

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

Abstract

The bacterial Wood-Ljungdahl pathway for CO₂-reductive acetogenesis is important for the nutritional mutualism occurring between wood-feeding insects and their hindgut microbiota. A key step in this pathway is the reduction of CO₂ to formate, catalyzed by the enzyme formate dehydrogenase (FDH). Putative selenocysteine- (Sec) and cysteine- (Cys) containing paralogs of hydrogenase-linked FDH (FDH_H) have been identified in the termite gut acetogenic spirochete, *Treponema primitia*, but knowledge of their relevance in the termite gut environment remains limited. In this study, we designed degenerate PCR primers for FDH_H genes (*fdhF*) and assessed *fdhF* diversity in insect gut bacterial isolates and the gut microbial communities of termites and roaches. The insects examined herein represent the wood-feeding termite families *Termopsidae*, *Kalotermitidae*, and *Rhinotermitidae* (phylogenetically “lower” termite taxa), the wood-feeding roach family *Cryptocercidae* (the sister taxon to termites), and the omnivorous roach family *Blattidae*. Sec and Cys FDH_H variants were identified in every wood-feeding insect but not the omnivorous roach. Of 68 novel phlotypes obtained from inventories, 66 affiliated phylogenetically with enzymes from *T. primitia*. These formed two sub-clades (37 and 29 phlotypes) almost completely comprised of Sec-containing and Cys-containing enzymes, respectively. A gut cDNA

inventory showed transcription of both variants in the termite *Zootermopsis nevadensis* (family *Termopsidae*). The results suggest FDH_H enzymes are important for the CO_2 -reductive metabolism of uncultured acetogenic treponemes and imply that the trace element selenium has shaped the gene content of gut microbial communities in wood-feeding insects.

Introduction

Xylophagy, the ability to feed exclusively on lignocellulose, in termites and wood-feeding roaches results from obligate nutritional mutualisms with their hindgut microbes (Breznak, 1982; Breznak and Brune, 1994; Inoue *et al.*, 2000). Studies on phylogenetically lower wood-feeding termites and wood-feeding roaches in the genus *Cryptocercus* have pointed to commonalities in their gut microbiota, in particular the presence of unique cellulolytic protozoa, as the basis for similarities in nutritional physiology (Honigberg, 1970; Inoue *et al.*, 2000). The process of lignocellulose degradation, elucidated in lower termites, is stepwise (Odelson and Breznak, 1983; Breznak and Switzer, 1986; Pester and Brune, 2007), and is applicable to *Cryptocercus* based on radiotracer measurements of hindgut carbon flow (Breznak and Switzer, 1986). Lignocellulose-derived polysaccharides are first hydrolyzed to glycosyl units and then fermented into acetate, H₂, and CO₂ by cellulolytic protozoa. The H₂ generated by this activity can approach saturation levels (Ebert and Brune, 1997; Pester and Brune, 2007). The final step in hindgut fermentation is CO₂-reductive acetogenesis, which outcompetes methanogenesis as the H₂ sink in the guts of these wood-feeding insects (Odelson and Breznak, 1983; Breznak and Switzer, 1986; Brauman *et al.*, 1992; Pester and Brune, 2007).

Acetogenesis in wood-feeding termites is mediated by anaerobic bacteria (Odelson and Breznak, 1983; Breznak and Switzer, 1986) and is estimated to contribute up to 1/3 of the acetate used by the insect host as its carbon and energy source (Odelson and Breznak, 1983). H₂ flux measurements in termite guts have shown that acetogens catalyze rapid and efficient turnover of H₂ in a system that is essentially optimized for acetate production (Pester and

Brune, 2007). The pathway by which these bacteria metabolize H₂ and CO₂ to acetate is the Wood-Ljungdahl pathway for CO₂ reductive acetogenesis (Ljungdahl, 1986). The enzymology underlying the four reductions culminating in the reductive fixation of CO₂ to acetate has been largely elucidated in the model actogen, *Moorella thermoacetica*, a member of the phylum *Firmicutes*, which comprises the majority of acetogens (Wood and Ljungdahl, 1991; Drake, 1994; Drake *et al.*, 2006; Drake *et al.*, 2007). The enzyme formate dehydrogenase (FDH) is one of two enzymes in the pathway critical for both H₂ turnover and autotrophic carbon fixation in acetogens (Drake, 1994; Drake *et al.*, 2002; Vorholt and Thauer, 2002). FDH catalyzes the reduction of CO₂ to formate with H₂ (or its equivalent) in the first step within the methyl branch of the Wood-Ljungdahl pathway (Drake *et al.*, 2006).

