19	0	-	-
Co4-III	-	-	-	-	0.4	0.41	-	-	0.17	0.01	-	-
Co4-IV	-	-	0.01	-	0.39	0.42	-	-	0.17	0.01	-	-
Co5-I	-	-	0.03	-	0.39	0.33	-	-	0.25	0.01	-	-
Co5-II	-	-	-	-	0.45	0.38	-	-	0.16	-	-	-
Co5-III	-	-	-	-	0.43	0.4	-	-	0.17	-	-	-
Co5-IV	-	-	-	-	0.41	0.4	-	-	0.18	-	-	-
Co6-II	-	-	-	-	0.36	0.38	0.05	-	0.2	-	-	-
Co6-III	-	-	-	-	0.35	0.35	0.08	-	0.21	-	-	-
Co6-IV	-	-	-	-	0.41	0.36	0.06	-	0.17	-	-	-
Cn1-I	-	-	0.01	-	0.39	0.32	0.02	-	0.26	0.01	-	-
Cn2-I	-	-	0.02	-	0.42	0.29	0.01	-	0.25	0.01	-	-
Cn3-I	-	-	-	-	0.36	0.33	0.01	-	0.3	-	-	-
Cn4-I	-	-	0.01	-	0.43	0.39	-	-	0.18	-	-	-
Cn5-I	-	-	0.02	-	0.42	0.32	-	-	0.24	-	-	-
Cn6-I	-	-	0.02	-	0.39	0.34	-	-	0.23	0.01	-	-
Cn7-I	-	-	0.05	-	0.1	0.36	0.32	-	0.14	-	0.03	-
Ec1-I	0.01	-	0.04	0.01	0.21	0.45	0.13	0.01	0.14	0	0.01	-
Ec2-I	0.01	-	0.03	-	0.19	0.49	0.15	0.01	0.12	0	0	-
Ec3-I	-	-	0.02	-	0.16	0.5	0.18	-	0.11	0.02	0.01	-
Ec4-I	-	-	0.04	-	0.12	0.53	0.21	-	0.09	0	0.02	-
Ec5-I	0.02	-	0.04	0.01	0.2	0.43	0.11	-	0.17	-	0.01	-
Ec5-II	0.02	-	0.04	-	0.02	0.46	0.32	0.01	0.04	-	0.08	-
Ec5-III	0.02	-	0.04	0.01	0.15	0.44	0.16	0.01	0.15	-	0.01	-
Ec5-IV	0.02	-	0.04	0.01	0.01	0.5	0.22	0.01	0.15	-	0.02	-
Ec6-I	0.03	0.01	0.06	0.02	0.26	0.38	0.07	0.01	0.16	0.01	-	-
Ec6-II	0.03	0.01	0.07	0.02	0.29	0.38	0.04	0.01	0.15	-	-	-
Ec6-III	0.02	0.01	0.05	0.02	0.33	0.42	0.14	-	0.01	-	-	-
Rp1-I	-	-	-	-	0.01	0.08	-	-	0.53	0.21	0.15	0.02
Rp2-I	-	-	-	-	0.03	0.14	-	-	0.65	0.11	0.07	-
Rp3-I	-	-	-	-	0.03	0.15	-	-	0.69	0.1	0.01	-

\*12:0 = lauric acid, 14:1 = myristoleic acid, 14:0 = myristic acid, 15:0 = pentadecanoic acid, 16:1 =

palmitoleic acid, 16:0 = palmitic acid; cyc17 = cyclopropyl-heptadecanoic acid, 17:0 = heptadecanoic acid, 18:1 = oleic acid, 18:0 = stearic acid, cyc19 = cyclopropyl-nonadecanoic acid, 19:0 = nonadecanoic acid. Relative abundances are calculated from TIC peak areas of FAMES as the fraction of total quantified fatty acids.

**Table 5.3.** Measured  $\delta D$  values of fatty acids and culture media water.

Culture	n <sup>†</sup>	Fatty Acid*										Medium
		16:1	$\sigma$	16:0	$\sigma$	cycl7	$\sigma$	18:1	$\sigma$	18:0	$\sigma$	$\delta D_w$
Co1-I	1	-356	-	-343	-	-	-	-338	-	-	-	-68.6
Co1-II	1	-281	-	-266	-	-	-	-260	-	-	-	44.6
Co1-III	1	-219	-	-198	-	-	-	-193	-	-	-	130.3
Co1-IV	1	-174	-	-148	-	-	-	-140	-	-	-	218.3
Co2-I	1	-362	-	-355	-	-	-	-342	-	-	-	-68.6
Co2-II	1	-291	-	-270	-	-	-	-269	-	-	-	44.6
Co2-III	1	-240	-	-212	-	-	-	-215	-	-	-	130.3
Co2-IV	1	-176	-	-143	-	-	-	-131	-	-	-	218.3
Co3-I	1	-344	-	-322	-	-	-	-319	-	-	-	-68.6
Co3-II	1	-269	-	-246	-	-	-	-239	-	-	-	44.6
Co3-III	1	-191	-	-170	-	-	-	-167	-	-	-	130.3
Co3-IV	1	-154	-	-130	-	-	-	-123	-	-	-	218.3
Co4-I	1	38	-	92	-	-	-	76	-	-	-	-68.6
Co4-II	1	93	-	149	-	-	-	129	-	-	-	44.6
Co4-III	1	129	-	187	-	-	-	167	-	-	-	130.3
Co4-IV	1	190	-	263	-	-	-	234	-	-	-	218.3
Co5-I	2	64	1.7	109	3.0	-	-	77	0.8	-	-	-64.3
Co5-II	3	176	1.0	219	1.4	-	-	205	1.6	-	-	41.1
Co5-III	3	235	2.5	282	1.2	-	-	266	2.8	-	-	121.1
Co5-IV	3	296	3.6	331	4.2	-	-	326	5.0	-	-	214.1
Co6-II	3	106	1.3	166	1.1	-	-	128	4.1	-	-	41.1
Co6-III	3	163	0.7	226	1.4	-	-	184	0.7	-	-	121.1
Co6-IV	3	238	1.6	302	3.2	-	-	258	1.8	-	-	214.1
Cn1-I	4	-298	3.5	-294	3.5	-	-	-287	8.2	-	-	-68.3
Cn2-I	2	-137	0.1	-101	1.4	-	-	-110	2.1	-	-	-65.5
Cn3-I	2	-124	1.9	-124	3.4	-	-	-109	4.1	-	-	-68.1
Cn4-I	4	-12	2.2	26	3.0	-	-	8	5.8	-	-	-64.4
Cn5-I	4	71	2.1	127	11.8	-	-	101	1.8	-	-	-68.5
Cn6-I	2	51	1.7	89	1.5	-	-	62	0.7	-	-	-68.6
Cn7-I	2	-35	2.6	-3	0.5	-11	1.9	-34	2.2	-	-	-68.6
Ec1-I	2	-176	2.7	-178	0.3	-160	2.8	-173	2.2	-	-	-61.9
Ec2-I	2	-196	4.3	-190	3.3	-166	1.4	-187	4.6	-	-	-62.2
Ec3-I	4	-124	3.8	-120	5.8	-112	3.7	-108	5.2	-	-	-68.1
Ec4-I	2	-23	2.2	-12	3.1	-7	0.7	-5	0.5	-	-	-62.4
Ec5-I	2	-197	0.0	-180	3.3	-178	2.7	-183	2.7	-	-	-60.0
Ec5-II	2	-122	10.4	-128	1.1	-121	1.3	-122	0.6	-	-	49.9
Ec5-III	2	-68	1.8	-44	1.4	-52	0.7	-44	0.5	-	-	152.0
Ec5-IV	2	30	0.4	57	0.2	41	1.2	50	3.2	-	-	314
Ec6-I	2	-152	0.6	-143	0.0	-139	1.4	-121	0.4	-	-	-60.0
Ec6-II	2	-98	0.3	-83	1.6	-116	2.9	-61	0.7	-	-	49.9
Ec6-III	2	-58	3.4	-34	0.5	-70	4.9	-20	7.1	-	-	152.0
Rp1-I	2	-	-	-87	1.4	-	-	-77	4.0	-37	0.5	-53.6
Rp2-I	2	-169	3.6	-185	1.4	-	-	-173	2.1	-157	1.4	-53.6
Rp3-I	2	-	-	-220	0.9	-	-	-229	1.2	-208	1.2	-53.6

