

HYDROGENASES AND HYDROGEN
SENSORS IN THE SYMBIOTIC
MICROBIAL COMMUNITIES OF WOOD-
FEEDING TERMITES

Thesis by

Nicholas R. Ballor

In Partial Fulfillment of the Requirements for the

degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2011

(Defended August 16, 2010)

© 2011

Nicholas R. Ballor

All Rights Reserved

ACKNOWLEDGEMENTS

I am grateful to everyone who has helped me, in any way, to reach this climactic point in my education. I do not know where to begin. The list is endless. It seems any attempt at thanks is destined for oversight and inadequacy, but here goes.

I can't say I thought I'd be working with termites when I left Michigan Tech. But, here I am, a fan of a biological system among the most fascinating and relevant microbial symbioses and bioreactors known in nature. I am thankful to Jared Leadbetter for allowing me to join his lab and discover this little world. It has been a wonderful journey. I am also thankful for his holistic philosophy that one's legacy encompasses more than his publication record.

A big "thank you" to my co-workers at Kimberly-Clark, especially Dave Fell and Amy Weinheimer, for inspiring me to go a bit deeper into the rabbit hole. And to the folks at the University of South Carolina and Montana State University's Center for Biofilm Engineering, especially Robin Gerlach, for allowing me to try my hand at academic research.

I am indebted to Donald Lueking for taking me under his wing at Michigan Tech. His confidence in me, and the amazing opportunities he entrusted to me, was instrumental to mobilizing me as a scientist. I am so blessed to have known this professor at the peak of his outstanding teaching career in molecular biochemistry. I am also thankful to Susan Bagley and John Adler from Michigan Tech. They were a tremendous help as I ventured to transition from the world of chemical engineering to the biological sciences. I also want to give a nod to the chemical engineering department at Michigan Tech. for helping me

weed my way through the process of forming a solid intellectual foundation and selflessly making education a first priority.

My labmates throughout the years have helped me through the daily grind. My professional development has been a community endeavor. I am grateful to all who had the patience to help me along the way.

Of course, I would be remiss not to thank my parents and family. They have been tremendously supportive through the years. They have always encouraged me to go beyond my comfort zone and to place high value on reason and education.

I am also grateful to the many friends who have helped me along the way.

This research was partially supported by an NSF Graduate Student Research Fellowship.

ABSTRACT

The termite gut is an ideal ecosystem for studying hydrogen ecophysiology. Hydrogen is central to the obligate mutualism between termites and their gut microbes and is turned over at rates as high as $33 \text{ m}^3 \text{ H}_2$ per m^3 hindgut volume daily and maintained near saturation in some species. Acetogenic bacteria use hydrogen to produce up to 1/3 of the total flux of the termite's primary carbon and energy source, acetate. We have taken a three-fold approach to investigate the hydrogen ecophysiology of the termite gut. In our first approach (Chapter 2) we completed a bioinformatic analysis of [FeFe] hydrogenase-like (H domain) proteins encoded in the genomes of three termite gut treponemes. Treponemes are among the most highly represented groups of gut bacteria. The remarkable diversity of H domain proteins encoded accentuates the importance of hydrogen to their physiology. Moreover, they encoded a poorly understood class hydrogen sensing H domain proteins and thereby present a unique opportunity for their further study. In our second approach (Chapters 3 and 4) we analyzed molecular inventories prepared from termite gut microbiomes of a class of [FeFe] hydrogenases found highly represented in a termite hindgut metagenome. The libraries of peptide sequences clustered with one another in a manner congruent with termite host phylogeny suggesting co-evolution. Interestingly, we observed that higher termite guts may harbor higher sequence diversity than lower termites. In our third approach (Chapter 5) we used microfluidic digital PCR to identify bacteria in the gut of *Reticulitermes tibialis* encoding [FeFe] hydrogenases. The majority of the 16S rRNA gene phlotypes observed to co-amplify with hydrogenase sequences were treponemal, and the only observed instances of the same 16S rRNA-hydrogenase gene pair co-amplifying in multiple microfluidic chambers corresponded to treponemal phlotypes. Therefore, treponemes may be an important or predominant bacterial group encoding an important family of [FeFe] hydrogenases in the termite gut. The above results provide support for an important role for treponemes in mediating hydrogen metabolism in the termite gut and accentuate the intimacy and stability of the association termites have maintained over the course of their evolution with their gut microbial communities.

TABLE OF CONTENTS

Acknowledgements	iii
Abstract	v
Table of Contents	vi
List of Figures	viii
List of Tables	xi
Chapter 1: Background	1-1
Hydrogen in the Termite Gut.....	1-1
Hydrogenases.....	1-4
Levels of Physiological Resolution to Study H Domain Proteins.....	1-10
Termite Species.....	1-12
Termite Gut Treponeme Isolates.....	1-15
References.....	1-17
Chapter 2: Genomic Analysis Reveals Multiple [FeFe] Hydrogenases and Hydrogen Sensors Encoded by Treponemes from the Hydrogen Rich Termite Gut	2-1
Abstract.....	2-1
Introduction.....	2-2
Methods.....	2-4
Results.....	2-6
Discussion.....	2-20
References.....	2-28
Chapter 3: A Phylogenetic Analysis of [FeFe] Hydrogenase Gene Diversity in the Hydrogen Metabolizing Guts of Lower Termites and Roaches Reveals Unique, Ecosystem-Driven, Adaptations and Similarity of <i>Cryptocercus</i> and Lower Termite Gut Communities	3-1
Abstract.....	3-1
Introduction.....	3-3
Methods.....	3-5
Results.....	3-10
Discussion.....	3-14
References.....	3-23
Appendix.....	3-31
Chapter 4: Analysis of [FeFe] Hydrogenase Sequences from the Hydrogen Rich Guts of Higher Termites Reveals Correlations Between Gut Ecosystem Parameters and Sequence Community Composition	4-1
Abstract.....	4-1
Introduction.....	4-3
Methods.....	4-5

Results	4-8
Discussion	4-15
References	4-22
Appendix	4-30

Chapter 5: Microfluidic Digital PCR Reveals that Treponemes May be a Predominant Genus of Bacteria Encoding an Important Family of [FeFe] Hydrogenases in the Gut of *Reticulitermes tibialis*..... 5-1

Abstract	5-1
Introduction	5-3
Methods	5-5
Results	5-15
Discussion	5-24
References	5-28
Appendix	5-38

LIST OF FIGURES

<i>Number</i>	<i>Page</i>
1-1	General scheme underlying symbiosis between anaerobes and wood-feeding lower termites 1-2
1-2	Representative biologically produced hydrogen partial pressures 1-3
1-3	Distribution of hydrogen partial pressures in the gut of <i>Reticulitermes flavipes</i> 1-5
1-4	H domain and its conserved sequence signatures..... 1-6
1-5	Model for the sensory [NiFe] hydrogenases of <i>Alcaligenes eutrophus</i> and <i>Rhodobacter capsulatus</i> 1-9
1-6	Gene organization and domain composition of multimeric [FeFe] hydrogenases 1-11
1-7	Phylogram of termite families and wood roaches 1-13
1-8	Higher and lower termite gut structures..... 1-15
1-9	Phase-contrast microscopy images of <i>T. primitia</i> and <i>T. azotonutricium</i> 1-16
2-1	Domain architectures representative of each family of H domain proteins observed in <i>Treponema primitia</i> strains ZAS-1 and ZAS-2 and <i>T. azotonutricium</i> ZAS-9 2-10
2-2	Phylogeny of H domain peptide sequences from putative hydrogen sensor and [FeFe] hydrogenase proteins 2-12
2-3	Phylogeny of H domain peptide sequences from putative [FeFe] hydrogenase proteins..... 2-14
2-4	Phylogeny of H domain peptide sequences from [FeFe] hydrogenase Family 6 and putative H ₂ sensor proteins 2-16
2-5	Multimeric hydrogenase gene clusters 2-18
3-1	Collector's curves for lower termite samples..... 3-12

3-2	Collector's curves for roach samples.....	3-13
3-3	Phylogram of Family 3 [FeFe] hydrogenases cloned from the guts of lower termites.....	3-15
3-4	Phylogram of Family 3 [FeFe] hydrogenases cloned from the guts of an Adult and Nymph <i>C. punctulatus</i> samples	3-16
3-5	Maximum likelihood tree of all cloned Family 3 [FeFe] hydrogenase sequences	3-17
3-6	Unifrac jackknife clustering of all cloned Family 3 [FeFe] hydrogenase sequences	3-18
3-7	Unifrac principle components analysis of all cloned Family 3 [FeFe] hydrogenase sequences.....	3-19
3-S1	Alignment used to design degenerate primers to amplify Family 3 [FeFe] hydrogenases.....	3-34
4-1	Collector's curves.....	4-10
4-2	Phylogram for Family 3 [FeFe] hydrogenases cloned from the guts of higher termites.....	4-12
4-3	Phylogram comparing Family 3 [FeFe] hydrogenases cloned from higher termites to sequences cloned previously from <i>C. punctulatus</i> and lower termites	4-13
4-4	Unifrac jackknife analysis of Family 3 [FeFe] hydrogenase sequences cloned from higher termites, lower termites, and <i>C. punctulatus</i>	4-14
4-5	Unifrac principle component analysis of Family 3 [FeFe] hydrogenase sequences cloned from the guts of higher termites, lower termites, and <i>C. punctulatus</i>	4-16
4-6	Unifrac principle component analysis of Family 3 [FeFe] hydrogenase sequences cloned from higher termites in this study.....	4-17
5-1	Mitochondrial cytochrome oxidase II (COII) phylogeny of the termite sample used in this study.....	5-8

5-2	Representative heat map for a digital PCR chip panel from which amplicons were retrieved for analysis.....	5-17
5-3	Phylograms of 16S rRNA gene sequences (left panel) and putative Family 3 [FeFe] hydrogenase peptide sequences (right panel) corresponding to products co-amplified in dPCR..	5-19
5-4	Phylogram of non-treponemal 16S rRNA gene sequences co-amplified with Family 3 [FeFe] hydrogenase genes in dPCR....	5-23
5-S1	Alignment used to design degenerate primers for use in digital PCR	5-42
5-S2	Alignment used in the design of locked nucleic acid (LNA) probes for use in digital PCR.....	5-43

LIST OF TABLES

<i>Number</i>	<i>Page</i>
2-1 FeFe hydrogenase-like proteins observed in the genomes of three termite gut isolates	2-8
2-2 Putative hydrogenase-like proteins encoded by three termite gut spricochetes, a termite hindgut metagenome and several reference bacteria	2-9
3-1 Quantifying hydrogenase clone library diversity	3-11
3-S1 Sequences cloned	3-32
4-1 Quantifying hydrogenase clone library diversity	4-9
4-S1 Sequences cloned	4-31
5-1 Primers and probes used.....	5-7
5-2 Digital PCR chip wells from which amplicons were retrieved...	5-18
5-3 <i>Reticulitermes</i> environmental genomovars proposed in this study.....	5-22
5-S1 Sequences cloned and proposed <i>Reticulitermes</i> environmental genomovars (REG).....	5-39

Chapter 1

BACKGROUND

Hydrogen in the Termite Gut

Hydrogen is of central importance to the symbiotic community residing in the termite gut (1, 5, 15, 35, 36, 41). This obligate mutualism, in many instances comprising microbes from all three domains of life, enables termites to derive carbon and energy from wood (2-4, 6, 11, 12, 17, 19, 20, 35, 36, 56, 61, 62).

The general scheme underlying the symbiosis between wood-feeding termites and their gut microbes is shown in Figure 1-1. Termites ingest wood and microbial symbionts then ferment its component polysaccharides to produce primarily acetate, carbon dioxide, and hydrogen (1, 5, 12, 19, 35-37, 41, 54, 56, 62). Acetate is absorbed by the termite and used for energy and biosynthesis (37). The carbon dioxide and hydrogen produced in this initial fermentation are used by bacteria in reductive acetogenesis to produce more acetate— up to 1/3 of the total pool in the gut (1, 5, 27, 37, 41). With these high rates of reductive acetogenesis, the termite gut is “the smallest and most efficient natural bioreactor currently known” (41). Only a small portion of the hydrogen is used by methanogens to produce methane (1, 25, 41). The total daily production of hydrogen in this environment can be up to 9-33 m³ H₂ per m³ hindgut volume (41). Hydrogen partial pressures have been measured in the termite gut that exceed those measured for any other biological system (15, 18, 41, 45, 47, 49, 53). Figure 1-2 presents a comparison of partial pressures of hydrogen measured in representative microbial communities.

A complex matrix of microenvironments characterized by different hydrogen concentrations are maintained in the termite gut (6-8, 15, 25, 26, 41). Hydrogen partial

Figure 1-1.

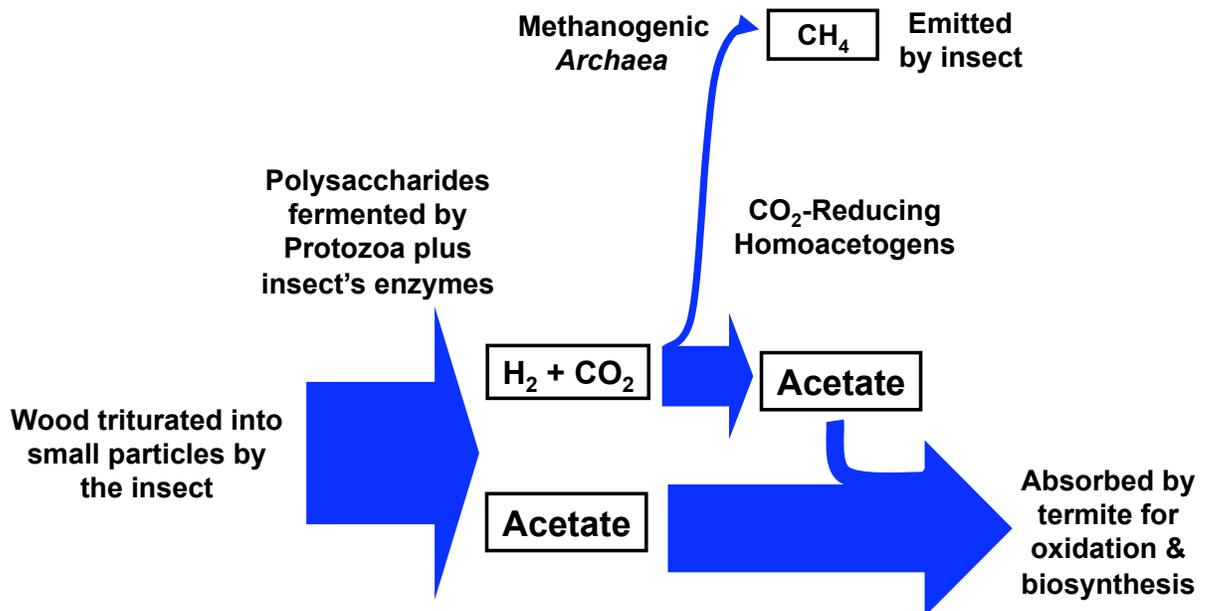


Figure 1-1. General scheme underlying symbiosis between anaerobes and wood-feeding lower termites. Termites consume wood and break it down into small particles. The particles are degraded and their component polysaccharides fermented by the symbiotic microbial community in the termite gut (12, 19, 56, 62). The acetate produced in this fermentation is absorbed by the termite and used for respiration and biosynthesis (37). The H₂ and CO₂ formed in the initial fermentation is used primarily by homoacetogenic bacteria in reductive acetogenesis to produce more acetate (1, 5, 41). A small fraction of the H₂ and CO₂ is used by methanogenic archaea to produce methane that is emitted by the insect (41, 52, 54). Figure was kindly provided through a private communication by Jared R. Leadbetter.

Figure 1-2.

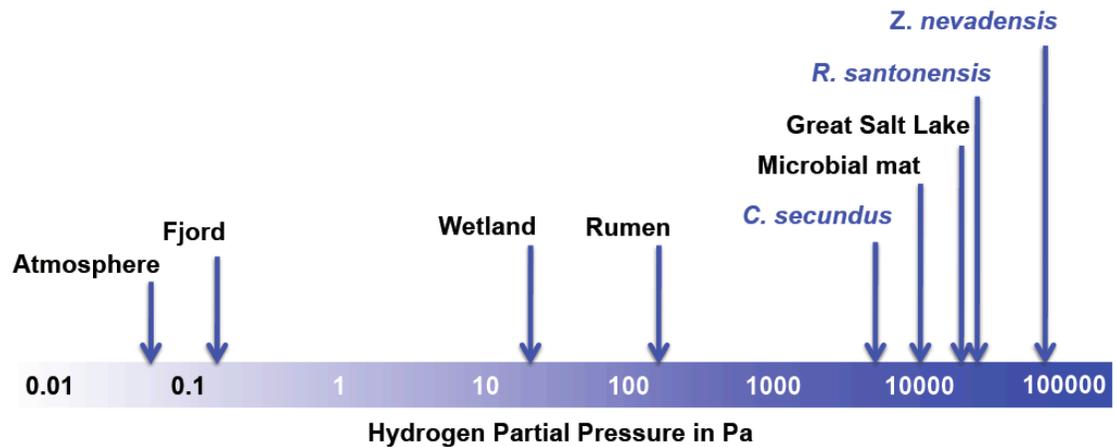


Figure 1-2. Representative biologically produced hydrogen partial pressures. The partial pressure of hydrogen in the atmosphere is given as a reference. Hydrogen pressures were measured in the bubble gas of a natural wetland (53); in a surface layer of intertidal mats dominated by *Lyngbya* spp. (18); in the hindguts of termites, shown in blue (41); in the open waters of Saanich Inlet, an anoxic fjord in British Columbia (47); in the rumen of a steer (49); and in the pycnocline of the Great Salt Lake in Utah (45).

pressures in the guts of termites have been found to vary dramatically with position, see Figure 1-3 (15, 41). Hydrogen partial pressure reaches a maximum in the hindgut paunch, see Figures 1-3A, and decreases as you approach the axial extremities of the gut, see Figure 1-3B (7). Hydrogen partial pressures also vary radially being highest at the center of the gut and decreasing symmetrically to near zero at the epithelium(7).

The work presented in the following chapters addresses questions remaining unanswered about genes encoding proteins that eubacteria use for producing, consuming, and monitoring hydrogen in the termite. The objective has been to advance our understanding of the nature of the symbiosis in the termite gut. Study of the rich diversity of hydrogenase-like proteins (H domain proteins) endemic to the termite gut is a unique opportunity to provide substantial contributions to our understanding of these proteins.

Hydrogenases

H domain proteins are used to make, break, or sense hydrogen. Hydrogenases catalyze the following reaction:

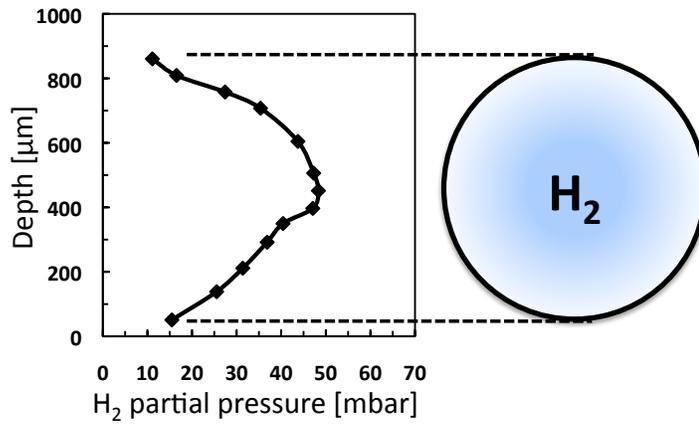


There are four major classes of hydrogenases – all named according to the metal composition of their catalytic sites: the evolutionarily related nickel iron (NiFe) hydrogenases and nickel iron selenium (NiFeSe) hydrogenases, and two evolutionarily distinct classes called [FeFe] hydrogenases and metal-free hydrogenases, for a review see Schwartz *et al.* (46).

The structures of two [FeFe] hydrogenases, CpI from *Clostridium pasteurianum* (42) and the heterodimeric [FeFe] hydrogenase from *Desulfovibrio vulgaris* (32), have been solved, and the catalytic site, or H cluster, of CpI from *C. pasteurianum* is shown in Figure 1-4.

Figure 1-3.

A.



B.

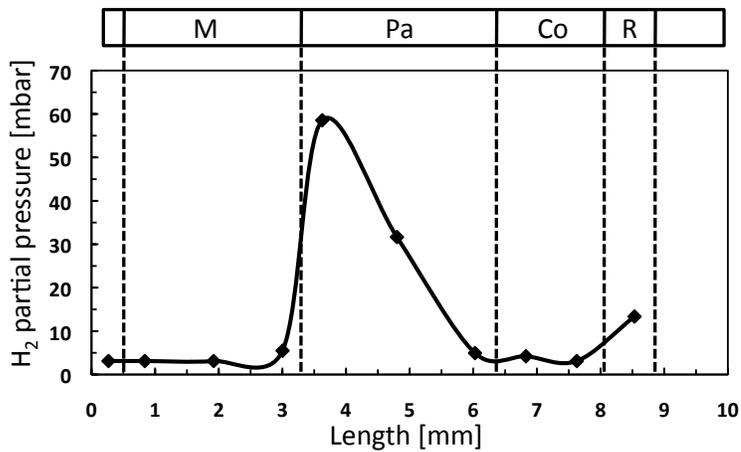
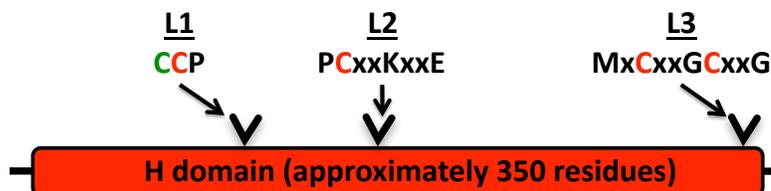


Figure 1-3. Distribution of hydrogen partial pressures in the gut of *Reticulitermes flavipes*. (A) Radial distribution of hydrogen in the hindgut paunch of *Reticulitermes flavipes* with an image illustrating how hydrogen concentrations are maximal at the center of the hindgut paunch (dark blue) and diminish symmetrically (fading blue) toward the epithelium. (B) Distribution of hydrogen partial pressures along the axis of the gut of *Reticulitermes flavipes*. M – midgut; Pa – hindgut paunch; Co – colon; R – rectum. Measurements were taken using a microsensors. Figure is based upon a figure created by Ebert and Brune (15), which has been borrowed with permission.

Figure 1-4.

A.



B.

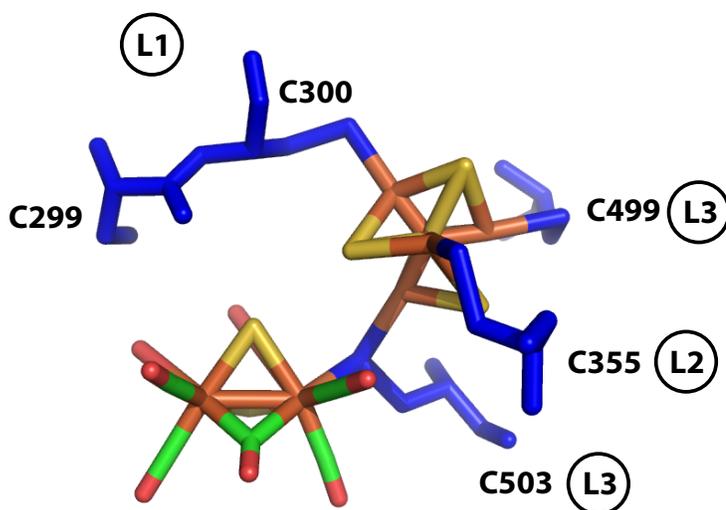


Figure 1-4. H domain and its conserved sequence signatures. (A) [FeFe] hydrogenases can be identified by the three conserved sequence signatures in their H domain (31). Each signature contains cysteine residues essential for catalysis (31). The cysteines in red are involved in coordinating the [4Fe-4S] cluster or bridging the cluster to the 2Fe cluster (31). The cysteine in green is believed to act as an acid/base in catalysis (31). (B) The H domain of *C. pasteurianum*. The cysteines of the H domain sequence signatures are in blue and the name of the sequence signature to which each belongs is indicated. The image was prepared using MacPyMOL and structure 1feh from the PDB database. C299 may participate in catalysis as an acid/base. C300, C499, C355 and C503 coordinate the [4Fe-4S] cluster domain (31). C503 bridges the [4Fe-4S] cluster to the 2Fe cluster domain (31). The 2Fe cluster is coordinated by CO and CN ligands and the two atoms are bridged by a carbon monoxide atom (31).

The cluster is made of a diatomic cluster of two iron atoms bridged by a cysteine to a [4Fe-4S] iron sulfur cluster (32, 42). These two iron atoms interact directly with hydrogen and are the namesake of this class of hydrogenases (42). Three proteins, or maturases, HydE, HydG and HydF, are necessary for assembling the H cluster of [FeFe] hydrogenases (24).

An analysis of hydrogenase-like proteins encoded in a termite gut metagenome sequence revealed that the vast majority are H domain proteins (60). It is unclear why only one of the over 100 hydrogenases identified was a [NiFe] hydrogenase. This may be a consequence of the high specific molar activities of [FeFe] hydrogenases (16). The metagenome paper provided initial experimental evidence that the termite gut microbial community is a rich reservoir of H domain proteins.

The multitude and diversity of [FeFe] hydrogenases observed in the termite gut metagenome enabled the definition of families of H domain proteins based upon phylogenetic and primary sequence character analyses. This was the first effort to classify these proteins based upon evolutionary relationships. Because of their relevance to the termite hindgut and evolutionary significance, these family designations have been used in the following chapters. Classifications based upon sequence characteristics have also been proposed by others (31, 57).

H domain proteins can be identified by three conserved sequence signatures, see Figure 1-4 (31, 57). Sequences containing these signatures are likely to be H domain proteins (31, 57).

The signatures contain cysteine residues that are essential to the H domain coordinating the H cluster (31, 32, 42).

H domain proteins contain domains that augment the function of the H domain in catalysis. H domain proteins typically have several iron sulfur cluster coordinating sites that mediate electron transfer (31, 32, 42, 57). The most common iron sulfur clusters are [Fe-S], [2Fe-2S], and [4Fe-4S] (31, 57). H domain proteins commonly contain two closely spaced, consecutive [4Fe-4S] clusters near their N-terminus (31, 57). Iron sulfur clusters are usually coordinated by cysteine residues, but it is common to find near the N-terminus of a hydrogenase that the 1st cysteine of a [4Fe-4S] has been replaced by a histidine (31, 57). Some H domain proteins contain domains not involved in electron transfer that, instead, may couple behavioral or transcriptional modifications to hydrogen levels.

Prior to the sequencing of a termite gut metagenome, few (43, 48, 60, 61) H domain proteins were proposed to contain domains normally implicated in cell signalling, for a review see Schwartz *et al.* (46). The only characterized hydrogen sensors are [NiFe] hydrogenase homologues including, most notably, those from *Alcaligenes eutrophus* and *Rhodobacter capsulatus* (9). These two proteins, as shown in Figure 1-5, are components of two-component regulatory systems involved in transcriptional regulation (14, 29). The *Nasutitermes* termite hindgut metagenome paper reported a multitude of H domains fused with domains usually implicated in signal transduction (60). These domains include the PAS domain that, in bacterial systems, is most often found in sensors of two-component regulatory systems, as in the *A. eutrophus* and *R. capsulatus* sensory hydrogenases (55). A response regulator receiver domain typically found in proteins participating in phosphorelays or two component regulatory systems (39, 50, 51) was also observed in some sequences. The final sensory domain observed was the methyl-accepting chemotaxis protein domain that may function in regulating bacterial swimming behavior (59). The

Figure 1-5.

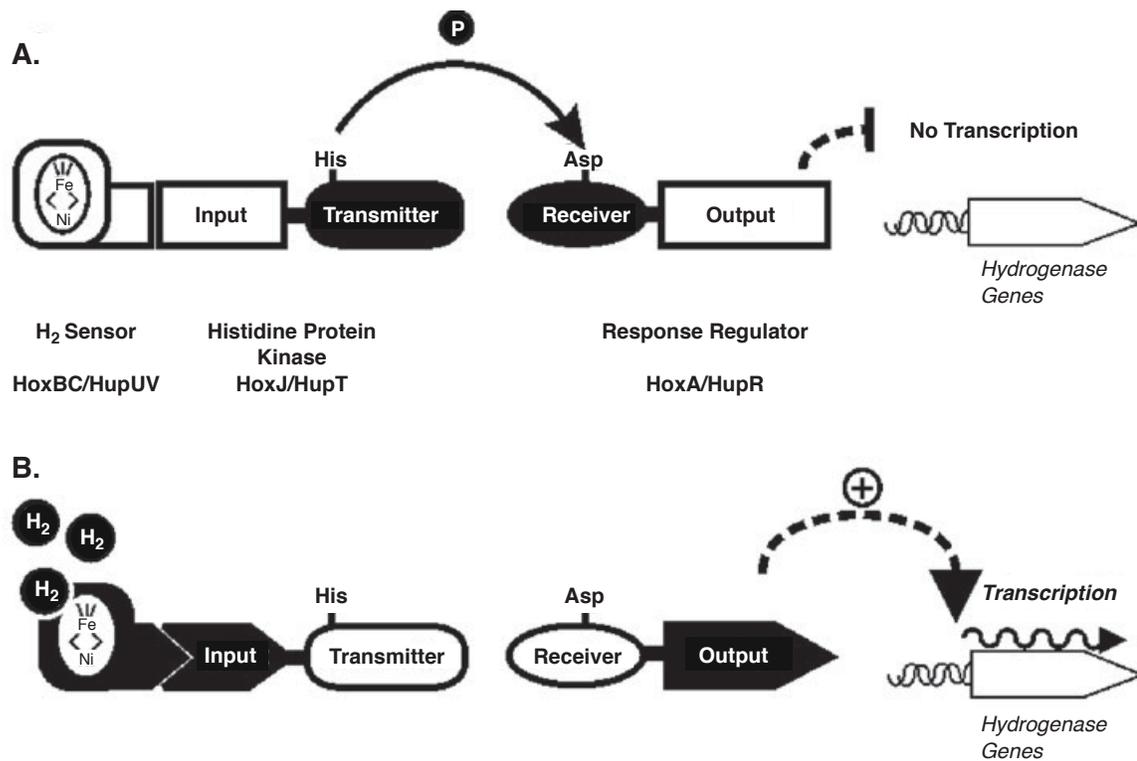


Figure 1-5. Model for the sensory [NiFe] hydrogenases of *Alcaligenes eutrophus* and *Rhodobacter capsulatus*. The sensory [NiFe] hydrogenases are part of a two-component regulatory system. In the absence of hydrogen (A) the sensing system suppresses transcription of target genes, and in the presence of hydrogen (B) transcription of target genes is activated. The input module is a PAS domain. The output module is a transcriptional activator. The transmitter and receiver modules are the canonical histidine-kinase and response regulator receiver modules of two-component regulatory systems. In the presence of hydrogen, the [NiFe] hydrogenase oxidizes hydrogen. The PAS domain then senses a change in electron potential and communicates this signal to the transmitter domain. Figure is taken from Schwartz with permission (46).

discovery of this domain in a putative H domain protein was intriguing because chemotaxis toward hydrogen has not been demonstrated experimentally. Shaw *et al.* have recently reported a PAS domain containing H domain protein from *Thermoanaerobacterium saccharolyticum* and some clostridia encode similar proteins and Posewitz *et al.* have reported proteins in *Halothermothrix orenii* with a region sharing homology simultaneously with PAS and histidine kinase domains (10, 43, 48). The discovery of a multitude of putative sensory H domain proteins in a termite's gut metagenome supports the hypothesis that termites are a rich reservoir of novel [FeFe] hydrogenase homologues and that their study may provide insight into the functional diversity and evolution of this class of proteins.

Some [FeFe] hydrogenases have heteromeric quaternary structures (31, 57, 58). The best studied multimeric [FeFe] hydrogenases are the trimeric complex from *Thermotoga maritima* and the tetrameric complexes from *Thermoanaerobacter tengcongensis*, and *Desulfovibrio fructosovorans*, see Figure 1-6 (13). These complex hydrogenases are believed to couple the oxidation or reduction of NAD(P)(H) to hydrogen production or consumption, respectively.

Levels of Physiological Resolution to Study H Domain Proteins

H domain proteins may be studied at three interdependent levels of molecular resolution: The level of individual genes or proteins, the level of individual cells or cell genomes, and the level of an entire symbiotic microbial community or metagenome. A gene or protein based analysis of H domain proteins provides the highest level of resolution facilitating an understanding of function and evolution on the molecular level. A higher level of complexity and lower level of molecular resolution may be sought through the study of the

Figure 1-6.

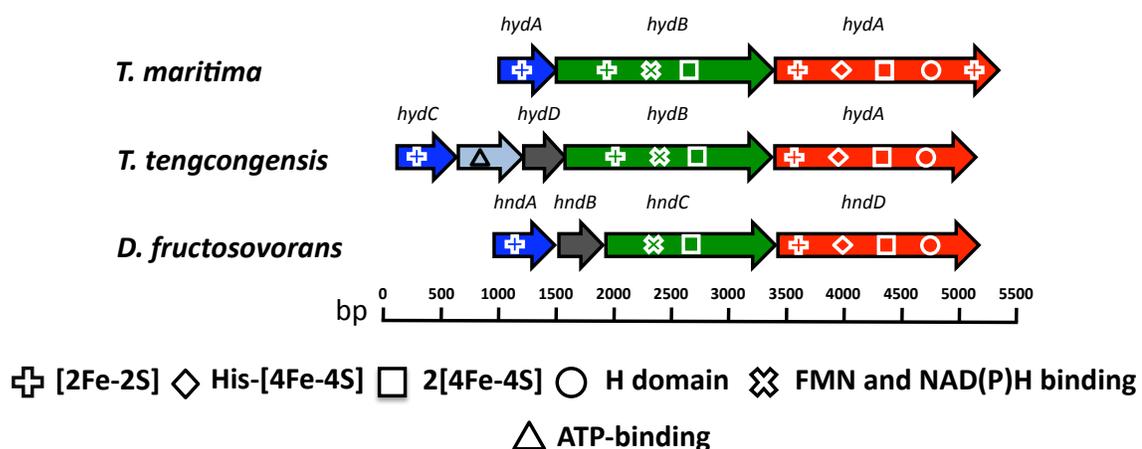


Figure 1-6. Gene organization and domain composition of multimeric [FeFe] hydrogenases. Each arrow represents a gene and homologous genes share the same color. Domain symbols are listed in proper order but are not intended to represent precise locations. Domains were identified using the Pfam server. Domains represented in the figure are: 2[4Fe-4S] – F cluster made up of two adjacent Fer4 domains, PF00037; [2Fe-2S] – PF00111, [2Fe-2S] iron-sulfur cluster binding domain; ATP-binding – PF02518, Histidine kinsae-, DNA gyrase B-, and HSP90-like ATPase; FMN and NAD(P)H binding – PF01512, Respiratory-chain NADH dehydrogenase 51 Kd subunit; H domain – PF02906 and PF02256, iron only hydrogenase large subunit, C-terminal domain, and iron only hydrogenase small subunit; His-[4Fe-4S] – a [4Fe-4S] cluster with the first coordinating cysteine replaced with a histidine.

physiological context of H domain proteins through the study of the genomes of single cells or metabolic and behavioral responses to hydrogen. At the lowest level of molecular resolution H domain proteins are investigated across an entire bacterial community in its native setting, the termite gut. This last level of resolution introduces the most complexity because it accounts for all interactions, environmental and biological, that occur *in situ*. Each level of resolution complements the understanding of the termite gut symbiosis made accessible by the others. The metagenome sequence of a termite (60) has provided an initial glimpse into hydrogenase-like protein function and phylogeny through complementary analyses at the community-wide and individual gene levels of resolution. The studies reported in the following chapters provide examples of insights obtained using all three levels of resolution to advance our understanding of the function and distribution of H domain proteins in the termite gut.

Termite Species

Over 281 genera comprising at least 2600 species of termites are known (23, 28). Based upon evolutionary relationships, these termites may be divided into six families, see Figure 1-7 (28). Members of Termitidae, the largest family of termite species, are referred to as “higher termites.” All other termites are “lower termites.”

The most commonly referenced distinction between higher and lower termites is that the latter have protozoa in their hindgut and the former do not (11). Higher termites also have a more segmented gut structure, see Figure 1-8 (6, 33, 34). The family Cryptocercidae, or wood roaches, is believed to represent the most recent common ancestor of all termites (20, 27, 30). The six families of termites provide a unique opportunity to investigate the representation of H domain proteins across evolutionarily distinct symbiotic communities.

Figure 1-7.

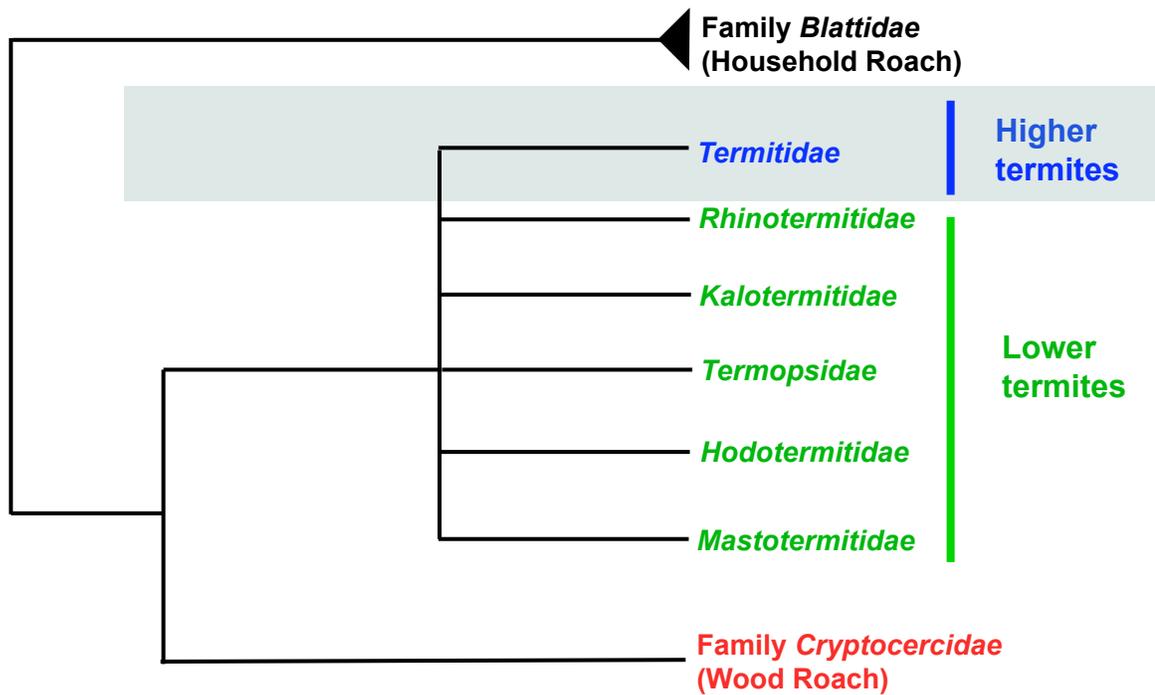


Figure 1-7. Phylogram of termite families and wood roaches. Tree based upon phylogenetic analyses reported by Inward et al. (21, 22).