Culture and gene-inventory studies of termite gut acetogens indicate spirochetes, rather than firmicutes, are the predominant acetogenic bacteria in wood-feeding termite guts (Leadbetter *et al.*, 1999; Graber and Breznak, 2004; Graber *et al.*, 2004; Pester and Brune, 2006; Warnecke *et al.*, 2007). Despite the importance of acetogenic spirochetes in termite nutrition, only two isolates have ever been obtained; they remain, to this day, the sole examples of chemolithoautotrophy in the phylum *Spirochaetes* (Leadbetter *et al.*, 1999). The H₂-utilizing acetogenic spirochete *Treponema primitia* str. ZAS-2, isolated from the hindgut of the lower wood-feeding termite *Zootermopsis angusticollis*, is one these isolates.

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

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

Phylogenetic analysis of the FDH genes in *T. primitia* indicated they are *fdhF* paralogs that encode hydrogenase-linked FDH enzymes (FDH_H), similar to those used for formate oxidation during sugar fermentation in *Gammaproteobacteria* such as *Escherichia coli* (Pecher *et al.*, 1985; Zinoni *et al.*, 1986). The result was noteworthy as *T. primitia* FDH_H enzymes were expected to group with the well-characterized FDH for acetogenesis, an NADPH-linked tungsten containing selenoprotein from the classic acetogen, *Moorella thermoacetica* (Thauer, 1972; Yamamoto *et al.*, 1983; Pierce *et al.*, 2008). This led to the suggestion that the use of hydrogenase-linked FDH enzymes to directly access H₂ for CO₂-reductive metabolism may be an adaptation of *T. primitia* to life its H₂-rich gut environment.

Study of FDH_H in *T. primitia* (Matson *et al.*, 2010) imply that it has both evolved mechanisms to deal with changing selenium availability and adapted its metabolism to take advantage of high H₂ levels in the gut of its host termite, *Z. angusticollis*. Yet the extent to which *T. primitia* reflects general characteristics of the gut microbial community in *Zootermopsis* and other wood-feeding insects remains unknown. Here, using novel degenerate *fdhF* primers, we investigated FDH_H diversity in insect-gut isolates, the gut microbial communities of three wood-feeding lower termite species (*Zootermopsis nevadensis*, *Reticulitermes hesperus*, *Incisitermes minor*), wood-feeding roaches (*C. punctulatus*), and an omnivorous roach (*Periplaneta americana*). Together these insect species represent 4 of 5 wood-feeding basal phylogenetic taxa within the insect superorder *Dictyoptera*, comprised of termites, roaches, and mantids (Grimaldi and Engel, 2005). Insights into Sec/Cys FDH variant evolution in gut communities are highlighted and the likely importance of FDH_H enzymes to CO₂-reductive acetogenesis is discussed.

Materials and Methods

Microbial strains

Microbial isolates *Treponema primitia* str. ZAS-1 (DSM 12426), *Treponema primitia* str. ZAS-2 (DSM 12427), and *Acetonema longum* str. APO-1 (DSM 6540) were grown in anaerobic YACo medium under a headspace of 80% H₂ + 20% CO₂ as described previously (Kane and Breznak, 1991; Leadbetter *et al.*, 1999; Lilburn *et al.*, 2001). *Treponema azotonutricium* str. ZAS-9 (DSM 13862) was cultivated in a similar medium (Leadbetter *et al.*, 1999; Lilburn *et al.*, 2001). *Buttiauxiella* sp. str. SN-1 was isolated from the homogenized gut of a common garden snail (*Helix aspersa*), collected at the California

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

Insect collection

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

Hindgut nucleic acid extraction

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

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

***fdhF* primer design**

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

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

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

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

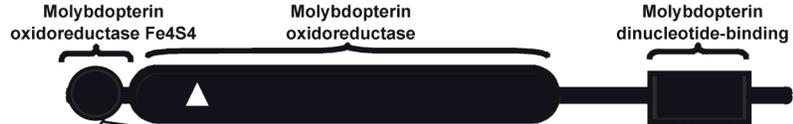
¹Forward primers indicated by monikers, ‘F1’, ‘F2’, or ‘F3’ in primer name; reverse primers by ‘R1’, ‘R2’ in primer name.

²Canonical nucleotides symbols are used. I, inosine.

Figure 2.1. Degenerate *fdhF* PCR primer design.