\* Fatty acid structures for corresponding abbreviations are listed in Table S1. Tabulated values are the average  $\delta D$  values for replicate analyses, in permil. Values of  $s$  are calculated from replicate analyses.

† Number of replicate measurements for fatty acids.

**Table 5.4.** Coefficients for regression of  $R_I$  on  $R_w$  and their standard errors (SE). Intercepts and their standard errors are  $\times 10^6$ .

Cultures	16:1					16:0					18:1				
	Slope	SE	Intercept	SE	R <sup>2</sup>	Slope	SE	Intercept	SE	R <sup>2</sup>	Slope	SE.	Intercept	SE	R <sup>2</sup>
Co1-I,II,II,IV	0.64	0.03	7.22	4.77	1.00	0.69	0.03	2.83	4.77	1.00	0.70	0.02	1.72	3.48	1.00
Co2-I,II,II,IV	0.64	0.02	6.25	3.21	1.00	0.73	0.01	-5.70	2.13	1.00	0.72	0.05	-3.60	8.88	0.99
Co3-I,II,II,IV	0.69	0.06	3.04	10.25	0.98	0.69	0.05	5.97	8.62	0.99	0.70	0.04	4.92	7.50	0.99
Co4-I,II,II,IV	0.52	0.04	85.80	6.30	0.99	0.58	0.06	85.07	10.57	0.98	0.54	0.05	88.72	8.85	0.98
Co5-I,II,II,IV	0.83	0.07	46.55	12.18	0.99	0.80	0.09	58.07	14.58	0.98	0.89	0.10	40.34	0	0.98
Co6-II,II,IV	0.76	0.03	47.99	4.45	1.00	0.79	0.02	53.88	2.91	1.00	0.75	0.03	53.98	4.62	1.00
Ec5-I,II,III,IV	0.60	0.02	37.83	3.12	0.99	0.65	0.04	31.70	7.03	0.99	0.63	0.03	34.08	4.38	0.99
Ec6-I,II,III	0.44	0.03	67.19	5.28	0.99	0.52	0.02	57.99	3.42	0.99	0.48	0.04	67.41	6.42	0.99