Figure 1-8.

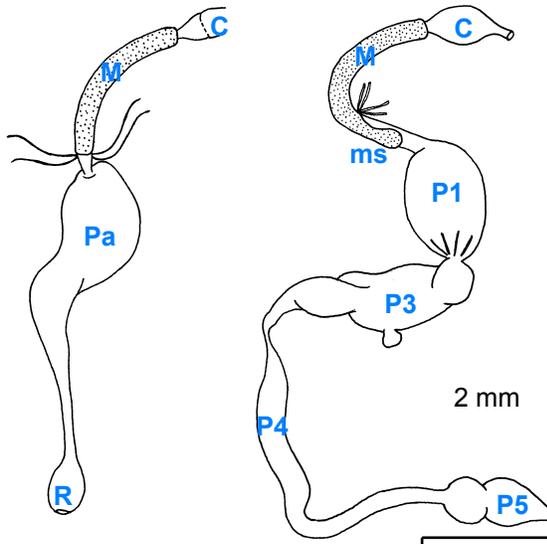


Figure 1-8. Higher and lower termite gut structures. (A) Lower termite, *Reticulitermes*, gut and (B) a higher termite, *Cubitermes*, gut. Pa – Hindgut paunch; R – rectum; C – crop; M – midgut; ms – mixed segment; P1-P5 – proctodeal segments. Images are taken from Brune with permission (6).

The hindgut metagenome has been sequenced for a higher termite, *Nasutitermes*, from Costa Rica (60). Much of the work reported in the following chapters investigates H domain proteins in an untapped and potentially rich reservoir of their sequence diversity, namely lower termites. We have furthered our understanding of factors that may influence the distribution and evolution of hydrogenases in gut communities by cross-comparing representative sets of sequences across termite species. This has advanced our understanding of H domain proteins at the community level of resolution.

Termite Gut Treponeme Isolates

In 1999, Jared Leadbetter was the first to isolate treponemes from the hindgut of a termite, *Zootermopsis angusticolis* (17, 27, 30). Treponemes are helical shaped bacteria belonging to the phylum Spirochaete. Treponemes are among the most abundant groups of bacteria in the guts of termites, constituting up to 50% of the total prokaryotes in some species (6, 40). They may also be a major producer of acetate by reductive acetogenesis (38, 40, 44).

In the following chapter, I will present the results of a bioinformatic analysis of the hydrogenases encoded in the genomes of *Treponema azotonutricium* ZAS-9 and *T. primitia* ZAS-1 and ZAS-2, shown in Figure 1-9. *Treponema primitia* is a hydrogen consuming acetogen (17, 27). *T. azotonutricum* is not an acetogen and, therefore, not believed to be a substantial consumer of hydrogen; rather, it produces hydrogen (17).

These isolates represent a unique opportunity to study hydrogenases in species having distinct and complimentary hydrogen physiologies. Investigating the hydrogenases of these strains has contributed to our understanding of these enzymes at the single cell genome and individual gene or protein levels of resolution.

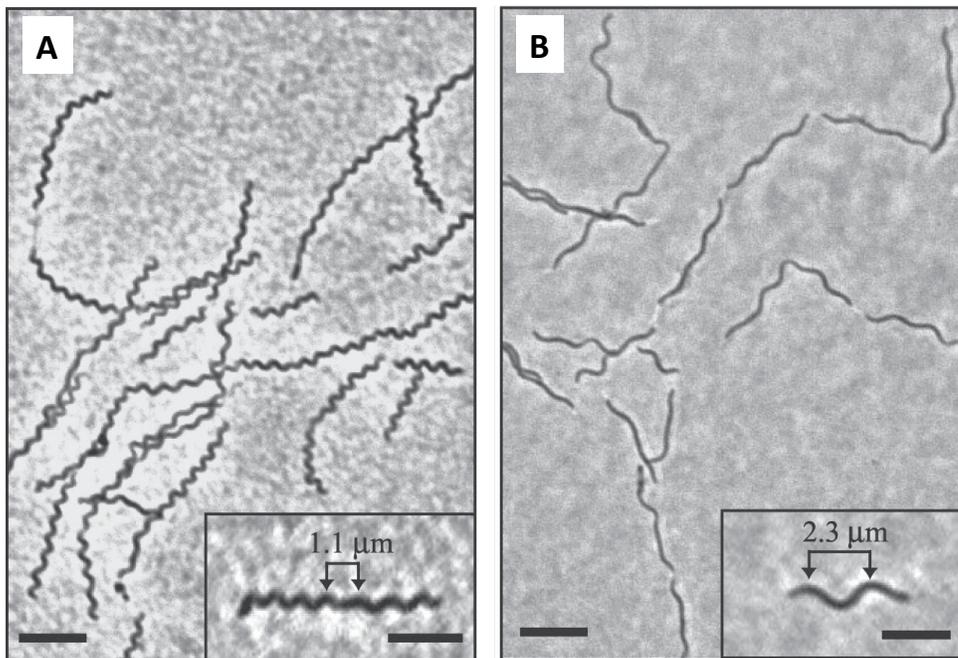
Figure 1-9.

Figure 1-9. Phase-contrast microscopy images of *T. primitia* and *T. azotonutricum*. Insets show single cells of each strain. (A) *T. primitia* ZAS-2, (B) *T. azotonutricum* ZAS-9. Bars, 5 μm for images and 2.5 μm for insets. Images taken from Graber *et al.* with permission (17).

References

1. **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.
2. **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, symbioses, ecology*. Kluwer Academic Publishers, Boston.
3. **Breznak, J. A.** 1982. Intestinal microbiota of termites and other xylophagous insects. *Annu. Rev. Microbiol.* **36**:323-343.
4. **Breznak, J. A.** 1975. Symbiotic Relationships Between Termites and Their Intestinal Microbiota. *Symp. Soc. Exp. Biol.* :559-580.
5. **Breznak, J. A., and J. M. Switzer.** 1986. Acetate Synthesis from H₂ plus CO₂ by Termite Gut Microbes. *Appl. Environ. Microbiol.* **52**:623-630.
6. **Brune, A.** 2006. Symbiotic Associations Between Termites and Prokaryotes. *Prokaryotes* **1**:439-474.
7. **Brune, A.** 1998. Termite guts: the world's smallest bioreactors. *Trends in Biotechnol.* **16**:16-21.
8. **Brune, A., and M. Friedrich.** 2000. Microecology of the termite gut: structure and function on a microscale. *Curr. Opin. Microbiol.* **3**:263-269.
9. **Buhrke, T., O. Lenz, A. Porthun, and B. Friedrich.** 2004. The H₂-sensing complex of *Ralstonia eutropha*: interaction between a regulatory [NiFe] hydrogenase and a histidine protein kinase. *Mol. Microbiol.* **51**:1677-1699.

10. **Calusinska, M., T. Happe, B. Joris, and A. Wilmotte.** 2010. The surprising diversity of clostridial hydrogenases: a comparative genomic perspective. *Microbiology* **156**:157-1588.
11. **Cleveland, L. R.** 1923. Correlation Between the Food and Morphology of Termites and the Presence of Intestinal Protozoa. *Am. J. Hyg.* **3**:444-461.
12. **Cleveland, L. R.** 1923. Symbiosis between Termites and Their Intestinal Protozoa. *Proc. Natl. Acad. Sci. U.S.A.* **9**:424-428.
13. **De Luca, G., M. Asso, J.-P. Bélaïch, and Z. Dermoun.** 1998. Purification and Characterization of HndA Subunit of NADP-Reducing Hydrogenase from *Desulfovibrio fructosovorans* Overproduced in *Escherichia coli*. *Biochemistry* **37**:2660-2665.
14. **Dischert, W., P. M. Vignais, and A. Colbeau.** 1999. The synthesis of *Rhodobacter capsulatus* HupSL hydrogenase is regulated by the two-component HupT/HupR system. *Mol. Microbiol.* **34**:995-1006.
15. **Ebert, A., and A. Brune.** 1997. Hydrogen Concentration Profiles at the Oxidic-Anoxic Interface: a Microsensor Study of the Hindgut of the Wood-Feeding Lower Termite *Reticulitermes flavipes* (Kollar). *Appl. Environ. Microbiol.* **63**:4039-4046.
16. **Frey, M.** 2002. Hydrogenases: Hydrogen-Activating Enzymes. *ChemBioChem* **3**:153-160.
17. **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.

18. **Hoehler, T. M., B. M. Bebout, and D. J. Des Marais.** 2001. The role of microbial mats in the production of reduced gases on the early Earth. *Nature* **412**:324-327.
19. **Hungate, R. E.** 1939. Experiments on the Nutrition of *Zootermopsis*. III. The anaerobic carbohydrate dissimilation by the intestinal protozoa. *Ecology* **20**:230-245.
20. **Hungate, R. E.** 1943. Quantitative Analysis on the Cellulose Fermentation by Termite Protozoa. *Ann. Entomol. Soc. Am.* **36**:730-9.
21. **Inward, D., G. Beccaloni, and P. Eggleton.** 2007. Death of an order: a comprehensive molecular phylogenetic study confirms that termites are eusocial cockroaches. *Biol. Lett.* **3**:331-335.
22. **Inward, D. J. G., A. P. Volger, and P. Eggleton.** 2007. A comprehensive phylogenetic analysis of termites (Isoptera) illuminates key aspects of their evolutionary biology. *Mol. Phylogenet. Evol.* **44**:953-967.
23. **Kambhampati, S., and P. Eggleton.** 2000. Taxonomy and Phylogeny of Termites, p. 1-23. *In* T. Abe, D. E. Bignell, and M. Higashi (ed.), *Termites: evolution, sociality, symbioses, ecology*. Kluwer Academic Publishers, Boston.
24. **Leach, M. R., and D. B. Zamble.** 2007. Metallocenter assembly of the hydrogenase enzymes. *Curr. Opin. Chem. Biol.* **11**:159-165.
25. **Leadbetter, J. R.** 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**:3620-31.

26. **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.
27. **Leadbetter, J. R., T. M. Schmidt, J. R. Graber, and J. A. Breznak.** 1999. Acetogenesis from H₂ Plus CO₂ by Spirochetes from Termite Guts. Science **283**:686-689.
28. **Legendre, F., M. F. Whiting, C. Bordereau, E. M. Canello, T. A. Evans, and P. Grandcolas.** 2008. The phylogeny of termites (Dictyoptera: Isoptera) based on mitochondrial and nuclear markers: Implications for evolution of the worker and pseudergate castes, and foraging behaviors. Mol. Phylogenet. Evol. **48**:615-627.
29. **Lenz, O., M. Bernhard, T. Buhrke, E. Schwartz, and B. Friedrich.** 2002. The Hydrogen-Sensing Apparatus in *Ralstonia eutropha*. J. Mol. Microbiol. Biotechnol. **4**:255-262.
30. **Lilburn, T. G., K. S. Kim, N. E. Ostrom, K. R. Byzek, J. R. Leadbetter, and J. A. Breznak.** 2001. Nitrogen Fixation by Symbiotic and Free-Living Spirochetes. Science **292**:2495-2498.
31. **Meyer, J.** 2007. [FeFe] hydrogenases and their evolution: a genomic perspective. Cell. Mol. Life Sci. **64**:1063-1084.
32. **Nicolet, Y., C. Piras, P. Legrand, C. E. Hatchikian, and J. C. Fontecilla-Camps.** 1999. *Desulfovibrio desulfuricans* iron hydrogenase: the structure shows unusual coordination to an active site Fe binuclear center. Structure **7**:13-23.

33. **Noirot, C.** 2001. The Gut of Termites (Isoptera) Comparative Anatomy, Systematics, Phylogeny. II. -- Higher Termites (Termitidae). *Ann. Soc. Entomol. Fr.* **37**:431-471.
34. **Noirot, C.** 1995. The Gut of Termites (Isoptera). Comparative Anatomy, Systematics, Phylogeny. I. Lower Termites. *Ann. Soc. Entomol. Fr.* **31**:197-226.
35. **Odelson, D. A., and J. A. Breznak.** 1985. Cellulase and Other Polymer-Hydrolyzing Activities of *Trichomitopsis termopsidis*, a Symbiotic Protozoan from Termites. *Appl. Environ. Microbiol.* **49**:622-626.
36. **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.
37. **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.
38. **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.
39. **Parkinson, J. S., and E. C. Kofoid.** 1992. Communication Modules in Bacterial Signaling Proteins. *Annu. Rev. Genet.* **26**:71-112.
40. **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**:347-352.

41. **Pester, M., and A. Brune.** 2007. Hydrogen is the central free intermediate during lignocellulose degradation by termite gut symbionts. *ISME J.* **1**:551-565.
42. **Peters, J. W., W. N. Lanzilotta, B. J. Lemon, and L. C. Seefeldt.** 1998. X-ray Crystal Structure of the Fe-Only Hydrogenase (CpI) from *Clostridium pasteurianum* to 1.8 Angstrom Resolution. *Science* **282**:1853-1858.
43. **Posewitz, M. C., D. W. Mulder, and J. W. Peters.** 2008. New frontiers in hydrogenase structure and biosynthesis. *Curr. Chem. Bio.* **2**:178-199.
44. **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.
45. **Schink, B., F. S. Lupton, and J. G. Zeikus.** 1983. Radioassay for Hydrogenase Activity in Viable Cells and Documentation of Aerobic Hydrogen-Consuming Bacteria Living in Extreme Environments. *Appl. Environ. Microbiol.* **45**:1491-1500.
46. **Schwartz, E., and B. Friedrich.** 2006. The H₂-Metabolizing Prokaryotes. *Prokaryotes* **2**:496-563.
47. **Scranton, M. I., P. C. Novelli, and P. A. Loud.** 1984. The distribution and cycling of hydrogen gas in the waters of two anoxic marine environments. *Limnol. Oceanogr.* **29**:993-1003.
48. **Shaw, A. J., D. A. Hogsett, and L. R. Lynd.** 2009. Identification of the [FeFe]-hydrogenase responsible for hydrogen generation in *Thermoanaerobacterium*

- saccharolyticum and demonstration of increased ethanol yield via hydrogenase knockout. *J. Bacteriol.* **20**:6457-64.
49. **Smolenski, W. J., and J. A. Robinson.** 1988. In situ rumen hydrogen concentrations in steers fed eight times daily, measured using a mercury reduction detector. *FEMS Microbiol. Ecol.* **53**:95-100.
50. **Stock, A. M., V. L. Robinson, and P. N. Goudreau.** 2000. Two-Component Signal Transduction. *Annu. Rev. Biochem.* **69**:183-215.
51. **Stock, J. B., A. J. Ninfa, and A. M. Stock.** 1989. Protein Phosphorylation and Regulation of Adaptive Responses in Bacteria. *Microbiol. Rev.* **53**:450-490.
52. **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, ecology*. Kluwer Academic Publishers, Boston.
53. **Sugimoto, A., and N. Fujita.** 2006. Hydrogen concentrations and stable isotopic composition of methane in bubble gas observed in a natural wetland. *Biogeochemistry* **81**:33-44.
54. **Sugimoto, A., T. Inoue, I. Tayasu, L. Miller, S. Takeichi, and T. Abe.** 1998. Methane and hydrogen production in a termite-symbiont system. *Ecol. Res.* **13**:241-257.
55. **Taylor, B. L., and I. B. Zhulin.** 1999. PAS Domains: Internal Sensors of Oxygen, Redox Potential, and Light. *Microbiol. Mol. Biol. Rev.* **63**:479-506.

56. **Trager, W.** 1934. The Cultivation of a Cellulose-Digesting Flagellate, *Trichomonas termopsidis*, and of Certain Other Termite Protozoa. Biol. Bull. **66**:182-190.
57. **Vignais, P. M., and B. Billoud.** 2007. Occurrence, Classification, and Biological Function of Hydrogenases: An Overview. Chem. Rev. **107**:4206-4272.
58. **Vignais, P. M., B. Billoud, and J. Meyer.** 2001. Classification and phylogeny of hydrogenases. FEMS Microbiol. Rev. **25**:455-501.
59. **Wadhams, G. H., and J. P. Armitage.** 2004. Making Sense of it All: Bacterial Chemotaxis. J. Mol. Cell Biol. **5**:1024-1037.
60. **Warnecke, F., P. Luginbühl, N. Ivanova, M. Ghassemian, T. H. Richardson, J. T. Stege, M. Cayouette, A. C. McHardy, G. Djordjevic, N. Aboushadi, R. Sorek, S. G. Tringe, M. Podar, H. G. Martin, V. Kunin, D. Dalevi, J. Madejska, E. Kirton, D. Platt, E. Szeto, A. Salamov, K. Barry, N. Mikhailova, N. C. Kyrpides, E. G. Matson, E. A. Ottesen, X. Zhang, M. Hernández, C. Murillo, L. G. Acosta, I. Rigoutsos, G. Tamayo, B. D. Green, C. Chang, E. M. Rubin, E. J. Mathur, D. E. Robertson, P. Hugenholtz, and J. R. Leadbetter.** 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. Nature **450**:560-569.
61. **Yamin, M. A.** 1979. Cellulolytic activity of an axenically-cultivated termite flagellate, *Trichomitopsis termopsidis*. J. Gen. Microbiol. **113**:417-20.
62. **Yamin, M. A.** 1981. Cellulose Metabolism by the Flagellate *Trichonympha* from a Termite Is Independent of Endosymbiotic Bacteria. Science **211**:58-59.

Chapter 2

GENOMIC ANALYSIS REVEALS MULTIPLE [FeFe] HYDROGENASES AND HYDROGEN SENSORS ENCODED BY TREPONEMES FROM THE HYDROGEN RICH TERMITE GUT

Abstract

H₂ is an important free intermediate in the breakdown of wood by termite gut microbial communities, reaching concentrations in some species exceeding those measured for any other biological system. We have completed a bioinformatic analysis of the hydrogenases encoded in the genomes of three termite gut treponeme isolates: hydrogenotrophic, homoacetogenic *Treponema primitia* strains ZAS-1 and ZAS-2, and the hydrogen producing, sugar fermenting *T. azotonutricium* ZAS-9. These spirochetes encoded 4, 8, and 5 [FeFe] hydrogenase-like proteins, identified by their H domains, respectively, but no other recognizable hydrogenases. The [FeFe] hydrogenases represented many sequence families previously defined in an analysis of termite gut metagenomic data (Warnecke, F., *et al.* 2007. *Nature* 450:560-569). Each strain encoded both putative [FeFe] hydrogenase enzymes and evolutionarily related hydrogen sensor/transducer proteins likely involved in phosphorelay and methylation pathways, and possibly even chemotaxis. A new family of [FeFe] hydrogenases is proposed that may form a multimeric complex with formate dehydrogenase to provide reducing equivalents for CO₂-reductive acetogenesis in *T. primitia*. The many and diverse [FeFe] hydrogenase-like proteins encoded by termite gut treponemes accentuates the importance of H₂ to the ecophysiology of spirochetes in and the fermentation of lignocellulose by wood-feeding termite hindgut communities.

Introduction

The role of termites in global carbon cycling is well established (9, 72). Hydrogen plays a prominent role in this degradation of lignocellulosic biomass by wood feeding termites (6, 10, 18, 48, 49). In wood-feeding lower termites, hydrogen is produced by several protozoal species and is a major product of their cellulose and xylan fermentation (85, 86). Several hydrogenase genes have been cloned from the hindgut protozoa of *Coptotermes formosanus*, and one encoded enzyme, originating from the protist largely responsible for cellulose decomposition, preferentially catalyzed H₂ evolution in biochemical analyses (31). Before H₂ escapes the system, most of this gas is consumed by CO₂-reducing homoacetogenic bacteria (35, 50), and to a lesser extent, methanogenic archaea (6, 33). The flux and standing concentrations of gut H₂ has lead to the development of the concept of this energy rich gas being the central free intermediate in the conversion of plant biomass in wood feeding termites (56). In some species, hydrogen concentrations approach saturation and are among the highest measured for any biological system (18, 27, 56, 65, 68, 70, 73). Moreover, daily productions as high as 33 m³ H₂ per m³ gut volume have been reported (56). The environment is also spatially complex, comprising a matrix of microenvironments characterized by different concentrations of hydrogen (12, 13, 18, 33, 34, 56).

The importance of hydrogen to the termite gut is further highlighted by the abundance of H domain containing [FeFe] hydrogenase-like proteins that were revealed in an analysis of a termite gut metagenome (81). They represented a broad diversity of putative functions, including putative [FeFe] hydrogenase-like hydrogen sensors, which remains an uncharacterized and poorly understood class of proteins (21, 77, 78). The

overwhelming majority of the hydrogen-turnover enzymes identified in that study were [FeFe] hydrogenases (81).

Termites rely upon a complex symbiosis with their respective gut microbial communities to derive carbon and energy from lignocellulosic biomass (7, 8, 11, 16, 17, 48). The primary product of this symbiosis is acetate that the termites use for biosynthesis and energy (50). The fermentation of polysaccharides produces primarily acetate, carbon dioxide, and hydrogen (29, 30, 49, 50). Most of this carbon dioxide and hydrogen is used by bacteria in reductive acetogenesis to produce up to 1/3 (6, 10) of the total acetate pool in the gut. This is why hydrogenases are so important to this environment. A small fraction of the hydrogen is either used by methanogens or released to the atmosphere (6, 35).

Treponemes are among the most abundant groups of bacteria in termite guts (37, 53, 54, 81). They may be the primary agent of reductive acetogenesis (55, 63). The first isolation of termite gut spirochetes was reported in 1999 (35). Among the strains isolated were organisms later characterized as novel species, *Treponema primitia* strains ZAS-1 and ZAS-2 and *Treponema azotonutricium* strain ZAS-9 (23, 35, 36). They represent contrasting hydrogen physiologies. *T. primitia* consumes hydrogen during reductive acetogenesis, and *T. azotonutricium* produces hydrogen during the fermentation of sugars (23). The genomes of *T. primitia* ZAS-2 and *T. azotonutricium* ZAS-9 have recently been sequenced and closed (Genbank accessions tprim_26881 and tazo_31594, respectively). Here we report a bioinformatic analysis of hydrogenase-like proteins from the sequenced genomes of these spirochetes. The objective was to better understand the

genes underlying the hydrogen physiologies of these isolates and to identify potential adaptations to their unique, H₂-rich environment.

Methods

Sequencing and Annotation. The details of the genome sequencing and closure of *T. primitia* ZAS-2 and *T. azotonutricium* ZAS-9 are being reported elsewhere (Genebank accessions tprim_26881 and tazo_31594, respectively). *Treponema primitia* strain ZAS-1 was grown under standard conditions for this isolate (35) and its genome was kindly partially sequenced via 454 pyrosequencing by the Steven Quake lab at Stanford University (sequence available in a local database).

Identification of genes for putative H₂ metabolism. A Hidden Markov Model from Pfam (4, 19) was used with HMMER (19) to search for nickel-dependent hydrogenases (PF00374) within each genome's putative proteome. [FeFe] hydrogenases were identified within each genome database with IPR004108 from the Blocks (57) server using MAST (3). Sequences containing the three sequence signatures characteristic of the H domain of [FeFe] hydrogenases (44), corresponding to blocks 3, 4, and 6 of IPR004108, were collected for further analysis. Sequences sharing high sequence identity with each of the three chaperones, HydE, HydG, and HydF, necessary for the assembly of the H cluster of [FeFe] hydrogenases (5, 32, 58) were identified using BLAST. Homologs to [FeFe] hydrogenases from the *T. primitia* and *T. azotonutricium* strains were identified within the termite gut metagenome sequence database at the JGI IMG/M (41, 42) server using BLAST searches.

Phylogenetic Analysis. The ARB software environment was used for phylogenetic analyses (38). Sequence alignments were prepared using DIALIGN (45) on the MobyL

server (47). Trees were routinely constructed in sets of three, corresponding to distance matrix (Fitch), maximum parsimony (Phylip PROTPARS), and maximum likelihood (Phylip PROML) methods. The sequence database used within ARB contained 183 publically available protein sequences harboring H domains. Many of the [FeFe] hydrogenase sequences were chosen from those highlighted in reviews by Meyer (44) or Vignais (77). A number of sequences were identified by BLAST searches against the NCBI GenBank non-redundant protein sequences database. The database also included four protist [FeFe] hydrogenase sequences from the gut of *Coptotermes formosanus* (31). 84 sequences from the 123 identified as containing H domains in the termite gut metagenome database were of sufficient length to be included in the analysis. The following sequences comprised the outgroup used to construct Figure 2-2: *Caenorhabditis elegans* (NP_498092), *Homo sapiens* (NP_036468, NP_071938), *Kluyveromyces lactis* (CAA49833), *Oryza sativa* (XP_469746), *Saccharomyces cerevisiae* (NP_014159), *Schizosaccharomyces pombe* (NP_588309). The following sequences comprise the outgroup used to construct Figure 2-3: *Chlamydomonas moewusii* (Q56UD8), *Chlorella fusca* (Q8VX03), *Holomastigotoides mirabile* (AB331669), *Pseudotriconympha grasii* (AB331668, AB331667), *Scenedesmus obliquus* (Q9AU60, Q9AR66), *Trichomonas vaginalis* (Q27096, Q27094, XP_001305709, XP_001310180, XP_001328981, XP_001322682, XP_001580286), uncultured parabasalid (AB331669). The following sequences comprise the outgroup used to construct Figure 2-4: *Caenorhabditis elegans* (NP_498092), *Homo sapiens* (NP_036468, NP_071938), *Kluyveromyces lactis* (CAA49833), *Oryza sativa* (XP_469746), *Saccharomyces cerevisiae* (NP_014159), *Schizosaccharomyces pombe*

(NP_588309), *Entamoeba histolytica* (Q51EJ9, Q50YQ4), *Giardia lamblia* (EAA39802), and *Spiroplasma bickhamii* (Q9GTP1). Phylogenetic analyses were completed using only the H domain region of each peptide sequence, as defined by a filter used to select appropriate residues from sequence alignments. This subset of amino acids corresponded to the roughly 310 span of residues: *C. pasteurianum* (P29166) E207-K515 (Total Length = 308) and *D. vulgaris* (YP_010987) E83-V394 (Total Length = 311).

Sequence and gene cluster analyses. Pfam (4) and InterProScan (46) were used to identify previously characterized domain sequences within each protein. Sequences were not analyzed further if they lacked domains established by precedent to be essential for functionality, see reviews by Meyer and Vignais (44, 77). The Prediction of Protein Subcellular Localization for bacterial sequences (PSORTb v.2.0) program (22) was used to predict protein subcellular localization. For gene cluster analysis, all genomes were uploaded to the SEED server (52) using RAST (2).

Nomenclature. Genes were named following the convention proposed by Vignais (78). Wherever possible, hydrogenases were also classified into termite gut community associated families as defined by Warnecke *et al.* (81); that is, according to their phylogenetic position. Where no family membership was clear, a new family was proposed.

Results

Hydrogenase-like genes and associated maturases identified. The closed genome sequences of *Treponema primitia* ZAS-2 and *T. azotonutricium* ZAS-9, and the partial genome sequence of *T. primitia* ZAS-1, were inspected for the presence of candidate genes that might be associated with H₂ metabolism. No obvious homologs of known

NiFe or NiFeSe hydrogenases were identified. In contrast, each genome encoded a number of proteins containing putative H domain modules similar to those from known [FeFe] hydrogenases. *T. primitia* ZAS-1, *T. primitia* ZAS-2 and *T. azotonutricium* ZAS-9 each encoded 4, 8 and 5 such proteins, respectively (Table 2-1). Within the H domain of [FeFe] hydrogenases there are three unique, highly conserved, sequence regions (or sequence signatures) that coordinate the H cluster (44, 77). The H cluster is an iron-sulfur cluster that functions as the site of catalysis in [FeFe] hydrogenases. With respect to these conserved regions, the putative H domains encoded by the three treponemes fell into two groups (Table 2-1), one with canonical sequence signatures and the other with at least one key residue diverging from the conserved sequence, a topic discussed further, below. The genomes of the three treponemes also encoded “H cluster assembly” proteins (maturases) similar to the chaperones HydE, HydG, and HydF, which are known to be relevant to the expression of functional [FeFe] hydrogenases (32, 58).

Domain architecture and predicted subcellular localization. In an initial effort to deduce possible functions for each of the proteins possessing H domains, we examined their domain architectures. The treponeme genome sequences collectively encoded a set of H domain containing proteins having domain architectures representing most of those observed in a termite gut metagenome (Table 2-2, Figure 2-1). Only one domain structure was not represented, and it had a low representation in the metagenome (81). All sequences encoded a predicted 2[4Fe-4S] cluster immediately N-terminal to the H domain (Figure 2-1). We predict that all proteins encoding canonical H domain sequence signatures (44) (3, 5, and 2, respectively, in *T. primitia* strains ZAS-1 and ZAS-2, and *T. azotonutricium* ZAS-9; Table 2-1) function as [FeFe] hydrogenase enzymes. For all such

Table 2-1. FeFe hydrogenase-like proteins observed in the genomes of three termite gut isolates.

Strain	Gene Name	Gene No.	Family ^b	Functionality	Cellular Localization	L1 ^c	L2 ^c	L3 ^c
<i>T. primitia</i> ZAS-1	<i>HydA1</i>	1128 ^a	FDH-Linked	Enzyme	Cytoplasmic or Periplasmic	CCP	PCxxKxxE	MxCxxGCxxG
<i>T. primitia</i> ZAS-1	<i>HydA3</i>	2678 ^a	6	Enzyme	Cytoplasmic or Periplasmic	CCP	PCxxKxxE	MxCxxGCxxG
<i>T. primitia</i> ZAS-1	<i>HndA1</i>	2692 ^a	7	Enzyme	Cytoplasmic or Periplasmic	CCP	PCxxKxxE	MxCxxGCxxG
<i>T. primitia</i> ZAS-1	<i>HydA2</i>	1488 ^a	5	Sensor	Cytoplasmic Membrane	PCP	PCxxKxxE	LxCxxGCxxG
<i>T. primitia</i> ZAS-2	<i>HydA1</i>	TREPR_3094	FDH-Linked	Enzyme	Cytoplasmic or Periplasmic	CCP	PCxxKxxE	MxCxxGCxxG
<i>T. primitia</i> ZAS-2	<i>HydA5</i>	TREPR_0390	6	Enzyme	Cytoplasmic or Periplasmic	CCP	PCxxKxxE	MxCxxGCxxG
<i>T. primitia</i> ZAS-2	<i>HndA1</i>	TREPR_3594	3	Enzyme	Cytoplasmic or Periplasmic	CCP	PCxxKxxE	MxCxxGCxxG
<i>T. primitia</i> ZAS-2	<i>HndA2</i>	TREPR_3288	7	Enzyme	Cytoplasmic or Periplasmic	CCP	PCxxKxxE	MxCxxGCxxG
<i>T. primitia</i> ZAS-2	<i>HndA3</i>	TREPR_3300	7	Enzyme	Cytoplasmic or Periplasmic	CCP	PCxxKxxE	MxCxxGCxxG
<i>T. primitia</i> ZAS-2	<i>HydA2</i>	TREPR_3122	4	Sensor	Cytoplasmic or Periplasmic	[AV] CP	[PS] CxxKxxE	LxCxxGCxxG
<i>T. primitia</i> ZAS-2	<i>HydA3</i>	TREPR_3283	4	Sensor	Cytoplasmic or Periplasmic	[AV] CP	[PS] CxxKxxE	LxCxxGCxxG
<i>T. primitia</i> ZAS-2	<i>HydA4</i>	TREPR_1589	5	Sensor	Cytoplasmic Membrane	PCP	PCxxKxxE	LxCxxGCxxG
<i>T. azotonutricium</i> ZAS-9	<i>HydA3</i>	TREAZ_2002	6	Enzyme	Cytoplasmic or Periplasmic	CCP	PCxxKxxE	MxCxxGCxxG
<i>T. azotonutricium</i> ZAS-9	<i>HndA</i>	TREAZ_3250	3	Enzyme	Cytoplasmic or Periplasmic	CCP	PCxxKxxE	MxCxxGCxxG
<i>T. azotonutricium</i> ZAS-9	<i>HydA1</i>	TREAZ_2458	5	Sensor	Cytoplasmic Membrane	PCP	PCxxKxxE	LxCxxGCxxG
<i>T. azotonutricium</i> ZAS-9	<i>HydA2</i>	TREAZ_2238	10	Sensor	Cytoplasmic Membrane	PCP	PCxxKxxE	LxCxxGCxxG
<i>T. azotonutricium</i> ZAS-9	<i>HydA4</i>	TREAZ_2003	8	Sensor	Cytoplasmic	CCH	PCxxKxxE	SxCxxSxCxxG

^aGene numbers for *T. primitia* ZAS-1 are arbitrarily assigned based upon feature identifiers assigned by RAST (2).

^bFamily numbers are taken from Warnecke *et al.* (81). The FDH-Linked family of [FeFe] hydrogenases has been proposed in the present study.

^cConserved sequence signatures observed in the H domain of all known [FeFe] hydrogenases (44). Variations from the canonical sequences are given in bold.

Table 2-2. Putative hydrogenase-like proteins encoded by three termite gut spricochetes, a termite hindgut metagenome and several reference bacteria.

	<i>T. primitia</i> ZAS-1	<i>T. primitia</i> ZAS-2	<i>T. azotonutricium</i> ZAS-9	<i>T. denticola</i>	<i>B. hyodysenteriae</i>	<i>M. thermoacetica</i>	Gut metagenome (see Warnecke, 2007)
Putative FeFe Hydrogenases							
Family 3 ^{a,c}	0	1	1	0	1	0	43
Family 6 ^a	1	1	1	1	1	0	17
Family 7 ^{a,c}	1	2	0	0	0	0	6
FDH-Linked ^b	1	1	0	0	0	1	0
N/A ^f	0	0	0	1	0	1	0
Putative Hydrogen Sensors							
Family 4 ^{a,d}	0	2	0	0	0	0	26
Family 5 ^{a,e}	1	1	1	0	0	0	4
Family 8 ^{a,d}	0	0	1	0	0	2	6
Family 10 ^{a,e}	0	0	1	0	0	0	4
NiFe Hydrogenases							
N/A ^f	0	0	0	0	0	1	2
Total	4	8	5	2	2	5	108

^aFamilies have been defined by Warnecke *et al.* (81).

^bThe family of FDH-Linked hydrogenases has been defined in this chapter.

^cFamilies 3 and 7 share similar domain architectures.

^dFamilies 4 and 8 share similar domain architectures.

^eFamilies 5 and 10 share similar domain architectures, as defined by Warnecke *et al.* (81).

^fProteins for which a family designation could not be unambiguously defined or for which a [FeFe] hydrogenase family designation would not be relevant.

^g*T. primitia* strains ZAS-1 and ZAS-2 and *M. thermoacetica* are homoacetogens. *T. azotonutricium* ZAS-9 is primarily a hydrogen producing bacterium. *B. hyodysenteriae* and *T. denticola* are both well studied pathogens.

Figure 2-1.

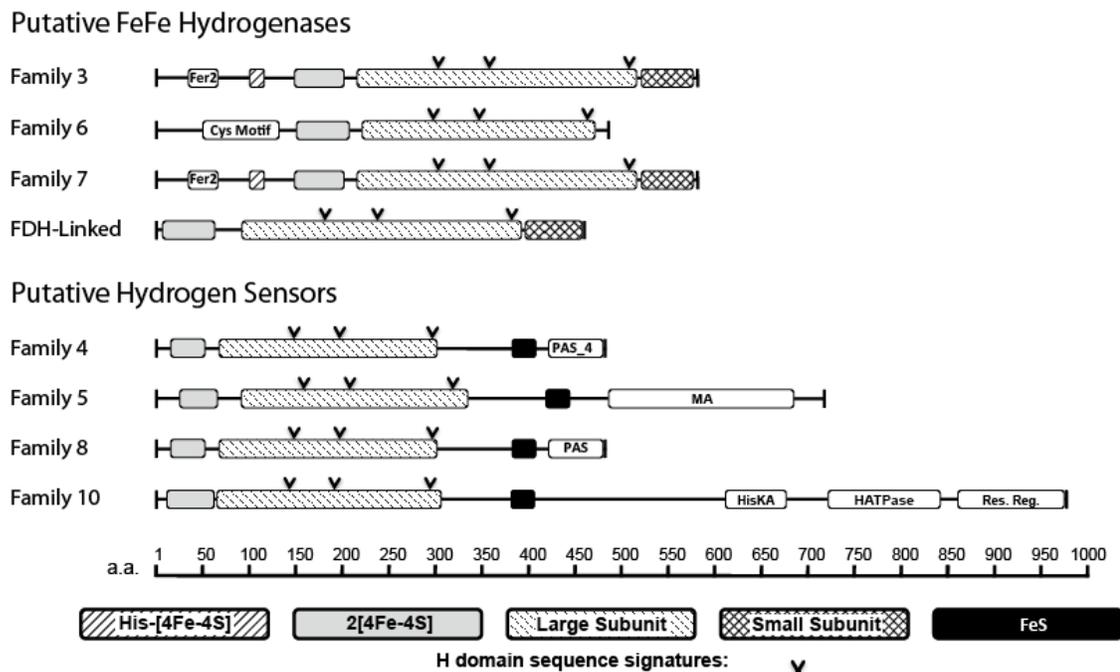


Figure 2-1. Domain architectures representative of each family of H domain proteins observed in *Treponema primitia* strains ZAS-1 and ZAS-2 and *T. azotonutricium* ZAS-9. Domains were identified using the Pfam server (4). Following initial detection, iron sulfur cluster domain regions were annotated manually from alignments as spanning from the first to the last coordinating cysteine. The following domains were observed: 2[4Fe-4S] – F cluster made up of two adjacent Fer4 domains, PF00037; Cys Motif – eight Cys residues occurring in three runs, CC, Cx2C, Cx2Cx4Cx3C; Fer2 – PF00111, [2Fe-2S] iron-sulfur cluster binding domain; FeS – PF04060, putative Fe-S cluster; HATPase – PF02518, Histidine kinase-, DNA gyraseB-, and HSP90-like ATPase; His-[4Fe-4S] – a [4Fe-4S] cluster with the first (N-terminal most) coordinating cysteine replaced with a histidine; HisKA – PF00512, His Kinase A (phosphoacceptor) domain; Large Subunit – PF02906, iron only hydrogenase large subunit, C-terminal domain; MA – PF00015, methyl-accepting chemotaxis protein (MA) signaling domain; PAS – PF00989, PAS domain; PAS_4 – PF08448, PAS_4 domain, a part of the PAS domain clan; REC – PF00072, response regulator receiver domain; Small Subunit – PF02256, iron hydrogenase small subunit. The first residue of each conserved H domain signature, as defined by Meyer, is indicated in the Large Subunit domains by arrowheads (44).

sequences an additional domain or a conserved sequence motif was present at the N-terminal flank of the 2[4Fe4S] cluster (Figure 2-1). These modules consisted of either a duo of domains comprising a [2Fe-2S] cluster followed by a histidine-coordinated [4Fe-4S] cluster, or a single cysteine-rich motif. This motif, previously reported in a number of putative [FeFe] hydrogenase sequences (44, 81), has been proposed to coordinate an iron-sulfur cluster.

