Chapter 2

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

Abstract

H_2 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 CO_2-reductive acetogenesis in *T. primitia*. The many and diverse [FeFe] hydrogenase-like proteins encoded by termite gut treponemes accentuates the importance of H_2 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 H_2 evolution in biochemical analyses (31). Before H_2 escapes the system, most of this gas is consumed by CO_2-reducing homoacetogenic bacteria (35, 50), and to a lesser extent, methanogenic archaea (6, 33). The flux and standing concentrations of gut H_2 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 m^3 H_2 per m^3 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 spriochetes 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 (Genebank 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, H_2-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 H_2 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), *Pseudotrichonympha 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 *Spironucleus barkhanus* (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 H_2 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]
T. primitia ZAS-1	*HydA1*	1128[a]	FDH-Linked	Enzyme	Cytoplasmic or Periplasmic	CCP	PCxxKxxE	MxCxxGCxxG
T. primitia ZAS-1	*HydA3*	2678[a]	6	Enzyme	Cytoplasmic or Periplasmic	CCP	PCxxKxxE	MxCxxGCxxG
T. primitia ZAS-1	*HndA1*	2692[a]	7	Enzyme	Cytoplasmic or Periplasmic	CCP	PCxxKxxE	MxCxxGCxxG
T. primitia ZAS-1	*HydA2*	1488[a]	5	Sensor	Cytoplasmic Membrane	**P**CP	PCxxKxxE	**L**xCxxGCxxG
T. primitia ZAS-2	*HydA1*	TREPR_3094	FDH-Linked	Enzyme	Cytoplasmic or Periplasmic	CCP	PCxxKxxE	MxCxxGCxxG
T. primitia ZAS-2	*HydA5*	TREPR_0390	6	Enzyme	Cytoplasmic or Periplasmic	CCP	PCxxKxxE	MxCxxGCxxG
T. primitia ZAS-2	*HndA1*	TREPR_3594	3	Enzyme	Cytoplasmic or Periplasmic	CCP	PCxxKxxE	MxCxxGCxxG
T. primitia ZAS-2	*HndA2*	TREPR_3288	7	Enzyme	Cytoplasmic or Periplasmic	CCP	PCxxKxxE	MxCxxGCxxG
T. primitia ZAS-2	*HndA3*	TREPR_3300	7	Enzyme	Cytoplasmic or Periplasmic	CCP	PCxxKxxE	MxCxxGCxxG
T. primitia ZAS-2	*HydA2*	TREPR_3122	4	Sensor	Cytoplasmic or Periplasmic	[**AV**]CP	[**PS**]CxxKxxE	**L**xCxxGCxxG
T. primitia ZAS-2	*HydA3*	TREPR_3283	4	Sensor	Cytoplasmic or Periplasmic	[**AV**]CP	[**PS**]CxxKxxE	**L**xCxxGCxxG
T. primitia ZAS-2	*HydA4*	TREPR_1589	5	Sensor	Cytoplasmic Membrane	**P**CP	PCxxKxxE	**L**xCxxGCxxG
T. azotonutricium ZAS-9	*HydA3*	TREAZ_2002	6	Enzyme	Cytoplasmic or Periplasmic	CCP	PCxxKxxE	MxCxxGCxxG
T. azotonutricium ZAS-9	*HndA*	TREAZ_3250	3	Enzyme	Cytoplasmic or Periplasmic	CCP	PCxxKxxE	MxCxxGCxxG
T. azotonutricium ZAS-9	*HydA1*	TREAZ_2458	5	Sensor	Cytoplasmic Membrane	**P**CP	PCxxKxxE	**L**xCxxGCxxG
T. azotonutricium ZAS-9	*HydA2*	TREAZ_2238	10	Sensor	Cytoplasmic Membrane	**P**CP	PCxxKxxE	**L**xCxxGCxxG
T. azotonutricium ZAS-9	*HydA4*	TREAZ_2003	8	Sensor	Cytoplasmic	CC**H**	PCxxKxxE	**S**xCxx**S**CxxG

[a]Gene numbers for T. primitia ZAS-1 are arbitrarily assigned based upon feature identifiers assigned by RAST (2).