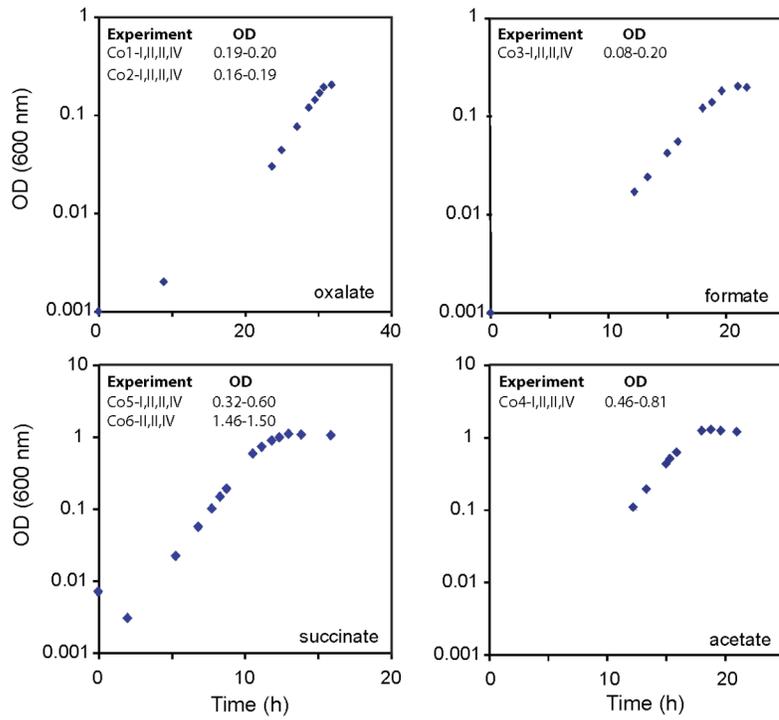
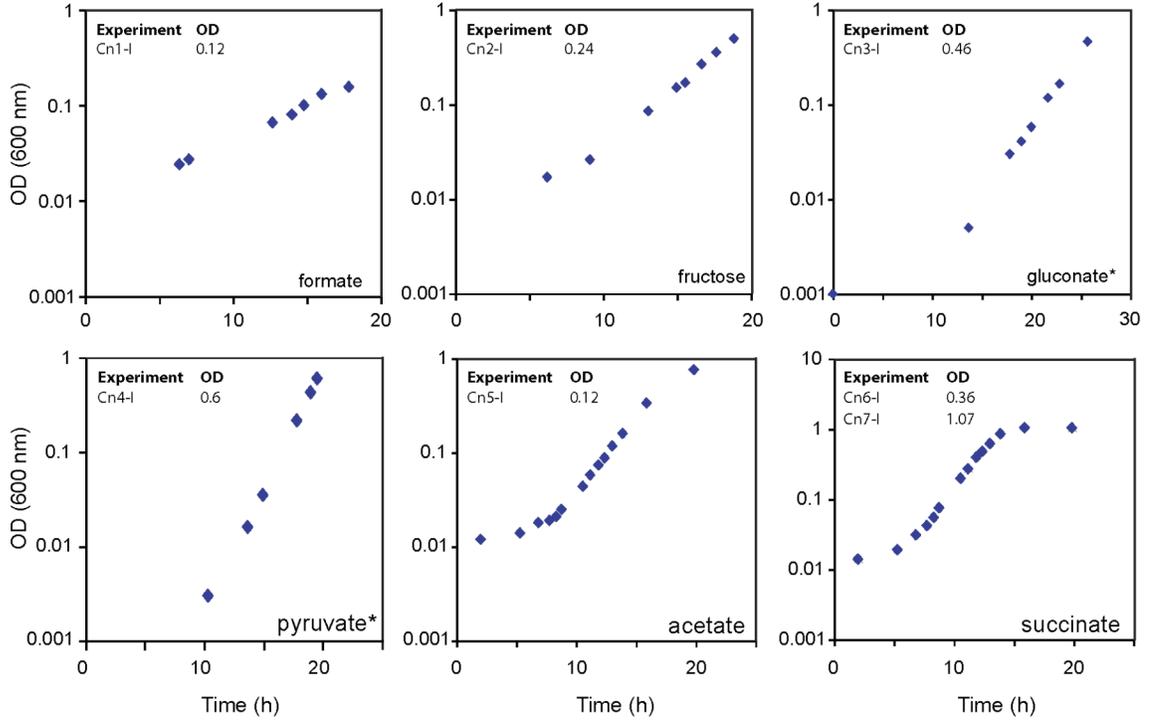
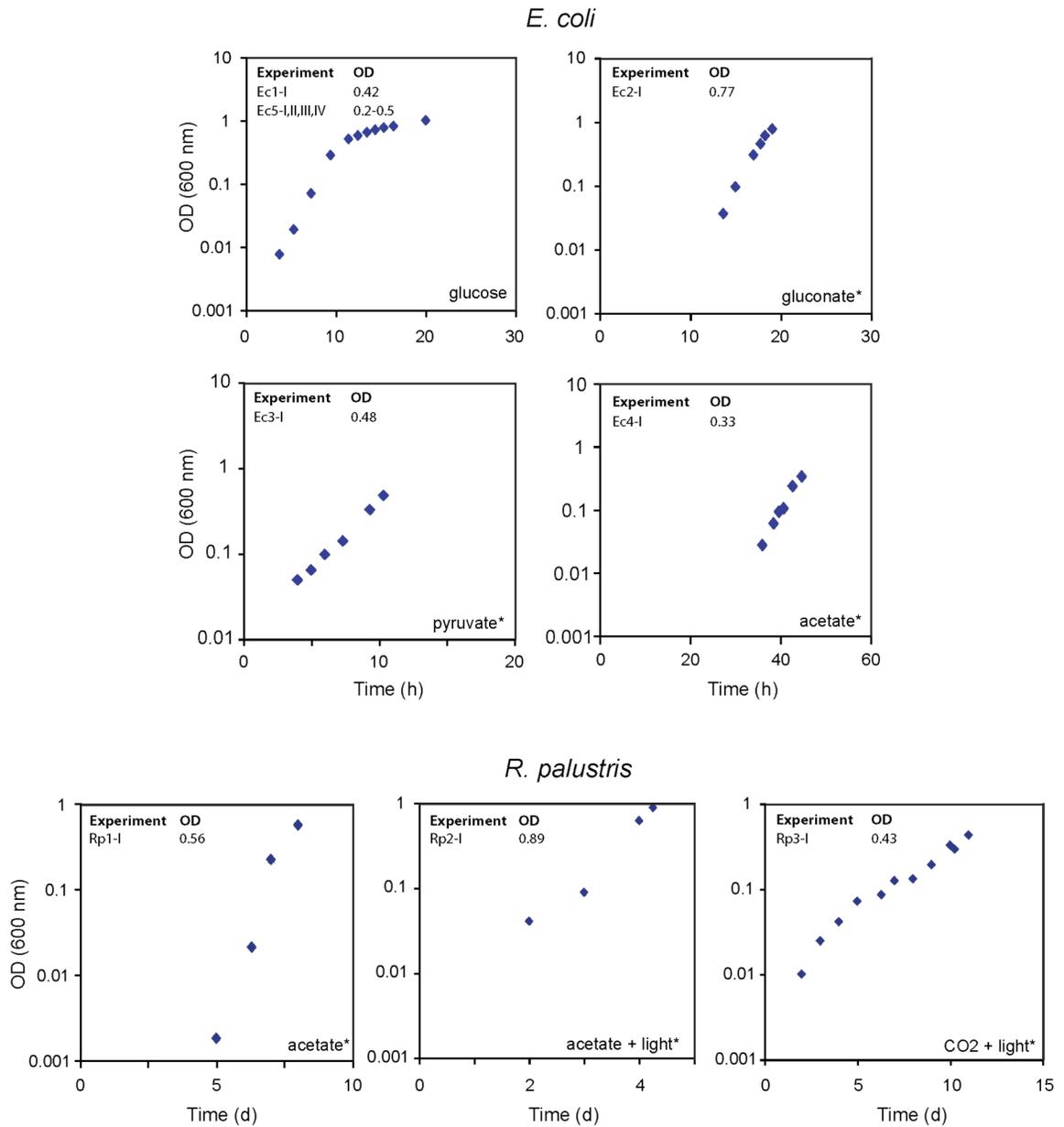
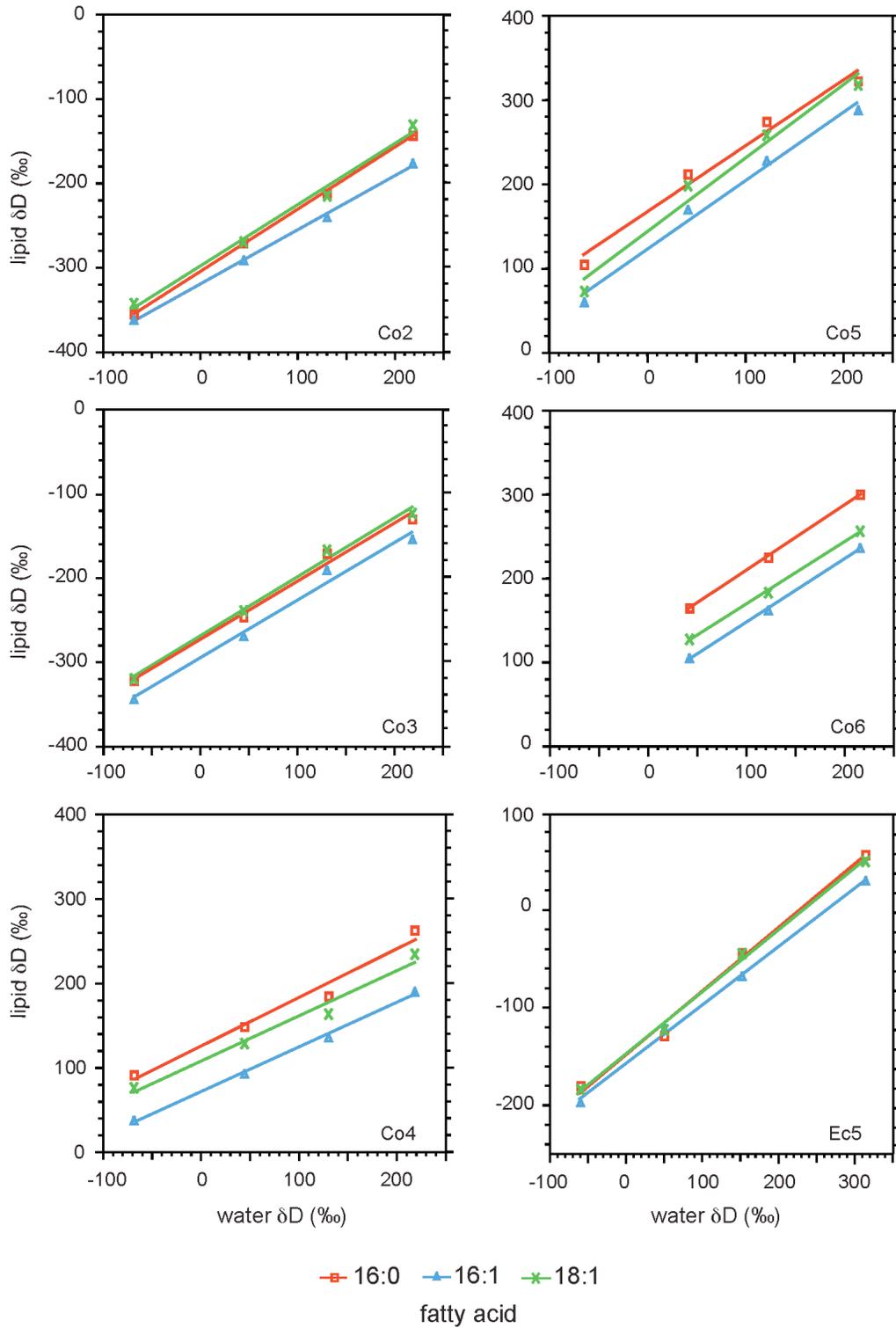
*C. oxalaticus**C. necator*