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

A



B

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

***fdhF* amplification and cloning**

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

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

COII amplification

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

RFLP analysis, sequencing, and diversity assessment

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

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

Phylogenetic analysis

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

RNA secondary structure prediction

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

Accession numbers

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

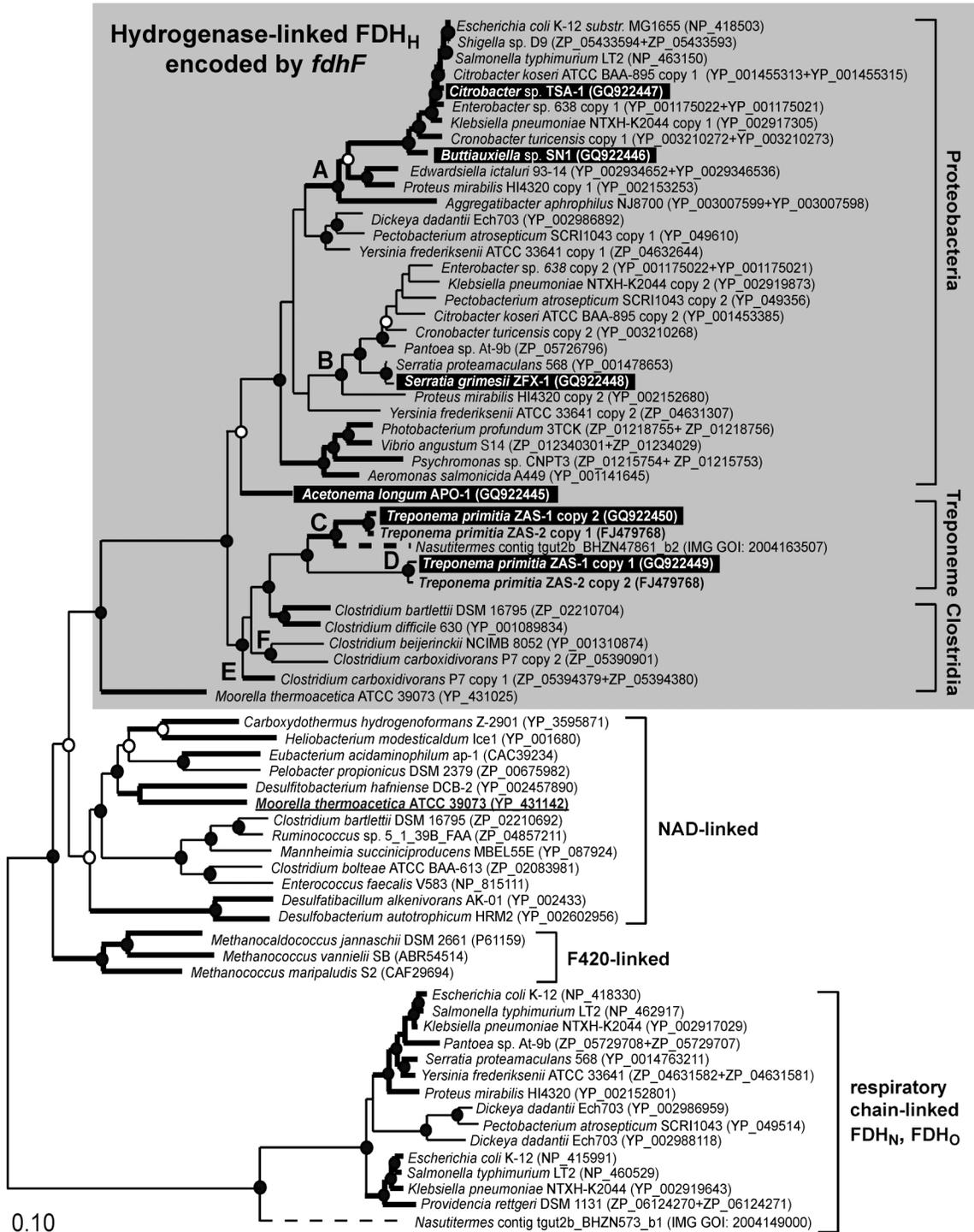
Results

***fdhF* primers amplify phylogenetically diverse FDH_H genes from pure cultures**

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

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

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



Convergent evolution of *fdhF*_{Sec} and *fdhF*_{Cys} in *Proteobacteria*, *Treponemes*, and *Firmicutes*