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

[c]Conserved sequence signatures observed in the H domain of all known [FeFe] hydrogeneases (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.

	T. primitia ZAS-1	*T. primitia* ZAS-2	*T. azotonutricium* ZAS-9	*T. denticola*	*B. hyodysenteriae*	*M. thermoacetica*	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

[a]Families have been defined by Warnecke *et al*. (81).

[b]The family of FDH-Linked hydrogenases has been defined in this chapter.

[c]Families 3 and 7 share similar domain architectures.

[d]Families 4 and 8 share similar domain architectures.

[e]Families 5 and 10 share similar domain architectures, as defined by Warnecke *et al*. (81).

[f]Proteins for which a family designation could not be unambiguously defined or for which a [FeFe] hydrogenase familiy designation would not be relavant.

[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.

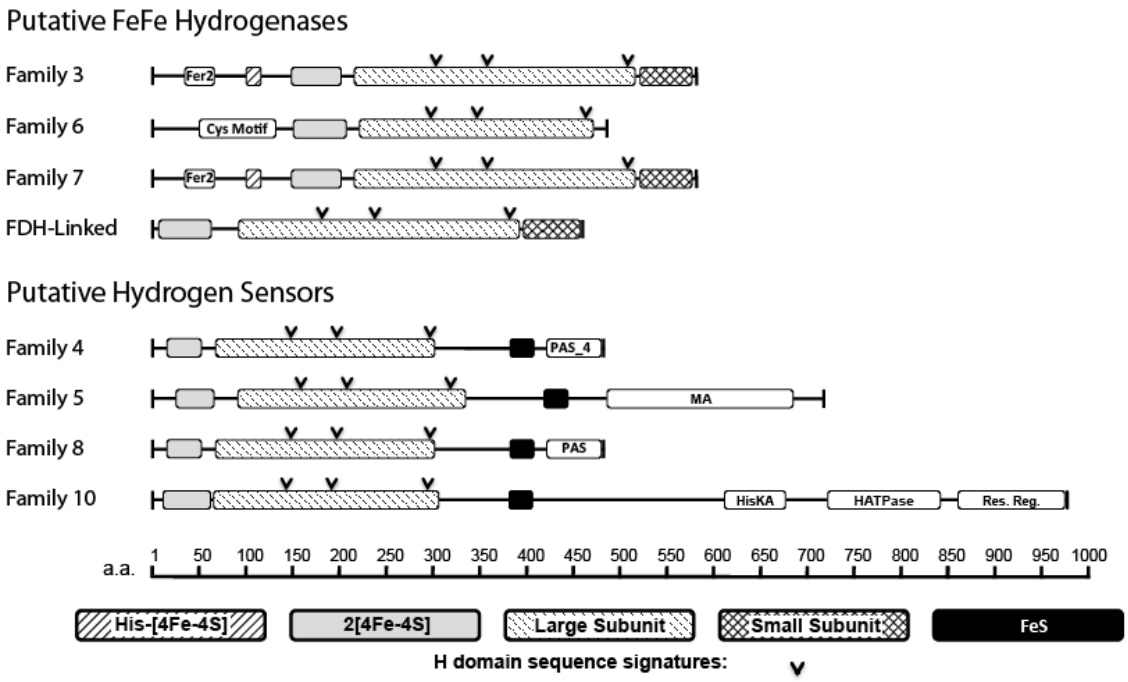


Figure 2-1. Domain architectures representative of each family of H domain proteins observed in *Treponema primitia* strains ZAS-1 and ZAS-2 and *T. azotonutricium* ZAS-9. Domains were identified using the Pfam server (4). Following initial detection, iron sulfur cluster domain regions were annotated 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 H_2-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 H_2 sensor proteins. All putative H_2 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 H_2 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.