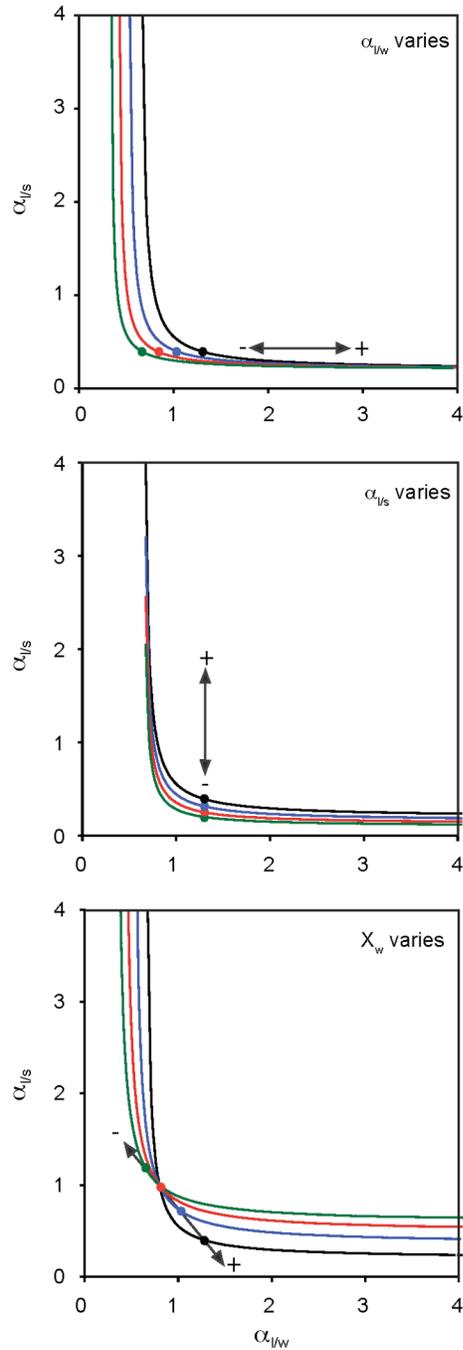
Fig. 5.4. continued on next page



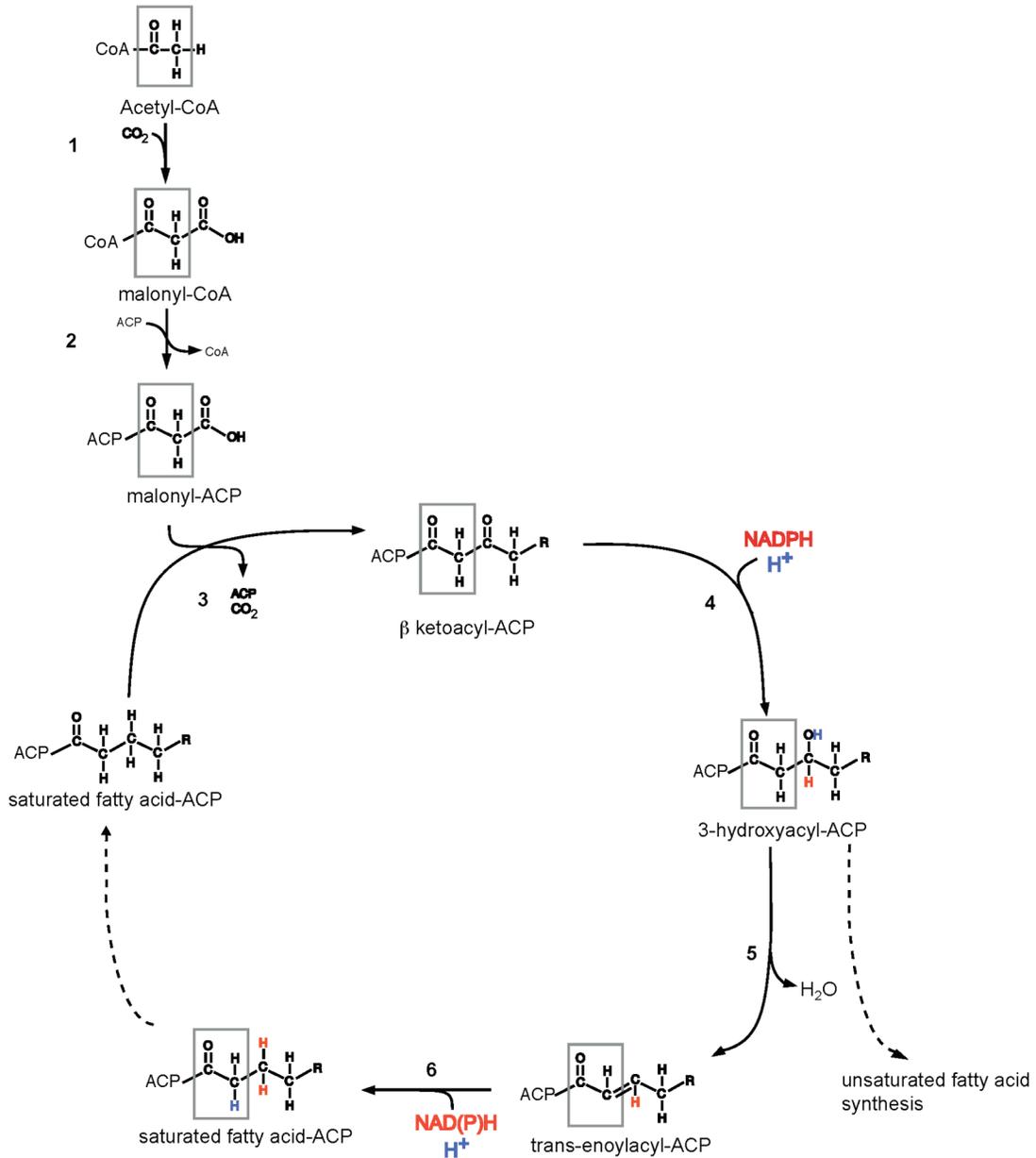
**Fig. 5.4.** Representative growth curves for *C. oxalaticus*, *C. necator*, *E. coli*, and *R. palustris* cultures on selected substrates. OD<sub>600nm</sub> values at harvest are listed. For most substrates, plotted growth data are from a single culture grown to stationary phase to define the growth curve, but not then analyzed. Panels containing substrates marked with asterisks show the growth curve from a culture that was harvested for isotopic analysis, generally in mid-log phase.



**Fig. 5.5.** Relationship between fatty acid and water  $\delta D$  values for *C. oxalaticus* and *E. coli* cultures. The slope of each regression curve is equivalent to  $X_w \alpha_{l/w}$ . Culture numbers are labeled in the lower right of each plot.