The recovery of genes for both Sec and Cys variants of FDH_H from the acetogenic spirochete isolate, *T. primitia* str. ZAS-1, was noteworthy as this proved the presence of both genes, hereafter referred to as *fdhF*_{Sec} and *fdhF*_{Cys}, in all acetogenic spirochete isolates to date (Leadbetter *et al.*, 1999). Search of the NCBI database resulted in the discovery of several other distantly related organisms possessing dual *fdhF*_{Sec} and *fdhF*_{Cys} variants. These were *Enterobacteriaceae* belonging to the phylum γ -*Proteobacteria* (i.e., isolates in the genera *Citrobacter*, *Cronobacter*, *Enterobacter*, *Klebsiella*, *Pantoea*, and *Proteus*) and a *Firmicute*, *Clostridium carboxidovorans*. FDH_H in enteric bacteria such as *E. coli* is used in the direction of H₂ production from formate oxidation during sugar fermentation (Sawers, 1994). In contrast, the FDH_H in the solvent producing acetogen *C. carboxidovorans* may catalyze the opposite reaction (i.e., formate production from H₂ + CO₂, as has been implicated for *T. primitia* str. ZAS-2 during acetogenic growth) (Liou *et al.*, 2005; Matson *et al.*, 2010).

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

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

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

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

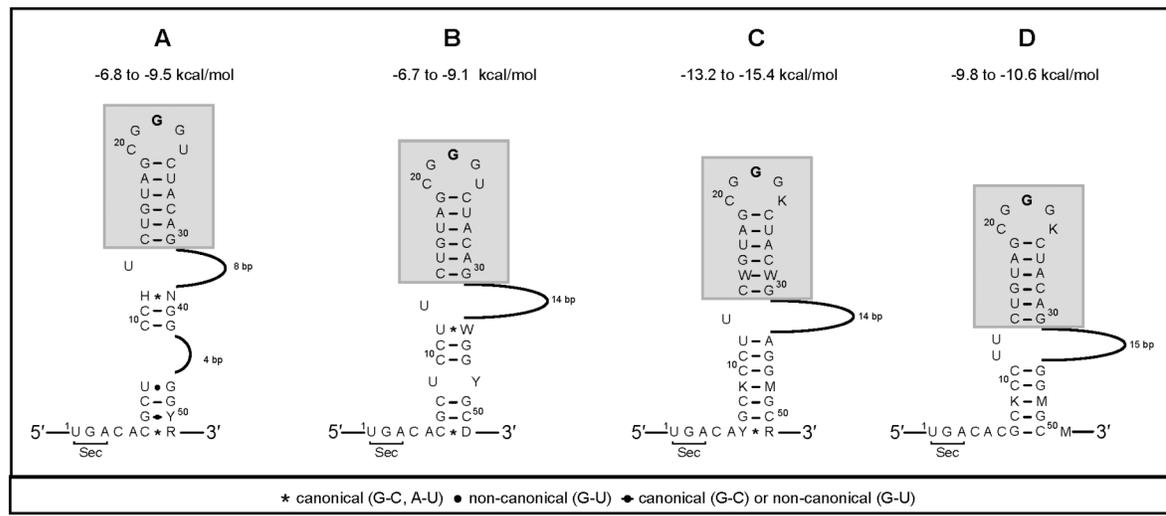
Polymerase chain reactions (PCR) containing gut community DNA from termites or the wood-feeding roach consistently yielded *fdhF* amplicons. In contrast, PCR with gut DNA from the *P. americana* did not amplify *fdhF* after repeated attempts. Clone inventories for *fdhF*, each comprised of 80-136 clones,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