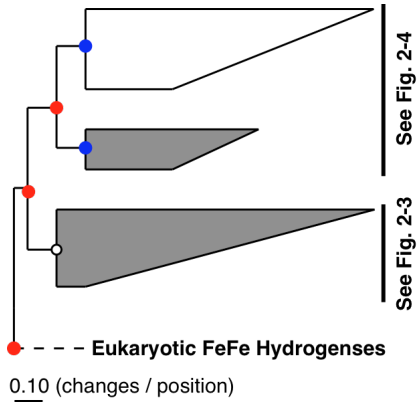


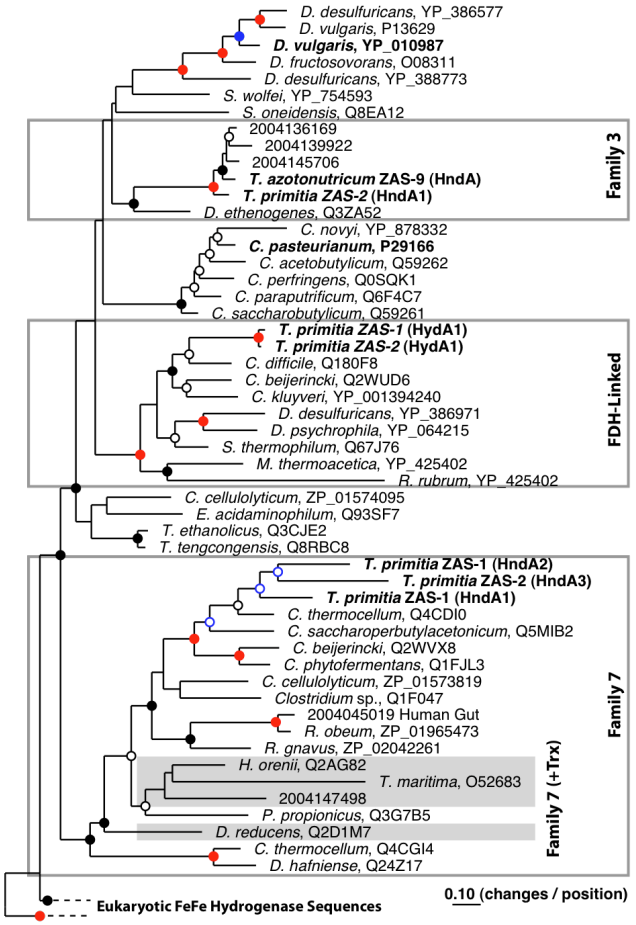
Figure 2-2. Phylogeny of H domain peptide sequences from putative hydrogen sensor and [FeFe] hydrogenase proteins. The tree was calculated using a maximum liklihood (Phylip ProML, 100 bootstraps) method with 150 unambiguously aligned amino acids. The clade containing hydrogen sensors is shown in white 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 H_2 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------ ---------- ---------- ----PVGAIY
2004139922                      SAHC------ ---------- ---------- ----PVGAIY
2004145706                      AAHC------ ---------- ---------- ----PVGAIY
T. azotonutricium ZAS-9 (HndA)  AAHC------ ---------- ---------- ----PVGAIY
T. primitia ZAS-2 (HndA1)       AAHC------ ---------- ---------- ----PVGAIY
D. ethenogenes, Q3ZA52          ILVC------ ---------- ---------- ----PVGAIK
T. primitia ZAS-1 (HydA1)       VQKCKSYVSL IDHGPEMYNK KREERMLPET VR-EPLFAAH
T. primitia ZAS-2 (HydA1)       VQKCKSYVSL IDHGPEMYNK KREERMLPET VR-EPLFAAH
C. difficile, Q180F8            VQTCKSYASV IDEGFEFLQE KKQEREIPES IN-EPIFAAY
C. beijerincki, Q2WUD6          IQVCNSYGFE NRENSHLIEE KRRDRGVLES VK-EPVFAAF
C. kluyveri, YP_001394240       IQICKGYYSI YDDVATPVSK KLFDRGLLDN VD-EPLFAAY
D. desulfuricans, YP_386971     VQQCSAFASF YEQHPACIAE KKRERGLFVS EA-APLFAAW
D. psychrophila, YP_064215      VQICSGYDSD LMSYATGRGK RLQNRGMLAT VV-EPLFAAH
S. thermophilum, Q67J76         VQVCSSYGSI WDDGLTPREQ KLAERGLLPS VK-EPLFAAW
M. thermoacetica, YP_425402     VQICSAYASP YTTSPETMAA KNRERRLLPA aapEPLFAAY
R. rubrum, YP_425402            VATCAAFDSI FDAFPTPRPV RLKRRGLPGS 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.