All sequences having non-canonical H domain sequence signatures (Table 2-1) contained an [FeS] cluster C-terminal to the H domain followed by a putative signaling domain(s) (Figure 2-1), and are herewith considered to be putative H₂-sensors falling within three functional groups. One group contained a methyl-accepting chemotaxis protein (MA) domain; another a PAS domain; and the third a triad of domains comprised of a histidine kinase, ATPase, and response regulator domain.

PSORTb gave an unambiguous prediction of cellular localization for only three groups of putative hydrogenase and H₂ sensor proteins. All putative H₂ sensors with either MA or the domain triad (above) are predicted to localize to the cytoplasmic membrane. Sequences containing a PAS domain (Figure 2-1, Table 2-1) and classified as Family 8, *sensu* Warnecke et al. (81), are unambiguously predicted to be cytoplasmic. All other putative sensors and enzymes were ambiguously predicted to be periplasmic (localization scores = 4.48 ea.) and cytoplasmic (localization scores = 5.41 ea.).

Phylogenetic analysis of putative hydrogenases and H₂ sensors. Phylogenetic analysis was performed on all seventeen H domains identified from the genome sequences of the three treponemal strains. The H domains clustered within one or the other of two coherent clades (Figure 2-2). Each of these coherent clades was analyzed separately and

Figure 2-2.

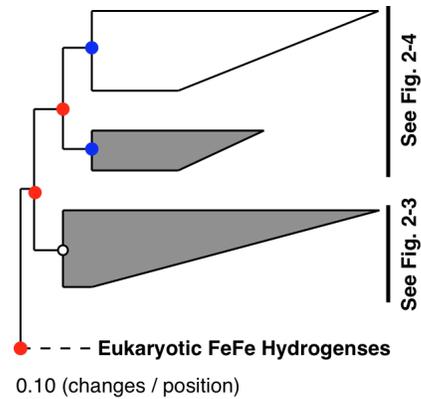


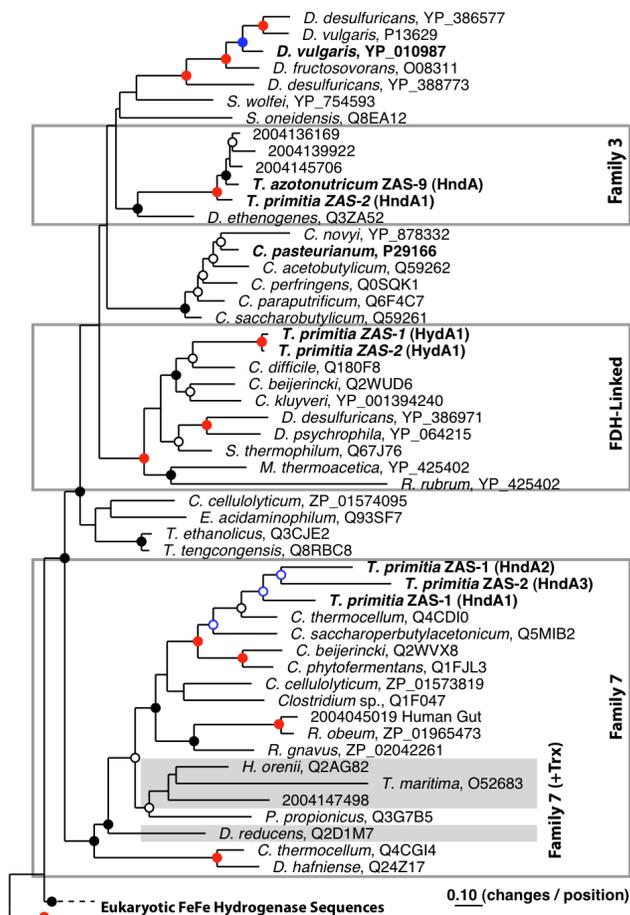
Figure 2-2. Phylogeny of H domain peptide sequences from putative hydrogen sensor and [FeFe] hydrogenase proteins. The tree was calculated using a maximum likelihood (Phylip ProML, 100 bootstraps) method with 150 unambiguously aligned amino acids. The clade containing hydrogen sensors is shown in white and clades containing [FeFe] hydrogenases are shown in grey. Open circles mark groupings also supported by either parsimony (Phylip PROTPARS, 100 bootstraps) or distance matrix (Fitch) methods. Closed circles mark grouping supported by all three methods. Coloring of the circles reflects the magnitude of the corresponding bootstrap values for both maximum parsimony (100 bootstraps) and maximum likelihood (100 bootstraps) methods: black = one of the bootstrap values is below 50%; blue = both bootstrap values are over 50%; and red = both bootstrap values are over 85%. All outgroup sequences are listed in the methods section.

in greater detail (Figures 2-3A & 2-4). Based on the phylogenetic position of their H-domains, most of the H-domain encoding proteins were classified as belonging to any one of several different termite gut community associated, iron-only hydrogenase families (Table 2-1, Figures 2-3A & 2-4), established previously (81). However, 2 proteins (HydA1 from *T. primitia* ZAS-1 and HydA1 from *T. primitia* ZAS-2) fell within a clade that did not contain or group closely with any sequence from the termite gut metagenome database (81) and, therefore, could not be designated as belonging to a pre-defined family. The number of hydrogenase-like proteins representative of each sequence family were compared, see Table 2-2, between the termite gut isolates, two anaerobic treponemes, a termite gut metagenome sequence, and a canonical acetogen. The classification of these as likely being biochemically associated with hydrogenase-linked formate dehydrogenases is described below. Each sequence within this clade contained an insertion 28 amino acids in length falling within the H domain region (Figure 2-3B). This insert was not observed in other sequences in our database and was not included in phylogenetic calculations. Phylogenetic analysis revealed that the H-domains associated with Family 7 proteins may be divided into two sub-families, based on the presence or absence of a thioredoxin-like [2Fe-2S] cluster in the parent protein. The cluster is absent from the Family 7 proteins identified in the treponemal isolates. Lastly, phylogenetic analysis revealed that the H-domains of the putative H₂ sensors all cluster together and likely comprise a radiation evolutionarily derived from putatively enzymatic Family 6 [FeFe] hydrogenases (Figure 2-2 & 2-4).

Gene cluster analysis. Many of the putative [FeFe] hydrogenase encoding sequences were observed to occur within gene clusters on their respective genomes implying a

Figure 2-3.

A.



B.

2004136169	AAHC-----	-----	PVGAII
2004139922	SAHC-----	-----	PVGAII
2004145706	AAHC-----	-----	PVGAII
T. azotonutricum ZAS-9 (HndA)	AAHC-----	-----	PVGAII
T. primitia ZAS-2 (HndA1)	AAHC-----	-----	PVGAII
D. ethenogenes, Q3ZA52	ILVC-----	-----	PVGAII
T. primitia ZAS-1 (HydA1)	VQKCKSYVSL	IDHGFPEMVK	KREERMLPET VR-EPLFAAH
T. primitia ZAS-2 (HydA1)	VQKCKSYVSL	IDHGFPEMVK	KREERMLPET VR-EPLFAAH
C. difficile, Q180F8	VQTKSYASV	IDEGFELQE	KKQEREIPES IN-EPIFAAY
C. beijerincki, Q2WUD6	IQVCNSYGF	NRENSHLIEE	KRRDRGVLES VK-EPVFAAF
C. kluyveri, YP_001394240	IQICKGYYSI	YDDVATPVSK	KLFDRGLLDN VD-EPLFAAY
D. desulfuricans, YP_386971	VQCCSAPASF	YEQHPACIAE	KKRERGLFVS EA-APLFAAW
D. psychrophila, YP_064215	VQICSGYSD	LMSYATGRGK	RLQNRGLLAT VV-EPLFAAH
S. thermophilum, Q67J76	VQVCSSYGS	WDDGLTPREQ	KLAERGLLPS VK-EPLFAAW
M. thermoacetica, YP_425402	VQICSAVASP	YTTSPETMAA	KNRERLLPA aapEPLFAAY
R. rubrum, YP_425402	VATCAAFDSI	FDAPPTPRPV	RLKRRGLPVS LK-EPLFAAH

Figure 2-3. Phylogeny of H domain peptide sequences from putative [FeFe] hydrogenase proteins. **A.)** Refer to Figure 2-2. The tree was calculated using a maximum likelihood (Phylip ProML) method with 250 unambiguously aligned amino acids. Open circles mark groupings also supported by either parsimony (Phylip PROTPARS, 1000 bootstraps) or distance matrix (Fitch) methods. Closed circles mark grouping supported by all three methods. Circle colors have the same meaning as described in Figure 2-2. Family names are taken from Warnecke *et al.* (81). Family names were assigned to the boxed regions based upon homology to sequences observed in the gut metagenome of a higher termite (81). The “FDH-linked” [FeFe] hydrogenase clade is so named because of a close genomic proximity to formate dehydrogenase. Within the Family 7 clade, see sequences highlighted in grey, is a group of sequences containing a C-terminal Trx-like iron-sulfur cluster. GenBank accession numbers for each protein are listed following the name of its origin species. Numbers beginning in 2004 are IMG gene object identifiers for sequences taken from a termite gut metagenome (81). *Treponema azotonutricium* ZAS-9 and *Treponema primitia* strains ZAS-1 and ZAS-2 sequences are in bold. Also in bold are the two [FeFe] hydrogenase sequence for which structures are available. For clarity, the tree does not contain all enzymatic [FeFe] hydrogenase sequences in the ARB database used in analyses. All outgroup sequences are listed in the methods section. **B.)** An alignment of sequences taken from the tree depicted in part A. of this figure. Each member of the FDH-Linked family of sequences contained a 28 amino acid insert not observed in other sequences from the database.

Figure 2-4.

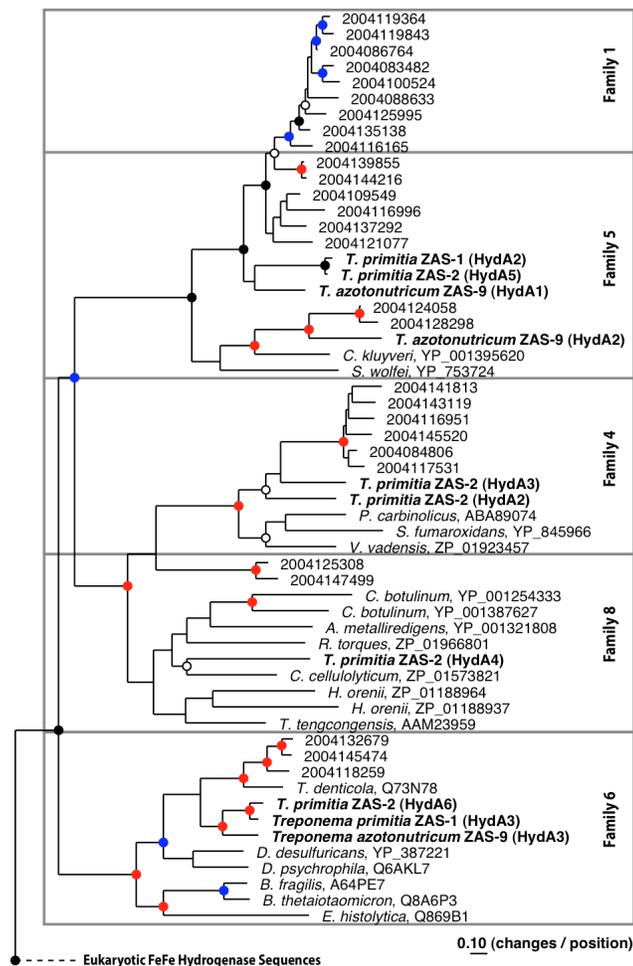
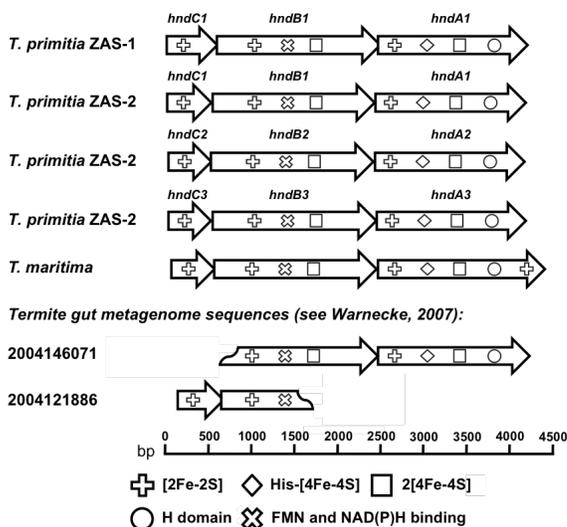


Figure 2-4. Phylogeny of H domain peptide sequences from [FeFe] hydrogenase Family 6 and putative H₂ sensor proteins. Refer to Figure 2-2. The tree was calculated using a maximum likelihood (Phylip ProML) method with 238 unambiguously aligned amino acids. Open circles mark groupings also supported by either parsimony (Phylip PROTPARS, 100 bootstraps) or distance matrix (Fitch) methods. Closed circles mark grouping supported by all three methods. Circle colors have the same meaning as described in Figure 2-2. Family names are taken from Warnecke *et al.* (81). Family names were assigned to the highlighted regions based upon homology to sequences observed in the hindgut metagenome of a higher termite (81). GenBank accession numbers for each protein are listed following the name of its origin species. Numbers beginning with the number 2004 correspond to IMG gene object identifiers. *Treponema azotonutricum* ZAS-9 and *Treponema primitia* strains ZAS-1 and ZAS-2 sequences are in bold. For clarity, the tree does not contain all enzymatic [FeFe] hydrogenase sequences in the ARB database used in extensive analyses. All outgroup sequences are listed in the methods section.

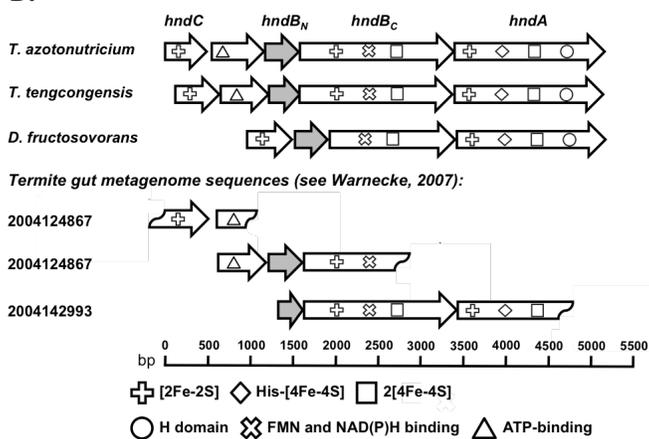
multimeric quaternary structure (Figures 2-5A, 2-5B & 2-5C). Each Family 3 or 7 putative [FeFe] hydrogenase from *T. primitia* strains ZAS-1 and ZAS-2 fell within gene clusters implying a hetero-trimeric quaternary structure (Figure 2-5A). Several known trimeric [FeFe] hydrogenases have a C-terminal thioredoxin-like domain in the H domain containing protein (44, 76, 77). This pattern was not observed in the treponemes, and no obvious gene was observed nearby in the genome that might serve to compensate for this absence. This C-terminal thioredoxin-like module/domain was also entirely absent in homologous genes occurring within similar contexts in genomes available on the JGI IMG/M (41, 42) server. The Family 3 putative [FeFe] hydrogenase encoded by *T. azotonutricium* ZAS-9 occurred within a gene cluster implying a hetero-trimeric quaternary structure (Figure 2-5B). Gene configurations similar to the trimeric and tetrameric [FeFe] hydrogenases identified in the treponeme genomes were observed across sequence reads in the termite gut metagenome (81), see Figures 2-5A & 2-5B. As introduced above, the genomes of *T. primitia* strains ZAS-1 and ZAS-2 each encoded a putative [FeFe] hydrogenase that did not cluster phylogenetically with previously established termite gut community families. Examination of their gene contexts suggested that they might be biochemically coupled in function to the recently described (43) cysteine variant of hydrogenase-linked formate dehydrogenases (FDH H, or FdhF; Figure 2-5C and Table 2-1) in *T. primitia*. The genes for these two proteins are each proximal to two genes encoding 16Fe ferredoxin-like proteins sharing homology with HycB from *Escherichia coli* and CooF from *Rhodospirillum rubrum*. A similar gene configuration was observed in the genome of *Clostridium difficile* (Figure 2-5C).

Figure 2-5.

A.



B.



C.

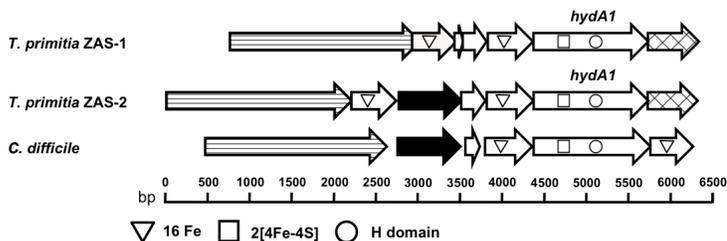


Figure 2-5. Multimeric hydrogenase gene clusters. Arrows represent genes and the symbols within each arrow represent encoded protein domains. Domain symbols are listed in proper order but are not intended to represent precise locations. Homologous genes without annotated domains share the same shading or patterning. [FeFe] hydrogenase subunit gene symbols are provided above the treponeme strain genes. A wavy line at the 5' or 3' end of a gene indicates an incomplete sequence. Domains were identified using the Pfam server (4). Domains represented in the figure are: 16Fe – a ferredoxin-like protein sharing homology to HycB from *E. coli*; 2[4Fe-4S] – iron-sulfur cluster made up of two adjacent Fer4 domains, PF00037; [2Fe-2S] – PF00111, [2Fe-2S] iron-sulfur cluster binding domain; ATP-binding – PF02518, Histidine kinsae-, DNA gyrase B-, and HSP90-like ATPase; FMN and NAD(P)H binding – PF01512, Respiratory-chain NADH dehydrogenase 51 Kd subunit; H domain – PF02906 and PF02256, iron only hydrogenase large subunit, C-terminal domain, and iron only hydrogenase small subunit; His-[4Fe-4S] – a [4Fe-4S] cluster with the first coordinating cysteine replaced with a histidine. **A.)** 2004146071, 2004121886 are IMG gene object identifiers for a gene within sequence reads taken from a termite gut metagenome sequence database (81). **B.)** 2004124867, 2004124867, and 2004142993 are IMG gene object identifiers for a gene within sequence reads taken from a termite gut metagenome sequence database (81). **C.)** The formate dehydrogenase gene is filled with a horizontal line pattern. Hypothetical genes are colored white.

Discussion

A previous metagenomic analysis of a subset of the gut community from a Costa Rican, wood-feeding “Higher” termite had identified a large number of genes for novel [FeFe] hydrogenases and novel [FeFe] hydrogenase-like sensor proteins (81). That study had also posited that the majority of these were encoded by not yet cultivated spirochetes. Here, we see a similar pattern mirrored and extended in our analysis of the genomes of hydrogen-metabolizing spirochetes isolated from a wood-feeding, dampwood termite from California. The three *Treponema* species analyzed in this study are intriguing because they 1) represent both hydrogen-consuming and producing physiotypes, 2) are members of one of the more abundant bacterial phyla (Spirochetes) generally observed in a variety of termite gut ecosystems, and 3) encounter, as members of this ecosystem, the most H₂-rich environments found anywhere in Nature. The isolates encoded a large number and broad diversity of H domain containing proteins (Tables 2-1 & 2-2). These included both putatively enzymatic [FeFe] hydrogenases and putative H₂ sensors.

The number and the variety of H-domains represented in these genomes is at the upper end for those typically observed in the sequenced genomes of bacteria (15, 44). This, taken together with the non-observation of other types of hydrogenases in these genomes (Table 2-1), accentuates and underscores the relevance of analyzing environments and isolates in which H₂ has been demonstrated to be important. It is not yet clear *why* iron-only hydrogenases would be the most abundant, and thus possibly the most dominant, types of hydrogenases operating in the termite gut environment (31, 81). Mechanistically, others have shown that [FeFe] hydrogenases may have higher specific molar activities than the other varieties of hydrogenases (21). The millimolar amounts of ferrous iron

bioavailable in the guts of several termites (21, 79) might also have an influence on what hydrogenases are at play in these systems. No doubt, many other possible factors might be relevant and at play, and it will become interesting in the future to learn more about what has shaped the hydrogenase landscape in termite gut ecosystems.

The [FeFe] hydrogenase-like proteins encoded by the treponemes analyzed here comprise both putative enzymes and putative H₂-sensors. That each strain encoded examples of both types suggests that having them in concert may be relevant to H₂ processing and competition within the termite gut ecosystem. Curiously, the phylogeny of the H-domain of the putative sensor domains suggests that they have evolved as a radiation after a duplication and subsequent modification from a “Family 6” hydrogenase in the past (Figure 2-2 & 2-4).

Putative multimeric [FeFe] hydrogenases. Family 3 and 7 [FeFe] hydrogenases from the H₂-consuming *T. primitia* strains ZAS-1 and ZAS-2 fell within gene clusters similar to that of the trimeric [FeFe] hydrogenase from *T. maritima* (76) (Figure 2-5A). Interestingly, *T. maritima* is a predominantly hydrogen-producing, anaerobic, fermentative hyperthermophile (28). H₂ is known to inhibit its growth, although it is consumed in a non-energy coupled detoxification reaction, also observed in *Pyrococcus furiosus*, to produce H₂S in the presence of S⁰ (20, 28). Thus, it is unclear whether its hydrogenase functions primarily in hydrogen production, or in some cryptic consumption capacity. We postulate that the observed adjacent HydB and HydC genes form a complex with the hydrogenase. These accessory genes are homologous to the NuoF and NuoE subunits of the *E. coli* NADH: ubiquinone oxidoreductase (Complex I) (82), respectively, that together function in Complex I to oxidize NADH and transfer the

electrons along to other subunits. Therefore, these accessory proteins may serve to form a diaphorase moiety interacting with the H domain containing subunit (HydA) to couple hydrogen turnover to the oxidation or reduction of NAD(P)(H).

The Family 3 hydrogenase encoded by *T. azotonutricium* ZAS-9 fell within a gene cluster similar to those of the tetrameric [FeFe] hydrogenases from *Thermoanaerobacter tengcongensis* (40) and *Desulfovibrio fructosovorans* (71) (Figure 2-5B). *T. tengcongensis* and many *Desulfovibrio* species produce hydrogen as a fermentation product (25, 75, 84). *D. fructosovorans* can also couple energy production to hydrogen consumption, but the growth of *T. tengcongensis* is inhibited by H₂ (51, 84). The tetrameric [FeFe] hydrogenase of *D. fructosovorans* is known to be an NADP-reducing hydrogenase (40) and the *T. tengcongensis* enzyme may have a similar function, though it has not yet been biochemically characterized (71). The accessory domains comprising this complex may, just as described above for the Family 3 and 7 proteins from *T. primitia* (above), form a diaphorase moiety to enable the coupling of co-factor oxidation or reduction to hydrogen production or consumption, respectively. In support of the more general relevance of the multimeric [FeFe] hydrogenases to termite gut communities, analysis of sequence reads suggested that homologous trimeric and tetrameric complexes are present in a termite hindgut metagenome (Figure 2-5A & 2-5B).

It was surprising to us to find that *T. primitia* ZAS-1 did not encode a Family 3 [FeFe] hydrogenase. This family was found in both of the other treponeme strains and was the most highly represented family observed in the termite hindgut metagenome. Its absence from the genome of this hydrogen consuming strain was further supported by the lack of

any amplification product using degenerate primers targeting Family 3 [FeFe] hydrogenases (data not shown). The strain did, however, encode a Family 7 trimeric [FeFe] hydrogenase. This may imply that Family 7 hydrogenases have a physiological role similar to that of Family 3, and fulfill this function in *T. primitia* ZAS-1 (see also Figure 2-5A).

Putative FDH-linked [FeFe] hydrogenases. Many of the hydrogenase genes identified in this study fall within phylogenetic clusters established during an earlier analysis of termite metagenomic sequence data (81), and several represent the first alleles identified from any cultured organism for their respective clusters. However, a few hydrogenases encoded by these spirochete pure cultures were not represented by any alleles identified in that earlier study. For example, both of the *T. primitia* strains encoded an [FeFe] hydrogenase gene whose locus is in close proximity to that for a formate dehydrogenase (FDH; Figure 2-5C). Several homologs encoded by other bacteria clustered phylogenetically with these hydrogenases (Figure 2-3A), and all contained a unique insert of 28 amino acids in length at a conserved location. This stretch of amino acid residues was not found in other hydrogenases in our databases; moreover, it was filtered out during phylogenetic analyses, thus serving as independent support for the cluster. We hypothesize that, together, these FDH and hydrogenase genes operate in a formate hydrogen lyase-like complex, whereby the generation of formate from carbon dioxide and H₂ would be the first step of the methyl-branch of the Wood-Ljungdahl pathway of reductive acetogenesis (60, 67). *T. azotonutricium*, which does not encode any obvious FDH genes and is not an H₂ + CO₂ acetogen, does not encode one of these particular hydrogenase homologs.

The FDH-linked [FeFe] hydrogenase genes in the two *T. primitia* strains were found proximal to two genes encoding putative 16Fe ferredoxin-like proteins. These may serve to shuttle electrons between the hydrogenase and FDH subunits of the formate hydrogen lyase complex. Their shared homology with HycB from *E. coli* provides further support for this hypothesis because this protein is believed to shuttle electrons between the FdhF subunit (FDH-H) and the NiFe hydrogenase subunit of a formate-hydrogen lyase complex (64). Interestingly, these ferredoxin-like proteins are also homologous to CooF of *Rhodospirillum rubrum*. CooF is believed to have an analogous electron shuttling function, only in this case it is between a carbon monoxide dehydrogenase (CooS) and a NiFe hydrogenase (1).

E. coli is not a homoacetogen, and it operates its formate hydrogen lyase complex in the direction of formate oxidation to generate hydrogen, as does *Eubacterium acidaminophilum*, which also encodes its FDH gene in close proximity to a NiFe hydrogenase gene (1, 24). NiFe hydrogenases, absent in the treponemes analyzed here, are entirely distinct phylogenetically from the [FeFe] hydrogenases that are the focus of this study (78). Thus, we propose that the treponemes encode a novel formate hydrogen lyase-like complex, one that operates with an iron-only hydrogenase, and in the reductive direction (43). Interestingly, homologs of these FDH-linked [FeFe] hydrogenase alleles were also found in other *bona fide* acetogens (Figure 2-3A), including *Moorella thermoacetica*, and other strains that encode genes associated with the Wood-Ljungdahl pathway. For example, the gut pathogen *Clostridium difficile*, which may be a cryptic acetogen (61), encodes a homolog of the FDH-linked [FeFe] hydrogenase gene in close proximity to an FDH gene (Figure 2-5C), and does not encode a NiFe hydrogenases

homolog. Thus, it may turn out that there is a more widespread role for formate-hydrogen lyases in homoacetogenic and acetoclastic metabolic pathways than is currently now recognized.

Putative hydrogen sensors. The sensory [FeFe] hydrogenase-like proteins encoded by the three spirochete strains could be divided into two groups, those likely involved in two-component regulatory systems or phosphorelays, and those likely involved in methylation cascades, perhaps modulating changes in real time cell behavior such as bacterial chemotaxis. To date, little is known about the function of sensory hydrogenase-like proteins in biology, and those that have been examined comprise NiFe hydrogenase-like moieties (78), not the [FeFe] hydrogenase H domain-like moiety observed in these termite gut treponemes. Previously, [FeFe] hydrogenase-like sensory proteins have been reported (59, 69, 81, 83), but these remain poorly studied. These sensors were found to be especially abundant in the termite gut metagenomic analysis (81); however, much less could be deduced about the modular structure and gene environment of the genes encoding those domains, due to the shrapnel based nature of that study. Thus, the cultured treponemes analyzed in this study become excellent candidates for examining the possible roles and functions of putative H₂ sensor proteins in gene regulation and cell behavior.

The H-domains of the putative H₂ sensor proteins fell within 4 phylogenetic clusters. The H-domains corresponding to Families 4 and 8 each were associated with a PAS domain in the C-terminal region of their respective protein sequences. A previously described and biochemically characterized NiFe hydrogenases-like H₂ sensor also encodes a PAS domain (14). Bacterial PAS domains, also referred to as LOV domains for their role in

sensing light, oxygen, or voltage, are usually found in sensor proteins of two-component regulatory systems (26, 74). Shaw *et al.* have recently reported a PAS domain containing H domain protein from *Thermoanaerobacterium saccharolyticum* and Posewitz *et al.* have reported proteins in *Halothermothrix orenii* with a region sharing homology simultaneously with PAS and histidine kinase domains (59, 69). PAS domain containing H domain proteins have also been observed in clostridia (15).

The H-domains corresponding to Family 10 had an arrangement of domains at their C-termini similar to the same region of the RcsC signal receptor protein from *E. coli* (39, 62). This suggests that this hydrogenase-like protein operates in a phosphorelay that alters gene transcription in response to H₂ (39). RcsC is known to be a cytoplasmic membrane protein; here PSORTb unambiguously predicts that the Family 10 H₂ sensor is also a cytoplasmic membrane protein.

Sequences belonging to H domain Family 5 contained a methyl-accepting chemotaxis protein domain (MA) at their C-termini. These can be postulated to modulate changes in swimming behavior in response to H₂ gradients (80, 81), although to our knowledge H₂-taxis has not yet been demonstrated in any bacterium. Alternatively, MA domain containing proteins have been found to influence gene regulation – see Box 4 in Wadhams' review (80). Methyl-accepting chemotaxis proteins are typically membrane-bound, and each MA domain containing protein from the treponemes was predicted by PSORTb to localize to the cytoplasmic membrane. Each of the three treponeme strains encoded an H domain protein belonging to Family 5.

Hydrogen sensors from termite gut microbes appear to have arisen as a late radiation from within the [FeFe] hydrogenase enzyme line of descent (Figure 2-2 & 2-4). A similar

pattern has been observed for the sensor proteins having moieties with homology to NiFe hydrogenases (77), suggesting that in each case, the sensory variants have arisen after a gene duplication with subsequent modification and radiation into a new niche.

Conclusions. H₂ is a central metabolite during the degradation of organic materials and in the physiologies of the symbiotic microbial communities residing in all termites examined (13, 18, 56, 66). Treponemes are among the most abundant bacterial groups comprising the gut communities in many termites. The 17 [FeFe] hydrogenases and hydrogenase-like proteins identified here in the genome sequences of three termite gut treponemes underscore the importance of H₂ to their tiny ecosystem. It is intriguing that these strains encode putative [FeFe] hydrogenase-like hydrogen sensors, a function only recently proposed for H domain containing proteins (81). This suggests that these spirochetes may have the ability to change their gene expression in response to (and perhaps even their physical positions along) hydrogen gradients encountered within the gut. Hydrogen and other chemical and pH gradients have previously been elucidated in termite guts (13). It has previously been suggested that perhaps it might be the ability of highly motile, homoacetogenic spirochetes to better position themselves between their sources of H₂ and their competitors that might help explain their otherwise enigmatic outcompetition for this electron donor with methanoarchaea (35). Thus, the current genome sequence results provide another dimension to our understanding, as well as avenues for future exploration of H₂ metabolism in high flux, H₂-rich environments.

References

1. **Andrews, S. C., B. C. Berks, J. McClay, A. Ambler, M. A. Quail, P. Golby, and J. R. Guest.** 1997. A 12-cistron *Escherichia coli* operon (*hyf*) encoding a putative proton-translocating formate hydrogenlyase system. *Microbiology* **143**:3633-3647.
2. **Aziz, R., D. Bartels, A. Best, M. DeJongh, T. Disz, R. Edwards, K. Formsma, S. Gerdes, E. Glass, M. Kubal, F. Meyer, G. Olsen, R. Olson, A. Osterman, R. Overbeek, L. McNeil, D. Paarmann, T. Paczian, B. Parrello, G. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, and O. Zagnitko.** 2008. The RAST Server: Rapid Annotations Using Sybsystems Technology. *BMC Genomics* **9**:75-90.
3. **Bailey, T. L., and M. Gribskov.** 1998. Combining evidence using p-values: Application to sequence homology searches. *Bioinformatics* **14**:48-54.
4. **Bateman, A., E. Birney, L. Cerruti, R. Durbin, L. Etwiller, S. R. Eddy, S. Griffiths-Jones, K. L. Howe, M. Marshall, and E. L. L. Sonnhammer.** 2002. The Pfam Protein Families Database. *Nucleic Acids Res.* **30**:276-280.
5. **Böck, A., P. W. King, M. Blokesch, and M. C. Posewitz.** 2006. Maturation of Hydrogenases. *Adv. Appl. Microbiol.* **51**:1-71.
6. **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.
7. **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, symbioses, ecology. Kluwer Academic Publishers, Boston.

8. **Breznak, J. A.** 1982. Intestinal microbiota of termites and other xylophagous insects. *Annu. Rev. Microbiol.* **36**:323-343.
9. **Breznak, J. A., and B. A.** 1994. Role of microorganisms in the digestion of lignocellulose by termites. *Annu. Rev. Entymology* **39**:453-487.
10. **Breznak, J. A., and J. M. Switzer.** 1986. Acetate Synthesis from H₂ plus CO₂ by Termite Gut Microbes. *Appl. Environ. Microbiol.* **52**:623-630.
11. **Brune, A.** 2006. Symbiotic Associations Between Termites and Prokaryotes. *Prokaryotes* **1**:439-474.
12. **Brune, A.** 1998. Termite guts: the world's smallest bioreactors. *Trends in Biotechnol.* **16**:16-21.
13. **Brune, A., and M. Friedrich.** 2000. Microecology of the termite gut: structure and function on a microscale. *Curr. Opin. Microbiol.* **3**:263-269.
14. **Buhrke, T., O. Lenz, A. Porthun, and B. Friedrich.** 2004. The H₂-sensing complex of *Ralstonia eutropha*: interaction between a regulatory [NiFe] hydrogenase and a histidine protein kinase. *Mol. Microbiol.* **51**:1677-1699.
15. **Calusinska, M., T. Happe, B. Joris, and A. Wilmotte.** 2010. The surprising diversity of clostridial hydrogenases: a comparative genomic perspective. *Microbiology* **156**:157-1588.
16. **Cleveland, L. R.** 1923. Correlation Between the Food and Morphology of Termites and the Presence of Intestinal Protozoa. *Am. J. Hyg.* **3**:444-461.

17. **Cleveland, L. R.** 1923. Symbiosis between Termites and Their Intestinal Protozoa. Proc. Natl. Acad. Sci. U.S.A. **9**:424-428.
18. **Ebert, A., and A. Brune.** 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.
19. **Eddy, S. R.** 1998. Profile hidden Markov models. Bioinformatics **14**:755-763.
20. **Fiala, G., and K. O. Stetter.** 1986. *Pyrococcus friosus* sp. nov. represents a novel genus of marine heterotrophic archaeobacteria growing optimally at 100°C. Arch. Microbiol. **145**:56-61.
21. **Frey, M.** 2002. Hydrogenases: Hydrogen-Activating Enzymes. ChemBioChem **3**:153-160.
22. **Gardy, J., M. Laird, F. Chen, S. Rey, C. Walsh, M. Ester, and F. Brinkman.** 2005. PSORTb v.2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. Bioinformatics **21**:617-23.
23. **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.
24. **Graentzdoerffer, A., D. Rauh, A. Pich, and J. R. Andreessen.** 2003. Molecular and biochemical characterization of two tungsten- and selenium-containing formate dehydrogenases from *Eubacterium acidaminophilum* that are associated with components of an iron-only hydrogenase. Arch. Microbiol. **179**:116-130.

25. **Hatchikian, C. E., M. Chaigneau, and J. Le Gall (ed.).** 1976. Analysis of gas production by growing culture of three species of sulfate-reducing bacteria. E. Goltze KG, Göttingen, Germany.
26. **Hefti, M. H., K.-J. François, S. D. De Vries, and J. Vervoort.** 2004. The PAS Fold. *Eur. J. Biochem.* **271**:1198-1208.
27. **Hoehler, T. M., B. M. Bebout, and D. J. Des Marais.** 2001. The role of microbial mats in the production of reduced gases on the early Earth. *Nature* **412**:324-327.
28. **Huber, R., T. A. Langworthy, H. König, M. Thomm, C. R. Woese, U. B. Sleytr, and K. O. Stetter.** 1986. *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C. *Arch. Microbiol.* **144**:324-333.
29. **Hungate, R. E.** 1939. Experiments on the Nutrition of Zootermopsis. III. The anaerobic carbohydrate dissimilation by the intestinal protozoa. *Ecology* **20**:230-245.
30. **Hungate, R. E.** 1943. Quantitative Analysis on the Cellulose Fermentation by Termite Protozoa. *Ann. Entomol. Soc. Am.* **36**:730-9.
31. **Inoue, J.-I., K. Saita, T. Kudo, S. Ui, and M. Ohkuma.** 2007. Hydrogen Production by Termite Gut Protists: Characterization of Iron Hydrogenases of Parabasalian Symbionts of the Termite *Coptotermes formosanus*. *Eukaryotic Cell* **6**:1925-1932.
32. **Leach, M. R., and D. B. Zamble.** 2007. Metallocenter assembly of the hydrogenase enzymes. *Curr. Opin. Chem. Biol.* **11**:159-165.