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

***fdhF* genes in gut microbial communities encode both Sec and Cys FDH_H variants.**

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

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



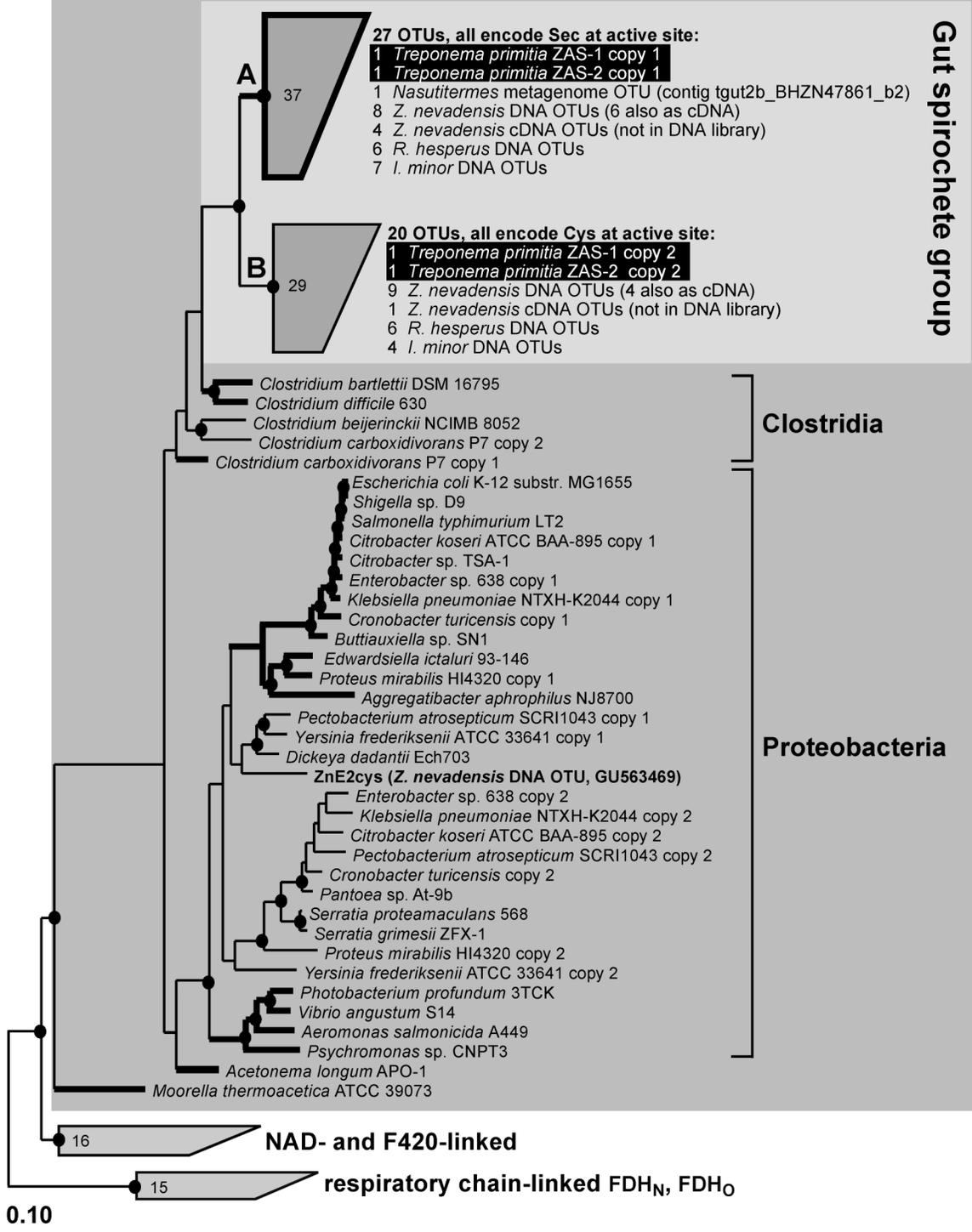
Phylogenetic analysis of termite gut FDH_H sequences