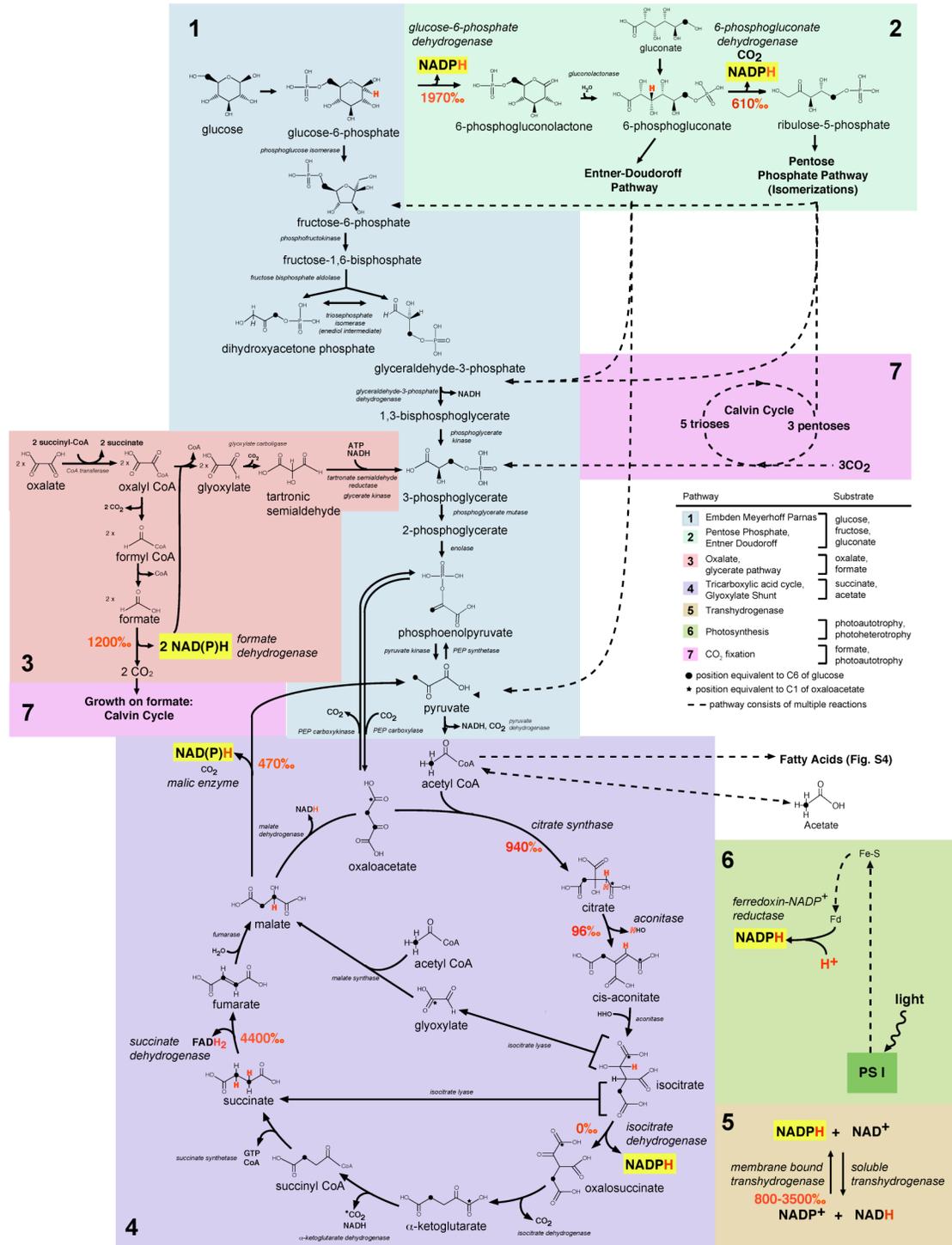
**Fig. 5.6.** Fractionation curves for hypothetical sets of cultures that differ only in a single parameter ( $\alpha_{l/s}$ ,  $\alpha_{l/w}$ , or  $X_w$ ). The plotted curves reflect 20% incremental changes in the specified parameter and arrows indicate the direction of change for an entire curve. The effects can be treated as independent, such that the result of changing 2 parameters can be estimated by vector addition. Filled circles mark  $X_w = 0.5$ . To account for curves that shift up and to the right (e.g., those in Fig. 5.3), both  $\alpha_{l/s}$  and  $\alpha_{l/w}$  must simultaneously change.



**Fig. 5.7.** Fatty acid biosynthetic pathway highlighting cellular sources of H (refs. 1–5). (1) Acetyl-CoA carboxylase, which is regulated to control carbon flux into lipids. (2) Malonyl CoA-ACP transacylase. (3) β-ketoacyl ACP synthase (FabB, FabF, FabH). FabH controls biosynthesis initiation and fatty acid composition based on acyl-CoA specificity, whereas FabB and FabF catalyze subsequent rounds of elongation by condensing malonyl-ACP with acyl-ACP. (4) β-ketoacyl ACP reductase (FabG). (5) β-hydroxyacyl ACP dehydratase (FabZ, FabA). (6) Enoyl ACP reductase (FabI, FabK, FabL).

1. Marrakchi H, Zhang Y, Rock CO (2002) Mechanistic diversity and regulation of Type II fatty acid synthesis. *Biochem Soc Trans* 30:1050–1055.