33. **Leadbetter, J. R.** 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**:3620-31.
34. **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.
35. **Leadbetter, J. R., T. M. Schmidt, J. R. Graber, and J. A. Breznak.** 1999. Acetogenesis from H₂ Plus CO₂ by Spirochetes from Termite Guts. *Science* **283**:686-689.
36. **Lilburn, T. G., K. S. Kim, N. E. Ostrom, K. R. Byzek, J. R. Leadbetter, and J. A. Breznak.** 2001. Nitrogen Fixation by Symbiotic and Free-Living Spirochetes. *Science* **292**:2495-2498.
37. **Lilburn, T. G., T. M. Schmidt, and J. A. Breznak.** 1999. Phylogenetic diversity of termite gut spirochaetes. *Environ. Microbiol.* **1**:331-345.
38. **Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Förster, I. Brettske, S. Gerber, A. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. König, T. Liss, R. Lüssmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N. Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode, and K. Schleifer.** 2004. ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**:1363-1371.
39. **Majdalani, N., and S. Gottesman.** 2006. The Rcs Phosphorelay: A Complex Signal Transduction System. *Annu. Rev. Microbiol.* **59**:379-405.

40. **Malki, S., I. Saimmaime, G. De Luca, M. Rousset, Z. Dermoun, and J.-P. Belaich.** 1995. Characterization of an Operon Encoding an NADP-Reducing Hydrogenase in *Desulfovibrio fructosovorans*. *J. Bacteriol.* **177**:2628-2636.
41. **Markowitz, V. M., N. N. Ivanova, E. Szeto, K. Palaniappan, K. Chu, D. Dalevi, I.-M. A. Chen, Y. Grechkin, I. Dubchak, I. Anderson, A. Lykidis, K. Mavromatis, P. Hugenholtz, and N. C. Kyrpides.** 2008. IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Res.* **36**:D534-D538.
42. **Markowitz, V. M., E. Szeto, K. Palaniappan, Y. Grechkin, K. Chu, I.-M. A. Chen, I. Dubchak, I. Anderson, A. Lykidis, K. Mavromatis, N. N. Ivanova, and N. C. Kyrpides.** 2008. The integrated microbial genomes (IMG) system in 2007: data content and analysis tool extensions. *Nucleic Acids Res.* **36**:D528-D533.
43. **Matson, E. G., X. Zhang, and J. R. Leadbetter.** 2010. Selenium controls transcription of paralogous formate dehydrogenase genes in the termite gut acetogen, *Treponema primitia*. *Environ. Microbiol.*
44. **Meyer, J.** 2007. [FeFe] hydrogenases and their evolution: a genomic perspective. *Cell. Mol. Life Sci.* **64**:1063-1084.
45. **Morgenstern, B.** 1999. DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. *Bioinformatics* **15**:211-218.
46. **Mulder, N., and R. Apweiler.** 2007. InterPro and InterProScan: tools for protein sequence classification and comparison. *Methods Mol. Biol.* **396**:59-70.

47. **Néron, B., H. Ménager, C. Maufrais, N. Joly, T. Pierre, and C. Letondal.** 2008. Presented at the Bio Open Source Conference, Toronto.
48. **Odelson, D. A., and J. A. Breznak.** 1985. Cellulase and Other Polymer-Hydrolyzing Activities of *Trichomitopsis termopsidis*, a Symbiotic Protozoan from Termites. *Appl. Environ. Microbiol.* **49**:622-626.
49. **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.
50. **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.
51. **Olliver, B., R. Cord-Ruwisch, E. C. Hatchikian, and J. L. Garcia.** 1988. Characterization of *Desulfovibrio fructosovorans* sp. nov. . *Arch. Microbiol.* **149**:447-450.
52. **Overbeek, R., T. Begley, R. Butler, J. Choudhuri, H. Chuang, M. Cohoon, V. de Crécy-Lagard, N. Diaz, T. Disz, R. Edwards, M. Fonstein, E. Frank, S. Gerdes, E. Glass, A. Goesmann, A. Hanson, D. Iwata-Reuyl, R. Jensen, N. Jamshidi, L. Krause, M. Kubal, N. Larsen, B. Linke, A. McHardy, F. Meyer, H. Neuweger, G. Olsen, R. Olson, A. Osterman, V. Portnoy, G. Pusch, D. Rodionov, C. Rückert, J. Steiner, R. Stevens, I. Thiele, O. Vassieva, Y. Ye, O. Zagnitko, and V. Vonstein.** 2005. The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* **33**:5691-702.

53. **Paster, B. J., and F. E. Dewhirst.** 2000. Phylogenetic foundation of spirochetes. *J. Mol. Microbiol. Biotechnol.* **2**:341-4.
54. **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**:347-352.
55. **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.
56. **Pester, M., and A. Brune.** 2007. Hydrogen is the central free intermediate during lignocellulose degradation by termite gut symbionts. *ISME J.* **1**:551-565.
57. **Pietrokovski, S., J. G. Henikoff, and S. Henikoff.** 1998. Exploring protein homology with the Blocks server. *Trends . Genet.* **14**:162-163.
58. **Posewitz, M. C., P. W. King, S. L. Smolinski, L. Zhang, M. Seibert, and M. L. Ghirardi.** 2004. Discovery of Two Novel Radical *S*-Adenosylmethionine Proteins Required for the Assembly of an Active [Fe] Hydrogenase. *J. Biol. Chem.* **279**:25711-25720.
59. **Posewitz, M. C., D. W. Mulder, and J. W. Peters.** 2008. New frontiers in hydrogenase structure and biosynthesis. *Curr. Chem. Bio.* **2**:178-199.
60. **Ragsdale, S. W.** 2008. Enzymology of the Wood-Ljungdahl Pathway of Acetogenesis. *Ann. N. Y. Acad. Sci.* **1125**:129-136.
61. **Rieu-Lesme, F., C. Dauga, G. Fonty, and J. Dore.** 1998. Isolation from the Rumen of a New Acetogenic Bacterium Phylogenetically Closely Related to *Clostridium difficile*. *Anaerobe* **4**:89-94.

62. **Rogov, V. V., N. Y. Rogova, F. Bernhard, A. Koglin, F. Löhr, and V. Dötsch.** 2006. A New Structural Domain in the *Escherichia coli* RcsC Hybrid Sensor Kinase Connects Histidine Kinase and Phosphoreceiver Domains. *J. Mol. Biol.* **364**:68-79.
63. **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.
64. **Sauter, M., R. Böhm, and A. Böck.** 1992. Mutational analysis of the operon (*hyc*) determining hydrogenase 3 formation in *Escherichia coli*. *Mol. Microbiol.* **6**:1523-1532.
65. **Schink, B., F. S. Lupton, and J. G. Zeikus.** 1983. Radioassay for Hydrogenase Activity in Viable Cells and Documentation of Aerobic Hydrogen-Consuming Bacteria Living in Extreme Environments. *Appl. Environ. Microbiol.* **45**:1491-1500.
66. **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.
67. **Schwartz, E., and B. Friedrich.** 2006. The H₂-Metabolizing Prokaryotes. *Prokaryotes* **2**:496-563.
68. **Scranton, M. I., P. C. Novelli, and P. A. Loud.** 1984. The distribution and cycling of hydrogen gas in the waters of two anoxic marine environments. *Limnol. Oceanogr.* **29**:993-1003.

69. **Shaw, A. J., D. A. Hogsett, and L. R. Lynd.** 2009. Identification of the [FeFe]-hydrogenase responsible for hydrogen generation in *Thermoanaerobacterium saccharolyticum* and demonstration of increased ethanol yield via hydrogenase knockout. *J. Bacteriol.* **20**:6457-64.
70. **Smolenski, W. J., and J. A. Robinson.** 1988. In situ rumen hydrogen concentrations in steers fed eight times daily, measured using a mercury reduction detector. *FEMS Microbiol. Ecol.* **53**:95-100.
71. **Soboh, B., D. Linder, and R. Hedderich.** 2004. A multisubunit membrane-bound [NiFe] hydrogenase and an NADH-dependent Fe-only hydrogenase in the fermenting bacterium *Thermoanaerobacter tengcongensis*. *Microbiology* **150**:2451-2463.
72. **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, ecology*. Kluwer Academic Publishers, Boston.
73. **Sugimoto, A., and N. Fujita.** 2006. Hydrogen concentrations and stable isotopic composition of methane in bubble gas observed in a natural wetland. *Biogeochemistry* **81**:33-44.
74. **Taylor, B. L., and I. B. Zhulin.** 1999. PAS Domains: Internal Sensors of Oxygen, Redox Potential, and Light. *Microbiol. Mol. Biol. Rev.* **63**:479-506.
75. **Traoré, A. S., E. C. Hatchikian, J. P. Belaich, and J. Le Gall.** 1981. Microcalorimetric studies of the growth of sulfate-reducing bacteria: energetics of *Desulfovibrio vulgaris* growth. *J. Bac.* **145**:191-199.

76. **Verhagen, M. F. J. M., T. O'Rourke, and M. W. W. Adams.** 1999. The hypothermophilic bacterium, *Thermotoga maritima*, contains an unusually complex iron-hydrogenase: amino acid sequence analyses versus biochemical characterization. *Biochim. Biophys. Acta* **1412**:212-229.
77. **Vignais, P. M., and B. Billoud.** 2007. Occurrence, Classification, and Biological Function of Hydrogenases: An Overview. *Chem. Rev.* **107**:4206-4272.
78. **Vignais, P. M., B. Billoud, and J. Meyer.** 2001. Classification and phylogeny of hydrogenases. *FEMS Microbiol. Rev.* **25**:455-501.
79. **Vu, A. T., N. C. Nguyen, and J. R. Leadbetter.** 2004. Iron reduction in the metal-rich guts of wood-feeding termites. *Geobiology* **2**:239-247.
80. **Wadhams, G. H., and J. P. Armitage.** 2004. Making Sense of it All: Bacterial Chemotaxis. *J. Mol. Cell Biol.* **5**:1024-1037.
81. **Warnecke, F., P. Luginbühl, N. Ivanova, M. Ghassemian, T. H. Richardson, J. T. Stege, M. Cayouette, A. C. McHardy, G. Djordjevic, N. Aboushadi, R. Sorek, S. G. Tringe, M. Podar, H. G. Martin, V. Kunin, D. Dalevi, J. Madejska, E. Kirton, D. Platt, E. Szeto, A. Salamov, K. Barry, N. Mikhailova, N. C. Kyrpides, E. G. Matson, E. A. Ottesen, X. Zhang, M. Hernández, C. Murillo, L. G. Acosta, I. Rigoutsos, G. Tamayo, B. D. Green, C. Chang, E. M. Rubin, E. J. Mathur, D. E. Robertson, P. Hugenholtz, and J. R. Leadbetter.** 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**:560-569.

82. **Weidner, U., S. Geier, A. Ptock, T. Friedrich, H. Leif, and H. Weiss.** 1993. The Gene Locus of the Proton-translocating NADH: Ubiquinone Oxidoreductase in *Escherichia coli*. *J. Mol. Biol.* **233**:109-122.
83. **Wu, M., Q. Ren, A. Durkin, S. Daugherty, L. Brinkac, R. Dodson, R. Madupu, S. Sullivan, J. Kolonay, D. Haft, W. Nelson, L. Tallon, K. Jones, L. Ulrich, J. Gonzalez, I. Zhulin, F. Robb, and J. Eisen.** 2005. Life in hot carbon monoxide: the complete genome of *Carboxythermus hydrogenoformans* Z-2901. *PLOS Genetics* **1**:563-574.
84. **Xue, Y., Y. Xu, Y. Liu, Y. Ma, and P. Zhou.** 2001. *Thermoanaerobacter tengcongensis* sp. nov., a novel anaerobic, saccharolytic, thermophilic bacterium isolated from a hot spring in Tengcong, China. *Int. J. Syst. Evol. Microbiol.* **51**:1335-1341.
85. **Yamin, M. A.** 1979. Cellulolytic activity of an axenically-cultivated termite flagellate, *Trichomitopsis termopsidis*. *J. Gen. Microbiol.* **113**:417-20.
86. **Yamin, M. A.** 1981. Cellulose Metabolism by the Flagellate *Trichonympha* from a Termite Is Independent of Endosymbiotic Bacteria. *Science* **211**:58-59.

*Chapter 3*A PHYLOGENETIC ANALYSIS OF [FeFe] HYDROGENASE GENE DIVERSITY IN THE HYDROGEN METABOLIZING GUTS OF LOWER TERMITES AND ROACHES REVEALS UNIQUE, ECOSYSTEM-DRIVEN, ADAPTATIONS AND SIMILARITY OF *CRYPTOCERCUS* AND LOWER TERMITE GUT COMMUNITIES**Abstract**

Hydrogen is an important free intermediate in the breakdown of wood by termite gut microbial communities, reaching concentrations in some species exceeding those measured for any other biological system. We have designed and utilized degenerate primers for the study of [FeFe] hydrogenase evolution and representation in the gut ecosystems of roaches and lower termites. The primers target with specificity the largest group of enzymatic [FeFe] hydrogenases identified in a termite gut metagenome (Warnecke, F., *et al.* 2007. *Nature* 450: 560-569). Sequences were cloned from the guts of lower termites, *Incisitermes minor*, *Zootermopsis nevadensis*, and *Reticulitermes hesperus*, and two roaches, *Cryptocercus punctulatus* and *Periplaneta americana*. All termite and *Cryptocercus* sequences were phylogenetically distinct from non-termite associated hydrogenases available in public databases. This may be a consequence of unique adaptations to their respective ecosystems. The abundance of unique sequence OTUs cloned, as many as 21 from each species, highlights the physiological importance of hydrogen to the gut ecosystems of wood feeding insects. The diversity of sequences observed may be reflective of multiple niches to which the enzymes have adapted. Sequences cloned from *Cryptocercus* and the lower termite samples, all wood feeding insects, clustered closely with one another in phylogenetic and Unifrac analyses to the exclusion of those from *P. americana*, an omnivorous roach. These results provide

evidence for the importance of hydrogen metabolism to the gut ecosystems of wood feeding insects. Moreover, they provide support for a close evolutionary relationship of lower termites to wood roaches and a common origin of their symbiotic microbial communities.

Introduction

Hydrogen plays a prominent role in the digestion of wood by termites (1, 6, 13, 39, 40, 44). Hydrogen concentrations in the guts of some termites can reach concentrations exceeding those measured for any other biological system (13, 16, 44, 46, 48, 49, 51). Turnover rates have been measured in some species at fluxes as high as 33 m³/m³ gut volume (44). The environment is also spatially complex, comprising a matrix of microenvironments characterized by different hydrogen concentrations (8, 9, 13, 24, 25, 44).

This hydrogen is produced during the fermentation of lignocellulosic polysaccharides by the symbiotic microbial community residing in the termite gut, particularly the protozoa (15, 17, 18, 40, 53, 57, 58). The termites are dependent upon this complex symbiosis for the degradation of wood (2-4, 7, 10, 11, 39). The primary product of this symbiosis is acetate, which the termites use as their primary carbon and energy source (41). Most of the hydrogen produced in the gut is used by CO₂-reducing bacteria to produce up to 1/3 of this acetate in reductive acetogenesis (1, 6, 26, 41, 44). Methanoarchaea consume only a small portion of this hydrogen (1, 24).

The role of termites in global carbon cycling is well established (50, 61). It is, therefore, of interest to further investigate factors influencing how the gut ecosystem processes hydrogen so efficiently. Indeed, the termite gut has been reported as the smallest, most efficient natural bioreactor degradation system known (44).

A rich diversity of hydrogenases were identified in the recently published *Nasutitermes* gut metagenome (56). The vast majority of the hydrogenases – over 99% – were classified as [FeFe] hydrogenases (56). *Nasutitermes* is a member of a group of termites known as higher termites, which are distinguishable from lower termites by their

characteristic lack of protozoa in their gut and by their more extensively segmented gut anatomy (10, 21, 22, 27). Chapter 2 reports a total of 17 [FeFe] hydrogenase-like genes in the genome sequences of three treponemes isolated from the gut of *Zootermopsis angusticolis* indicating that lower termites too may be a rich source of [FeFe] hydrogenase diversity.

Wood roaches, *Cryptocercus punctulatus*, are generally believed to share their most recent common ancestor with all termites (20, 27, 30). In fact, termites have been referred to as eusocial cockroaches (20). The gut ecosystem of *Cryptocercus* shares a number of characteristics with termites. For example, they and termites are dependent upon a complex mutualism with a microbial community in their gut to be able to derive nutrition and energy from wood (5, 23, 42, 53). The predominant microbes found in the cockroach gut are similar to those found in termites and, more specifically, protozoa are believed to play an important role in this symbiosis (5, 23, 53). It is for this reason, and as a consequence of evolutionary relationships (20), that *Cryptocercus* are most specifically similar to a group of termites classified as lower termites. Moreover, the *Cryptocercus* gut is anatomically similar to the lower termite gut (38).

The similarities of *Cryptocercus* to lower termites may extend to the metabolic activities of their gut microbial communities. Hydrogen concentration profiles have been quantified for the gut of the roach, *Blaberus* sp., and *Cryptocercus* may harbor similar profiles along its gut (28). The gut microbes of *Cryptocercus* are capable of carrying out reductive acetogenesis implying that hydrogen produced in the gut may be utilized in acetate genesis (6, 28).

Here we report a phylogenetic analysis of [FeFe] hydrogenase genes cloned from the guts of roaches and lower termites using degenerate primers. The objective was to better understand the diversity, adaptation, and evolution of the genes in these hydrogen-metabolizing ecosystems.

Methods

Termites. *Incisitermes minor* collection Pas1 termites were collected from a woodpile in Pasadena. *Reticulitermes Hesperus* collection ChiA2 and *Zootermopsis nevadensis* collection ChiA1 were collected at Chilao National Park in Southern California.

Termites were classified previously (43, 60) using insect mitochondrial cytochrome oxidase subunit II (COXII) gene sequences (43). The COXII genes were amplified directly from the DNA samples that hydrogenases were cloned from. COXII was amplified using the primers CI-J-1773 and B-tLys and cycling conditions described by Miura *et al.* (35) where FailSafe PremixD (Epicentre) and Expand High Fidelity Taq (Roche) were substituted for the polymerase and buffers, respectively. Sequences were edited and analyzed in the same manner as that described below for cloned [FeFe] hydrogenase sequences. Samples were identified as belonging to the genus of the termite harboring the COXII sequence to which they were found most near in phylogenetic analyses.

Cryptocercus punctulatus were kindly provided by Christina Nalepa (NC State University). The adult sample was from a roach collected at Mt. Collins, and the nymph sample was collected at the South Mountains. *Periplaneta americana* (HM208259) was collected on the Caltech campus and identified as belonging to the genus of the roach harboring the COXII sequence to which it was found most near in phylogenetic analyses (43).

DNA Extraction. DNA was extracted from whole dissected guts as described previously (33). DNA concentrations were quantified using the Hoefer DyNAQuant 200 fluorometer and DNA quantification system (Amersham Pharmacia Biotech) according to manufacturer instructions.

Primer Design. Degenerate primers for the amplification of [FeFe] hydrogenases classified as belonging to “Family 3” by Warnecke *et al.* (56) were designed manually from a multiple-sequence alignment, see Figure 3-S1 in the appendix to this chapter. Family 3 [FeFe] hydrogenases, first described by Warnecke *et al.* (56), were the most highly represented group of enzymatic hydrogenases observed in the *Nasutitermes* hindgut metagenome sequence and have also been observed in the genome sequences of treponemes isolated from the gut of *Zootermopsis angusticolis*, see Chapter 2. To highlight their physiological relevance, Family 3 [FeFe] hydrogenases were the only group of hydrogenases observed in the *Nasutitermes* gut metagenome whose *in situ* translation was verified by mass spectroscopy (56).

Sequences were aligned using ClustalX available on the PBIL network protein sequence analysis server (12). Included in the alignment were the two Family 3 [FeFe] hydrogenase sequences previously identified in the genome sequences of two treponemes isolated from a termite gut, see Chapter 2, and 9 Family 3 sequences identified in the gut metagenome sequence of *Nasutitermes* (56). The [FeFe] hydrogenases of *Desulfovibrio vulgaris* and *Clostridium pasteurianum* were included in the alignment because they are the best characterized [FeFe] hydrogenases (37, 45). Also included in the alignment were top BLAST hits identified using the termite gut treponeme Family 3 [FeFe] hydrogenase sequences identified in the genomes of the treponemes isolated from a termite gut as

queries against GeneBank. Sequences not having a termite origin were included in the alignment to identify regions conserved across a broad evolutionary range. Upon identifying these highly conserved regions, the consensus of the termite sequences in this region was used for primer design.

A functional primer set and optimal conditions for gene amplification were determined empirically. The primers amplify approximately 537 bp, or 51%, of the H domain (34, 54, 55) known to be highly conserved among all [FeFe] hydrogenases. The amplified region corresponds approximately to the regions spanning T330-I494 and A209-I373 in the [FeFe] hydrogenases from *C. pasteurianum* (P29166) and *D. vulgaris* (YP_010987), respectively. The sequences for the forward and reverse primers were WSICCCARCARATGATGG and CCIKRCAIGCCATIACYTC, respectively, where “I” represents inositol. The peptide sequences targeted by the primers are highlighted in Figure 3-S1 found in the appendix of this chapter.

Cloning. Primers were ordered from IDT DNA. Gene sequences were amplified from template DNA using Expand High Fidelity Taq Polymerase (Roche), FailSafe Premix D (Epicentre) and 0.1 or 10 ng of template DNA. The temperature cycling regimen was 5 min at 95°C, 35 x (30 s at 95°C, 30 s at 53°C, 1 min at 72°C), 10 min at 72°C, and final cooling to 4°C. It was necessary to use 50 cycles to successfully clone sequences from the *P. americana* sample.

Sequences amplified were cloned into TOP10 chemically competent *E. coli* (Invitrogen) using the TOPO TA cloning kit (Invitrogen) according to manufacturer instructions.

RFLP Analysis. 96 clones were randomly selected for analysis. Each clone was suspended in TE (Sigma) and used as a template for PCR. The cloned sequences were

amplified by PCR using T7 and T3 primers, NEB Taq Polymerase (New England Biolabs) and FailSafe Buffer H (Epicentre). The temperature cycling regimen was 5 min at 95°C, 25 x (30 s at 95°C, 30 s at 55°C, 1.5 min at 72°C), 10 min at 72°C followed, and final cooling to 4°C.

The products of each of these reactions were then subjected to digestion with HinPII and the resulting restriction fragment length polymorphism (RFLP) patterns were analyzed by agarose gel electrophoresis.

Sequencing. For each termite sample analyzed, cloned sequences representing each unique RFLP pattern observed were arbitrarily selected for sequencing. Plasmids were purified using a QIAprep Spin Miniprep Kit (Qiagen) and submitted to Davis Sequencing for sequencing. The sequences obtained were manually trimmed in SeqMan, available from DNA* as part of the Lasergene software suite, to remove the plasmid and degenerate primer sequences.

The identity of each sequence as a hydrogenase was verified by BLASTing it against GeneBank.

Phylogenetic Analysis. An operational taxonomic unit (OTU) was defined as those peptide sequences sharing a minimum of 97% sequence identity. Sequences were grouped into OTUs using the furthest-neighbor algorithm in DOTUR (47).

The ARB software environment (32) was used for phylogenetic analysis of hydrogenase sequences. Sequence alignments were prepared using DIALIGN on the Moby server (36). Trees were constructed using 173 unambiguously aligned amino acid positions with distance matrix (Fitch), maximum parsimony (Phylip PROTPARS), and maximum likelihood (PhylipPROML) treeing methods. The sequence database used within ARB

contained 183 publically available protein sequences harboring H domains. Many of the [FeFe] hydrogenase sequences were chosen from those highlighted in reviews by Meyer (34) or Vignais (54). A number of sequences were identified by BLAST searches against the NCBI GenBank non-redundant protein sequences database. The database also included four protist [FeFe] hydrogenase sequences from the gut of *Coptotermes formosanus* (19). 84 sequences of the 123 identified as containing H domains in the termite gut metagenome database were of sufficient length to be included in the analysis. The following sequences comprised the outgroup used to construct Figures 3-3, 3-4 and 3-5: *Pseudotrichonympha grassii* (AB331668); uncultured parabasilid (AB331670); *Holomastigotoides mirabile* (AB331669). The following Family 3 [FeFe] hydrogenase sequences reported elsewhere, were also used to construct Figures 3-3, 3-4 and 3-5: *Treponema primitia* strain ZAS-2 (HndA1, see Chapter 2); *Treponema azotonutricium* strain ZAS-9 (HndA, see Chapter 2); *Nasutitermes* sp. gut (2004084376, JGI gene object ID (56)).

Diversity and Sequence Richness Calculations. Chao1 sequence richness and Shannon diversity indices for each clone set were calculated using EstimateS version 8.0.0 for Macintosh computers, written and made freely available by Robert K. Colwell (<http://viceroy.eeb.uconn.edu/EstimateS>). OTUs and their respective sequence abundances were used as input for the program. To visualize the evenness and diversity of OTUs sequences represented in each clone library, collector's curves were constructed showing the number of sequences represented by each OTU.

Community Comparisons. Unifrac (31) was used for quantitative comparisons of [FeFe] hydrogenase sequence libraries cloned from each insect sample. A maximum

likelihood phylogenetic tree was prepared as described above. Each sequence library was designated as a unique environment in the environment file used as an input to Unifrac and the file was also used to input abundance weights. The environments were compared using the Unifrac jackknife and principle component analysis calculations. Normalized abundance weights were used in each calculation. The jackknife calculation was completed with 1000 samplings and using 75% of the OTUs contained in the smallest environment sample input as the minimum number of sequences to keep.

Results

[FeFe] Hydrogenases Cloned. At least 28 unique RFLP patterns were cloned from each termite sample, see Table 3-1. The sequences could be grouped into 16, 20 and 21 OTUs for the *Incisitermes minor*, *Reticulitermes hesperus*, and *Zootermopsis nevadensis* clone sets, respectively. Sequences representing 28 and 37 unique RFLP patterns were cloned from the *C. punctulatus* adult and nymph samples, respectively. Their corresponding sequences could be grouped into 15 and 17 OTUs, respectively. 14 RFLP patterns were cloned from the *P. americana* sample, which could be grouped into 8 OTUs.

Collector's cures are provided as Figures 3-1 and 3-2. The Shannon diversity index and Chao1 species richness index for each sample are listed in Table 3-1. A list of all sequences cloned in this study and their respective abundance weights is provided as Table 3-S1.

Phylogenetic analysis of cloned sequences. In phylogenetic analyses, all hydrogenases from the lower termite and *Cryptocercus* samples grouped within a single clade separate from all previously sequenced non-termite associated bacterial [FeFe] hydrogenases in the database, tree not shown. This clade contained both of the Family 3 [FeFe] hydrogenases previously identified, see Chapter 2, in the genomes of treponemes isolated

Table 3-1. Quantifying hydrogenase clone library diversity.

	RFLPs^a	OTUs^b	Chao1 Mean^c	Chao1 95% CI Lower Bound^c	Chao1 95% CI Upper Bound^c	Shannon Mean^d
<i>C. punctulatus</i> Adult	28	15	27.5	16.67	108.84	2
<i>C. punctulatus</i> Nymph	37	17	25.17	18.57	59.42	2.13
<i>I. minor</i>	28	16	17.13	16.13	26.02	2.46
<i>P. americana</i>	14	8 ^e	5.25	5.01	9.73	1.27
<i>R. hesperus</i>	32	20	26.75	21.55	49.45	2.12
<i>Z. nevadensis</i>	37	21	41.25	24.97	124.16	2.57

^aNumber of unique restriction fragment polymorphism patterns (RFLPs) observed.

^bNumber of operational taxonomic units (OTUs); calculated using the furthest-neighbor method and a 97% amino-acid sequence similarity cut-off.

^cChao1 species-richness index calculated using the classic method in EstimateS. OTUs representing Family 3 [FeFe] hydrogenases with their respective abundances were used as program inputs.

^dShannon diversity index calculated using EstimateS. OTUs representing Family 3 [FeFe] hydrogenases with their respective abundances were used as program inputs.

^eOnly 5 of these OTUs represented Family 3 [FeFe] hydrogenase sequences, see Table 3-S1, and were used in the calculation of diversity indices.

Figure 3-1.

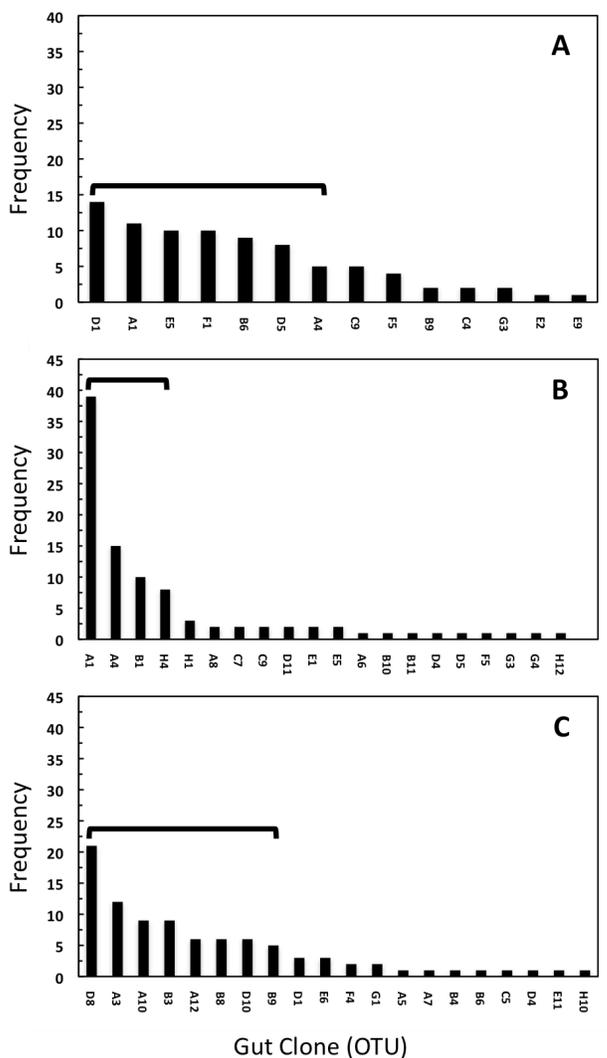


Figure 3-1. Collector's curves for lower termite samples. The horizontal brackets in each figure indicate the number of OTUs comprising 75% of all sequences cloned. Each bin represents an OTU calculated using the furthest-neighbor method in DOTUR (47) with a minimum of 97% amino-acid similarity used as a cut-off. A) *Incisitermes minor*, B) *Reticulitermes hesperus*, C) *Zootermopsis nevadensis*.

Figure 3-2.

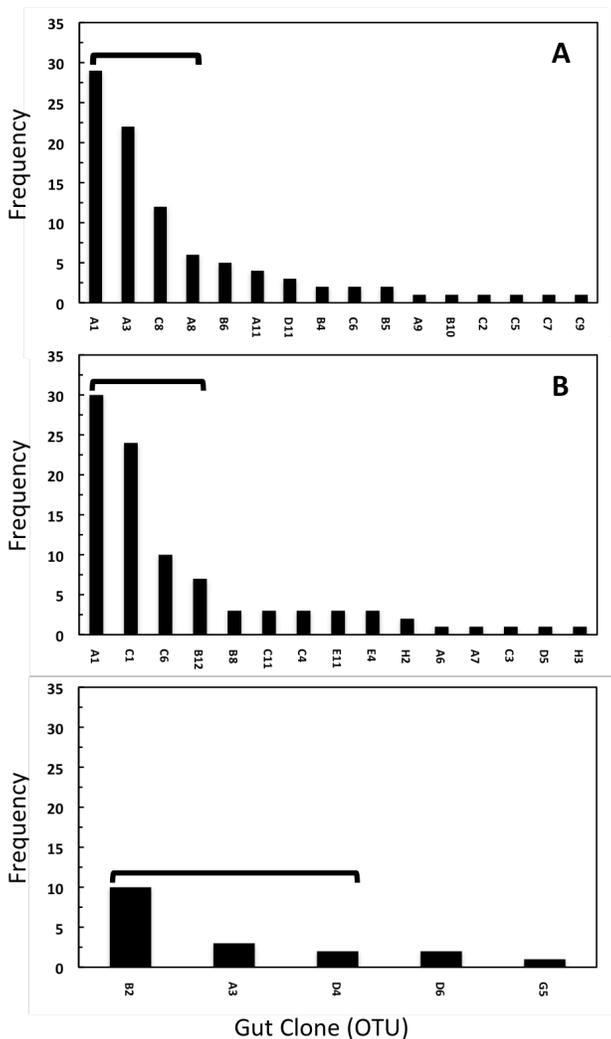


Figure 3-2. Collector's curves for roach samples. The horizontal brackets in each figure indicate the number of OTUs comprising 97% of all sequences cloned. Each bin represents an OTU calculated using the furthest-neighbor method in DOTUR (47) with a 97% amino-acid similarity cut-off. A) *Cryptocercus punctulatus* Nymph, B) *Cryptocercus punctulatus* Adult, C) *Periplaneta americana*.

from a lower termite gut, *T. azotonutricum* ZAS-9 and *T. primitia* ZAS-2. Three of the sequence OTUs from *P. americana* fell outside of this clade. The families, as defined by Warnecke *et al.*, into which these “outlying” sequences fell are provided as a footnote to Table 3-S1.

Phylogenetic analysis. Maximum likelihood trees for all termite, all roach, or the collective set of all Family 3 [FeFe] hydrogenase sequence OTUs cloned in this study are provided as Figures 3-3, 3-4, and 3-5, respectively. In an analysis of all Family 3 [FeFe] hydrogenase sequences cloned in this study, Family 3 hydrogenases taken from the genome sequences of *T. primitia* ZAS-2 and *T. azotonutricum* ZAS-9 each formed coherent clades with sequence OTUs from *Zootermopsis nevadensis*.

Community comparisons. Unifrac jackknife and principle component analyses were used to cluster the [FeFe] hydrogenase sequences cloned from each insect sample. Consistent with qualitative observations drawn from phylogenetic analyses, see Figures 3-4 and 3-5, the *P. americana* sequence community clustered to the exclusion of all other Family 3 [FeFe] hydrogenase sequences, see Figures 3-6 and 3-7. The *C. punctulatus* and lower termite samples clustered closely with each other.

Discussion

Sequence diversity and phylogeny. An analysis of cloned sequences representing the largest family of [FeFe] hydrogenases observed in a termite gut metagenome, and the only family verified by mass spectroscopy to be translated *in situ* (56), has revealed that the guts of lower termites and woodroaches are rich reservoirs of [FeFe] hydrogenase sequence diversity uniquely adapted to these small ecosystems. All sequence OTUs grouped together to the exclusion of all other sequences in our database, data not shown. This indicates that, as has been proposed for other environmental samples, the gut [FeFe]

Figure 3-3.

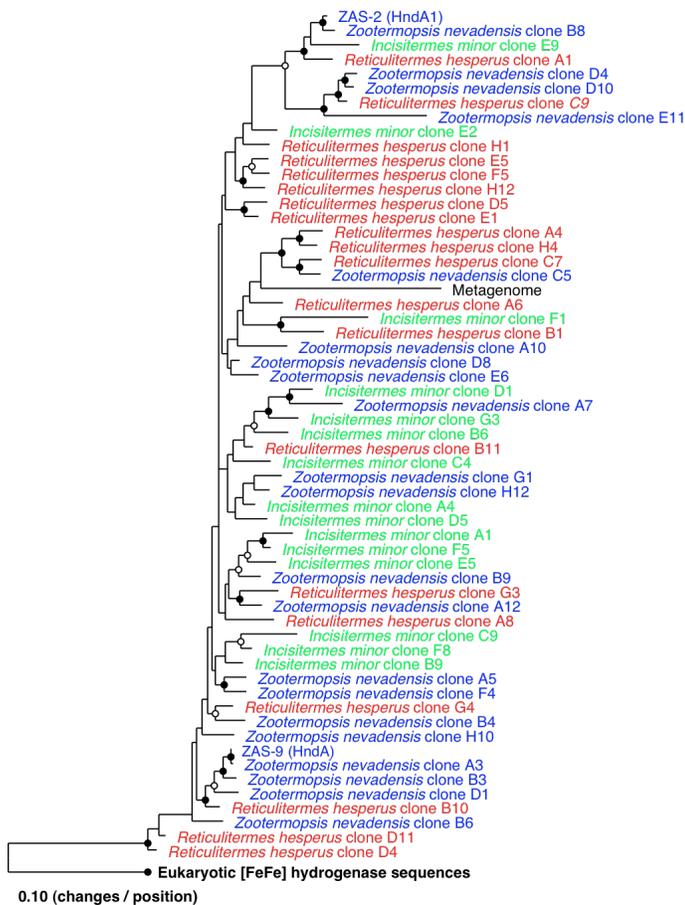


Figure 3-3. Phylogram of Family 3 [FeFe] hydrogenases cloned from the guts of lower termites. The tree was calculated using a maximum likelihood (Phylip ProML) method with 173 unambiguously aligned amino acid positions. Open circles designate groupings also supported by either parsimony (Phylip PROTPARS, 1000 bootstraps) or distance matrix (Fitch) methods. Closed circles designate groupings supported by all three methods. Hydrogenase sequences taken from *T. primitia* ZAS-2 and *T. primitia* ZAS-9 are labeled as ZAS-2 (HndA1) and ZAS-9 (HndA), respectively. The sequence labeled as “Metagenome” corresponds to the sequence with the gene identifier 2004084376 taken from a termite hindgut metagenome sequence (56).

Figure 3-4.