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

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

Hydrogenase-linked FDH_H

Gut spirochete group

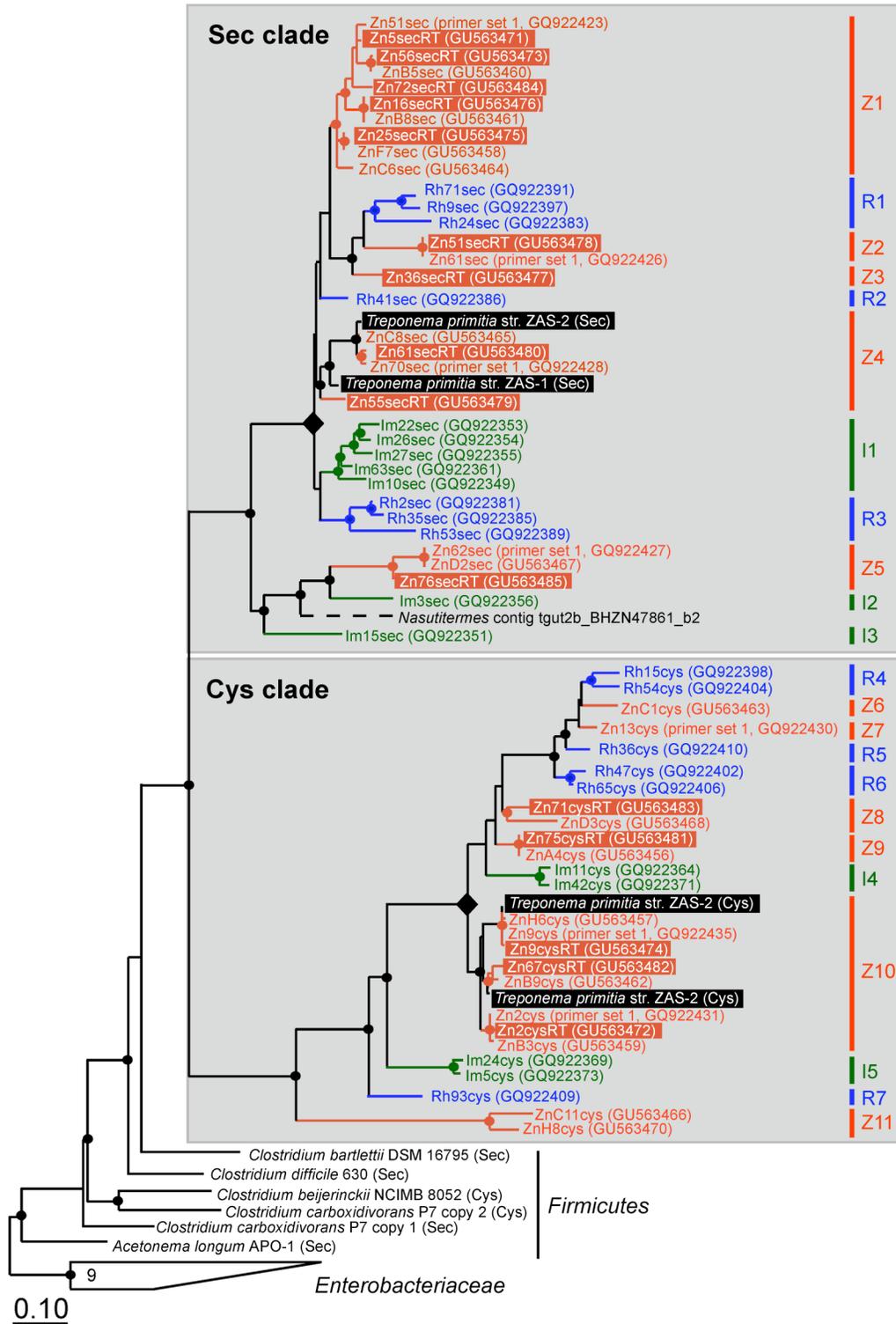


Within the Gut spirochete group, DNA and cDNA sequences could be consistently sub-grouped into one of two clades (Fig. 2.4, grouped clades A and B) based on whether a Sec or Cys amino acid is encoded at the position corresponding to the catalytically active Sec140 residue within the FDH_H from *E. coli*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

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

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

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



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

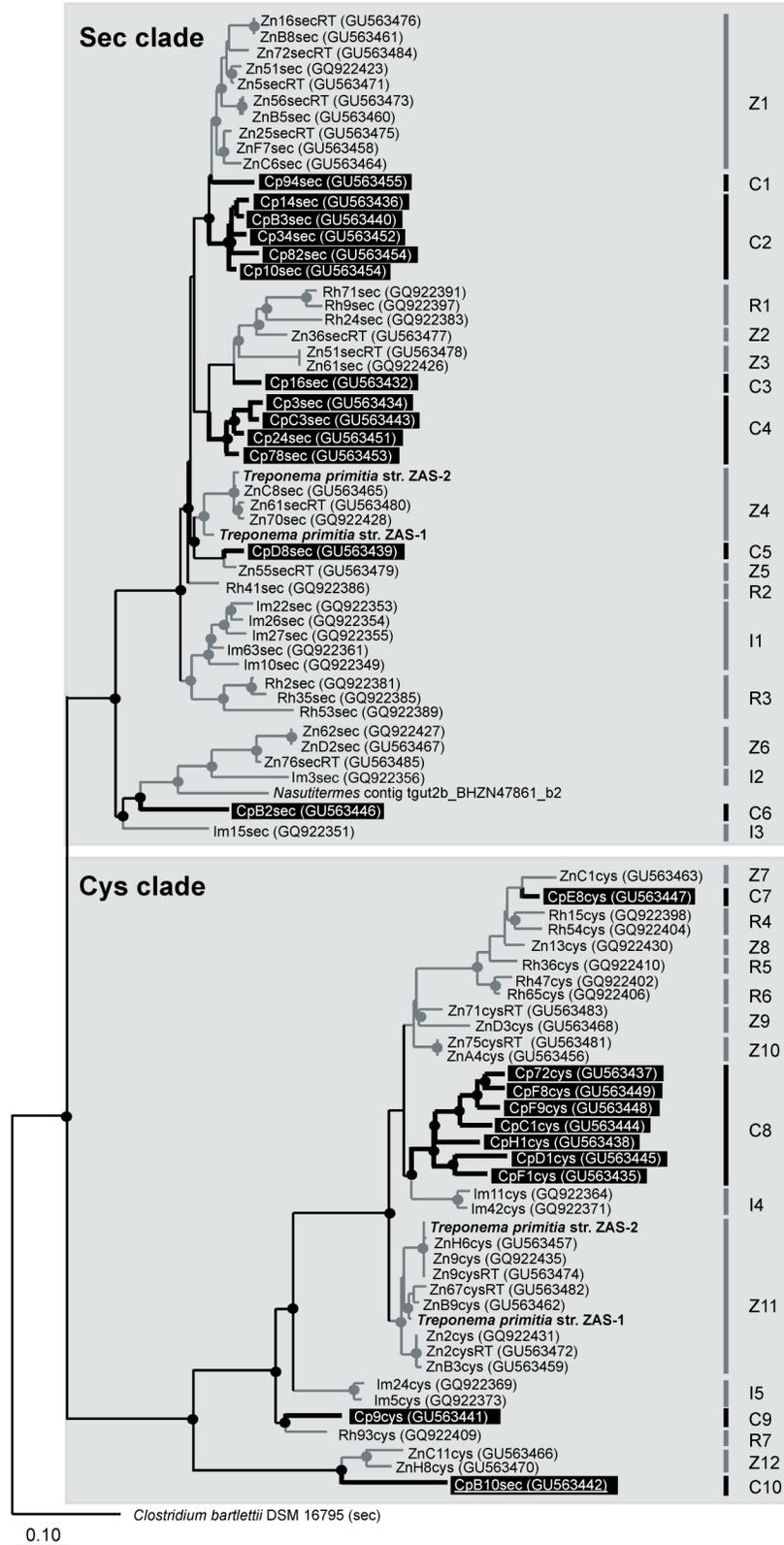
Phylogenetic analysis of wood-roach FDH_H sequences