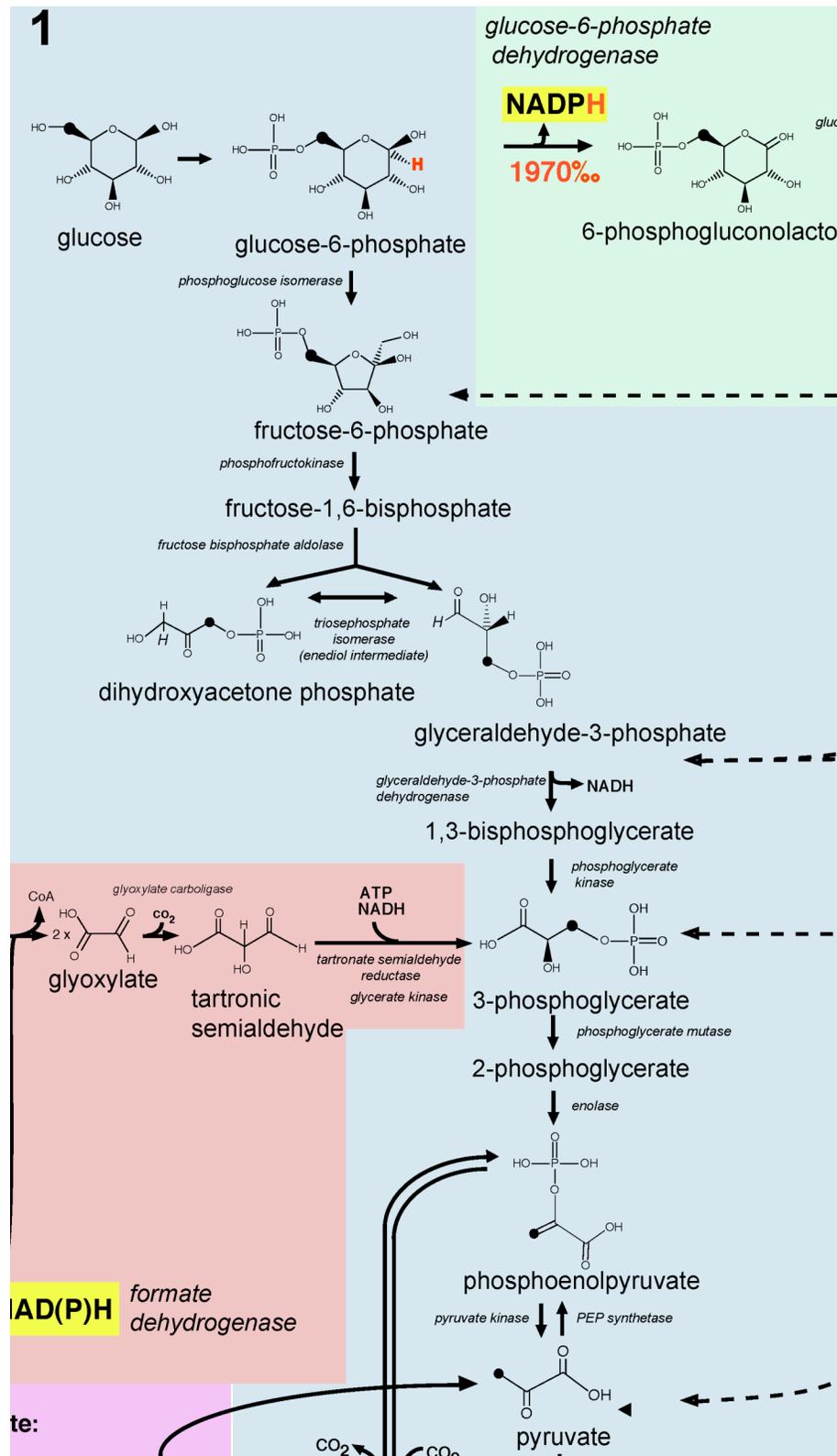
2. Magnuson K, Jackowski S, Rock CO, Cronan JE (1993) Regulation of fatty acid biosynthesis in *Escherichia coli*. *Microbiol Molec Biol Rev* 57:522–542.
3. Campbell JW, Cronan JE (2001) Bacterial fatty acid biosynthesis: Targets for antibacterial drug discovery. *Annu Rev Microbiol* 55:305–332.
4. Rock CO, Cronan JE (1996) *Escherichia coli* as a model for the regulation of dissociable (type II) fatty acid biosynthesis. *Biochim Biophys Acta* 1302:1–16.
5. White S, Zheng J, Zhang Y, Rock CO (2005) The structural biology of Type II fatty acid biosynthesis. *Annu Rev Biochem* 74:791–831.



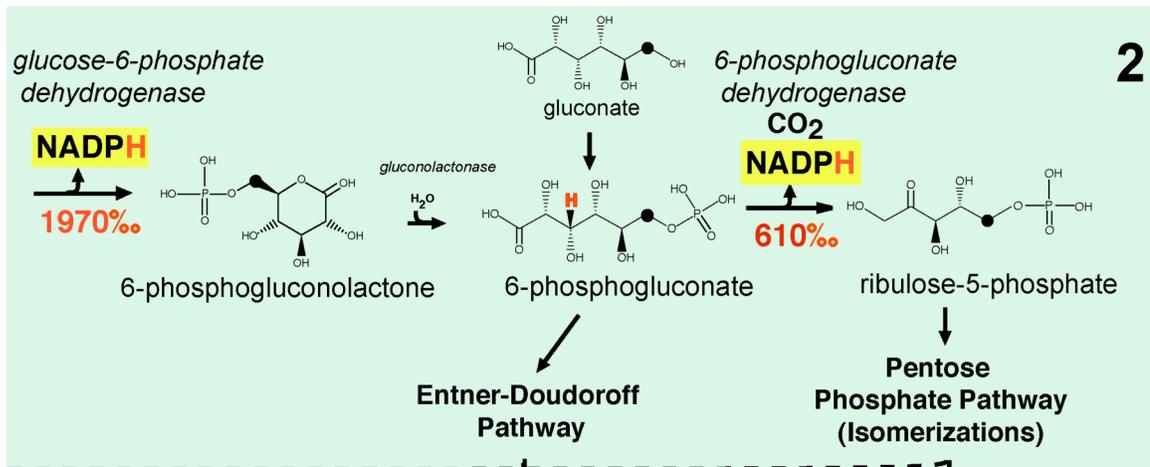
**Fig. 5.8.** Schematic summary of major central metabolic pathways highlighting the most important sources of NADPH reduction (refs. 1–5). (1) Embden–Meyerhoff–Parnas Pathway (glycolysis). (2) Oxidative pentose phosphate pathway and Entner–Doudoroff pathway (sugar degradation, glycolysis). (3) Oxalate degradation and assimilation via the glycerate pathway. (4) Tricarboxylic acid cycle and glyoxylate shunt. (5) Transhydrogenase. (6) Light reactions of photosynthesis. (7) Calvin cycle for CO<sub>2</sub>

fixation. Reactions producing NADPH or potentially having strong influence on the isotopic composition of NADPH are highlighted. D/H fractionations (in permil notation), corresponding to isotope effects  $^{H/K}$  or  $^{D(V/K)}$ , are indicated by red numbers (6–10). Filled circles denote positions equivalent to C-6 of glucose and the methyl group of acetate. Filled stars denote the position equivalent to C-1 of oxaloacetate.

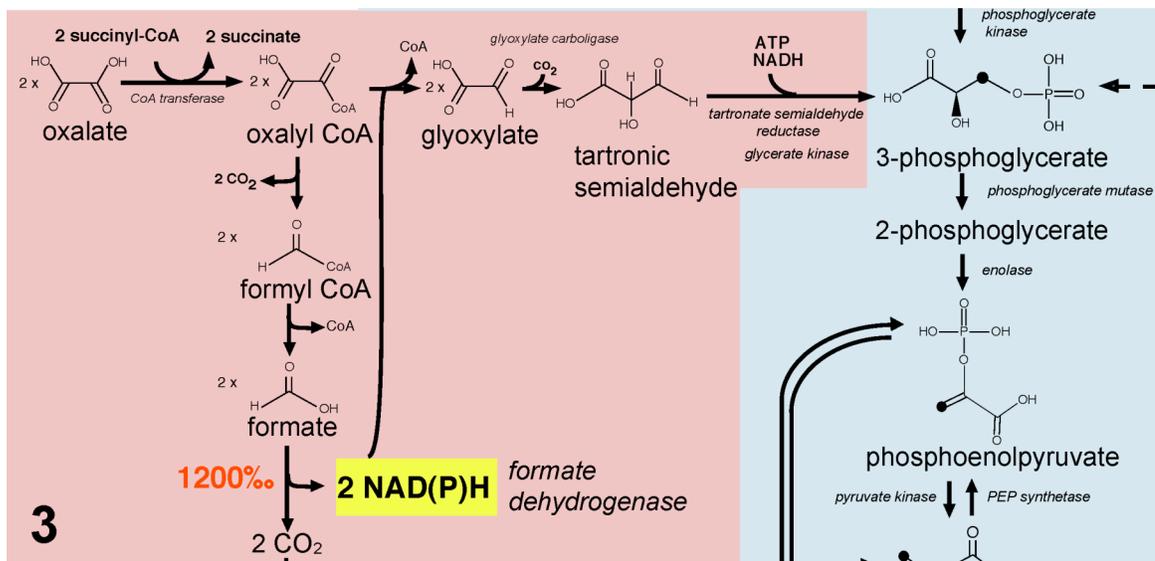
1. Gottschalk G (1986) *Bacterial Metabolism* (Springer-Verlag, New York).
2. Ingraham JL, Maaløe O, Neidhardt FC (1983) *Growth of the Bacterial Cell* (Sinauer, Sunderland, MA).
3. White D (2000) *The Physiology and Biochemistry of Prokaryotes* (Oxford Univ Press, New York).
4. Quayle JR (1961) Metabolism of C1 compounds in autotrophic and heterotrophic microorganisms. *Annu Rev Microbiol* 15:119–152.
5. Kornberg HL, Elsdén SR (1961) The metabolism of 2-carbon compounds by microorganisms. *Adv Enzymol Relat Subj Biochem* 23:401–470.
6. Retey J, et al. (1970) Stereochemical studies of the exchange and abstraction of succinate hydrogen on succinate dehydrogenase. *Eur J Biochem* 14:232–242.
7. Thomson JF, Nance SL, Bush KJ, Szczepanik PA (1966) Isotope and solvent effects of deuterium on aconitase. *Arch Biochem Biophys* 117:65–74.
8. O'Leary MH (1989) Multiple isotope effects on enzyme-catalyzed reactions. *Annu Rev Biochem* 59:377–401.
9. Jackson JB, Peake SJ, White SA (1999) Structure and mechanism of proton-translocating transhydrogenase. *FEBS Lett* 464:1–8.
10. Lenz H, et al. (1971) Stereochemistry of *si*-citrate synthase and ATP-citrate-lyase reactions. *Eur J Biochem* 24:207–215.