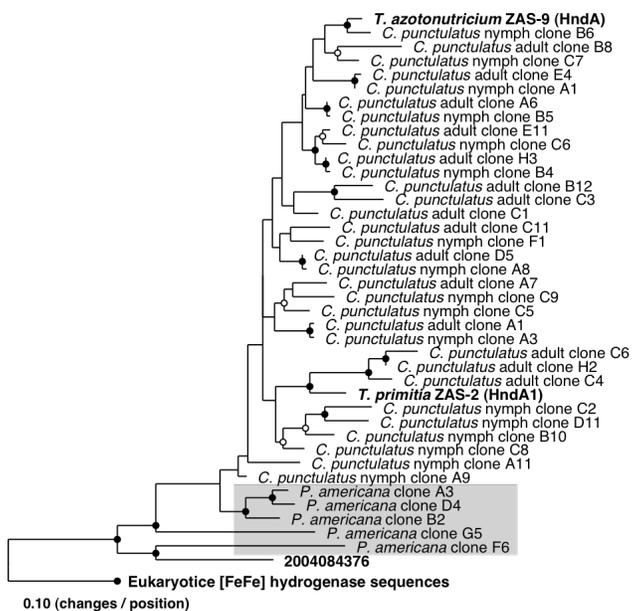


Figure 3-4. Phylogram of Family 3 [FeFe] hydrogenases cloned from the guts of an Adult and Nymph *C. punctulatus* samples. See Figure 3-3 caption for description of open and closed black circles and tree construction methods. All sequences cloned from *P. americana* are highlighted by a grey box.

Figure 3-5.

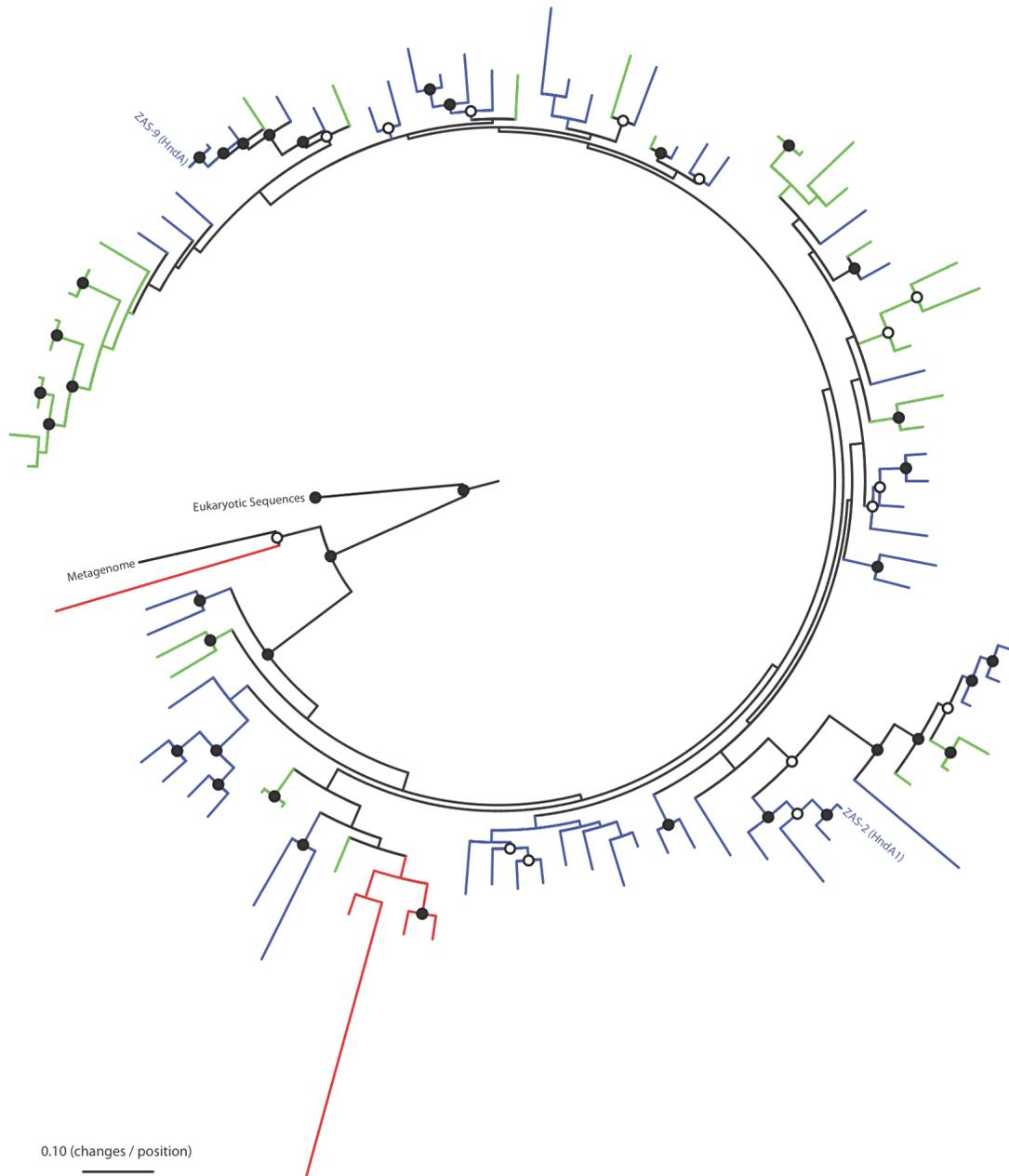


Figure 3-5. Maximum likelihood tree of all cloned Family 3 [FeFe] hydrogenase sequences. See Figure 3-3 caption for description of open and closed black circles and tree construction methods. Each leaf represents an OTU. Leaves and branches representing OTUs cloned from lower termites are in blue, from *C. punctulatus* are in green, and from *P. americana* are in red. Hydrogenase sequences taken from *T. primitia* ZAS-2 and *T. primitia* ZAS-9 are labeled as ZAS-2 (HndA1) and ZAS-9 (HndA), respectively. The sequence labeled as “Metagenome” corresponds to the sequence with the gene identifier 2004084376 taken from a termite hindgut metagenome sequence (56). Tree drawn using Phylip drawgram (14).

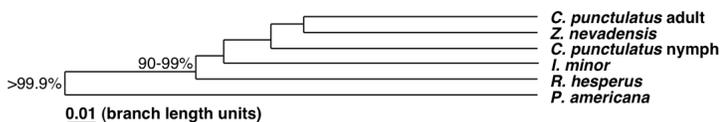
Figure 3-6.

Figure 3-6. Unifrac jackknife clustering of all cloned Family 3 [FeFe] hydrogenase sequences. The maximum-likelihood tree shown in Figure 3-5 and the OTUs with their respective abundance weights given in Table 3-S1 were used as inputs to Unifrac. The analysis was completed using normalized abundance weights, 1000 samplings, and keeping a number of sequences equal to 75% of the number of OTUs represented by the smallest sample analyzed (4 sequences). Each insect gut sample was designated as a unique environment in calculations. The numbers designate the percentage of samplings supporting a particular cluster.

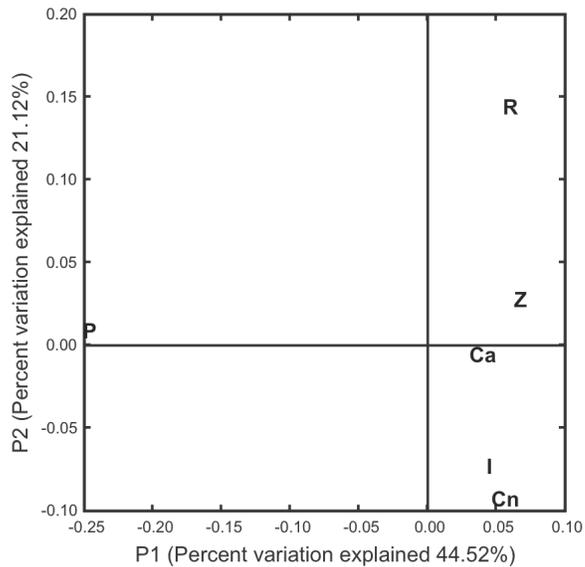
Figure 3-7.

Figure 3-7. Unifrac principle components analysis of all cloned Family 3 [FeFe] hydrogenase sequences. The maximum-likelihood tree shown in Figure 3-5 and the OTUs with their respective abundance weights given in Table 3-S1 were used as inputs to Unifrac. The analysis was completed using normalized abundance weights. Each insect gut sample was designated as a unique environment in calculations. Ca = *C. punctulatus* Adult, Cn = *C. punctulatus* Nymph, I = *Incisitermes minor*, P = *P. americana*, R = *Reticulitermes hesperus*, Z = *Zootermopsis nevadensis*.

hydrogenases are uniquely adapted to their respective ecosystems. Further analysis revealed a diversity of sequence clades, see Figures 3-3, 3-4 and 3-5. There is good reason to believe that this is only a portion of a much larger diversity present in the guts because only one of a total of 9 families of [FeFe] hydrogenases reported in the *Nasutitermes* gut metagenome sequence was targeted in this analysis (56). Each coherent clade may represent an adaptation to a niche or microenvironment shown previously to exist in the guts of termites and roaches and have a measurable influence on bacterial community structure (7, 9, 13, 28, 59). The sequence diversity shared among these samples may point toward metabolic similarities of their respective symbiotic gut microbial communities. The comparable lack of sequence diversity observed in *P. americana*, further emphasized by a necessary increase in PCR cycle number to be able to clone any sequences, may reflect a corresponding fundamental metabolic difference from the other communities sampled. This makes sense because *P. americana* is an omnivorous insect that does not consume wood, whereas the other insects sampled are wood feeders.

As shown in Figures 3-3 and 3-5, the Family 3 hydrogenase sequences from the genome sequences of *T. azotonutricium* ZAS-9 and *T. primitia* ZAS-1 each fall within coherent clades containing sequences from a *Zootermopsis* gut community. This is what one might expect because both of the treponemes were originally isolated from *Zootermopsis angusticolis* (26, 29). This provides further support for the relevance of these treponemes and their respective hydrogenases to the ecology of the termite gut ecosystem from which they were isolated.

Community cross-comparisons. There was a clear separation between the *P. americana* sequences and all other cloned [FeFe] hydrogenase sequences in phylogenetic analyses (see Figures 3-4 and 3-5). This is what one might predict based upon the gross nutritional differences between *P. americana*, which is an omnivorous household roach, and the *Cryptocercus* and lower termite samples, which are all wood feeders. Unifrac analyses (see Figures 3-6 and 3-7) provided quantitative support for this observed separation and pointed toward a close similarity of the sequence communities cloned from the *Cryptocercus* and lower termite samples. This similarity provides evidence for a close relationship between the gut communities of these wood-feeding insects. This is particularly interesting in light of previous proposals that lower termites are eusocial roaches descended from *Cryptocercus* (30). Moreover, it provides further support for hypotheses proposing a common origin for the gut communities of these two insects (42, 52).

Conclusions. Hydrogen plays a pivotal role in the digestion of wood by the gut symbiotic microbial communities of termites (1, 6, 13, 39, 40, 44). Our findings support this important role of hydrogen as a metabolic intermediate in wood degradation by establishing wood feeding insects as rich reservoirs of [FeFe] hydrogenase gene sequence diversity. Moreover, cloned sequences represent unique adaptations to their respective ecosystems. The non-wood feeding, cockroach *P. americana*, harbored comparatively few [FeFe] hydrogenase sequences. Clustering of the sequence communities of each sample provides support for a close similarity between the gut communities of the wood feeding insects sampled. This is in agreement with the currently accepted phylogenetic relationship between wood roaches and lower termites as well as proposals of a common

origin for the gut microbial communities of these insects (30, 42, 52). The rich variety of [FeFe] hydrogenases observed in each of the lower termite and *Cryptocercus* samples accentuates the important role of hydrogen as an intermediate in wood degradation by xylophagous insects.

References

1. **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.
2. **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, symbioses, ecology*. Kluwer Academic Publishers, Boston.
3. **Breznak, J. A.** 1982. Intestinal microbiota of termites and other xylophagous insects. *Annu. Rev. Microbiol.* **36**:323-343.
4. **Breznak, J. A.** 1975. Symbiotic Relationships Between Termites and Their Intestinal Microbiota. *Symp. Soc. Exp. Biol.* :559-580.
5. **Breznak, J. A.** 2006. Termite Gut Spirochetes, p. 421-444. *In* J. Radolf and S. Lukehart (ed.), *Pathogenic treponema: molecular and cellular biology*. Caister Academic Press.
6. **Breznak, J. A., and J. M. Switzer.** 1986. Acetate Synthesis from H₂ plus CO₂ by Termite Gut Microbes. *Appl. Environ. Microbiol.* **52**:623-630.
7. **Brune, A.** 2006. Symbiotic Associations Between Termites and Prokaryotes. *Prokaryotes* **1**:439-474.
8. **Brune, A.** 1998. Termite guts: the world's smallest bioreactors. *Trends in Biotechnol.* **16**:16-21.
9. **Brune, A., and M. Friedrich.** 2000. Microecology of the termite gut: structure and function on a microscale. *Curr. Opin. Microbiol.* **3**:263-269.

10. **Cleveland, L. R.** 1923. Correlation Between the Food and Morphology of Termites and the Presence of Intestinal Protozoa. *Am. J. Hyg.* **3**:444-461.
11. **Cleveland, L. R.** 1923. Symbiosis between Termites and Their Intestinal Protozoa. *Proc. Natl. Acad. Sci. U.S.A.* **9**:424-428.
12. **Combet, C., C. Blanchet, C. Geourjon, and G. Deléage.** 2000. NPS@: network protein sequence analysis. *Trends Biochem. Sci.* **25**:147-50.
13. **Ebert, A., and A. Brune.** 1997. Hydrogen Concentration Profiles at the Oxidic-Anoxic Interface: a Microsensor Study of the Hindgut of the Wood-Feeding Lower Termite *Reticulitermes flavipes* (Kollar). *Appl. Environ. Microbiol.* **63**:4039-4046.
14. **Felsenstein, J.** 1989. PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* **5**:164-166.
15. **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.
16. **Hoehler, T. M., B. M. Bebout, and D. J. Des Marais.** 2001. The role of microbial mats in the production of reduced gases on the early Earth. *Nature* **412**:324-327.
17. **Hungate, R. E.** 1939. Experiments on the Nutrition of Zootermopsis. III. The anaerobic carbohydrate dissimilation by the intestinal protozoa. *Ecology* **20**:230-245.
18. **Hungate, R. E.** 1943. Quantitative Analysis on the Cellulose Fermentation by Termite Protozoa. *Ann. Entomol. Soc. Am.* **36**:730-9.

19. **Inoue, J.-I., K. Saita, T. Kudo, S. Ui, and M. Ohkuma.** 2007. Hydrogen Production by Termite Gut Protists: Characterization of Iron Hydrogenases of Parabasalian Symbionts of the Termite *Coptotermes formosanus*. *Eukaryotic Cell* **6**:1925-1932.
20. **Inward, D., G. Beccaloni, and P. Eggleton.** 2007. Death of an order: a comprehensive molecular phylogenetic study confirms that termites are eusocial cockroaches. *Biol. Lett.* **3**:331-335.
21. **Inward, D. J. G., A. P. Volger, and P. Eggleton.** 2007. A comprehensive phylogenetic analysis of termites (Isoptera) illuminates key aspects of their evolutionary biology. *Mol. Phylogenet. Evol.* **44**:953-967.
22. **Kambhampati, S., and P. Eggleton.** 2000. Taxonomy and Phylogeny of Termites, p. 1-23. *In* T. Abe, D. E. Bignell, and M. Higashi (ed.), *Termites: evolution, sociality, symbioses, ecology*. Kluwer Academic Publishers, Boston.
23. **Klass, K.-D., C. Nalepa, and N. Lo.** 2008. Wood-feeding cockroaches as models for termite evolution (Insecta: Dictyoptera): *Cryptocercus* vs. *Parasphaeria boleiriana*. *Mol. Phylogenet. Evol.* **46**:809-817.
24. **Leadbetter, J. R.** 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**:3620-31.
25. **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.

26. **Leadbetter, J. R., T. M. Schmidt, J. R. Graber, and J. A. Breznak.** 1999. Acetogenesis from H₂ Plus CO₂ by Spirochetes from Termite Guts. *Science* **283**:686-689.
27. **Legendre, F., M. F. Whiting, C. Bordereau, E. M. Canello, T. A. Evans, and P. Grandcolas.** 2008. The phylogeny of termites (Dictyoptera: Isoptera) based on mitochondrial and nuclear markers: Implications for evolution of the worker and pseudergate castes, and foraging behaviors. *Mol. Phylogenet. Evol.* **48**:615-627.
28. **Lemke, T., T. v. Alen, J. H. P. Hackstein, and A. Brune.** 2001. Cross-Epithelial Hydrogen Transfer from the Midgut Compartment Drives Methanogenesis in the Hindgut of Cockroaches. *Appl. Environ. Microbiol.* **67**:4657-4661.
29. **Lilburn, T. G., K. S. Kim, N. E. Ostrom, K. R. Byzek, J. R. Leadbetter, and J. A. Breznak.** 2001. Nitrogen Fixation by Symbiotic and Free-Living Spirochetes. *Science* **292**:2495-2498.
30. **Lo, N., G. Tokuda, H. Watanabe, H. Rose, M. Slaytor, K. Maekawa, C. Bandi, and H. Noda.** 2000. Evidence from multiple sequences indicates that termites evolved from wood-feeding cockroaches. *Curr. Biol.* **10**:801-804.
31. **Lozupone, C., M. Hamady, and R. Knight.** 2006. UniFrac -- An online tool for comparing microbial community diversity in a phylogenetic context. *BMC Bioinformatics* **7**.
32. **Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Förster, I. Brettske, S. Gerber, A. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. König, T. Liss, R. Lüssmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N.**

- Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode, and K. Schleifer.** 2004. ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**:1363-1371.
33. **Matson, E. G., E. A. Ottesen, and J. R. Leadbetter.** 2007. Extracting DNA from the gut microbes of the termite (*Zootermopsis nevadensis*), *J. Vis. Exp.*
34. **Meyer, J.** 2007. [FeFe] hydrogenases and their evolution: a genomic perspective. *Cell. Mol. Life Sci.* **64**:1063-1084.
35. **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. *Ann. Entomol. Soc. Am.* **91**:515-523.
36. **Néron, B., H. Ménager, C. Maufrais, N. Joly, T. Pierre, and C. Letondal.** 2008. Presented at the Bio Open Source Conference, Toronto.
37. **Nicolet, Y., C. Piras, P. Legrand, C. E. Hatchikian, and J. C. Fontecilla-Camps.** 1999. *Desulfovibrio desulfuricans* iron hydrogenase: the structure shows unusual coordination to an active site Fe binuclear center. *Structure* **7**:13-23.
38. **Noirot, C.** 1995. The Gut of Termites (Isoptera). *Comparative Anatomy, Systematics, Phylogeny. I. Lower Termites.* *Ann. Soc. Entomol. Fr.* **31**:197-226.
39. **Odelson, D. A., and J. A. Breznak.** 1985. Cellulase and Other Polymer-Hydrolyzing Activities of *Trichomitopsis termopsidis*, a Symbiotic Protozoan from Termites. *Appl. Environ. Microbiol.* **49**:622-626.

40. **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.
41. **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.
42. **Ohkuma, M., H. Noda, Y. Hongoh, C. A. Nalepa, and T. Inoue.** 2009. Inheritance and diversification of symbiotic trichonymphid flagellates from a common ancestor of termites and the cockroach *Cryptocercus*. *Proc. Biol. Sci.* **276**:239-245.
43. **Ottesen, E. A.** 2009. The Biology and Community Structure of CO₂-Reducing Acetogens in the Termite Hindgut. California Institute of Technology, Pasadena.
44. **Pester, M., and A. Brune.** 2007. Hydrogen is the central free intermediate during lignocellulose degradation by termite gut symbionts. *ISME J.* **1**:551-565.
45. **Peters, J. W., W. N. Lanzilotta, B. J. Lemon, and L. C. Seefeldt.** 1998. X-ray Crystal Structure of the Fe-Only Hydrogenase (CpI) from *Clostridium pasteurianum* to 1.8 Angstrom Resolution. *Science* **282**:1853-1858.
46. **Schink, B., F. S. Lupton, and J. G. Zeikus.** 1983. Radioassay for Hydrogenase Activity in Viable Cells and Documentation of Aerobic Hydrogen-Consuming Bacteria Living in Extreme Environments. *Appl. Environ. Microbiol.* **45**:1491-1500.

47. **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.
48. **Scranton, M. I., P. C. Novelli, and P. A. Loud.** 1984. The distribution and cycling of hydrogen gas in the waters of two anoxic marine environments. *Limnol. Oceanogr.* **29**:993-1003.
49. **Smolenski, W. J., and J. A. Robinson.** 1988. In situ rumen hydrogen concentrations in steers fed eight times daily, measured using a mercury reduction detector. *FEMS Microbiol. Ecol.* **53**:95-100.
50. **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, ecology*. Kluwer Academic Publishers, Boston.
51. **Sugimoto, A., and N. Fujita.** 2006. Hydrogen concentrations and stable isotopic composition of methane in bubble gas observed in a natural wetland. *Biogeochemistry* **81**:33-44.
52. **Thorne, B. L.** 1990. A case for ancestral transfer of symbionts between cockroaches and termites. *Proc. Biol. Sci.* **241**:37-41.
53. **Trager, W.** 1934. The Cultivation of a Cellulose-Digesting Flagellate, *Trichomonas termopsidis*, and of Certain Other Termite Protozoa. *Biol. Bull.* **66**:182-190.
54. **Vignais, P. M., and B. Billoud.** 2007. Occurrence, Classification, and Biological Function of Hydrogenases: An Overview. *Chem. Rev.* **107**:4206-4272.

55. **Vignais, P. M., B. Billoud, and J. Meyer.** 2001. Classification and phylogeny of hydrogenases. *FEMS Microbiol. Rev.* **25**:455-501.
56. **Warnecke, F., P. Luginbühl, N. Ivanova, M. Ghassemian, T. H. Richardson, J. T. Stege, M. Cayouette, A. C. McHardy, G. Djordjevic, N. Aboushadi, R. Sorek, S. G. Tringe, M. Podar, H. G. Martin, V. Kunin, D. Dalevi, J. Madejska, E. Kirton, D. Platt, E. Szeto, A. Salamov, K. Barry, N. Mikhailova, N. C. Kyrpides, E. G. Matson, E. A. Ottesen, X. Zhang, M. Hernández, C. Murillo, L. G. Acosta, I. Rigoutsos, G. Tamayo, B. D. Green, C. Chang, E. M. Rubin, E. J. Mathur, D. E. Robertson, P. Hugenholtz, and J. R. Leadbetter.** 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**:560-569.
57. **Yamin, M. A.** 1979. Cellulolytic activity of an axenically-cultivated termite flagellate, *Trichomitopsis termopsidis*. *J. Gen. Microbiol.* **113**:417-20.
58. **Yamin, M. A.** 1981. Cellulose Metabolism by the Flagellate *Trichonympha* from a Termite Is Independent of Endosymbiotic Bacteria. *Science* **211**:58-59.
59. **Yang, H., D. Schmitt-Wagner, U. Stingl, and A. Brune.** 2005. Niche heterogeneity determines bacterial community structure in the termite gut (*Reticulitermes santonensis*). *Environ. Microbiol.* **7**:916-932.
60. **Zhang, X.** 2010. Formate dehydrogenase gene diversity in lignocellulose-feeding insect gut microbial communities. California Institute of Technology, Pasadena.
61. **Zimmerman, P. R., J. P. Greenberg, S. O. Wandiga, and P. J. Crutzen.** 1982. Termites: A Potentially Large Source of Atmospheric Methane, Carbon Dioxide, and Molecular Hydrogen. *Science* **218**:563-565.

Appendix

Table 3-S1. Cloned sequences

Table 3-S2. Alignment used in primer design

Table 3-S1. Sequences cloned.

Phylotype ^a	Genotype ^a	Number ^b	% ^c
<i>Incisitermes minor</i> collection Pas1			
A1	A1	11	13
A4	A4, F11	5	6
B6	B6, B8, D11, G12	9	11
B9	B9	2	2
C4	C4	2	2
C9	C9	5	2
D1	D1, H11, G7, G2	14	16
D5	D5, E1, H8	8	9
E2	E2	1	1
E5	E5, H5	10	12
E9	E9	1	1
F1	F1, H12	10	12
F5	F5	4	5
F8	F8	1	1
G3	G3	2	2
Total:		85	
<i>Reticulitermes hesperus</i> collection ChiA2			
A1	A1, F2, A2, F11, A9, E3, C2	39	41
A4	A4, D1, D6, G2, H2	15	16
A6	A6	1	1
A8	A8	2	2
B1	B1	10	10
B10	B10	1	1
B11	B11	1	1
C7	C7, F12	2	2
C9	C9, A10	2	2
D11	D11	2	2
D4	D4	1	1
D5	D5	1	1
E1	E1	2	2
E5	E5	2	2
F5	F5	1	1
G3	G3	1	1
G4	G4	1	1
H1	H1	3	3
H12	H12	1	1
H4	H4	8	8
Total:		96	
<i>Zootermopsis nevadensis</i> collection ChiA1			
A10	A10, E8, C4, H11	9	10
A12	A12, G7, E3, F10, G5	6	6
A3	A3, G6, E5	12	13
A5	A5	1	1
A7	A7	1	1
B3	B3	9	10
B4	B4	1	1
B6	B6	1	1
B8	B8, E7, C1	6	6
B9	B9	5	5
C5	C5	1	1
D1	D1, E9	3	3
D10	D10, E12, F8	6	6
D4	D4	1	1
D8	D8, E10, F1, H2, C10	21	23
E11	E11	1	1
E6	E6	3	3
F4	F4	2	2
G1	G1	2	2
H10	H10	1	1
H12	H12	1	1
Total:		93	

Continuing Table 3-S1.

	Phylotype ^a	Genotype ^a	Number ^b	% ^c	
<i>Cryptocercus punctulatus</i> nymph	A1	A1, B1, B2	29	32	
	A11	A11	4	4	
	A3	A3, F8, F9, G8, E7, F11	22	24	
	A8	A8, B12, D12, D8	6	7	
	A9	A9	1	1	
	B10	B10	1	1	
	B4	B4, E4	2	2	
	B5	B5, C12	2	2	
	B6	B6, E5, H7	5	5	
	C2	C2	1	1	
	C5	C5	1	1	
	C6	C6	2	2	
	C7	C7	1	1	
	C8	C8, D2	12	13	
	C9	C9	1	1	
	D11	D11, E2, G4	3	3	
	F1	F1	1	1	
	Total:		94		
<i>Cryptocercus punctulatus</i> adult	A1	A1, C8, G1, H8, D9, G8, E5, F10	30	32	
	A6	A6	1	1	
	A7	A7	1	1	
	B12	B12	7	7	
	B8	B8	3	3	
	C1	C1	24	26	
	C11	C11, H7	3	3	
	C3	C3	1	1	
	C4	C4	3	3	
	C6	C6	10	11	
	D5	D5	1	1	
	E11	E11, G9	3	3	
	E4	E4	3	3	
	H2	H2	2	2	
	H3	H3	1	1	
		Total:		93	
	<i>Periplaneta americana</i>	A3	A3, B11, G10	3	11
B2		B2, A8, G9, H8, F2, A9, E8, C1	10	37	
B3 ^d		B3	1	4	
C6 ^e		C6, G4, H3, E7, F5, F4, H7	7	26	
D4		D4, G11	2	7	
D6		D6	2	7	
F12 ^d		F12	1	4	
G5		G5	1	4	
	Total:		27		

^aOTUs calculated using the furthest-neighbor method in DOTUR (47) with a 97% amino-acid similarity cut-off.

^bNumber of cloned sequences grouped within each OTU.

^cPercent of cloned sequences represented by each OTU.

^dSequences that could not be classified as belonging to any of the sequence families defined by Warnecke *et al.* (56).

^eSequence classified as belonging to the Family 7 hydrogenase sequence family defined by Warnecke *et al.* (56).

	330	340	350	360	540	550
ZAS-9	ADMIPNFSTAKS	SPQQMMGAMI	KAYWAEKAGVDP		YHFVEVMACRGG	GCIGGGG
ZAS-2	TDMIPNFSTAKS	SPQQMMGAMI	KAYWAEKAGVNP		WQFVEVMACRGG	GCVGGGG
^a 2005586165	PDMIDNFSTAKS	SPQQMMGAMI	KAYWAKKAGIAP		YHFVEVMACRGG	GCVAGGG
^a 2005580432	NDMIPNFSTAKS	SPQQMMGAMI	KAYWAEKAGVNP		YHFVEVMACRGG	GCVAGGG
^a 2005575223	NDMIPNFSTAKS	SPQQMMGAMI	KAYWAEKAGVDP		YHFIEVMACRGG	GCIAGGG
^a 2005586899	TDMIPNFSTAKS	SPQQMMGAMI	KAYWANKAGVNP		YHFIEVMACRGG	GCIAGGG
^a 2005580141	PDMIDNFSTAKS	SPQQMMGAMI	KAYWAKKAGVAP		YHFVEVMACRGG	GCVAGGG
^a 2005576126	ADMIPNFSTAKS	SPQQMMGAMI	KAYWAKKAGVAP		YHFVEVMACRGG	GCVAGGG
^a 2005576125	PDMINNFSTAKS	SPQQMMGAMI	KAYWAGKAGVDP		YHFVEVMACRGG	GCIGGGG
^a 2005568632	TDMIPNFSTAKS	SPQQMMGAMI	KAYWADKAGISP		YHFVEVMACRGG	GCIAGGG
^a 2005562460	ADMIPNFSTAKS	SPQQMMGAMI	KAYWAGKAAVDP		YHFVEVMACRGG	GCIAGGG
<i>D. vulgaris</i>	PELLPHFSTCKSP	IGMNGALAKTY	GAERMKYDP		YHFIEYMACPGG	CVCGGG
<i>C. pasteurianum</i>	PELLNNLSSAKSP	QQIFGTASKTY	YPSISGLDP		YHFIEVMACHGG	CVNNGG
^b YP_4611421	PDLLGHLSTCKSP	QQMFGALAKTY	YAQVSGIDP		YAFIEVMCCPGG	CVAGGG
^b YP_0012136911	PELAPNVSSAKSP	QQMFGAVCKTY	YAEKAGIDP		YHFIEIMACPGG	CVGGGG
^b YP_3073381	PELAPNVSSAKSP	QQMFGAVCKTY	YAEKAGIDP		YHFIEIMACPGG	CVGGGG
^b ZP_011896821	PSYLEHISSCKSP	QQMFGALAKTY	YPENNGIDP		YHFIEIMACPGG	CVGGGG
^b ZP_015740951	HDFIENLSSCKSP	QQMFGAIAKSY	YPTKADVDP		YHFIEVMGCEGG	CINGGG
^b ZP_022037121	PELAPNVSSAKSP	QQMFGAVCKTY	YAEKSGIDP		YHFIEIMACPGG	CVGGGG
^b NP_6225461	PEFIDNLSTCKSP	HMMMGALVKS	SYAEKKGLDP		YHFIEVMGCPGG	CIMGGG

Figure 3-S1. Alignment used to design degenerate primers to amplify Family 3 [FeFe] hydrogenases. The alignment was prepared using ClustalX on the PBIL network protein sequence analysis server (12). *C. pasteurianum* = [FeFe] hydrogenase from *C. pasteurianum* (45), *D. vulgaris* = [FeFe] hydrogenase from *D. vulgaris* (37), ZAS-2 = Family 3 [FeFe] hydrogenase from *T. primitia* ZAS-2, ZAS-9 = Family 3 [FeFe] hydrogenase from *T. azotonutricium* ZAS-9. ^aIMG Gene Object Identifier, ^bGenBank accession number.

Chapter 4

ANALYSIS OF [FeFe] HYDROGENASE SEQUENCES FROM THE HYDROGEN RICH GUTS OF HIGHER TERMITES REVEALS CORRELATION BETWEEN GUT ECOSYSTEM PARAMETERS AND SEQUENCE COMMUNITY COMPOSITION

Abstract

Hydrogen is the central free intermediate in the degradation of wood by termite gut microbes and can reach concentrations exceeding those measured for any other biological system. Degenerate primers targeting the largest family of [FeFe] hydrogenases observed in a termite gut metagenome (Warnecke, F., *et al.* 2007. *Nature* 450: 560-569) have been used to explore the evolution and representation of these enzymes in termites. Sequences were cloned from the guts of the higher termites *Amitermes* sp. Cost010, *Amitermes* sp. JT2, *Gnathamitermes* sp. JT5, *Microcerotermes* sp. Cost008, *Nasutitermes* sp. Cost003, and *Rhyncotermes* sp. Cost004. Each gut sample harbored a more rich and evenly distributed population of hydrogenase sequences than observed previously in the guts of lower termites and *C. punctulatus* (see Chapter 3). This accentuates the physiological importance of hydrogen to higher termite gut ecosystems and may reflect an increased metabolic burden imposed by a lack of gut protozoa. The sequences were phylogenetically distinct from previously sequenced [FeFe] hydrogenases. Phylogenetic and Unifrac comparisons revealed congruence between host phylogeny and hydrogenase sequence library clustering patterns. This may reflect the combined influences of the stable intimate relationship of gut microbes with their host and environmental alterations in the gut that have occurred over the course of termite evolution. Interestingly, host feeding habits were similarly observed to correlate with sequence library clustering in Unifrac. These results accentuate the physiological importance of hydrogen to termite

gut ecosystems and imply that gut microbes of wood feeding insects may have “co-evolved” with their hosts.

Introduction

Hydrogen plays a pivotal role in the digestion of wood by termites (3, 8, 16, 38, 39, 42). Concentrations in the guts of some species can reach concentrations exceeding those measured for any other biological system (16, 20, 42, 43, 46, 48, 49). The turnover of the gas in the gut has been measured in some species at rates as high as 33 m³ H₂ per m³ gut volume per day (42). The environment is also spatially complex, comprising a matrix of microenvironments characterized by different hydrogen concentrations (10, 11, 16, 27, 28, 42).

This hydrogen is produced during the fermentation of lignocellulosic polysaccharides by the symbiotic microbial community residing in the termite gut (19, 21, 22, 38, 39, 52, 56, 57). The termites are dependent upon this complex symbiosis for the degradation of wood (4-6, 9, 14, 15, 38). The primary product of this symbiosis is acetate, which the termites use as their primary carbon and energy source (40). The majority of the hydrogen in the gut is used by bacteria in reductive acetogenesis to produce up to 1/3 of this acetate (3, 8, 29, 40, 42). A small portion of the hydrogen in the gut is used by methanogenic archaea (3, 27, 42).

Termites can be classified as belonging to one of two phylogenetic groups, higher termites or lower termites (25). Higher termites characteristically lack protozoa in their guts, which are abundant in the guts of lower termites, and have more highly segmented gut structures than lower termites (14, 36, 37). Of the over 2600 known species of termites, over 70% are higher termites (25, 55). They represent the largest and most diverse group of termites (24, 55). Yet, most of what we know about termite gut microbes comes from work done with lower termites and comparatively little work has

been done with the communities of higher termites (4, 5, 7, 9). The primary reason for this is that it was believed until recently that the gut microbes of higher termites played only a minor role in wood digestion (47, 51, 53). This changed with the recent publication of the gut metagenome of a higher termite where it was found that the gut community encodes genes for reductive acetogenesis, polysaccharide degradation, and an abundance of [FeFe] hydrogenases, all pointing in the direction of a more active role in wood degradation (53). This previously under-acknowledged role for the gut microbes has also found support in the findings of Toduda and Watanabe (51).

Wood feeding insects have shared a stable and intimate mutualism with their respective gut microbial communities over the course of their evolution (54). The composition of these communities has been shown to vary substantially with host feeding habits (35, 45, 50). Interestingly, a study on the distribution of formyltetrahydrofolate synthetase (FTHFS) genes in the guts of higher termites has provided evidence that feeding habits have an important influence on community composition (41). Moreover, it has been proposed that the gut microbes of lower termites and *Cryptocercus* may “co-evolve” with their respective hosts (1, 12, 13, 17).

Here we report a phylogenetic analysis of [FeFe] hydrogenase genes cloned from the guts of higher termites. The objective was to better understand the diversity, adaptation, and evolution of the genes in these hydrogen-metabolizing ecosystems. Moreover, the influence of host ecosystem variations on the hydrogenase sequence composition of their associated microbial communities was investigated through cross-comparisons with sequence libraries reported previously for lower termite and wood-roach samples (see Chapter 3).

Methods

Termites. *Nasutitermes* sp. Cost003 and *Rhyncotermites* sp. Cost004 were collected in the INBIO forest preserve in Guápiles, Costa Rica. Cost003 was collected at a height of 1.2 m from a *Psidium guajaba* tree and was believed to be feeding on deadwood. Cost004 was collected from a nest located under a Bromeliad. Feeding trails leading from this nest to a pile of decaying wood and plant material suggested litter feeding. *Microcerotermes* sp. Cost008 was collected from the base of a palm tree about 100 m from the beach at Cahuita National Park in Costa Rica, and appeared to be feeding on the palm tree. *Amitermes* sp. Cost010 was collected from the roots of dead sugar cane plants at a plantation in Costa Rica. *Amitermes* sp. JT2 and *Gnathamitermes* sp. JT5 were collected from subterranean nests at Joshua Tree National Park (Permit#: JOTR:2008-SCI-002).

Termites were identified in a previous study (41) using insect mitochondrial cytochrome oxidase subunit II (COXII) gene sequences and morphology. The COXII genes were amplified directly from the DNA samples that hydrogenases were cloned from. COXII was amplified using the primers CI-J-1773 and B-tLys and cycling conditions described by Miura *et al.* (34) FailSafe PremixD (Epicentre) and Expand High Fidelity Taq (Roche) were substituted for the polymerase and buffers, respectively. Sequences were edited and analyzed in a manner analogous to that described below for cloned [FeFe] hydrogenase sequences. Samples were identified as belonging to the genus of the termite harboring the COXII sequence to which they were found most near in phylogenetic analyses.