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

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

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

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



Discussion

In this study, novel degenerate primers were designed to amplify genes for hydrogenase-linked formate dehydrogenase (FDH_H) enzymes encoded by *fdhF* genes. These primers were applied to nucleic acids from bacterial isolates and insect gut microbial communities. To our knowledge, this study represents the first examination of FDH_H diversity in any environment. The results show that (i) the presence of previously unknown *fdhF* genes in three enteric *Gammaproteobacteria*, an H₂ + CO₂ acetogenic treponeme, and an H₂ + CO₂ firmicute, (ii) genes for both Sec and Cys FDH_H variants are broadly represented in the gut communities of three phylogenetically distant lower, wood-feeding termites and a wood-feeding roach, a member of the extant sister taxon to termites, (iii) genes for both Sec and Cys FDH_H variants are transcribed by gut microbial communities, and (iv) nearly all gut sequences phylogenetically group with Sec and Cys FDH_H in cultured acetogenic treponemes.

Previous direct investigation of CO₂-reductive acetogenesis in wood-feeding insect gut microbial communities has demonstrated the unambiguous, H₂ dependent conversion of ¹⁴CO₂ to ¹⁴C-formate (Breznak and Switzer, 1986). While these results point to the importance of formate dehydrogenase for gut acetogenesis, the details underlying the production of formate from H₂ and CO₂ remain largely unknown. A study of the termite gut acetogenic spirochete *T. primitia* str. ZAS-2 (Matson *et al.*, 2010) indicated that it possesses two formate dehydrogenase enzymes that are likely hydrogenase-linked and encoded by *fdhF* homologs. The present survey of *fdhF* gene diversity in wood-feeding insect guts not only underscores observations in *T. primitia*, but suggests that many uncultured acetogenic

treponemes possess genes for FDH_H . This assertion is supported by the finding that all but two gut phylotypes grouped with FDH_H from acetogenic treponeme isolates to the exclusion of all other organisms based on phylogeny (Fig. 2.6) and the presence of a distinguishing sequence character (Appendix, Table 2.6). Analysis of a recently completed draft genome for *T. primitia* str. ZAS-2 (Genbank No. CP001843) confirms the absence of additional formate dehydrogenase gene homologs, demonstrating the relevance of *fdhF* for CO_2 -reductive acetogenesis in these spirochetes. As recent studies have since reinforced the long-standing view that spirochetes are responsible for much of CO_2 -reductive metabolism in termite guts (Salmassi and Leadbetter, 2003; Pester and Brune, 2006), we hypothesize that FDH_H enzymes may be important for acetogenesis in H_2 rich termite hindgut environments.