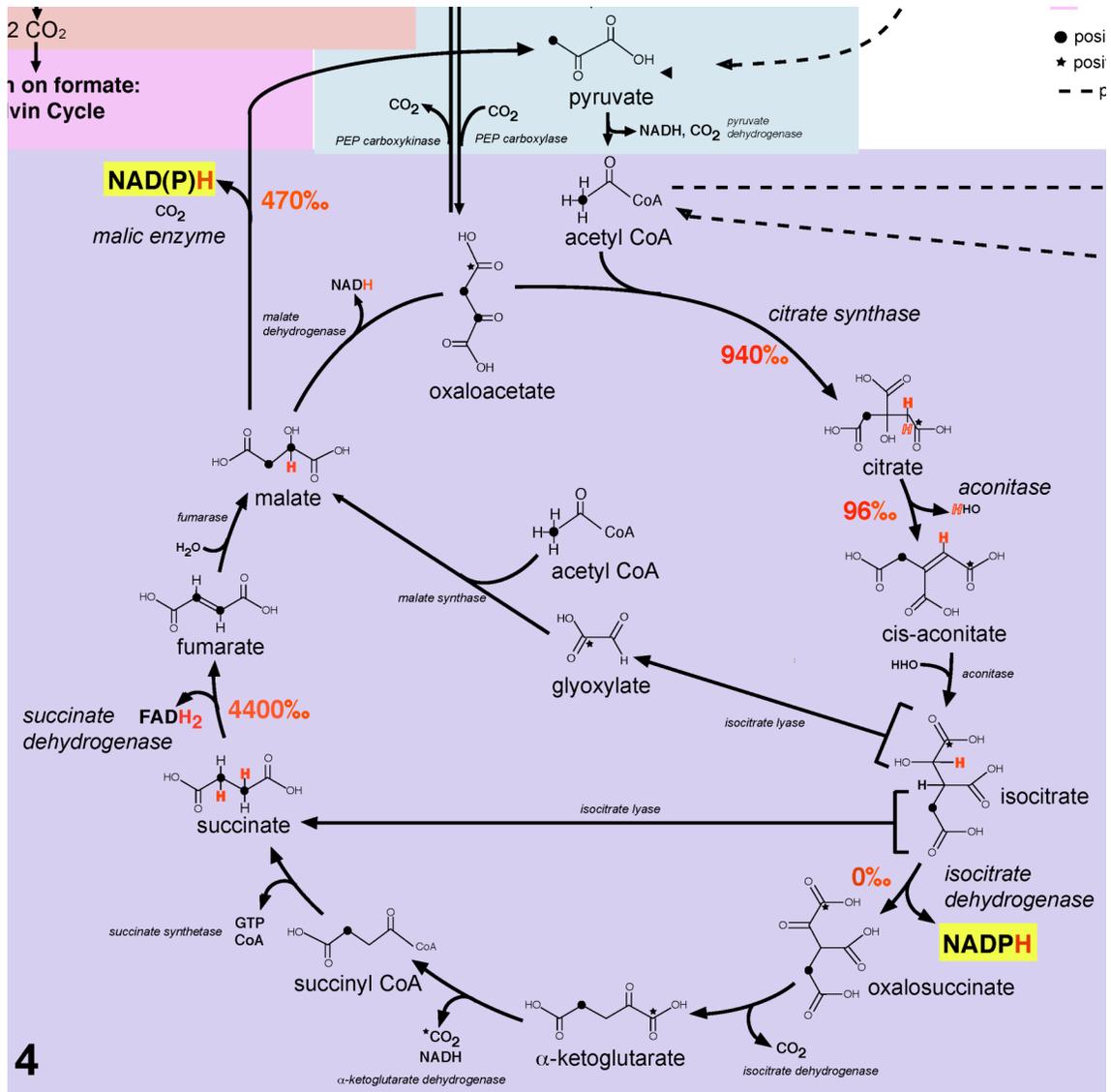
**Fig. 5.9.** Panel detail of pathway 1 (Fig. 5.8): Embden–Meyerhoff–Parnas Pathway (glycolysis).



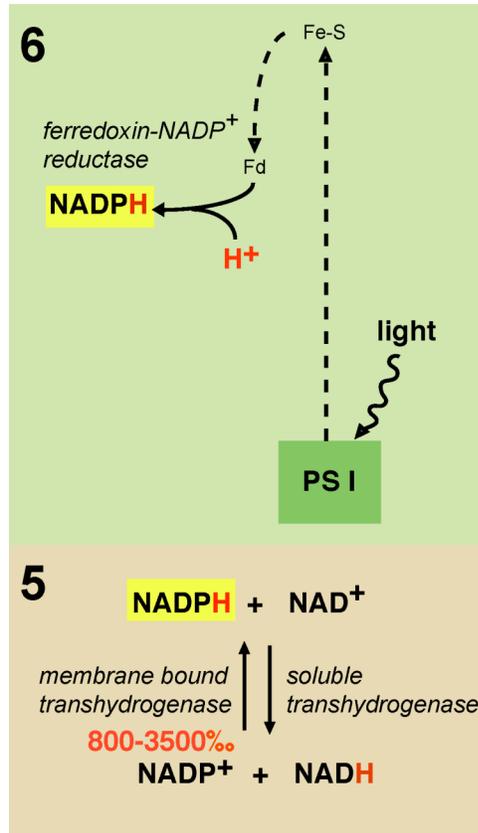
**Fig. 5.10.** Panel detail of pathway 2 (Fig. 5.8): Oxidative pentose phosphate pathway and Entner-Doudoroff pathway (sugar degradation, glycolysis).