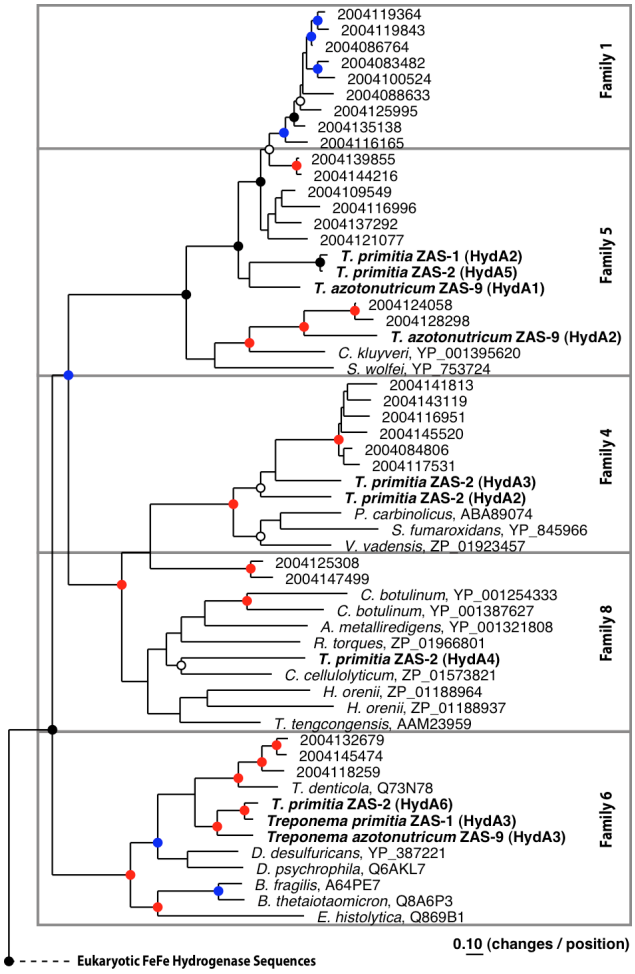


Figure 2-4. Phylogeny of H domain peptide sequences from [FeFe] hydrogenase Family 6 and putative H_2 sensor proteins proteins. Refer to Figure 2-2. The tree was calculated using a maximum likelihood (Phylip ProML) method with 238 unambiguously aligned amino acids. Open circles mark groupings also supported 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 azotonutricium* 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.