DNA Extraction and Cloning. DNA was extracted from whole dissected guts and quantitated as described previously (33) and in Chapter 3. Degenerate primers designed

in Chapter 3 for the specific amplification of Family 3 [FeFe] hydrogenases were used for the cloning of gut sequences as described there. Family 3 [FeFe] hydrogenases, first defined by Warnecke *et al.*, were the most highly represented group of enzymatic hydrogenases observed in the *Nasutitermes* hindgut metagenome sequence (53). Family 3 [FeFe] hydrogenases were the only group of hydrogenases observed in the *Nasutitermes* hindgut metagenome whose *in situ* translation was verified by MS (53). The degenerate primer sequences, which were ordered from IDT DNA, were WSICCARCARATGATGG and CCIKRCAIGCCATIACYTC for the forward and reverse primers, respectively, where “I” represents inositol.

RFLP Analysis and Sequencing. For each termite gut, 96 clones were randomly selected for RFLP analysis as described previously in Chapter 3. Sequences representing each unique RFLP pattern observed were arbitrarily selected and submitted for sequencing, as described previously in Chapter 3. The sequences obtained were manually trimmed in SeqMan, available from DNA* as part of the Lasergene software suite, to remove the plasmid and degenerate primer sequences.

The identity of each sequence as a hydrogenase was verified using by BLASTing it against GeneBank. Sequences that did not have hydrogenases as the top hits were not included in further analyses. Also, sequences that in subsequent analyses aligned poorly with other cloned hydrogenase sequences were re-sequenced and analyzed manually for frame-shift mutations if they continued to align poorly or contain internal stop codons. Frame-shift mutations were identified and manually corrected at the DNA level for three clones based upon sequence alignments and careful inspection of sequencer trace files, see footnotes to Table 4-S1 in this chapter’s appendix.

Phylogenetic Analysis. An operational taxonomic unit (OTU) was defined as those peptide sequences sharing a minimum of 97% sequence identity. Sequences were grouped into OTUs using the furthest-neighbor algorithm in DOTUR (44).

The ARB software environment (32) was used for phylogenetic analysis of hydrogenase sequences, which was completed as described previously in Chapter 3. Cloned sequences and their OTUs used in these analyses are listed in Table 4-S1 in the appendix to this chapter. Trees were constructed using 173 unambiguously aligned amino acid positions with distance matrix (Fitch), maximum parsimony (Phylip PROTPARS), and maximum likelihood (PhylipPROML) treeing methods. The following sequences comprised the outgroup used to construct Figures 4-2 and 4-3: *Pseudotrichonympha grassii* (AB331668); uncultured parabasilid (AB331670); *Holomastigotoides mirabile* (AB331669), *Pseudotrichonympha grassii* (AB331667), *Treponema primitia* ZAS-1 (HndA1, accession), *T. primitia* ZAS-2 (HndA2, accession), *T. primitia* ZAS-2 (HndA3, accession), *T. primitia* ZAS-1 (HydA1, accession). The following Family 3 [FeFe] hydrogenase sequences reported in Chapter 2, were also used to construct Figures 4-2 and 4-3: *Treponema primitia* strain ZAS-2 (HndA1, Chapter 2); *Treponema azotonutricium* strain ZAS-9 (HndA, Chapter 2).

Diversity and Sequence Richness Calculations. Chao1 sequence richness and Shannon diversity indices for each clone set were calculated using EstimateS version 8.0.0 for Macintosh computers, written and made freely available by Robert K. Colwell (<http://viceroy.eeb.uconn.edu/EstimateS>). OTUs and their respective sequence abundances were used as inputs to the program.

Community Comparisons. Unifrac (31) was used for quantitative comparisons of the higher termite [FeFe] hydrogenase sequences with each other or with those from lower termites and *C. punctulatus* reported previously in Chapter 3. Maximum likelihood trees were constructed according to the methods described above and subsequently used as the input for Unifrac analyses. 173 unambiguously aligned amino acids were used in treeing calculations. Each termite or *C. punctulatus* sequence library was designated as a unique environment. The number of cloned sequences represented by each OTU was input to Unifrac to be used for calculating abundance weights. The environments were compared using the Unifrac jackknife and principle component analyses. Normalized abundance weights were used in all calculations. The jackknife calculation was completed with 1000 samplings and using 75% of the OTUs contained in the smallest environment sample as the minimum number of sequences to keep.

Results

Sequences cloned. Hydrogenase sequences representing as many as 44 sequence OTUs were cloned from each of the higher termites, see Table 4-1. Table 4-S1 in the chapter's appendix lists all clones, and their corresponding OTUs, analyzed in this study. The collector's curves for each sequence library are provided as Figure 4-1. Microcerotermes was the only sample having 75% of all cloned sequences distributed among less than 7 OTUs. The Shannon diversity index and the Chao1 species richness index for each sequence library are listed in Table 4-1.

Phylogenetic analysis. In phylogenetic analyses comparing the cloned sequences to publically available [FeFe] hydrogenase sequences in our database, all but one (see

Table 4-1. Quantifying hydrogenase clone library diversity.

	RFLPs ^a	OTUs ^b	Chao1 Mean ^c	Chao1 95% CI Lower Bound ^c	Chao1 95% CI Upper Bound ^c	Shannon Mean ^d
<i>Amitermes</i> sp. Cost010	60	31	45	35.02	79.77	2.97
<i>Amitermes</i> sp. JT5	33	22	32.67	24.18	74.18	2.65
<i>Gnathamitermes</i> sp. JT5	44	30 ^e	40.29	32.75	68.49	3.05
<i>Microcerotermes</i> sp. Cost008	36	21	29.1	22.84	56.57	2.33
<i>Nasutitermes</i> sp. Cost003	38	25	43	29.54	96.38	2.69
<i>Rhyncotermes</i> sp. Cost004	54	44	68.05	52.73	110.24	3.53

^aNumber of unique restriction fragment polymorphism patterns (RFLPs) observed.

^bNumber of operational taxonomic units (OTUs); calculated using the furthest-neighbor method and a 97% amino-acid sequence similarity cut-off.

^cChao1 species-richness index calculated using the classic method in EstimateS. OTUs representing Family 3 [FeFe] hydrogenases with their respective abundances were used as program inputs.

^dShannon diversity index calculated using EstimateS. OTUs representing Family 3 [FeFe] hydrogenases with their respective abundances were used as program inputs.

^eOne of these OTUs represented Family 7 [FeFe] hydrogenase sequences (see Table 4-S1 in the chapter's appendix) and was not used in the calculation of the diversity indices.

footnote to Table 4-S1 in the chapter's appendix) formed a single large clade to the exclusion of all non-termite bacterial sequences, data not shown. Within this clade were included Family 3 [FeFe] hydrogenase sequences from a *Nasutitermes* gut metagenome (53) and from the genome sequences of two treponemes isolated from *Zootermopsis angusticolis*, *T. primitia* ZAS-2 and *T. azotonutricum* ZAS-9 (see Chapter 2), data not shown. A maximum likelihood tree for all of the cloned [FeFe] hydrogenase sequences is provided as Figure 4-2.

Upon a cursory inspection of phylogenetic groupings, the hydrogenase sequences appeared to cluster in a manner roughly congruent with the phylogeny of their hosts. For example, both *Amitermes* samples tended to group with each other in phylogenetic analyses. *Gnathamitermes* and *Amitermes* were the only higher termite samples analyzed in this study whose COII sequences formed a tight, coherent clade with each other in phylogenetic analyses (41). Interestingly, hydrogenase sequences from these samples tended to group with one another as well. Moreover, hydrogenase sequences from a given termite sample tended to cluster with each other.

Sequence library cross-comparisons. A maximum likelihood tree comparing all of the Family 3 hydrogenases cloned from the higher termite samples to those cloned previously from *C. punctulatus* and lower termite gut samples (see Chapter 3) is provided as Figure 4-3. There is a clear separation between the higher termite sequences and those from *C. punctulatus* and lower termites. The latter two groups of sequences appear to intermingle with each other in the tree. This apparent congruence between the phylogenetic clustering of the cloned hydrogenases and that of their respective hosts was much more striking in the Unifrac jackknife clustering of the samples, see Figure 4-4. In this

Figure 4-2.

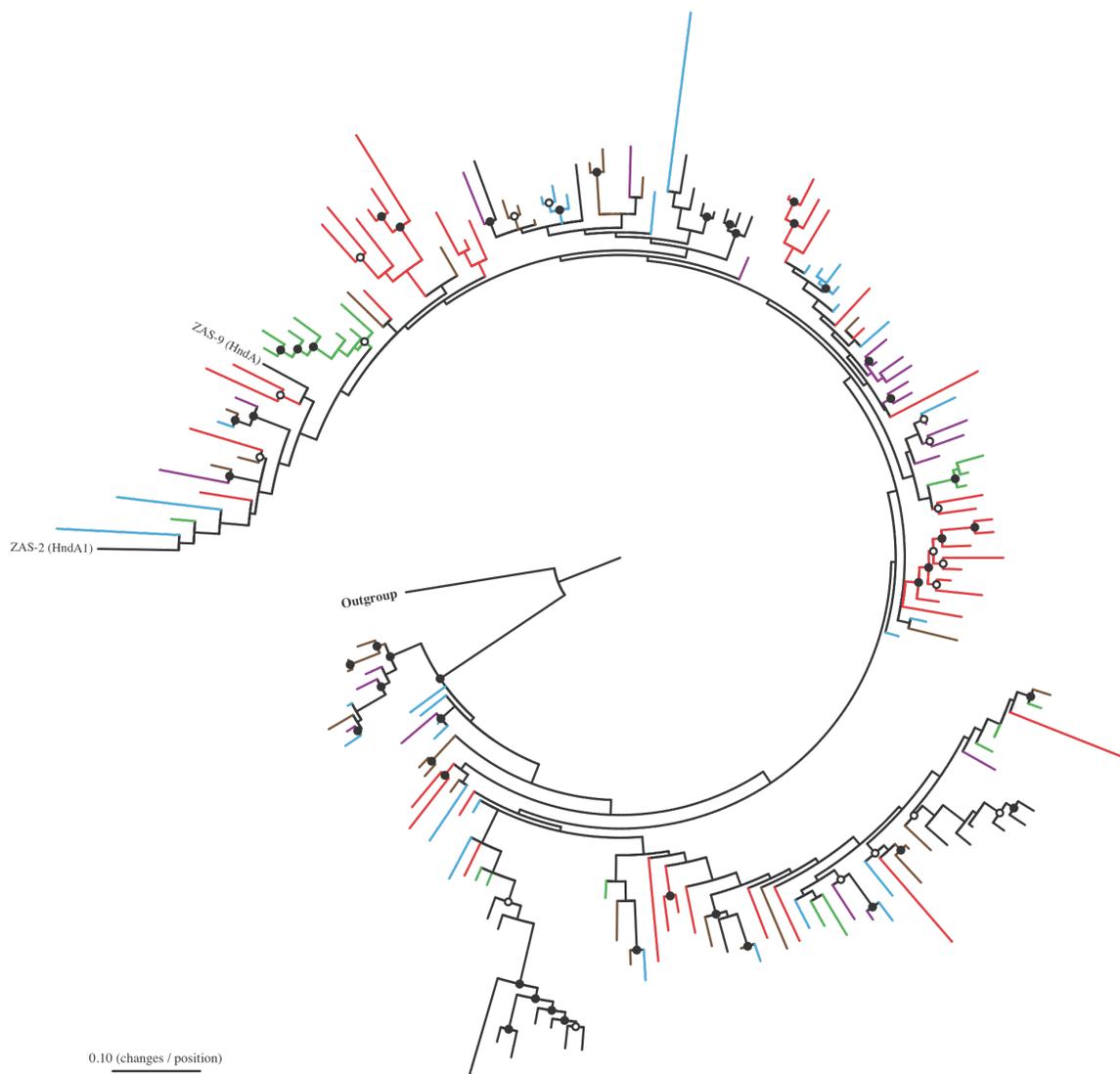


Figure 4-2. Phylogram for Family 3 [FeFe] hydrogenases cloned from the guts of higher termites. The tree was calculated using a maximum likelihood (Phylip ProML) method with 173 unambiguously aligned amino acid positions. Open circles designate groupings also supported by either parsimony (Phylip PROTPARS, 1000 bootstraps) or distance matrix (Fitch) methods. Closed circles designate groupings supported by all three methods. Each leaf represents an OTU. Leaves and branches representing OTUs cloned from *Amitermes* sp. Cost010 = blue; *Amitermes* sp. JT2 = purple; *Gnathamitermes* sp. JT5 = brown; *Microcerotermes* sp. JT5 = green; *Nasutitermes* sp. Cost003 = black; *Rhyncotermes* sp. Cost004 = red. Hydrogenase sequences taken from *T. primitia* ZAS-2 and *T. primitia* ZAS-9 are labeled as ZAS-2 (HndA1) and ZAS-9 (HndA), respectively. Tree drawn using Phylip drawgram (18).

Figure 4-3.

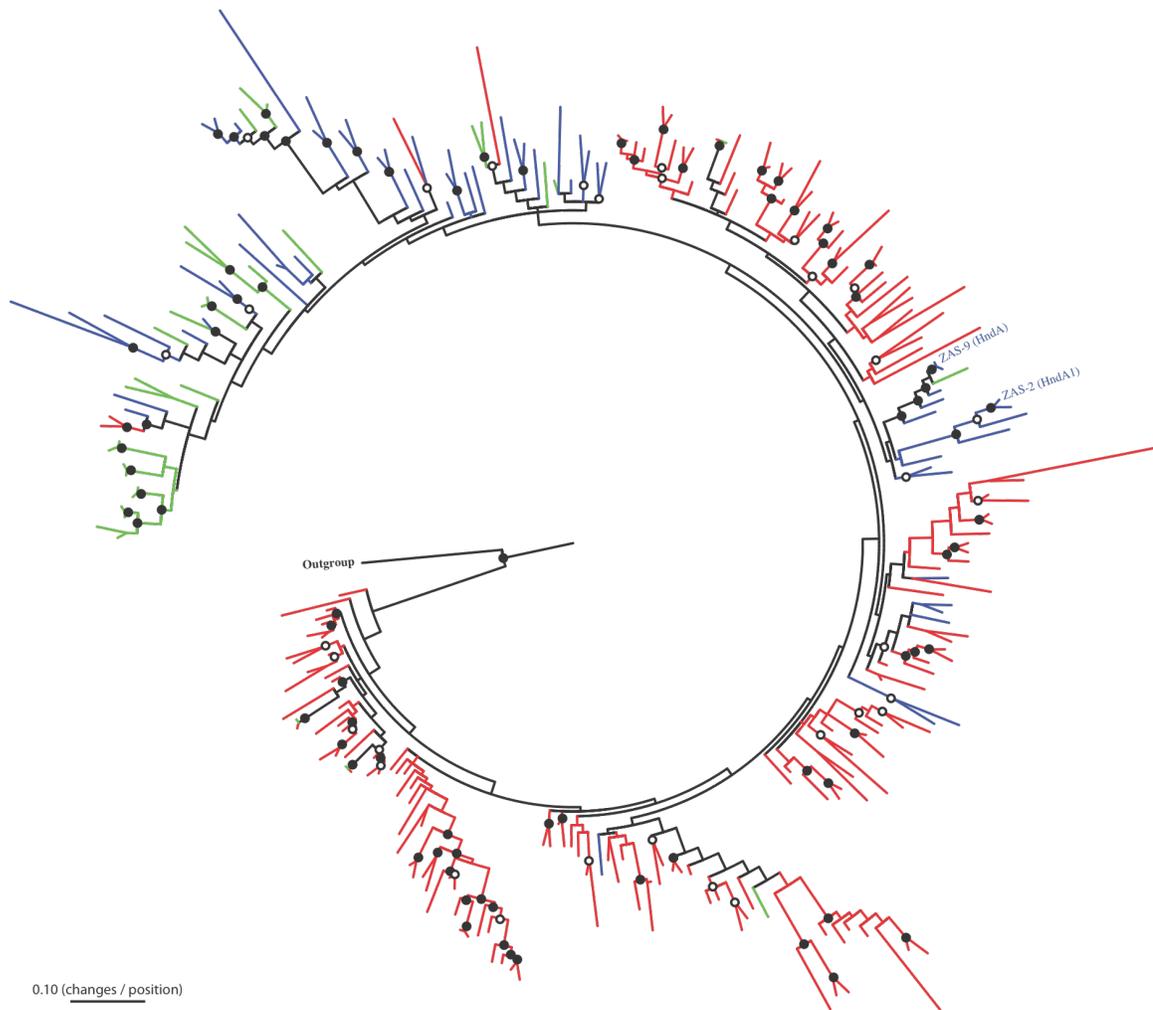


Figure 4-3. Phylogram comparing Family 3 [FeFe] hydrogenases cloned from higher termites to sequences cloned previously from *C. punctulatus* and lower termites. See Figure 4-2 caption for description of open and closed black circles and tree construction methods. Each leaf represents an OTU. Leaves and branches representing OTUs cloned from lower termites are in blue, from *C. punctulatus* are in green, and from higher termites are in red. Hydrogenase sequences taken from *T. primitia* ZAS-2 and *T. primitia* ZAS-9 are labeled as ZAS-2 (HndA1) and ZAS-9 (HndA), respectively. Tree drawn using Phylip drawgram (18).

Figure 4-4.

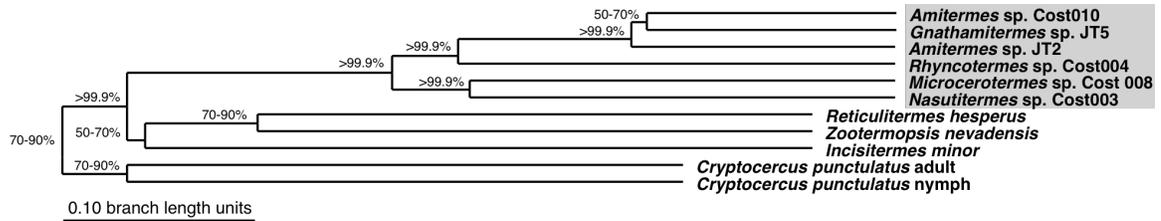


Figure 4-4. Unifrac jackknife analysis of Family 3 [FeFe] hydrogenase sequences cloned from higher termites, lower termites, and *C. punctulatus*. The maximum-likelihood tree shown in Figure 4-3 and the OTUs with their respective abundance weights listed in Table 4-S1 of the appendix to this chapter and taken from Table 3-S1 in the appendix to Chapter 3 were used as inputs to Unifrac. The analysis was completed using normalized abundance weights, 1000 samplings, and keeping a number of sequences equal to 75% of the number of OTUs represented by the smallest sample analyzed. Each insect sample was designated as a unique environment. The grey box highlights all higher termite environments. The numbers designate the percentage of samplings supporting a particular cluster.

analysis, the clustering of the hydrogenase sequences was congruent with the phylogeny of their respective hosts reported by Legendre *et al.* and Inward *et al.* (23, 24, 30). This clustering was further supported by the Unifrac PCA analysis of the sequences, see Figure 4-5. In the PCA analysis, there is a distinguishable separation between sequences from each of the three groups representing higher termites, lower termites, and *C. punctulatus*. Principle component 1, which accounted for the separation of higher termites from lower termites and *C. punctulatus*, explained 34.87% of the variation.

A Unifrac principle component analysis of the [FeFe] hydrogenase sequences cloned from higher termites is provided as Figure 4-6. Sequences from *Amitermes* sp. Cost010, *Amitermes* sp. JT5, and *Gnathamitermes* sp. JT5 clustered together. These termite samples are unique from the others because of their close phylogenetic relationship to one another, as discussed above, and because they were collected from sub-terranean nests. These samples could be distinguished from the others according to principle component 1, which explained 30.68% of the variation.

Discussion

High [FeFe] hydrogenase sequence diversity in higher termites. The abundance of [FeFe] hydrogenases cloned from the guts of higher termites, representing as many as 45 OTUs in the case of *Rhyncotermes* sp. Cost004, emphasizes the physiological importance of these enzymes to these complex ecosystems. Moreover, these cloned sequences, with the exception of one, belong to the largest family of [FeFe] hydrogenase sequences observed in a higher termite gut metagenome. There is good reason to believe that this is only a sampling of a much larger diversity because only one of a total of 9 families reported in the *Nasutitermes* gut metagenome sequence was targeted in this analysis. Interestingly, all of the higher termite sequences grouped with one another to the

Figure 4-5.

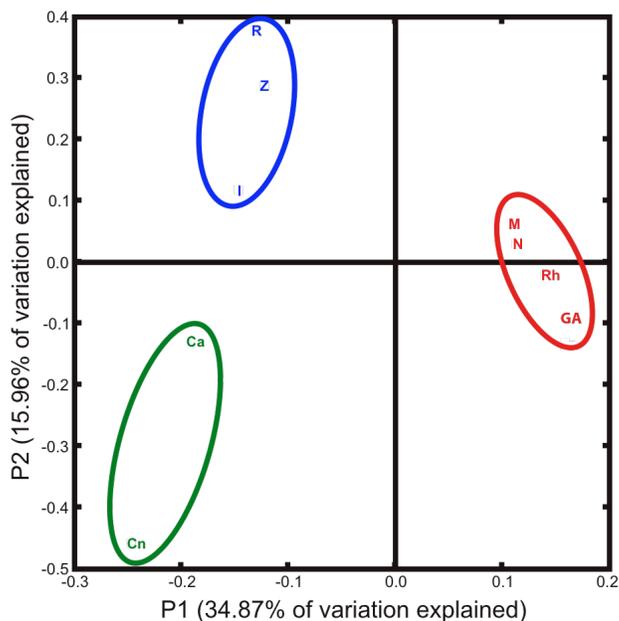


Figure 4-5. Unifrac principle component analysis of Family 3 [FeFe] hydrogenase sequences cloned from the guts of higher termites, lower termites, and *C. punctulatus*. The maximum-likelihood tree shown in Figure 4-3 and the OTUs with their respective abundance weights given in Table 4-S1 of the appendix to this chapter and taken from Table 3-S1 in the appendix to Chapter 3 were used as inputs to Unifrac. Principle components were calculated using normalized abundance weights. Each termite or *C. punctulatus* sample was designated as a unique environment. Higher termite environments are in red, lower termite environments are in blue, and *C. punctulatus* environments are in green. P1 = principle component 1, P2 = principle component 2. Ca = *C. punctulatus* adult, Cn = *C. punctulatus* nymph, GA = a cluster of samples comprising *Amitermes* sp. Cost010, *Amitermes* sp. Cost003, and *Gnathamitermes* sp. JT5, I = *Incisitermes minor* isolate collection Pas1, M = *Microcerotermes* sp. Cost008, N = *Nasutitermes* sp. Cost003, R = *Reticulitermes Hesperus* collection ChiA2, Rh = *Rhyncotermes* sp. Cost004, Z = *Zootermopsis nevadensis* collection ChiA1.

Figure 4-6.

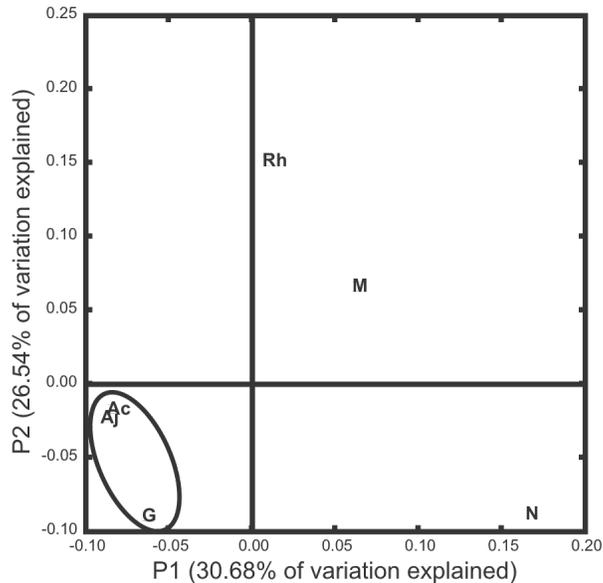


Figure 4-6. Unifrac principle component analysis of Family 3 [FeFe] hydrogenase sequences cloned from higher termites in this study. The maximum-likelihood tree shown in Figure 4-2 and the OTUs with their respective abundance weights given in Table 4-S1 of the appendix to this chapter were used as inputs to Unifrac. Principle components were calculated using normalized abundance weights. Each termite sample was designated as a unique environment. Environments representing sub-terranean termites are in purple, those representing all other higher termites are in red. P1 = principle component 1, P2 = principle component 2. Ac = *Amitermes* sp. Cost010, Aj = *Amitermes* sp. JT2, G = *Gnathamitermes* sp. JT5, M = *Microcerotermes* sp. Cost008, N = *Nasutitermes* sp. Cost003, Rh = *Rhyncotermes* sp. Cost004.

exclusion of all other non-termite [FeFe] hydrogenase sequences in our database. This may imply unique adaptations of these sequences to the termite gut ecosystem. Similar community-wide adaptations of [FeFe] hydrogenase sequences from unique ecosystems has been reported previously as reported in Chapter 3 and elsewhere (2).

Higher termites characteristically lack protozoa in the gut (14). Lower termites and *C. punctulatus* have an abundance of protozoa in their guts that are largely responsible for the fermentation of lignocellulosic polysaccharides and the concomitant production of most of the hydrogen in the termite gut (4, 7, 9, 15, 26, 42, 52). The absence of protozoa in higher termite guts may introduce important selective forces on bacteria unique to these ecosystems including a greater burden to produce and consume hydrogen. As one might expect then, the hydrogenases cloned from the higher termites tended to have a more even distribution and broader sequence diversity than sequences cloned from *C. punctulatus* or lower termites, compare Table 4-1 and Figure 4-1 from this study to Table 3-1 and Figures 3-1 and 3-2 from Chapter 3.

Congruence of [FeFe] hydrogenase and host phylogeny. [FeFe] hydrogenases cloned from closely related termites had a tendency to group with one another in phylogenetic analyses, see Figure 4-2. For example, sequences from both *Amitermes* gut samples tended to group together despite their being collected from locations separated by a great distance – California and Costa Rica. Sequence OTUs from a particular termite tended to group with one another rather than with sequences from other termites. In a phylogenetic analysis of the COII sequences used for molecular characterization of the termite samples, *Gnathamitermes* sp. JT5 and *Amitermes* sp. JT2 were found to be the most closely related of any of the higher termites used in this study (41). Correspondingly,

there was a tendency for sequences from *Gnathamitermes* sp. JT5 to group with those from the *Amitermes* sp. samples. As one would expect, sequences taken from the genomes *T. primitia* ZAS-2 and *T. azotonutricium* ZAS-9, each isolated from the gut of a lower termite, did not group strongly with any of the sequences cloned from the higher termites, see Figure 4-2.

This congruence was further supported by phylogenetic comparisons of the higher termite sequences to lower termite and *Cryptocercus* sequences cloned previously. In the maximum likelihood tree shown in Figure 4-3, there is a clear segregation of the higher termite hydrogenase sequences from those of *Cryptocercus* and lower termites. The lack of clear segregation of the lower termite sequences from those of *C. punctulatus* is in agreement with the close evolutionary relatedness of these insects (23, 24, 30). A Unifrac principle component analysis using the maximum likelihood tree shown in Figure 4-5 further supported these qualitative observations. The 1st principle component, explaining 34.87% of the variation, separated the higher termites from *Cryptocercus* and lower termites. The jackknife clustering of the [FeFe] hydrogenase communities mimicked previously proposed termite phylogenies remarkably (23, 24, 30).

The observed congruence between [FeFe] hydrogenase phylogeny and that of the host may imply that hydrogenases, and by extension their respective gut communities, have co-evolved in an intimate relationship with their host termites. This is in agreement with previous proposals of termite or *Cryptocercus* gut microbes having co-evolved with their host (1, 12, 13, 17). Perhaps more accurately, this observation may be explained as a consequence of the influence of environmental alterations in the gut, such as the presence

or lack of protozoa or various anatomical alterations, that have developed over the course of termite evolution.

Influence of host feeding habits. Unifrac principle component and jackknife clustering analyses of a maximum likelihood tree of all higher termite sequences, see Figures 4-6, revealed a close clustering of the *Amitermes* sp. samples and *Gnathamitermes* sp. JT5 samples. This clustering was apparent when the 1st and 2nd principle components, collectively explaining 57.22% of variation, were plotted against each other. In addition to sharing the close phylogenetic relationship discussed above, the *Amitermes* sp. and the *Gnathamitermes* sp. JT5 termite samples were all collected from sub-terranean galleries implying a grass- or soil-feeding diet and increased exposure to humics. Elizabeth Ottesen has reported a similar distinguishability between sub-terranean higher termites and other higher termites in her work on the FTHFS gene using the same higher termite samples used in this study (41). Previous studies have also shown that higher termites with different feeding habits have markedly different compositions of symbiotic bacteria in their guts (35, 45, 50). Feeding habits may be an important parameter, intimately associated with host phylogeny, influencing the [FeFe] hydrogenase sequence representation in the termite gut.

Conclusions. Termites are a rich reservoir of uniquely adapted [FeFe] hydrogenase gene diversity. The high representation of [FeFe] hydrogenases observed in the guts of higher termites accentuates the physiological importance of these ecosystems. The enzymes had a higher representation and more even population distributions than was observed previously in lower termites and a woodroach, see Chapter 3. This may be the

consequence of an increased metabolic burden on the gut bacteria in higher termites to metabolize hydrogen as a consequence of a lack of gut protozoa.

The congruence of [FeFe] hydrogenase sequence phylogeny with host phylogeny provides experimental support for the hypothesis that the gut microbial communities of termites and *Cryptocercus* have “co-evolved” with their host. This may reflect the combined influences of the stable, intimate relationship of gut microbes with their host and the environmental alterations in the gut, such as the presence or lack of protozoa or various anatomical or host nutritional alterations, that have occurred over the course of termite evolution. Unifrac analyses further revealed that long standing host-feeding preferences, a variable perhaps closely correlated with termite evolution, may have an important influence on the hydrogenase sequence population in the termite gut.

Surveying the representation of Family 3 [FeFe] hydrogenases has begun to shed light on the physiology and evolution of the gut microbial communities of termites.

References

1. **Berlanga, M., B. J. Paster, and R. Guerrero.** 2007. Coevolution of symbiotic spirochete diversity in lower termites. *Int. Microbiol.* **10**:133-139.
2. **Boyd, E. S., J. R. Spear, and J. W. Peters.** 2009. [FeFe] Hydrogenase Genetic Diversity Provides Insight into Molecular Adaptation in a Saline Microbial Mat Community. *Appl. Environ. Microbiol.* **75**:4620-4623.
3. **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.
4. **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, symbioses, ecology*. Kluwer Academic Publishers, Boston.
5. **Breznak, J. A.** 1982. Intestinal microbiota of termites and other xylophagous insects. *Annu. Rev. Microbiol.* **36**:323-343.
6. **Breznak, J. A.** 1975. Symbiotic Relationships Between Termites and Their Intestinal Microbiota. *Symp. Soc. Exp. Biol.* :559-580.
7. **Breznak, J. A., and B. A.** 1994. Role of microorganisms in the digestion of lignocellulose by termites. *Annu. Rev. Entymology* **39**:453-487.
8. **Breznak, J. A., and J. M. Switzer.** 1986. Acetate Synthesis from H₂ plus CO₂ by Termite Gut Microbes. *Appl. Environ. Microbiol.* **52**:623-630.
9. **Brune, A.** 2006. Symbiotic Associations Between Termites and Prokaryotes. *Prokaryotes* **1**:439-474.

10. **Brune, A.** 1998. Termite guts: the world's smallest bioreactors. *Trends in Biotechnol.* **16**:16-21.
11. **Brune, A., and M. Friedrich.** 2000. Microecology of the termite gut: structure and function on a microscale. *Curr. Opin. Microbiol.* **3**:263-269.
12. **Clark, J. W., S. Hossain, C. A. Burnside, and S. Kambhampati.** 2009. Coevolution between a cockroach and its bacterial endosymbiont: a biogeographical perspective. *Proc. R. Soc. Lond., B., Biol. Sci.* **268**:393-398.
13. **Clark, J. W., and S. Kambhampati.** 2003. Phylogenetic analysis of *Blattabacterium*, endosymbiotic bacteria from the wood roach, *Cryptocercus* (Blattodea: Cryptocercidae), including a description of three new species. *Mol. Phylogenet. Evol.* **26**:82-88.
14. **Cleveland, L. R.** 1923. Correlation Between the Food and Morphology of Termites and the Presence of Intestinal Protozoa. *Am. J. Hyg.* **3**:444-461.
15. **Cleveland, L. R.** 1923. Symbiosis between Termites and Their Intestinal Protozoa. *Proc. Natl. Acad. Sci. U.S.A.* **9**:424-428.
16. **Ebert, A., and A. Brune.** 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.
17. **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 Soil Invertebrates*, vol. 6. Springer-Verlag, Berlin Heidelberg.

18. **Felsenstein, J.** 1989. PHYLIP - Phylogeny Inference Package (Version 3.2). *Cladistics* **5**:164-166.
19. **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.
20. **Hoehler, T. M., B. M. Bebout, and D. J. Des Marais.** 2001. The role of microbial mats in the production of reduced gases on the early Earth. *Nature* **412**:324-327.
21. **Hungate, R. E.** 1939. Experiments on the Nutrition of Zootermopsis. III. The anaerobic carbohydrate dissimilation by the intestinal protozoa. *Ecology* **20**:230-245.
22. **Hungate, R. E.** 1943. Quantitative Analysis on the Cellulose Fermentation by Termite Protozoa. *Ann. Entomol. Soc. Am.* **36**:730-9.
23. **Inward, D., G. Beccaloni, and P. Eggleton.** 2007. Death of an order: a comprehensive molecular phylogenetic study confirms that termites are eusocial cockroaches. *Biol. Lett.* **3**:331-335.
24. **Inward, D. J. G., A. P. Volger, and P. Eggleton.** 2007. A comprehensive phylogenetic analysis of termites (Isoptera) illuminates key aspects of their evolutionary biology. *Mol. Phylogenet. Evol.* **44**:953-967.
25. **Kambhampati, S., and P. Eggleton.** 2000. Taxonomy and Phylogeny of Termites, p. 1-23. *In* T. Abe, D. E. Bignell, and M. Higashi (ed.), *Termites: evolution, sociality, symbioses, ecology*. Kluwer Academic Publishers, Boston.

26. **Klass, K.-D., C. Nalepa, and N. Lo.** 2008. Wood-feeding cockroaches as models for termite evolution (Insecta: Dictyoptera): *Cryptocercus* vs. *Parasphaeria boleiriana*. *Mol. Phylogenet. Evol.* **46**:809-817.
27. **Leadbetter, J. R.** 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**:3620-31.
28. **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.
29. **Leadbetter, J. R., T. M. Schmidt, J. R. Graber, and J. A. Breznak.** 1999. Acetogenesis from H₂ Plus CO₂ by Spirochetes from Termite Guts. *Science* **283**:686-689.
30. **Legendre, F., M. F. Whiting, C. Bordereau, E. M. Canello, T. A. Evans, and P. Grandcolas.** 2008. The phylogeny of termites (Dictyoptera: Isoptera) based on mitochondrial and nuclear markers: Implications for evolution of the worker and pseudergate castes, and foraging behaviors. *Mol. Phylogenet. Evol.* **48**:615-627.
31. **Lozupone, C., M. Hamady, and R. Knight.** 2006. UniFrac - An online tool for comparing microbial community diversity in a phylogenetic context. *BMC Bioinformatics* **7**.
32. **Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Förster, I. Brettske, S. Gerber, A. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. König, T. Liss, R. Lüssmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N.**

- Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode, and K. Schleifer.** 2004. ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**:1363-1371.
33. **Matson, E. G., E. A. Ottesen, and J. R. Leadbetter.** 2007. Extracting DNA from the gut microbes of the termite (*Zootermopsis nevadensis*), *J. Vis. Exp.*
34. **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. *Ann. Entomol. Soc. Am.* **91**:515-523.
35. **Miyata, R. N., N. Noda, H. Tamaki, T. 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. **Noirot, C.** 2001. The Gut of Termites (Isoptera) Comparative Anatomy, Systematics, Phylogeny. II. -- Higher Termites (Termitidae). *Ann. Soc. Entomol. Fr.* **37**:431-471.
37. **Noirot, C.** 1995. The Gut of Termites (Isoptera). Comparative Anatomy, Systematics, Phylogeny. I. Lower Termites. *Ann. Soc. Entomol. Fr.* **31**:197-226.
38. **Odelson, D. A., and J. A. Breznak.** 1985. Cellulase and Other Polymer-Hydrolyzing Activities of *Trichomitopsis termopsidis*, a Symbiotic Protozoan from Termites. *Appl. Environ. Microbiol.* **49**:622-626.

39. **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.
40. **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.
41. **Ottesen, E. A.** 2009. The Biology and Community Structure of CO₂-Reducing Acetogens in the Termite Hindgut. California Institute of Technology, Pasadena.
42. **Pester, M., and A. Brune.** 2007. Hydrogen is the central free intermediate during lignocellulose degradation by termite gut symbionts. *ISME J.* **1**:551-565.
43. **Schink, B., F. S. Lupton, and J. G. Zeikus.** 1983. Radioassay for Hydrogenase Activity in Viable Cells and Documentation of Aerobic Hydrogen-Consuming Bacteria Living in Extreme Environments. *Appl. Environ. Microbiol.* **45**:1491-1500.
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., 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**:6007-6017.