**Fig. 5.11.** Panel detail of pathway 3 (Fig. 5.8): Oxalate degradation and assimilation via the glycerate pathway.



**Fig. 5.12.** Panel detail of pathway 4 (Fig. 5.8): Tricarboxylic acid cycle and glyoxylate shunt.



**Fig. 5.13.** Panel detail of pathway 5 and 6 (Fig. 5.8): Transhydrogenase and Light reactions of photosynthesis.

## Conclusions

This thesis presents studies on two topics in environmental microbiology: formate dehydrogenase gene diversity in lignocellulose-feeding insect gut microbial communities and the biological determinants of hydrogen isotope composition in bacterial lipids. Studies made on formate dehydrogenase gene diversity have implications for autotroph microbiology, the global carbon budget, and the evolutionary biology of symbiotic gut microbes. Studies of the biology underlying bacterial lipid hydrogen isotope content enhance knowledge of the basic biology behind cell composition and the impacts microbes have on their surroundings. I discuss specific conclusions for the topics separately in the remainder of this section.

### *Topic I. Formate dehydrogenase gene diversity in lignocellulose-feeding insect gut microbial communities*

My research on formate dehydrogenase genes aims to clarify the microbial ecology of symbiotic acetogenic spirochetes inhabiting the guts of lignocellulose-feeding insects (Chapters 2-4). These bacteria perform an activity that provides significant benefit to the nutrition of their insect host and, by way of insect abundance, also impact the global carbon cycle. Most acetogenic spirochetes in lignocellulose-feeding insects belong to the genus *Treponema*. They are genetically diverse and likely occupy different environmental niches in termite guts. However, the nature of the niches and metabolisms associated with uncultured acetogenic spirochetes are unclear.

The work presented herein aims at revealing novel aspects of acetogenic spirochete microbial ecology. In this thesis, I describe studies in which I used the gene sequence for hydrogenase-linked formate dehydrogenase (FDH<sub>H</sub>), a key enzyme in the acetogenic metabolism of the termite gut isolate, *Treponema primitia*, to explore the diversity, evolution, and activity of uncultured acetogenic spirochetes.

In Chapter 2, I used novel degenerate primers to establish that FDH<sub>H</sub> genes are diverse, are encoded by uncultured acetogenic spirochetes, and can be broadly classified into selenium-dependent (Sec) and selenium-independent (Cys) enzyme clades in evolutionarily primitive wood-feeding insects (i.e, lower termites and a wood-feeding roach). Phylogenetic patterns imply that acetogenic spirochete communities existed in the last common ancestor of wood-feeding termites and roaches, and, moreover, harbored genes for both Sec and Cys enzyme variants. These results provide the first wide-scale evidence suggesting that selenium, a trace nutrient, may play a long-term role in shaping the genetic and metabolic capacities of diverse acetogenic spirochetes, and, as a consequence, impact acetogenesis, the primary H<sub>2</sub> sink in lignocellulose-feeding insect guts.

In Chapter 3, I present the diversity of FDH<sub>H</sub> genes in higher termites, the most species rich and numerically abundant group of termites on earth, and show evidence for major evolutionary shifts within gut communities during termite evolution. Phylogenetic analysis indicates only a single lineage of Sec FDH<sub>H</sub> is present in higher termites; all Cys clade and most Sec clade FDH<sub>H</sub> genes were apparently lost from the FDH<sub>H</sub> gene pool in

higher termite gut microbial communities. I also discovered multiple instances of selenium-independent  $\text{FDH}_H$  reinvention in two higher termite species. Finally, I identified a novel  $\text{FDH}_H$  group specific to termites that have lifestyles characterized by high soil exposure. Taken together,  $\text{FDH}_H$  phylogeny shows strong evolutionary trends that are consistent with an evolutionary bottleneck, convergent evolution, and recent symbiont invasion/acquisition having occurred in gut communities during the evolution of higher termites. I hypothesize that the extinction of cellulolytic protists in a progenitor higher termite may be an important determinant of gut community structure and gene content in extant higher termites.

In Chapter 4, I present a study in which I utilize gene inventory, high-throughput sequencing, and microfluidic digital PCR techniques to identify spirochetes responsible for significant proportions of  $\text{FDH}_H$  gene transcription in the gut microbial community of a lower wood-feeding termite. Transcriptional assessments indicate two *fdhF* phylotypes account for the majority of  $\text{FDH}_H$  gene transcription in the termite gut. I then use microfluidic digital PCR to discover the specific 16S rRNA ribotypes of organisms encoding important *fdhF* phylotypes. The results from this study (i) imply acetogenesis in termite guts may be largely driven by a few species of acetogenic spirochete and (ii) provide a framework for more targeted environmental transcription and single cell analyses of important uncultured termite gut acetogens.

Studies of formate dehydrogenase gene diversity described in this thesis provide a platform for future investigations of the microbiology underlying termite gut acetogenesis, a globally relevant process. I outline two possible studies below:

(i) The results from Chapter 2 suggest that variations in selenium concentration may influence transcription and acetogenesis rates in termite gut microbial communities. This hypothesis can be tested with filter paper feeding experiments (paper is dosed with different amounts of selenium and fed to termites), traditional  $^{14}\text{CO}_2$  fixation assays, and RNA-Seq or other transcriptomic techniques. As shown in Chapter 4, microfluidics can then be used to elucidate the ribotypes of the most transcriptionally active bacteria for each selenium treatment. 16S rRNA sequence data then enable further targeted studies that can include FISH and whole genome amplifications of important gut bacteria.

(ii) Studies of *T. primitia* and termite gut communities suggest selenium availability may be limited in termite guts. Possible influences of availability in guts include selenium concentration and redox state. With regard to concentration, the trace element content of woody biomass should be fairly depleted relative to other forms of plant biomass. This suggests that dietary selenium may pose a challenge for selenium utilizing termite gut microbes. Redox state may be another factor influencing availability. Selenium appears in nature in a variety of redox states ( $\text{Se}^{2-}$  to  $\text{Se}^{6+}$ ), some of which have very low biological availability (e.g., iron selenides, elemental Se). The presence of steep radial redox gradients in termite guts therefore implies selenium redox state, and thus bioavailability, may vary spatially and impact the activity (and genome content) of motile

microbes. A follow-up study could explore these hypotheses. Selenium levels and redox state in food substrates and termite hindguts could be determined with inductively coupled plasma mass spectroscopy and micro-X-ray absorption spectroscopy. These measurements should yield insight on why both Sec and Cys FDH gene variants are present in lower termite guts and why most higher termite guts only harbor Sec FDH variants.

*Topic II. Hydrogen isotope content of bacterial lipids*

The second topic of this thesis focuses on elucidating the biological basis for hydrogen isotope content in lipids. I show in Chapter 5 that lipid D/H varies systematically with different pathways of central metabolism in bacteria. I propose lipid D/H is controlled by NADPH, a key metabolite used for lipid biosynthesis, and the different pathways by which NADPH is synthesized in cells. This hypothesis can be tested with cultures of bacterial mutants in future studies. If such studies support NADPH production as a key determinant of lipid D/H, lipid D/H may constitute an isotopic marker for energy metabolism and prove as useful to microbiologists, geobiologists, and organic geochemists as  $^{13}\text{C}$ -based indicators of carbon fixation.