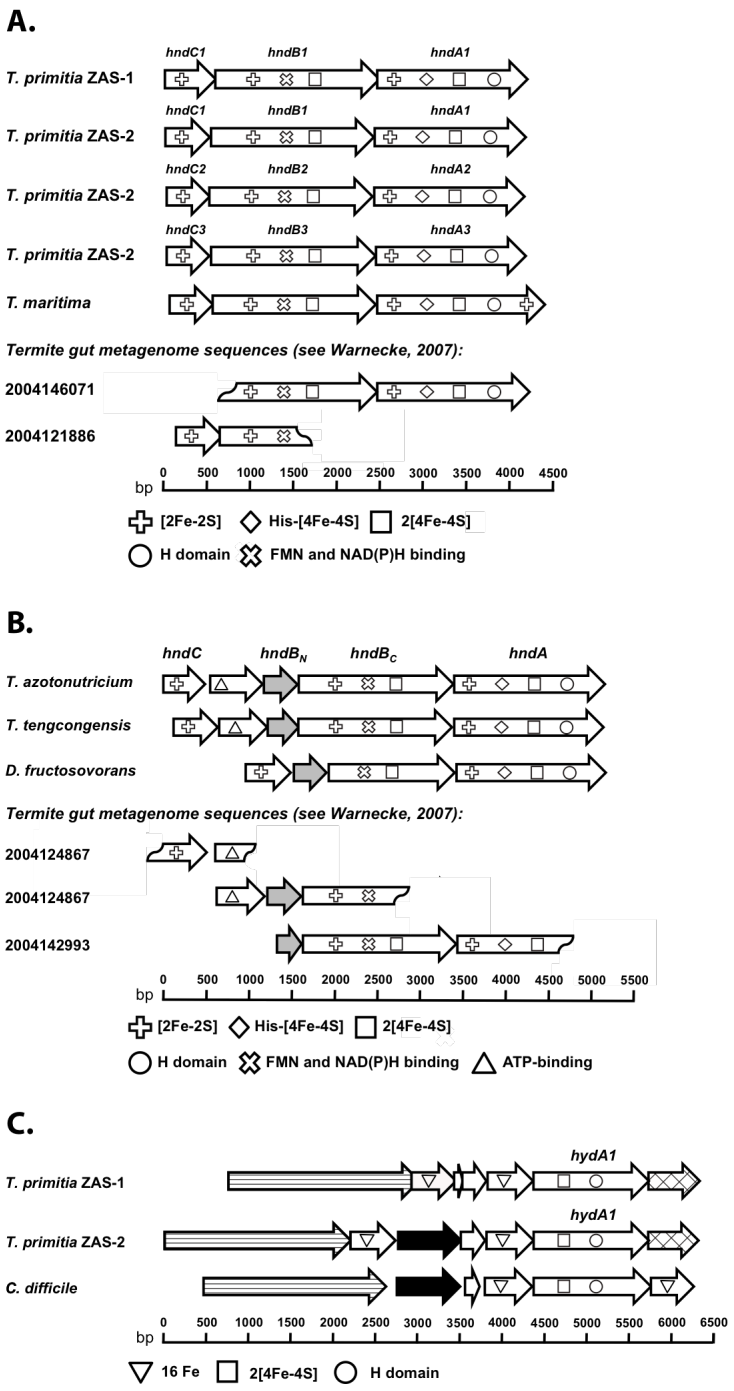

A.
hndC1
hndB1
hndA1
T. primitia ZAS-1
hndC1
hndB1
hndA1
T. primitia ZAS-2
hndC2
hndB2
hndA2
T. primitia ZAS-2
hndC3
hndB3
hndA3
T. primitia ZAS-2
T. maritima
Termite gut metagenome sequences (see Warnecke, 2007):
2004146071
2004121886
0 500 1000 1500 2000 2500 3000 3500 4000 4500
bp
[2Fe-2S] His-[4Fe-4S] 2[4Fe-4S]
H domain FMN and NAD(P)H binding
B.
hndC
hndB_N
hndB_C
hndA
T. azotonutricium
T. tengcongensis
D. fructosovorans
Termite gut metagenome sequences (see Warnecke, 2007):
2004124867
2004124867
2004142993
0 500 1000 1500 2000 2500 3000 3500 4000 4500 5000 5500
bp
[2Fe-2S] His-[4Fe-4S] 2[4Fe-4S]
H domain FMN and NAD(P)H binding ATP-binding
C.
hydA1
T. primitia ZAS-1
hydA1
T. primitia ZAS-2
C. difficile
0 500 1000 1500 2000 2500 3000 3500 4000 4500 5000 5500 6000 6500
bp
16 Fe 2[4Fe-4S] H domain

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 H_2-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 H_2 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 H_2 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 H_2-sensors. That each strain encoded examples of both types suggests that having them in concert may be relevant to H_2 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 H_2-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). H_2 is known to inhibit its growth, although it is consumed in a non-energy coupled detoxification reaction, also observed in *Pyrococcus furiosus*, to produce H_2S in the presence of S^0 (20, 28). Thus, it is unclear whether it's 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 oxidorecuctase (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 H_2 (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 H_2 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 $H_2 + CO_2$ 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 it's 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 hydrogeanses

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 H_2 sensor proteins in gene regulation and cell behavior.

The H-domains of the putative H_2 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 decribed and biochemically characterized NiFe hydrogenases-like H_2 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 H_2 (39). RcsC is known to be a cytoplasmic membrane protein; here PSORTb unambiguously predicts that the Family 10 H_2 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 H_2 gradients (80, 81), although to our knowledge H_2-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. H_2 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 H_2 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 H_2 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 H_2 metabolism in high flux, H_2-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 archaebacteria 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. Andreesen.** 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. Françoijs, 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 hypothermophylic 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. Occurence, 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 *Carboxydothermus 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.