46. **Scranton, M. I., P. C. Novelli, and P. A. Loud.** 1984. The distribution and cycling of hydrogen gas in the waters of two anoxic marine environments. *Limnol. Oceanogr.* **29**:993-1003.
47. **Slaytor, M.** 1992. Cellulose Digestion in Termites and Cockroaches: What Role do Symbionts Play. *Comp. Biochem. Physiol.* **103B**:775-784.
48. **Smolenski, W. J., and J. A. Robinson.** 1988. In situ rumen hydrogen concentrations in steers fed eight times daily, measured using a mercury reduction detector. *FEMS Microbiol. Ecol.* **53**:95-100.
49. **Sugimoto, A., and N. Fujita.** 2006. Hydrogen concentrations and stable isotopic composition of methane in bubble gas observed in a natural wetland. *Biogeochemistry* **81**:33-44.
50. **Tanaka, H., H. Aoyagi, S. Shina, Y. Dodo, T. Yoshimura, R. Nakamura, and H. Uchiyama.** 2006. Influence of the diet components on the symbiotic microorganisms community in hindgut of *Coptotermes formosanus* Shiraki. *Appl. Microbiol. Biotechnol.* **71**:907-917.
51. **Tokuda, G., and H. Watanabe.** 2007. Hidden cellulases in termites: revision of an old hypothesis. *Biol. Lett.* **3**:336-339.
52. **Trager, W.** 1934. The Cultivation of a Cellulose-Digesting Flagellate, *Trichomonas termopsidis*, and of Certain Other Termite Protozoa. *Biol. Bull.* **66**:182-190.
53. **Warnecke, F., P. Luginbühl, N. Ivanova, M. Ghassemian, T. H. Richardson, J. T. Stege, M. Cayouette, A. C. McHardy, G. Djordjevic, N. Aboushadi, R. Sorek, S. G. Tringe, M. Podar, H. G. Martin, V. Kunin, D. Dalevi, J.**

- Madejska, E. Kirton, D. Platt, E. Szeto, A. Salamov, K. Barry, N. Mikhailova, N. C. Kyrpides, E. G. Matson, E. A. Ottesen, X. Zhang, M. Hernández, C. Murillo, L. G. Acosta, I. Rigoutsos, G. Tamayo, B. D. Green, C. Chang, E. M. Rubin, E. J. Mathur, D. E. Robertson, P. Hugenholtz, and J. R. Leadbetter.** 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**:560-569.
54. **Wier, A., M. Dolan, D. Grimaldi, R. Guerrero, J. Wagensberg, and M. Margulis.** 2002. Spirochete and protist symbionts of a termite (*Mastotermes electrodominicus*) in Miocene amber. *Proc. Natl. Acad. Sci.* **99**:1410-1413.
55. **Wood, T., and R. Johnson.** 1986. The biology, physiology, and ecology of termites, p. 1-68. *In* S. B. Vinson (ed.), *Economic impact and control of insects*. Praeger, New York.
56. **Yamin, M. A.** 1979. Cellulolytic activity of an axenically-cultivated termite flagellate, *Trichomitopsis termopsidis*. *J. Gen. Microbiol.* **113**:417-20.
57. **Yamin, M. A.** 1981. Cellulose Metabolism by the Flagellate *Trichonympha* from a Termite Is Independent of Endosymbiotic Bacteria. *Science* **211**:58-59.

Appendix

Table 4-S1. Sequences cloned.

Table 4-S1. Sequences cloned.

Phylotype ^a	Genotype ^a	Number ^b	% ^c
<i>Amitermes</i> sp. Cost010			
A1	A1, C7, D12, H11, A3, B7, D7, F9, E6	18	19
A10	A10, A4, H5, C8, G2, G3	10	11
A8	A8, C1	4	4
A9	A9, G1	2	2
B11	B11	1	1
B12	B12, C6, F8	4	4
B3	B3	1	1
B5	B5	1	1
B6	B6, G9	2	2
B8	B8	1	1
B9	B9, C9, E11, F7	6	6
C10	C10	4	4
C12	C12, C4, D4, D10, F2	11	12
C2	C2	2	2
C3	C3	1	1
C5	C5	2	2
D1	D1	1	1
D11	D11	3	3
D3	D3	1	1
D5	D5	1	1
D6	D6	1	1
E4	E4, G7	4	4
E7	E7	1	1
E9	E9, H12	2	2
F1	F1	3	3
F10	F10	1	1
F12	F12	2	2
F3	F3	2	2
G4	G4	1	1
G8	G8	1	1
H10	H10	1	1
Total:		95	
<i>Amitermes</i> sp. JT2			
A1	A1, E1	5	5
A2	A2	1	1
A7	A7, C2, E3	8	8
A8	A8	2	2
B11	B11, G1	3	3
B2	B2, E9, H12	10	10
B5	B5, D10, E2	5	5
B6	B6	1	1
B7	B7	1	1
B8	B8	1	1
C11	C11	2	2
D2	D2	6	6
D6	D6	5	5
E10	E10	3	3
F12	F12	1	1
F2	F2	8	8
F3	F3	1	1
F4	F4	22	23
F5	F5, H1, G8	7	7
G3	G3	1	1
G7	G7	2	2
H5	H5	1	1
Total:		96	

Continuing Table 4-S1.

Phylotype ^a	Genotype ^a	Number ^b	% ^c
<i>Gnathamitermes</i> sp. JT5			
A10	A10, A3, G4	8	8
A11	A11	2	2
A2	A2, H11	2	2
A8	A8	4	4
B1	B1	1	1
B11	B11, D12, G6, E9	5	5
B12	B12, G12, B9, F8	13	14
B2	B2, H2	2	2
B5	B5	4	4
C1	C1, F11	10	11
C12	C12	4	4
C3	C3	1	1
C6	C6	1	1
C8	C8	1	1
D1 ^e	D1	2	2
D10	D10	1	1
D11	D11	7	7
D3	D3	1	1
E12	E12	1	1
E2	E2	5	5
E3	E3	1	1
E7	E7	1	1
F2	F2	1	1
F4	F4, H8	2	2
F5	F5	3	3
F6	F6, F7	6	6
G11	G11	1	1
G8	G8	2	2
H4	H4	2	2
H5	H5	1	1
Total:		95	
<i>Microcerotermes</i> sp. Cost008			
A12	A12, C3, E12, H7	5	5
A3	A3, B1, F4, H12, C12, C4	25	26
A4	A4	2	2
A8	A8	4	4
A9	A9	2	2
B7	B7	25	26
B8	B8	2	2
C1	C1	1	1
C10	C10 ^d	1	1
C5	C5, D11	2	2
C6	C6, D2	5	5
C9	C9	1	1
D5	D5, G1	5	5
D8	D8	1	1
E3	E3	1	1
E5	E5	2	2
F10	F10 ^d	1	1
F5	F5, G11	8	8
G3	G3	1	1
H11	H11	1	1
H8	H8	1	1
Total:		96	

Continuing Table 4-S1.

Phylotype ^a	Genotype ^a	Number ^b	% ^c
<i>Nasutitermes</i> sp. Cost003			
A11	A11, D5, G1	4	4
A12	A12	1	1
A6	A6	4	4
A9	A9, B5, D12, E1	10	11
B10	B10	2	2
B2	B2, C4, D4, D1, G12	19	21
B3	B3	1	1
B9	B9	1	1
C1	C1, C5	10	11
C2	C2	12	13
C8	C8, D9	5	6
D2	D2	1	1
E2	E2	1	1
E3	E3, D7	2	2
E4	E4	2	2
F1	F1	1	1
F2	F2	1	1
F5	F5	1	1
F8	F8	1	1
G10	G10	2	2
H1	H1	1	1
H11	H11	3	3
H2	H2	1	1
H5	H5	1	1
H6	H6	3	3
Total:		90	
<i>Rhynchotermes</i> sp. Cost004			
A1	A1	1	1
A12	A12	2	2
A2	A2	2	2
A3	A3	1	1
A4	A4	2	2
A5	A5	1	1
A6	A6	1	1
A7	A7, E8	4	4
A9	A9	5	5
B1	B1	4	4
B11	B11	3	3
B2	B2, C8, H6	6	6
B3	B3	1	1
B6	B6	1	1
B7	B7, H7	2	2
C11	C11	2	2
C12	C12	1	1
C4	C4, E12, H1	3	3
C5	C5, G5, H12	7	7
D10	D10, G11	7	7
D11	D11	1	1
D5	D5	2	2
D6	D6	2	2
E2	E2	2	2
E3	E3	1	1
E4	E4	3	3
E6	E6	2	2
F1	F1	1	1
F10	F10	1	1
F11	F11	1	1
F2	F2 ^d	6	6
F6	F6	1	1
F8	F8	1	1
F9	F9	1	1
G1	G1	1	1
G2	G2	1	1
G3	G3	1	1
G4	G4	1	1
G9	G9	1	1
H11	H11	1	1
H2	H2	2	2
H3	H3	1	1
H5	H5	1	1
H8	H8	2	2
Total:		93	

^aOTUs calculated using the furthest-neighbor method in DOTUR with a 97% amino-acid similarity cut-off.

^bNumber of cloned sequences grouped within each OTU.

^cPercent of cloned sequences represented by each OTU.

^dSequences containing frame-shift mutations that were “corrected” manually by the addition or subtraction of nucleotides at the DNA sequence level to allow for phylogenetic analyses using amino acid sequences.

^eSequence OUT grouping phylogenetically with sequences previously classified as Family 7 [FeFe] hydrogenases (53).

*Chapter 5*MICROFLUIDIC DIGITAL PCR REVEALS THAT TREPONEMES MAY BE A
PREDOMINANT GENUS OF EUBACTERIA ENCODING AN IMPORTANT
FAMILY OF [FeFe] HYDROGENEASES IN THE GUT OF *RETICULITERMES*
*TIBIALIS***Abstract**

Hydrogen is an important free intermediate in the degradation of wood by termites and is turned over at high fluxes and maintained at concentrations exceeding those measured for any other biological system. We have employed microfluidic digital PCR to identify bacteria encoding [FeFe] hydrogenase genes and, therefore, potentially participating in hydrogen metabolism in the gut of *Reticulitermes tibialis*. We successfully designed degenerate primers specifically targeting the largest group of [FeFe] hydrogenases observed in a termite hindgut metagenome. Nucleotide sequences gathered in previous molecular inventories were utilized in probe design. 27 16S rRNA – Family 3 [FeFe] hydrogenase gene pairs from putative single cell genomes were successfully co-amplified by multiplex PCR in microfluidic chambers and subsequently sequenced. 22 of the 16S rRNA phlotypes were treponemal, and of these 16 fell within the termite cluster of treponemes. All instances of the same 16S rRNA – hydrogenase gene pairings observed in multiple independent chambers, referred to as “*Reticulitermes* environmental genomovars, corresponded to treponeme phlotypes. The non-treponemal phlotypes fell within the β - and ϵ -Proteobacteria and Bacteroidetes phyla. All but 3 of the phlotypes grouped closely with phlotypes previously sequenced from the guts of *Reticulitermes* termites. A number of the hydrogenase gene peptide sequences grouped closely with sequences cloned from *Reticulitermes hesperus* in a previous study. Our

results provide compelling evidence that treponemes, particularly of the termite cluster, represent an important genus encoding Family 3 [FeFe] hydrogenases in the gut of *R. tibialis* that perhaps is comprised of members making an important contribution to bacterially mediated hydrogen metabolism.

Introduction

Hydrogen plays a pivotal role in the breakdown of lignocellulosic biomass by termites (3, 7, 15, 39, 40, 49). It can reach concentrations in some species exceeding those measured for any other biological system (15, 18, 49, 52, 55-57) and fluxes have been measured as high as $33 \text{ m}^3/\text{m}^3$ gut volume (49). Moreover, the gut ecosystem is spatially complex being comprised of numerous microenvironments formed as a consequence of the numerous chemical gradients extant in the gut, hydrogen being among the most prominent (9, 10, 15, 24, 25, 49).

This hydrogen is produced and turned-over as an important metabolic free intermediate in the breakdown of lignocellulosic biomass by the complex microbial community residing in the termite gut (17, 21, 22, 40, 58, 65, 66). Termites are entirely dependent upon this symbiosis, which can include representatives of all three domains of life, to derive energy and carbon from wood (4-6, 8, 11, 12, 39). Acetate, which serves as a termite's primary carbon and energy source, hydrogen and carbon dioxide are the primary products produced in this symbiosis (41). Only a small fraction of the hydrogen and carbon dioxide are emitted from the gut; indeed, these potential waste products are used by bacteria in reductive acetogenesis to produce up to 33% of the total acetate pool in the gut (3, 7, 26, 41, 49). This contributes substantially to the metabolic efficiency of the system and has led some to postulate that it is among the smallest and most efficient bioreactors found in nature (49). Only a small fraction of the hydrogen and carbon dioxide is lost from the system as methane produced by methanogens (3, 24).

The sequencing of a termite hindgut metagenome has revealed the presence of an abundance and striking diversity of hydrogenase-like proteins in the termite gut (61).

The vast majority of these were classified as [FeFe] hydrogenase-like proteins (61). Through for some of these sequences predictions as to the identity of the host of origin were made based upon nucleotide composition using Phylopythia (34, 61), it remains nebulous precisely what bacterial species encode these sequences or play an important role in hydrogen metabolism in the termite gut.

Traditionally, efforts to identify microbes filling a particular physiological niche in an environment have relied upon molecular inventories of a structural gene essential for a phenotype of interest. Numerous studies have taken this approach, for instance, to gain a better understanding of nitrogen metabolism (64, 68, 69), methanogenesis (14), or hydrogen metabolism (2, 16, 62, 63) in environmental samples. Several notable studies have taken this approach to analyze the diversity of microbes participating in reductive acetogenesis (48, 51) or nitrogen metabolism (38, 43, 64) in the termite gut. To infer a bacterium's phylogeny, these studies, by necessity, assume a reliable correlation between host phylogeny and the structural gene of interest. In some instances this may indeed be a reliable assumption, but it is notably not the case for [FeFe] hydrogenases (54, 59). Schmidt *et al.* noted in their 2010 study, "the presence of homologous multiple hydrogenases per organism and inconsistencies between 16S rRNA gene and [FeFe]-hydrogenase based phylogenies" (54). One might propose addressing this challenge by isolating bacteria from the termite gut and subsequently annotating hydrogenases encoded in their genome sequences, as has been done in the cases of *Treponema primitia* ZAS-1 and ZAS-2 and *Treponema azotonutricium* ZAS-9 in Chapter 2. But, on account of the resistance of termite gut bacteria to cultivation (23), one can not feasibly

investigate global community characteristics of the termite gut by isolating a large sampling of its component species.

Microfluidic digital PCR (dPCR) was developed by Ottesen *et al.* in 2006 and successfully used to co-amplify formyl-tetrahydrofolate synthetase (FTHFS) gene sequences and 16S rRNA gene sequences from single bacterial cells from a termite gut (45). This method allowed for the unambiguous phylogenetic identification of bacteria encoding FTHFS (45), an important marker for reductive acetogenesis (27, 28). Moreover, since a single microbial genome is used as a template in microfluidic digital PCR, the challenges of primer bias or chimera formation that plague the traditional molecular profiling techniques discussed above (1) were largely obviated.

Here we report the use of microfluidic digital PCR to unambiguously identify bacterial species encoding [FeFe] hydrogenases in the gut microbial community of *Reticulitermes tibialis*. Our objective was to use this metabolic marker as a means to identify bacteria that may contribute to hydrogen metabolism in the termite gut.

Methods

Sample collection and classification. *Reticulitermes tibialis* collection JT2 termites were collected from a single colony found in a fallen tree along the side of a road in Joshua Tree National Park (Permit#: JOTR:2008-SCI-002) in Southern California on December 15, 2009 at 3:00 PM. The GPS coordinates of the site were 34° 1' 7.2" N and 116° 10' 6.9" W. Termites were immediately used in experiments. During the 5 days duration of experiments, they were stored in a room-temperature glass aquarium with moist, sterile sand, and wood cut from the tree from which they were collected. The atmosphere of the chamber was maintained at 95% humidity.

The termites were classified using insect mitochondrial cytochrome oxidase subunit II (COII) gene sequences. The COII genes were amplified from the head of one of the termites used in a digital PCR experiment. The head was macerated with a sterile glass rod in a 2 ml tube with 50 μ l TRIS-EDTA (Sigma). The liquid fraction was transferred to another tube and incubated at 95°C for 10 min and subsequently diluted 10-fold in nuclease-free water. 3 μ l of this solution was used as a template in a 50 μ l PCR reaction. The template was amplified using primers COII-R and COII-F (see Table 5-1), each at a final concentration of 1 μ M, Expand High Fidelity Taq Polymerase (Roche), and FailSafe Premix D (Epicentre). The thermal cycling regimen was 94°C 3 min, (94°C 30s, 50°C 30s, 72°C 1.5 min) x 30, 72°C 10 min. Amplicons were cloned into One Shot Top10 chemical competent *Escherichia coli* cells (Invitrogen) using a QIAGEN PCR Cloning Kit following manufacturers' protocols. Colonies were submitted to GENEWIZ for sequencing. Sequences were edited manually using SeqMan, available from DNA* as part of the Lasergene software suite, and analyzed phylogenetically using the ARB software environment (32) in a manner analogous to that described below. The termites were classified as *Reticulitermes tibialis* (see Figure 5-1) and given the identifier "collection JT2."

Microbial Strains. Microbial isolate *Treponema primitia* str. ZAS-2 was grown in anaerobic YACo medium under a headspace of 80% H₂ + 20% CO₂, as described previously (26, 29). *Treponema azotonutricium* str. ZAS-9 was grown in a similar medium (26, 29).

Primer and Probe Design. Degenerate primers and probes for the detectable amplification of "Family 3" (61) [FeFe] hydrogenases in quantitative PCR (qPCR) and

Table 5-1. Primers and probes used.

Primer	Sequence^a	Target^b	Reference
386 F'	5' - CIC GIA TGA THA ARC ARG CIG G - 3'	Family 3 [FeFe] hydrogenase	this study
467 R'	5' - CCA TYT GRT GIG CIA YIG C - 3'	Family 3 [FeFe] hydrogenase	this study
1100R	5' - AGG GTT GCG CTC GTT G - 3'	16S	-
1492RL2D	5' - TAC GGY TAC CTT GTT ACG ACT T - 3'	Gen. Bac. 16S rRNA	(45)
357F	5' - CTC CTA CGG GAG GCA GCA G - 3'	Gen. Bac. 16S rRNA	-
COII-R	5' - GTT TAA GAG ACC AGT ACT TG - 3'	cytochrome oxidase II gene	(31)
COII-F	5' - ATG GCA GAT TAG TGC AAT GG - 3'	cytochrome oxidase II gene	(31)
LNA H2-1a ^c	5' - 56-FAM-CTT CCA TGA C-3BHQ_1 - 3'	Family 3 [FeFe] hydrogenase (probe)	this study
LNA H2-1b ^c	5' - 56-FAM-CTT CCA TAA C -3BHQ_1 - 3'	Family 3 [FeFe] hydrogenase (probe)	this study
Hex-1389Prb	5' - 5HEX-CTT GTA CAC ACC GCC CGT C-3BHQ_1 -3'	Gen. Bac. 16S rRNA (probe)	(45)

^a“I” represents inositol.

^bGen. Bac., general bacterial.

^cBoldface residues are locked nucleic acids.

Figure 5-1.

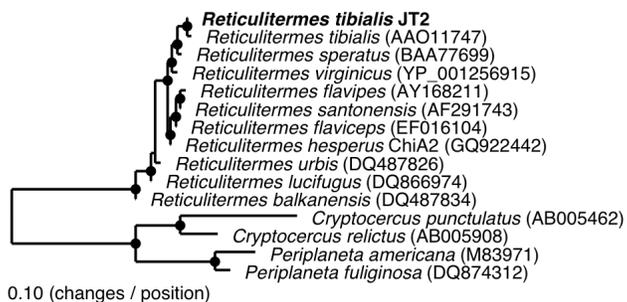


Figure 5-1. Mitochondrial cytochrome oxidase II (COII) phylogeny of the termite sample used in this study. Sequence accession numbers are listed in parentheses. The COII sequence cloned from the termite sample used in this study is in bold. The tree was calculated using a maximum likelihood (Phylip ProML) method with 225 unambiguously aligned amino acid positions. Closed circles designate groupings also supported by parsimony (Phylip PROTPARS, 1000 bootstraps) and distance matrix (Fitch) methods.

dPCR were designed manually from multiple sequence alignments, see Figures 5-S1 and 5-S2. Family 3 [FeFe] hydrogenases were the most highly represented group of enzymatic hydrogenases observed in the *Nasutitermes* gut metagenome sequence (61) and have also been observed, see Chapter 2, in the genome sequences of treponemes isolated from the gut of *Zootermopsis angusticolis*. They are the only group of hydrogenases whose *in situ* translation has been verified (61).

An alignment prepared using DIALIGN (36), available on the Mobyly Portal (37), of [FeFe] hydrogenase gene sequences cloned previously from *Reticulitermes hesperus* collection ChiA2 revealed a highly conserved nucleotide region, see Figure 5-S2 in the chapter appendix. Due to the short length of the region, 10 base pairs, it was necessary to use locked nucleic acids (LNA) (37) to design probes targeting it with sufficiently high melting temperatures to be use in qPCR or dPCR. The probes designed, see Table 5-1, target 89% of the hydrogenase sequences cloned in a gut microbe molecular inventory prepared from *R. hesperus*, see Chapter 3. Probes were manufactured by Integrated DNA Technologies.

For degenerate primer design, the peptide sequences for all bacterial [FeFe] hydrogenase genes previously cloned from termite guts and a wood roach (see Warnecke *et al.* (61) and Chapters 2, 3 and 4) were aligned using DIALIGN on the Mobyly Portal. A portion of the alignment is shown in Figure 5-S2 in the chapter appendix. Highly conserved regions identified in this alignment, and flanking the sequence targeted by the probes designed as described above, were used in the design of degenerate primer combinations for use in dPCR. A functional primer set and optimal conditions for gene amplification were determined empirically. All oligomers were ordered from Integrated DNA

Technologies. Initial screens for functionality of candidate primer sets and PCR conditions were done using qPCR. Genomic DNA purified from *Treponema primitia* ZAS-2 or *T. azotonutricium* ZAS-9 was used as template in these screens. The degenerate primers targeting Family 3 [FeFe] hydrogenases reported in Chapter 3 did not function well in qPCR. Successful primer sets and PCR conditions were defined as those facilitating successful quantitative discrimination between different template concentrations in a dilution series in a multiplex qPCR reaction with the primer and probe set targeting 16S rRNA gene sequences, see Table 5-1. The best conditions identified using qPCR, with minor modifications introduced following control dPCR experiments, are described below as those used for all dPCR experiments. The best degenerate primer set identified is listed in Table 5-1. The primers target the H domain known to be highly conserved among all [FeFe] hydrogenases (35, 60). The amplified region corresponds approximately to the regions spanning I393-V470 and I272-V349 in the [FeFe] hydrogenases from *Clostridium pasteurianum* (P29166) and *Desulfovibrio vulgaris* (YP_010987), respectively.

Template Preparation for Digital PCR. For each digital PCR run, a single *Reticulitermes tibialis* collection JT2 gut was dissected out of a worker termite and resuspended in 250 μ l of a “synthetic gut fluid” (SGF) solution. The SGF comprised 500 ng/ml bovine pancreas RNase (Roche), 10 mM Tris pH 8, 1 mM EDTA, 30 mM NaCl, and 60 mM KCl in water. The contents of the whole gut was suspended in this solution by pipetting up and down and crushing it against the sides of a centrifuge tube using a 1 ml tip. This suspension was used as the template for use in digital PCR. In control experiments, aliquots of *Treponema azotonutricium* ZAS-9 or *Treponema primitia* ZAS-

2 cells in mid-log phase growth were added to the SGF solution instead of a termite gut suspension. The crushed gut suspensions were allowed to stand for 30s to allow large particles to settle. The templates were diluted to working concentrations in SGF. These template solutions were then mixed with the PCR reaction mixture described below and immediately loaded onto a microfluidic digital array.

Digital PCR Protocol. Equipment for microfluidic digital PCR was purchased from Fluidigm Corporation and included a BioMark with its controlling software, a Nanoflex IFC controller with its controlling software, and custom microfluidic digital array 12.765P digital PCR devices. On-chip PCR reaction solutions contained iQ Multiplex Powermix (BioRad), 0.1% Tween-20 (Sigma), and 400 nM Rox standard (Quanta Biosciences). The primers and probes used are all listed in Table 5-1 and were used at the following concentrations: 154 nM each of 357F and 1492RL2D; 304 nM each of 386 F' and 467 R'; 450 nM each of LNA H2-1a and LNA H2-1b; 300 nM Hex-1389Prb.

Arrays were loaded and PCR was performed as recommended by Fluidigm. The thermal cycling protocol was 95°C 5 min, (95°C 15 s, 60°C 45 s) x 45, 60°C 10 min, 20°C 10 min. Data was analyzed using the Fluidigm Digital PCR Analysis program version 2.1.1, build 20090521.1140. The program data analysis parameters were set to the following to detect wells containing putative positive co-amplifications: target Ct range of 23-45, linear base correction, user data (global) Ct threshold method, a quality threshold of 0.65, and thresholds of 0.03 for FAM-MGB and 0.05 for VIC-MGB signals. FAM-MGB and VIC-MGB thresholds were selected such that the number of putative amplifications detected in a no template control panel run on a digital array were routinely on the order of 1.5% or less.

Sample Retrieval and Analysis. The thermal conducting silicon wafer was removed from each digital PCR array device using a device purchased for this purpose from Fluidigm. Pressure was released from each device after this process. Products were only retrieved from panels with less than 33% of the chambers containing putative 16S rRNA gene amplicons. Assuming a Poisson distribution of particles, this precaution should have assured that only 6% of the chambers contained multiple particles. Chambers containing putative successful co-amplifications were located with the aid of a dissecting microscope and pierced with the tip of a 26-gage needle. The needle was then mixed briefly in 10 μ l TRIS-EDTA (Sigma).

Retrieved samples were initially screened for the presence of an [FeFe] hydrogenase gene product by PCR. 2 μ l of each suspension were added as template to PCR reaction mixes comprised of 1.4 U Expand High Fidelity Taq (Roche) and FailSafe PCR Premix D (Epicentre), and 1 μ M each of primers 386F' and 467R', see Table 5-1. The thermal cycling program used was 95°C 5 min, (95°C 15 s, 60°C 45 s, 72°C 60 s) x 40. Reactions yielding a product detectable in gel electrophoresis were PCR purified using a QIAquick PCR Purification Kit (QIAGEN) following the manufacturer's protocol. 2 μ l of the resulting solutions were then used as template in PCR reaction mixes comprised of FailSafe PCR Premix D (Epicentre), 1 U Taq polymerase (Roche), and 500 nM each of primers 386 F' and 467 R', see Table 5-1. The thermal cycling regimen was 95°C 5 min, (95°C 15 s, 60°C 45 s, 72°C 60 s) x 15. The products of these reactions were cloned into One Shot[®] Top10 chemical competent *Escherichia coli* cells (Invitrogen) using the QIAGEN PCR Cloning Kit following manufacturer protocols.

All samples that yielded a detectable product in the above screen were further screened for the presence of a 16S rRNA gene product. 2 µl of each resuspension were added to PCR mixes containing iQMultiplex Powermix (BioRad) and 1 µM each of primers 1492RL2D and 357F, see Table 5-1. The thermal cycling regimen was 95°C 5 min, (95°C 15 s, 60°C 45 s, 72°C 60 s) x 30. Samples yielding a product detectable in gel electrophoresis were purified using a QIAquick PCR Purification kit (QIAGEN) following the manufacturer's protocol. 2 µl of the resulting suspensions were used as template in PCR reactions mixes comprised of FailSafe PCR Premix D (Epicentre), 1 U Taq polymerase (Roche), and 500 nM each of primers 1492RL2D and 357F, see Table 5-1. The thermal cycling regimen was 95°C 5 min, (95°C 15 s, 60°C 45 s, 72°C 60 s) x 15. The products of these reactions were cloned into One Shot Top10 chemical competent *Escherichia coli* cells (Invitrogen) using a TOPO TA Cloning Kit (Invitrogen) following manufacturer protocols. Clones corresponding to array chambers A1, A8, A11, A12, B5, B2, B6, B9, C7, and D9 were prepared using a QIAGEN PCR Cloning Kit (QIAGEN) following manufacturer protocols.

For each dPCR amplicon, 10 individual clones were selected at random for restriction fragment length polymorphism (RFLP) analysis. Each clone was suspended in TRIS-EDTA (Sigma) and used as a template for PCR. Sequences were amplified in PCR using T7 and T3 primers, NEB Taq Polymerase (New England Biolabs) and FailSafe PCR Premix H (Epicentre). Sequences cloned using the QIAquick PCR purification kit were amplified using SP6 and T7 primers. The temperature cycling program was 95°C 5 min, (95°C 30 s, 55°C 30s, 72°C 5 min) x 25, 72°C 10 min.

The products of each reaction were subjected to digestion with *HinPII* and the resulting RFLP patterns were analyzed by agarose gel electrophoresis. Cloned sequences representing unique RFLP patterns were arbitrarily selected for submission to GENEWIZ for colony sequencing.

Sequences were manually trimmed in SeqMan, available from DNA* as part of the Lasergene software suite, to remove plasmid and degenerate primer sequences. The 16S rRNA sequences were checked for chimeras using GreenGenes/Bellepheron (13) available on the GreenGenes website (13). A more rigorous screening for chimeras was not deemed necessary because digital PCR, by its very nature, dramatically reduces the likelihood of their formation because a single cell is used as the template in PCR reactions. The [FeFe] hydrogenase sequences were each BLASTed against the *Nasutitermes* hindgut metagenome sequence on the IMG/M server (33) to verify, by means of the top hits, their identity as Family 3 [FeFe] hydrogenase sequences.

The ARB software environment (32) was used for phylogenetic analysis of the 16S-rRNA gene and [FeFe] hydrogenase peptide sequences. Sequences cloned in this study and used for tree construction are listed in Table 5-S1 in the chapter appendix. Peptide sequences were aligned using DIALIGN (36), available on the Mobyly server (37). 16S rRNA gene sequences were aligned using the Silva aligner (53). If 16S rRNA gene nucleotide sequences or [FeFe] hydrogenase peptide sequences corresponding to amplicons from a dPCR array chamber shared less than 97% sequence identity, all sequences from the chamber were discarded as a precaution assuming that this may be an indication of either contamination or co-localization of multiple cells.

Nomenclature. All cloned dPCR products were given the prefix "Rt" designating the termite host of origin, *Reticulitermes tibialis* collection JT2, followed by an "H2" for all hydrogenase sequences or an "R" for all 16S rRNA gene sequences. A dash followed by a letter-number combination uniquely identifying the dPCR array chamber from which a given sequence was collected was then added to this prefix. This chamber designator was then appended on the right by a period and a number designating the identity of the specific clone out of the 10 prepared for each hydrogenase or 16S rRNA gene amplicon retrieved from a dPCR array chamber.

Reticulitermes environmental genomovars (REG), using terminology proposed by Ottesen *et al.* (45), were defined as sets of sequences co-amplified in dPCR for which the 16S rRNA gene nucleotide sequences and [FeFe] hydrogenase peptide sequences each share at least 97% identity. REGs were arbitrarily assigned a number identifying the REG followed by a period and another number identifying a unique dPCR chamber that sequences co-amplified in. The final "r" or "h" in each REG name serves to identify a sequence comprising the REG as a 16S rRNA gene sequence or an [FeFe] hydrogenase peptide sequence, respectively.

Results

Sequence amplification and retrieval. Efforts to employ the method of Ottesen *et al.* (44) used in the past in the design of degenerate primers with an appended probe binding site (70) for use in digital PCR failed. This method is typically employed by necessity when little is known about an environmental gene sequence of interest. This obstacle was overcome in this study by utilizing nucleotide sequences of Family 3 [FeFe] hydrogenases cloned previously from the gut of *R. hesperus* collection ChiA2, see

Chapter 3. These sequences allowed for the design of probes targeting termite gut Family 3 [FeFe] hydrogenases, as described in the methods section.

In control experiments, the Family 3 [FeFe] hydrogenase and 16S rRNA gene sequences of *T. azotonutricium* ZAS-9 and *Treponema primitia* ZAS-2 co-amplified in dPCR. In these control experiments, there were, on average (2 panels analyzed in each of two dPCR experiments), 9.5 16S rRNA gene amplification false positives and less than 1 [FeFe] hydrogenase false positive observed in array panels loaded with PCR mix containing no template.

A heat map for a representative panel from which PCR products amplified from *Reticulitermes tibialis* collection JT2 gut microbes were retrieved is shown in Figure 5-2. The digital array devices, panels, and chambers from which samples putatively amplified from a single cell were successfully retrieved and subsequently sequenced are listed in Table 5-2. Samples were retrieved from 16 panels distributed across four digital PCR array devices. The panels from which product was retrieved had an average of 180 ± 51 positive 16S rRNA gene signals, 11 ± 5 [FeFe] hydrogenase gene signals, and 4 ± 3 putative co-amplifications.

Sequence Analysis. Twenty-seven chambers yielded co-amplification products successfully sequenced and putatively corresponding to a single cell genome according to the criterion put forward in the methods section, see Table 5-S1 in the chapter appendix and Table 5-2. Of these, 22, or 81%, were classified as treponemal, see Figure 5-3. All of the treponemal 16S rRNA gene sequences fell within one of two groupings. Either sub-group 2 of known treponemes defined previously by Paster *et al.* (47), also referred to as “sub-cluster II” by Hongoh *et al.* (20), and the group of phylotypes referred to as

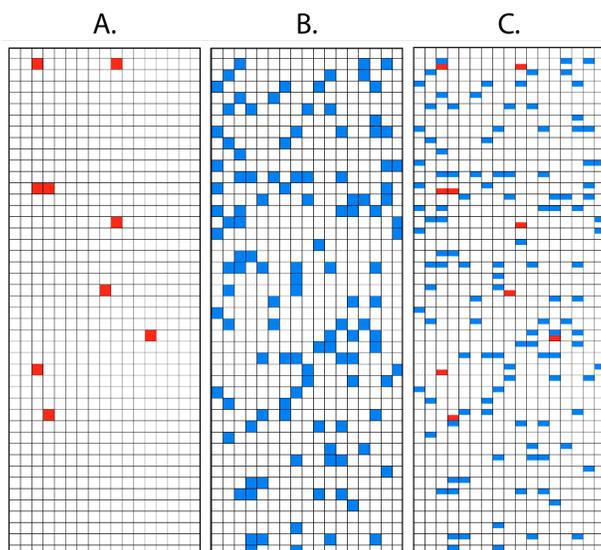
Figure 5-2.

Figure 5-2. Representative heat map for a digital PCR array panel from which amplicons were retrieved for analysis. The heat maps were constructed using the Fluidigm Digital PCR Analysis program version 2.1.1, build 20090521.1140. The images correspond to panel 6 of the device with serial number 1151065028. A.) Wells with positive FAM-MGB signals corresponding to putative Family 3 FeFe hydrogenase gene amplicons, B.) Wells with positive VIC-MGB signals corresponding to 16S rRNA gene amplicons, C.) An overlay of the heat maps represented in A. and B. used to identify wells corresponding to positive co-amplifications.

Table 5-2. Digital PCR array wells from which amplicons were retrieved.

Device Serial No.	Panel ^a	H ₂ ase signal ^b	16S signal ^c	Co-amp. ^d	Well ^e
1151065021	2	3	129	1	A1
1151065021	7	6	129	3	A8
1151065021	8	12	152	5	A11, A12
1151065021	9	8	253	4	B2, B5
1151065021	10	6	127	1	B6
1151065021	11	14	212	9	B9
1151065030	9	22	244	10	C7
1151065030	11	12	248	4	D9
1151065028	1	9	195	4	D11, E4
1151065028	3	12	231	4	E5
1151065028	6	9	119	2	E9
1151065028	8	7	106	2	E12
1151065028	12	9	205	3	F4, F2
1151065037	7	14	211	3	G6, G7
1151065037	8	21	185	7	G12, G10, H2
1151065037	10	16	135	9	H6, H11, H8, H10, H7

^aPanel (1 of 12) of the microfluidic digital PCR array.

^bNumber of wells having a positive FAM-MGB signal (see Methods), corresponding to a putative [FeFe] hydrogenase gene amplification.

^cNumber of wells having a positive VIC-MGB signal (see Methods), corresponding to a putative 16S rRNA gene amplification.

^dNumber of wells predicted to have successful 16S rRNA gene and [FeFe] hydrogenase gene amplifications.

^eThe names chosen for the wells from which amplicons were retrieved and subsequently yielded products putatively originating from a single cell genome (see Methods) and used in sequence analyses.

Figure 5-3.

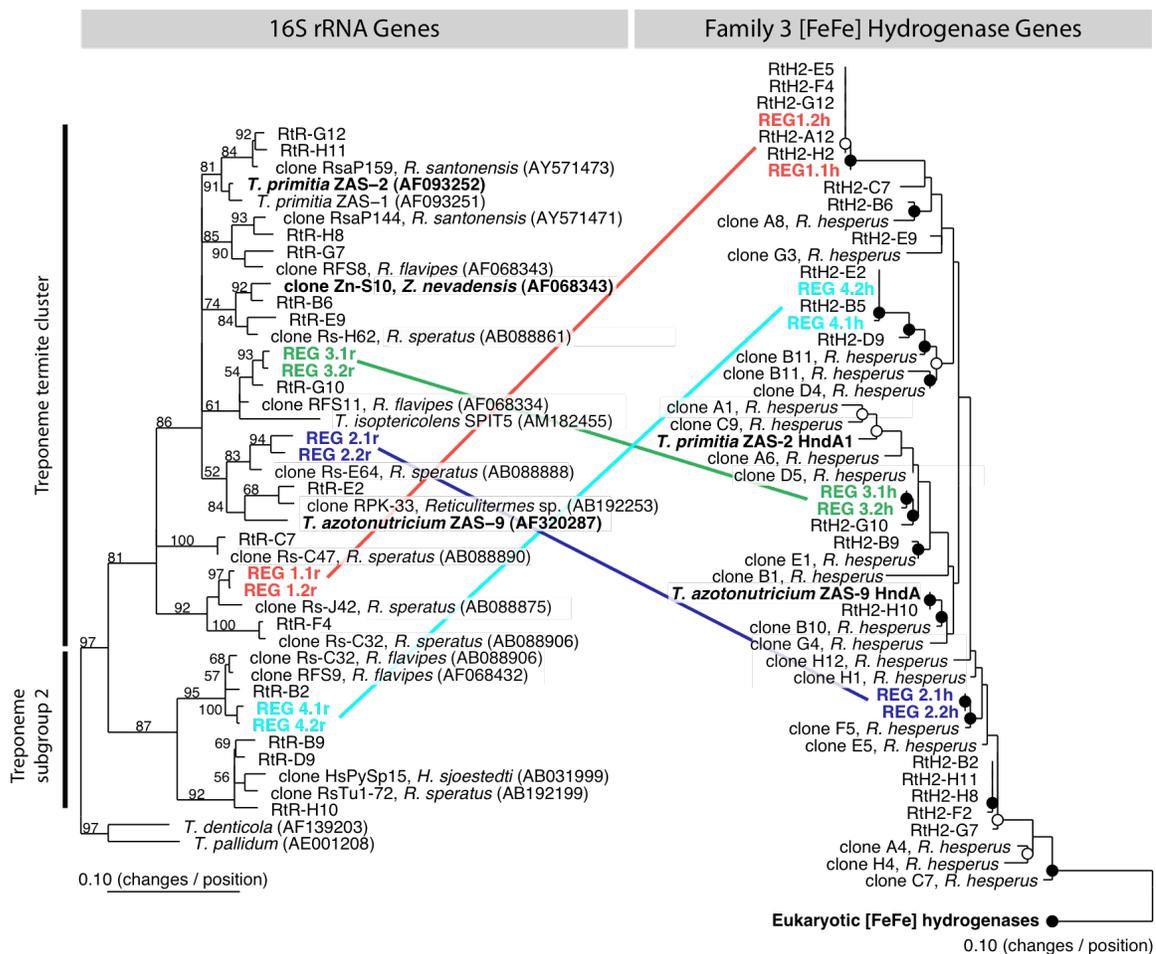


Figure 5-3. Phylograms of 16S rRNA gene sequences (left panel) and putative Family 3 [FeFe] hydrogenase peptide sequences (right panel) corresponding to products co-amplified in dPCR. The 16S rRNA gene sequence tree was calculated using the TreePuzzle algorithm (37) with 1,000 puzzling steps and 963 unambiguously aligned nucleotides. The numbers adjacent to branches are bootstrap values. Sequences from the genomes of termite gut treponeme isolates known to encode Family 3 [FeFe] hydrogenases in their genomes are in boldface, as is a 16S rRNA gene (clone Zn-S10) shown previously to co-amplify with a gene for FTHFS in dPCR (45). The hydrogenase peptide sequence tree was calculated using a maximum likelihood (Phylip ProML) method with 70 unambiguously aligned amino acids. On the hydrogenase tree, open circles designate groupings also supported by either parsimony (Phylip PROTPARS, 1000 bootstraps) or distance matrix (Fitch) methods. Closed circles on the hydrogenase tree designate groupings supported by all three methods. For each REG (see text for definition), its associated sequences share the same color and a line has been connecting its hydrogenase and 16S sequences across the two trees. Sequence accession numbers are listed in parentheses. Clone names of sequences taken from public databases that are derived from a termite gut are each listed followed by a comma and the name of the termite host of origin. *Reticulitermes hesperus* hydrogenase sequences are taken from Chapter 3.

“the termite cluster” by Lilburn *et al.* (30), or as “subcluster I” by Hongoh *et al.* (20). Most, 16 of 22 total, of the treponemal 16S rRNA gene phylotypes fell within the termite cluster. All but one, RtR-B6, of the treponemal phylotypes grouped with sequences cloned in previous studies from termites belonging to the genus *Reticulitermes*. These sequences shared an average of $98.2\% \pm 0.86\%$ (minimum of 96.5%), across 917 aligned nucleotides, with the *Reticulitermes* gut microbe sequences available in public databases to which they associated most closely in phylogenetic analyses. One 16S rRNA gene sequence grouped closely (98.1% sequence identity across 917 aligned nucleotides) in phylogenetic analyses with clone ZN-S10 from the gut of *Z. nevadensis* proposed previously results to originate from a bacterium also encoding a gene for FTHFS (45). The 4 REGs observed in this study, see Table 5-3, have been predicted to correspond to treponemes, see Figure 5-3.

Five of the 16S rRNA gene sequences fell outside the phylum Spirochaetes in phylogenetic analyses, see Figure 5-4. Three grouped with Beta-Proteobacteria, one with Bacteroidetes, and one with Epsilon-Proteobacteria. Three of these sequences shared an average of $98.2\% \pm 0.93\%$ (minimum of 97.4%), across 917 aligned nucleotides, with the *Reticulitermes* gut microbe sequences available in public databases to which they associated most closely in phylogenetic analyses. The other two sequences did not group with any sequences cloned previously from termites belonging to the genus *Reticulitermes*. There was not an unambiguous correlation between 16S rRNA gene and [FeFe] hydrogenase peptide sequence phylogenies, see Figures 5-3 and 5-4. In 10 instances, Family 3 [FeFe] hydrogenase peptide sequences grouped closely in phylogenetic analyses with sequences cloned previously, see Chapter 3, from the gut

Table 5-3. *Reticulitermes* environmental genomovars proposed in this study.

Clone Name	Gene	REG	Min. REG Seq. Id. (%) ^a
RsH2-A1	[FeFe] hydrogenase	REG 1.1h	98.5%
RsH2-E12	[FeFe] hydrogenase	REG 1.2h	98.5%
RsH2-A11	[FeFe] hydrogenase	REG 2.1h	98.5%
RsH2-A8	[FeFe] hydrogenase	REG 2.2h	98.5%
RsH2-H6	[FeFe] hydrogenase	REG 3.1h	98.5%
RsH2-H7	[FeFe] hydrogenase	REG 3.2h	98.5%
RsH2-D11	[FeFe] hydrogenase	REG 4.1h	98.5%
RsH2-G6	[FeFe] hydrogenase	REG 4.2h	98.5%
RsR-A1	16S	REG 1.1r	99.0%
RsR-E12	16S	REG 1.2r	99.0%
RsR-A11	16S	REG 2.1r	98.5%
RsR-A8	16S	REG 2.2r	98.5%
RsR-H6	16S	REG 3.1r	98.9%
RsR-H7	16S	REG 3.2r	98.9%
RsR-D11	16S	REG 4.1r	98.5%
RsR-G6	16S	REG 4.2r	98.5%

^aThe minimum percent identity shared by all sequences corresponding to a given set of hydrogenase peptide sequences or 16S gene sequences comprising a REG (see Table 5-S1 in the chapter appendix for a list of all sequences)

Figure 5-4.

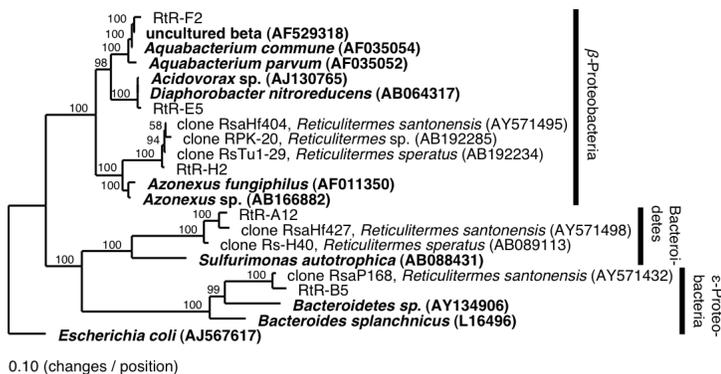


Figure 5-4. Phylogram of non-treponemal 16S rRNA gene sequences co-amplified with Family 3 [FeFe] hydrogenase genes in dPCR. The tree was calculated using the TreePuzzle algorithm (37) with 1,000 puzzling steps and 931 unambiguously aligned nucleotides. The numbers adjacent to branches are bootstrap values. 16S gene sequences from pure culture isolates are in boldface. Sequence accession numbers are listed in parentheses. Clone names of sequences taken from public databases that are derived from a termite gut are listed followed by a comma and the name of the termite host of origin.

microbial community of *Reticulitermes hesperus* collection ChiA2. In one instance, a hydrogenase peptide sequence, predicted to be treponemal in origin, was identical across 70 aligned amino acid residues with the Family 3 FeFe hydrogenase from *T. azotonutricium* ZAS-9.

Discussion

We have used microfluidic digital PCR to directly associate 16S rRNA gene sequences with Family 3 [FeFe] hydrogenase structural genes from the genomes of individual members of the gut microbial community of a lower termite. Hydrogen metabolism plays a prominent role in the degradation of wood in the termite gut (3, 7, 15, 39, 40, 49). By employing this microfluidic approach, we overcame many of the limitations inherent in traditional gene-inventory techniques for the characterization of bacterial genre or phyla filling a physiological niche in an environment.

Traditionally, to identify important microbial contributors to a particular physiological function in the environment, degenerate primers are designed targeting a gene important to a physiology of interest and environmental gene inventories are prepared using them (2, 14, 16, 62-64, 68, 69). The identity of the microbes from which the genes originate is then inferred based upon the phylogeny of the functional gene assuming a close correlation between the evolution of the gene and its cognate genome. This is not always a valid assumption, particularly in the case of [FeFe] hydrogenases whose phylogeny has been shown to be an unreliable predictor of the genus of its host of origin (54, 59). In fact, in this study we observed instances of very closely related [FeFe] hydrogenase gene sequences associating with 16S rRNA gene phylogtypes from different bacterial phyla, see Figures 5-3 and 5-4. This necessitates the use of methods such as microfluidic digital

PCR, a technique recently developed by Ottesen *et al.* (45), that allow for multiplex PCR using a single cell genome as template.

Employing the technique of microfluidic digital PCR, we have shown that treponemes may be an important, perhaps dominant, bacterial group encoding Family 3 [FeFe] hydrogenases in the gut of *Reticulitermes tibialis* collection JT2. This is important because Family 3 [FeFe] hydrogenase were the predominant group of hydrogenases observed in a termite hindgut metagenome and the only family for which translated products were detected *in situ* by mass spectroscopy (61). This family of [FeFe] hydrogenase was also observed in the genome sequences of two treponemes isolated from the gut of *Zootermopsis angusticolis*, see Chapter 2. Our results provide evidence consistent with an important role for treponemes in the metabolism of hydrogen, the “central free intermediate” in the degradation of wood by termites (49), in the gut microbial communities of termites. This is in agreement with findings supporting a similar contribution of treponemes to acetogenesis (48, 51), an important hydrogen sink in the termite gut ecosystem (3, 7, 26, 41, 49). In fact, one of the 16S rRNA gene phylotypes grouped closely with sequence Zn-S10 cloned from the gut of *Z. nevadensis*, shown previously in microfluidic digital PCR to co-amplify with an FTHFS gene, an important marker for acetogenesis (27, 28). Moreover, the majority of the 16S rRNA gene phylotypes fall within the termite cluster of treponemes defined by Lilburn *et al.* (30), which is the cluster into which most termite gut treponemal phylotypes have associated in previous 16S rRNA gene inventories (20, 30, 42, 67). The remaining treponemal 16S rRNA gene phylotypes associated with treponeme subgroup 2, as defined

by Paster *et al.* (47), and not subgroup 1, which is also consistent with the results of previous molecular inventories (20, 30, 42, 67).

In microfluidic digital PCR, there is always the possibility that false associations may be observed as a consequence of multiple cells or genomic fragments localizing to a single chamber. For this reason, we are suspicious of a 16S rRNA gene – [FeFe] hydrogenase gene co-localization in a dPCR chamber until the same co-localization is observed in another, independent, microfluidic chamber. Importantly, the only instances of such observed multiple co-localizations, referred to as REGs, corresponded to treponemal 16S rRNA gene phlotypes.

All but 3 of the 27 16S rRNA gene phlotypes observed grouped closely ($98.2 \pm 0.85\%$ sequence identity across 917 aligned nucleotides) with phlotypes cloned previously from *Reticulitermes* termites, rather than those from any other genus. This provides further evidence for previous proposals of the “co-evolution” of the symbiotic microbes in the termite gut with their host proposed in Chapter 4 and elsewhere (19, 30, 67).

16S rRNA gene phlotypes were observed that fell outside the phylum Spirochaetes. These other phyla included Epsilon-Proteobacteria, Bacteroidetes, and Beta-Proteobacteria. It would not be unexpected to observe bacteria encoding hydrogenases within these phyla (50, 59), but we view these co-localizations with caution because no REGs were observed falling within any of these phyla in this study.

On account of the short length of the [FeFe] hydrogenase sequences cloned, corresponding to 70 unambiguously aligned amino acid residues, it is difficult to draw precise conclusions with regard to their phylogeny, though two important observations are noteworthy. First, clone Rth2-H10, which co-localized in dPCR with a treponemal

16S rRNA gene sequence, shared 100% sequence identity across 70 aligned amino acid residues with the Family 3 [FeFe] hydrogenase from *T. azotonutricium* ZAS-9. This provides support for the relevance of the gene of this strain, isolated from the gut of *Zootermopsis angusticolis* (26), to termite gut ecosystems and their hydrogen metabolism. Second, several of the sequences grouped with those cloned from *R. hesperus* collection ChiA2 in a previous study, see Chapter 3. This provides support broader relevance of the cloned hydrogenase sequences cloned in this study to *Reticulitermes* termite gut ecosystems.

Conclusions. We have used the technique of microfluidic digital PCR to demonstrate directly that treponemes may be an important, perhaps dominant, bacterial group encoding Family 3 [FeFe] hydrogenases in the gut of *Reticulitermes tibialis* collection JT2. This is in agreement with studies supporting an important role for termite gut treponemes in acetogenesis (48, 51) and the high fraction that they comprise of all gut eubacteria (8, 46). It is now of interest to identify what bacterial phyla other families of [FeFe] hydrogenases may associate with, or to employ microfluidic techniques to sequence termite gut microbe genomes to elucidate the precise physiological context of hydrogenase genes in the termite gut. Work should also be done to further understand the *in situ* expression of hydrogenases in the termite gut.

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. **Boyd, E. S., J. R. Spear, and J. W. Peters.** 2009. [FeFe] Hydrogenase Genetic Diversity Provides Insight into Molecular Adaptation in a Saline Microbial Mat Community. *Appl. Environ. Microbiol.* **75**:4620-4623.
3. **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.
4. **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, symbioses, ecology*. Kluwer Academic Publishers, Boston.
5. **Breznak, J. A.** 1982. Intestinal microbiota of termites and other xylophagous insects. *Annu. Rev. Microbiol.* **36**:323-343.
6. **Breznak, J. A.** 1975. Symbiotic Relationships Between Termites and Their Intestinal Microbiota. *Symp. Soc. Exp. Biol.* :559-580.
7. **Breznak, J. A., and J. M. Switzer.** 1986. Acetate Synthesis from H₂ plus CO₂ by Termite Gut Microbes. *Appl. Environ. Microbiol.* **52**:623-630.
8. **Brune, A.** 2006. Symbiotic Associations Between Termites and Prokaryotes. *Prokaryotes* **1**:439-474.

9. **Brune, A.** 1998. Termite guts: the world's smallest bioreactors. Trends in Biotechnol. **16**:16-21.
10. **Brune, A., and M. Friedrich.** 2000. Microecology of the termite gut: structure and function on a microscale. Curr. Opin. Microbiol. **3**:263-269.
11. **Cleveland, L. R.** 1923. Correlation Between the Food and Morphology of Termites and the Presence of Intestinal Protozoa. Am. J. Hyg. **3**:444-461.
12. **Cleveland, L. R.** 1923. Symbiosis between Termites and Their Intestinal Protozoa. Proc. Natl. Acad. Sci. U.S.A. **9**:424-428.
13. **DeSantis, T. Z., P. Hugenholtz, N. Larsen, M. Rojas, E. L. Brodie, K. Keller, T. Huber, D. Dalevi, P. Hu, and G. L. Andersen.** 2006. Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. Appl. Environ. Microbiol. **72**:5069-5072.
14. **Dumont, M. G., and J. C. Murrell.** 2005. Community-level analysis: key genes of aerobic methane oxidation. Methods. Enzymol. **397**:413-427.
15. **Ebert, A., and A. Brune.** 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.
16. **Fang, H. H. P., T. Zhang, and C. Li.** 2006. Characterization of Fe-hydrogenase genes diversity and hydrogen-production population in an acidophilic sludge. Journal of Bacteriology **126**:357-364.

17. **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.
18. **Hoehler, T. M., B. M. Bebout, and D. J. Des Marais.** 2001. The role of microbial mats in the production of reduced gases on the early Earth. *Nature* **412**:324-327.
19. **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.
20. **Hongoh, Y., M. Ohkuma, and T. Kudo.** 2003. Molecular analysis of bacterial microbiota in the gut of the termite *Reticulitermes speratus* (Isoptera; Rhinotermitidae). *FEMS Microbiol. Ecol.* **44**:231-242.
21. **Hungate, R. E.** 1939. Experiments on the Nutrition of Zootermopsis. III. The anaerobic carbohydrate dissimilation by the intestinal protozoa. *Ecology* **20**:230-245.
22. **Hungate, R. E.** 1943. Quantitative Analysis on the Cellulose Fermentation by Termite Protozoa. *Ann. Entomol. Soc. Am.* **36**:730-9.
23. **Kudo, T.** 2009. Termite-Microbe Symbiotic System and Its Efficient Degradation of Lignocellulose. *Biosci. Biotechnol. Biochem.* **73**:2561-2567.

24. **Leadbetter, J. R.** 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**:3620-31.
25. **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.
26. **Leadbetter, J. R., T. M. Schmidt, J. R. Graber, and J. A. Breznak.** 1999. Acetogenesis from H₂ Plus CO₂ by Spirochetes from Termite Guts. *Science* **283**:686-689.
27. **Leaphart, A. B., M. J. Friez, and C. R. Lovell.** 2003. Formyltetrahydrofolate synthetase sequences from salt marsh plant roots reveal a diversity of acetogenic bacteria and other bacterial functional groups. *Appl. Environ. Microbiol.* **69**:693-696.
28. **Leaphart, A. B., and C. R. Lovell.** 2001. Recovery and analysis of formyltetrahydrofolate synthetase gene sequences from natural populations of acetogenic bacteria. *Appl. Environ. Microbiol.* **67**:1392-1395.
29. **Lilburn, T. G., K. S. Kim, N. E. Ostrom, K. R. Byzek, J. R. Leadbetter, and J. A. Breznak.** 2001. Nitrogen Fixation by Symbiotic and Free-Living Spirochetes. *Science* **292**:2495-2498.
30. **Lilburn, T. G., T. M. Schmidt, and J. A. Breznak.** 1999. Phylogenetic diversity of termite gut spirochaetes. *Environ. Microbiol.* **1**:331-345.

31. **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.* **23**:515-538.
32. **Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Förster, I. Brettske, S. Gerber, A. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. König, T. Liss, R. Lüssmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N. Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode, and K. Schleifer.** 2004. ARB: a software environment for sequence data. *Nucleic Acids Res.* **32**:1363-1371.
33. **Markowitz, V. M., N. N. Ivanova, E. Szeto, K. Palaniappan, K. Chu, D. Dalevi, I.-M. A. Chen, Y. Grechkin, I. Dubchak, I. Anderson, A. Lykidis, K. Mavromatis, P. Hugenholtz, and N. C. Kyrpides.** 2008. IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Res.* **36**:D534-D538.
34. **McHardy, A. C., H. G. Martin, A. Tsirigos, P. Hugenholtz, and I. Rigoutsos.** 2007. Accurate phylogenetic classification of variable-length DNA fragments. *Nat. Meth.* **4**:63.
35. **Meyer, J.** 2007. [FeFe] hydrogenases and their evolution: a genomic perspective. *Cell. Mol. Life Sci.* **64**:1063-1084.
36. **Morgenstern, B.** 1999. DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. *Bioinformatics* **15**:211-218.

37. **Néron, B., H. Ménager, C. Maufrais, N. Joly, T. Pierre, and C. Letondal.** 2008. Presented at the Bio Open Source Conference, Toronto.
38. **Noda, S., M. Ohkuma, R. Usami, K. KHorikoshi, and T. Kudo.** 1999. Culture-Independent Characterization of a Gene Responsible for Nitrogen Fixation in the Symbiotic Microbial Community in the Gut of the Termite *Neotermes koshunensis*. *Appl. Environ. Microbiol.* **65**:4935-4942.
39. **Odelson, D. A., and J. A. Breznak.** 1985. Cellulase and Other Polymer-Hydrolyzing Activities of *Trichomitopsis termopsidis*, a Symbiotic Protozoan from Termites. *Appl. Environ. Microbiol.* **49**:622-626.
40. **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.
41. **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.
42. **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.
43. **Ohkuma, M., S. Noda, and T. Kudo.** 1999. Phylogenetic Diversity of Nitrogen Fixation Genes in the Symbiotic Microbial Community in the Gut of Diverse Termites. *Appl. Environ. Microbiol.* **65**:4926-4934.
44. **Ottesen, E. A.** 2009. The Biology and Community Structure of CO₂-Reducing Acetogens in the Termite Hindgut. California Institute of Technology, Pasadena.

45. **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.
46. **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**:347-352.
47. **Paster, B. J., F. E. Dewhirst, W. G. Weisburg, L. A. Tordoff, G. J. Fraser, R. B. Hespell, T. B. Stanton, L. Zablen, L. Mandelco, and C. R. Woese.** 1991. Phylogenetic Analysis of Spirochetes. *Journal of Bacteriology* **173**:6101-6109.
48. **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.
49. **Pester, M., and A. Brune.** 2007. Hydrogen is the central free intermediate during lignocellulose degradation by termite gut symbionts. *ISME J.* **1**:551-565.
50. **Robson, R.** 2001. Biodiversity of hydrogenases, p. 9-32. *In* R. Cammack, M. Frey, and R. Robson (ed.), *Hydrogen as a Fuel: Learning from Nature*. Taylor & Francis Inc., London.
51. **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.
52. **Schink, B., F. S. Lupton, and J. G. Zeikus.** 1983. Radioassay for Hydrogenase Activity in Viable Cells and Documentation of Aerobic Hydrogen-Consuming

- Bacteria Living in Extreme Environments. *Appl. Environ. Microbiol.* **45**:1491-1500.
53. **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.
54. **Schmidt, O., H. L. Drake, and M. A. Horn.** 2010. Hitherto unknown [Fe-Fe]-Hydrogenase Gene Diversity in Anaerobes and Anoxic Enrichments from a Moderately Acidic Fen. *Appl. Environ. Microbiol.* **76**:2027-2031.
55. **Scranton, M. I., P. C. Novelli, and P. A. Loud.** 1984. The distribution and cycling of hydrogen gas in the waters of two anoxic marine environments. *Limnol. Oceanogr.* **29**:993-1003.
56. **Smolenski, W. J., and J. A. Robinson.** 1988. In situ rumen hydrogen concentrations in steers fed eight times daily, measured using a mercury reduction detector. *FEMS Microbiol. Ecol.* **53**:95-100.
57. **Sugimoto, A., and N. Fujita.** 2006. Hydrogen concentrations and stable isotopic composition of methane in bubble gas observed in a natural wetland. *Biogeochemistry* **81**:33-44.
58. **Trager, W.** 1934. The Cultivation of a Cellulose-Digesting Flagellate, *Trichomonas termopsidis*, and of Certain Other Termite Protozoa. *Biol. Bull.* **66**:182-190.
59. **Vignais, P. M., and B. Billoud.** 2007. Occurrence, Classification, and Biological Function of Hydrogenases: An Overview. *Chem. Rev.* **107**:4206-4272.

60. **Vignais, P. M., B. Billoud, and J. Meyer.** 2001. Classification and phylogeny of hydrogenases. *FEMS Microbiol. Rev.* **25**:455-501.
61. **Warnecke, F., P. Luginbühl, N. Ivanova, M. Ghassemian, T. H. Richardson, J. T. Stege, M. Cayouette, A. C. McHardy, G. Djordjevic, N. Aboushadi, R. Sorek, S. G. Tringe, M. Podar, H. G. Martin, V. Kunin, D. Dalevi, J. Madejska, E. Kirton, D. Platt, E. Szeto, A. Salamov, K. Barry, N. Mikhailova, N. C. Kyrpides, E. G. Matson, E. A. Ottesen, X. Zhang, M. Hernández, C. Murillo, L. G. Acosta, I. Rigoutsos, G. Tamayo, B. D. Green, C. Chang, E. M. Rubin, E. J. Mathur, D. E. Robertson, P. Hugenholtz, and J. R. Leadbetter.** 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**:560-569.
62. **Wawer, C., M. S. M. Jetten, and G. Muyzer.** 1997. Genetic Diversity and Expression of the [NiFe] Hydrogenase Large-Subunit Gene of *Desulfovibrio* spp. in Environmental Samples. *Appl. Environ. Microbiol.* **63**:4360-4369.
63. **Xing, D., N. Ren, and B. E. Rittmann.** 2008. Genetic Diversity of Hydrogen-Producing Bacteria in an Acidophilic Ethanol-H₂-Coproducting System, Analyzed Using the [Fe]-Hydrogenase Gene. *Appl. Environ. Microbiol.* **74**:1232-1239.
64. **Yamada, A., T. Inoue, S. Noda, Y. Hongoh, and M. Ohkhuma.** 2007. Evolutionary trend of phylogenetic diversity of nitrogen fixation genes in the gut community of wood-feeding termites. *Mol. Ecol.* **16**:3768-3777.
65. **Yamin, M. A.** 1979. Cellulolytic activity of an axenically-cultivated termite flagellate, *Trichomitopsis termopsisidis*. *J. Gen. Microbiol.* **113**:417-20.

66. **Yamin, M. A.** 1981. Cellulose Metabolism by the Flagellate *Trichonympha* from a Termite Is Independent of Endosymbiotic Bacteria. *Science* **211**:58-59.
67. **Yang, H., D. Schmitt-Wagner, U. Stingl, and A. Brune.** 2005. Niche heterogeneity determines bacterial community structure in the termite gut (*Reticulitermes santonensis*). *Environ. Microbiol.* **7**:916-932.
68. **Zehr, J. P., B. D. Jenkins, S. M. Short, and G. F. Steward.** 2003. Nitrogenase gene diversity and microbial community structure: a cross-comparison. *Environ. Microbiol.* **5**.
69. **Zehr, J. P., and L. A. McReynolds.** 1989. Use of degenerate oligonucleotides for amplification of the *nifH* gene from the marine cyanobacterium *Trichodesmium thiebautii*. *Appl. Environ. Microbiol.* **55**:2522-2526.
70. **Zhang, Y., D. Zhang, W. Li, J. Chen, Y. Peng, and W. Cao.** 2003. A novel real-time quantitative PCR method using attached universal template probe. *Nuc. Acids Res.* **31**:e123.

Appendix

Table S1. Sequences cloned.

Figure S1. Alignment used in primer design.

Figure S2. Alignment used in probe design.

Table 5-S1. Sequences cloned and proposed *Reticulitermes* environmental genomovars (REG).

Clone Name ^a	RFLP Abundance ^b	Min. RFLP Seq. Id. (%) ^c	Gene	REG ^d	Min. REG Seq. Id. (%) ^e	Accession
RsH2-A1.2	1	97.1%	[FeFe] hydrogenase			local ARB database
RsH2-A1.3	1	97.1%	[FeFe] hydrogenase			local ARB database
RsH2-A1.6	2	97.1%	[FeFe] hydrogenase	REG 1.1h	98.5%	local ARB database
RsH2-A1.7	2	97.1%	[FeFe] hydrogenase			local ARB database
RsH2-A11.1	5	98.5%	[FeFe] hydrogenase	REG 2.1h	98.5%	local ARB database
RsH2-A11.10	1	98.5%	[FeFe] hydrogenase			local ARB database
RsH2-A11.2	1	98.5%	[FeFe] hydrogenase			local ARB database
RsH2-A11.3	1	98.5%	[FeFe] hydrogenase			local ARB database
RsH2-A12.3	9	100.0%	[FeFe] hydrogenase			local ARB database
RsH2-A12.5	1	100.0%	[FeFe] hydrogenase			local ARB database
RsH2-A8.2	2	100.0%	[FeFe] hydrogenase			local ARB database
RsH2-A8.4	4	100.0%	[FeFe] hydrogenase	REG 2.2h	98.5%	local ARB database
RsH2-A8.9	1	100.0%	[FeFe] hydrogenase			local ARB database
RsH2-B2.1	8	98.5%	[FeFe] hydrogenase			local ARB database
RsH2-B2.5	1	98.5%	[FeFe] hydrogenase			local ARB database
RsH2-B5.1	9	-	[FeFe] hydrogenase			local ARB database
RsH2-B6.2	9	100.0%	[FeFe] hydrogenase			local ARB database
RsH2-B6.4	1	100.0%	[FeFe] hydrogenase			local ARB database
RsH2-B9.1	9	-	[FeFe] hydrogenase			local ARB database
RsH2-C7.10	10	-	[FeFe] hydrogenase			local ARB database
RsH2-D11.6	10	-	[FeFe] hydrogenase	REG 4.1h	98.5%	local ARB database
RsH2-D9.1	1	97.1%	[FeFe] hydrogenase			local ARB database
RsH2-D9.2	2	97.1%	[FeFe] hydrogenase			local ARB database
RsH2-D9.4	7	97.1%	[FeFe] hydrogenase			local ARB database
RsH2-E12.1	10	-	[FeFe] hydrogenase	REG 1.2h	98.5%	local ARB database
RsH2-E2.5	5	100.0%	[FeFe] hydrogenase			local ARB database
RsH2-E2.7	1	100.0%	[FeFe] hydrogenase			local ARB database
RsH2-E5.2	10	-	[FeFe] hydrogenase			local ARB database
RsH2-E9.1	10	-	[FeFe] hydrogenase			local ARB database
RsH2-F2.1	10	-	[FeFe] hydrogenase			local ARB database
RsH2-F4.1	10	-	[FeFe] hydrogenase			local ARB database
RsH2-G10.1	6	98.5%	[FeFe] hydrogenase			local ARB database
RsH2-G10.5	1	98.5%	[FeFe] hydrogenase			local ARB database
RsH2-G10.7	1	98.5%	[FeFe] hydrogenase			local ARB database
RsH2-G10.9	1	98.5%	[FeFe] hydrogenase			local ARB database
RsH2-G6.3	10	-	[FeFe] hydrogenase	REG 4.2h	98.5%	local ARB database
RsH2-G7.2	3	98.3%	[FeFe] hydrogenase			local ARB database
RsH2-G7.3	1	98.3%	[FeFe] hydrogenase			local ARB database
RsH2-H10.1	3	100.0%	[FeFe] hydrogenase			local ARB database
RsH2-H10.3	1	100.0%	[FeFe] hydrogenase			local ARB database
RsH2-H10.6	3	100.0%	[FeFe] hydrogenase			local ARB database
RsH2-H2.1	9	98.5%	[FeFe] hydrogenase			local ARB database
RsH2-H2.6	1	98.5%	[FeFe] hydrogenase			local ARB database
RsH2-H6.1	10	-	[FeFe] hydrogenase	REG 3.1h	98.5%	local ARB database
RsH2-H7.1	9	-	[FeFe] hydrogenase	REG 3.2h	98.5%	local ARB database
RsH2-H8.1	10	-	[FeFe] hydrogenase			local ARB database
RsH2-G12.1	10	-	[FeFe] hydrogenase			local ARB database
RsH2-H11.1	10	-	[FeFe] hydrogenase			local ARB database
RsR-A1.1	10	-	16S	REG 1.1r	99.0%	local ARB database

Continuing Table 5-S1.

Clone Name ^a	RFLP Abundance ^b	Min. RFLP Seq. Id. (%) ^c	Gene	REG ^d	Min. REG Seq. Id. (%) ^e	Accession
RsR-A11.2	5	97.8%	16S	REG 2.1r	98.5%	local ARB database
RsR-A11.4	4	97.8%	16S			local ARB database
RsR-A11.1	1	97.8%	16S			local ARB database
RsR-A12.1	8	97.8%	16S			local ARB database
RsR-A12.2	1	97.8%	16S			local ARB database
RsR-A12.3	1	97.8%	16S			local ARB database
RsR-A8.1	5	98.0%	16S	REG 2.2r	98.5%	local ARB database
RsR-A8.3	2	98.0%	16S			local ARB database
RsR-A8.8	1	98.0%	16S			local ARB database
RsR-B2.1	5	-	16S			local ARB database
RsR-B5.1	7	98.3%	16S			local ARB database
RsR-B5.9	1	98.3%	16S			local ARB database
RsR-B5.5	1	98.3%	16S			local ARB database
RsR-B5.8	1	98.3%	16S			local ARB database
RsR-B6.2	8	98.8%	16S			local ARB database
RsR-B6.9	1	98.8%	16S			local ARB database
RsR-B9.1	6	99.1%	16S			local ARB database
RsR-B9.5	3	99.1%	16S			local ARB database
RsR-B9.10	1	99.1%	16S			local ARB database
RsR-C7.4	10	-	16S			local ARB database
RsR-D11.1	6	98.3%	16S	REG 4.1r	98.5%	local ARB database
RsR-D11.3	1	98.3%	16S			local ARB database
RsR-D9.1	9	99.1%	16S			local ARB database
RsR-D9.2	1	99.1%	16S			local ARB database
RsR-E12.1	9	99.3%	16S	REG 1.2r	99.0%	local ARB database
RsR-E12.4	1	99.3%	16S			local ARB database
RsR-E2.2	9	-	16S			local ARB database
RsR-E5.1	9	-	16S			local ARB database
RsR-E9.1	8	98.8%	16S			local ARB database
RsR-E9.2	1	98.8%	16S			local ARB database
RsR-E9.10	1	98.8%	16S			local ARB database
RsR-F2.3	7	98.3%	16S			local ARB database
RsR-F2.2	1	98.3%	16S			local ARB database
RsR-F2.6	1	98.3%	16S			local ARB database
RsR-F4.1	6	98.7%	16S			local ARB database
RsR-F4.5	1	98.7%	16S			local ARB database
RsR-F4.8	1	98.7%	16S			local ARB database
RsR-G10.1	5	99.4%	16S			local ARB database
RsR-G10.2	1	99.4%	16S			local ARB database
RsR-G10.7	1	99.4%	16S			local ARB database
RsR-G10.8	1	99.4%	16S			local ARB database
RsR-G12.1	6	99.2%	16S			local ARB database
RsR-G12.9	1	99.2%	16S			local ARB database
RsR-G12.10	1	99.2%	16S			local ARB database
RsR-G12.6	1	99.2%	16S			local ARB database
RsR-G6.1	9	99.3%	16S	REG 4.2r	98.5%	local ARB database
RsR-G6.5	1	99.3%	16S			local ARB database
RsR-G7.1	8	98.8%	16S			local ARB database
RsR-G7.3	1	98.8%	16S			local ARB database

Continuing Table 5-S1.

Clone Name ^a	RFLP Abundance ^b	Min. RFLP Seq. Id. (%) ^c	Gene	REG ^d	Min. REG Seq. Id. (%) ^e	Accession
RsR-H10.2	8	98.8%	16S			local ARB database
RsR-H10.8	1	98.8%	16S			local ARB database
RsR-H11.1	5	98.7%	16S			local ARB database
RsR-H11.2	1	98.7%	16S			local ARB database
RsR-H11.5	1	98.7%	16S			local ARB database
RsR-H11.7	1	98.7%	16S			local ARB database
RsR-H11.9	1	98.7%	16S			local ARB database
RsR-H11.10	1	98.7%	16S			local ARB database
RsR-H2.1	9	-	16S			local ARB database
RsR-H6.1	8	99.0%	16S	REG 3.1r	98.9%	local ARB database
RsR-H6.5	1	99.0%	16S			local ARB database
RsR-H7.1	9	-	16S	REG 3.2r	98.9%	local ARB database
RsR-H8.1	10	-	16S			local ARB database

^aIn boldface are sequences selected as representative of each set of sequences.

^bNumber of RFLP patterns corresponding to each sequence. For each set of clones, this is out of a total of 10 clones selected at random from a clone library prepared from a digital PCR well. Some sequences have less than 10 total RFLP patterns reported because several of the clones were false positives.

^cThe minimum percent sequence (peptide for hydrogenases and nucleotide for 16S rRNA genes) identity shared among each clone set.

^d*Reticulitermes* environmental genomovar names.

^eThe minimum percent sequence identity shared by *all* hydrogenase gene (peptide sequence) or 16S gene (nucleotide sequence) clones corresponding to a particular REG.

Figure 5-S1.



Figure 5-S1. Alignment used to design degenerate primers for use in digital PCR.

The highlighted highly conserved regions were used in the design of degenerate primers. The following host designators were used: CA = *Cryptocercus punctulatus* adult, CN = *Cryptocercus punctulatus* nymph, N = *Nasutitermes* sp. Cost003, R = *Reticulitermes Hesperus* collection ChiA2, Z = *Zootermopsis nevadensis* collection ChiA1. The numbers-letter names following each host designator correspond to sample names taken from Chapters 3 and 4. ^aFamily 3 [FeFe] hydrogenases identified in the genomes of *Treponema primitia* ZAS-2 (HndA1), and *Treponema azotonutricium* ZAS-9 (HndA), ^bIMG Gene Object Identifier.

Figure 5-S2.

```

R_A1  GGCATCATGG AAGCGGCGGT
R_A2  GGCATCATGG AAGCGGCGGT
R_A4  GGCATCATGG AAGCTGCGGT
R_A6  GGCATTATGG AAGCGGCGGT
R_A8  GGCATTATGG AAGCGGCGGT
R_A9  GGCATCATGG AAGCGGCGGT
R_B1  GGCATCATGG AAGCAGCGGT
R_B10 GGCATCATGG AAGCAGCGGT
R_B11 GGCATCATGG AAGCGGCGGT
R_C2  GGCATTATGG AAGCGGCGGT
R_C7  GGCATTATGG AAGCGGCGGT
R_C9  GGTATTATGG AAGCGGCGGT
R_D1  GGCATCATGG AAGCTGCGGT
R_D11 GGCATCATGG AAGCGGCGGT
R_D4  GGCATTATGG AAGCGGCGGT
R_D5  GGCATCATGG AAGCGGCGGT
R_D6  GGCATCATGG AAGCTGCGGT
R_E1  GGCATTATGG AAGCGGCGGT
R_E3  GGCATCATGG AAGCGGCGGT
R_E5  GGCATCATGG AAGCGGCGGT
R_F11 GGCATCATGG AAGCGGCGGT
R_F12 GGCATTATGG AAGCGGCGGT
R_F2  GGCATCATGG AAGCGGCGGT
R_F5  GGCATCATGG AAGCGGCGGT
R_G2  GGCATCATGG AAGCTGCGGT
R_G3  GGCATCATGG AAGCTGCGGT
R_G4  GGCATCATGG AAGCGGCGGT
R_H1  GGCATTATGG AAGCGGCGGT
R_H12 GGTATCATGG AAGCGGCGGT
R_H2  GGCATCATGG AAGCTGCGGT
R_H4  GGCATCATGG AAGCTGCGGT

```

Figure 5-S2. Alignment used in the design of locked nucleic acid (LNA) probes for use in digital PCR. See Figure 5-S1 for a description of sequence names used in the alignment. The highlighted, highly conserved region was used in the design of LNA probes.