FDH_H was not an expected feature of the Wood-Ljungdahl pathway for several reasons. First, FDH_H in *E. coli* has only been shown to operate in the direction for formate oxidation (Zinoni *et al.*, 1986; Böhm *et al.*, 1990; Hakobyan *et al.*, 2005). Indeed, assays of formate dehydrogenase activity function in the oxidative direction (Ljungdahl and Andreesen, 1978), regardless of the reaction direction *in vivo*. Secondly, only one Wood-Ljungdahl pathway formate dehydrogenase has been biochemically characterized and it is not a hydrogenase-linked enzyme (Yamamoto *et al.*, 1983). The formate dehydrogenase in the classic *Firmicute* acetogen, *Moorella thermoacetica*, is a tungsten containing selenoprotein that uses electrons from the physiological electron donor NADPH to reduce CO_2 to formate (Thauer, 1972; Yamamoto *et al.*, 1983). We note that this enzyme was purified from cells grown under glucose-driven acetogenic conditions (Yamamoto *et al.*, 1983) and that a gene predicted to encode a hydrogenase-linked formate dehydrogenase in *M. thermoacetica* has

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

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

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

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

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

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

Acknowledgements

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

Appendix

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

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

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

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

Table 2.6. Amino acid alignment in the area of a characteristic amino acid indel (bold) found only in Gut spirochete group FDH_H sequences.

Figure 2.7. Mitochondrial cytochrome oxidase II phylogeny of insects.

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

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

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

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

¹ A 'c' before genome coordinates indicates complementary sequence

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

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

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

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

Reticulitermes hesperus ChiA2 (89, PS 2)

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

Incisitermes minor Pas1 (80, PS 2)

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

Cryptocercus punctulatus nymph (136, PS 2)

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

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

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

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

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

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

⁴ Number of unique Sec and Cys FDH_H phylotypes.

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

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

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

Im27sec	LSDQPGITLTMVPHHVL HEKDP AKQIHAYYVLGEDPAQSDPD
Im63sec	LSDQPGITLTMVPHHVL HEKDP AKKIRAYYIMGEDPAQSDPD
Im10sec	LSDQAGITLTGVPHQVL HETDP AKKIRAYYIMGEDPAQSDPD
Rh2sec	LPDQNGITLTVVPHQVL HEKDP TKQIHGYYIMGEDPVQSDPD
Rh35sec	LPDQNGITLTVVPHQVL HETDP AKKIHGYYIMGEDPVQSDPD
Rh53sec	LPAENGITLTVVGRVL HEKDP SKQIHAYYIMGEDPVQSDPD
Zn62sec	LSNKIGIPLTQVPHYVL HETE -EKKIRAYYIFGEDPAQSDPD
ZnD2sec	LSNKIGIPLTQVPHYVL HETE -EKKIRAYYIFGEDPAQSDPD
Zn76secRT	LSNKIGIPITQVPRYVL HEPE -EKKIRAYYIFGEDPAQSDPD
Im3sec	LPEKVGIPLTQVPHYVL HEPE -ERKIRAYYVFGEDPAQSDPD
CpB2sec	LSDKVGITLTKVPHHVL HEKG -AKKIHAYYIMGEDPAQSDPD
Im15sec	LSDKVGCPIPTHVPHRVL HEKDP AKRIHAYYIFGEDPAQSDPD
ZnClcys	LDNKVGIQLTRIPFV IHEKNP ANRIHAYYITGEDPAQSDPD
ZnHcys	LDNKVGIQLTRIPFV IHEKNP ANRIHAYYITGEDPAQSDPD
CpE8cys	LDNKVGIQLTRIPFV IHEQDP AKRIHAYYITGEDPAQSDPD
Rh15cys	LDNKVGIQLTRIAEFT IHQKDP AKRIHAYYITGEDPAQSDPD
Rh54cys	LDNKVGIQLTRIAEFT IHQADP AKRIHAYYITGEDPAQSDPD
Zn13cys	LDNKVGVQLTRIP ELVLHEKDP AKRIHAYYITGEDPAQSDPD
Rh36cys	LDNKVGIQLTRIP ELVIHEKDP AKRIHAYYITGEDPAQSDPD
Rh47cys	LDNKVGIQLTRIPFV LHEKDP AKRIHAYYITGEDPAQSDPD
Rh65cys	LDNKVGIQLTRIPFV IHEKDP AKRIHAYYITGEDPAQSDPD
Zn71cysRT	LDNKVGLQLTRVPEFV LQEKDP AKQIHAYYITGEDPAQSDPD
ZnD3cys	LDNKVGLQLTRVPEFV LHEKDP AKCIHAYYITGEDPAQSDPD
Zn75cysRT	LDNKVGIQLTRVPEFV LHEKDP KKQLHAYYITGEDPAQSDPD
ZnA4cys	LDNKVGIQLTRVPEFV LHEKDP KKQLHAYYITGEDPAQSDPD
Cp72cys	LDNKVGIQLTRVPEFV IHEKDP AKRIHAYYITGEDPAQSDPD
CpF8cys	LDNKVGIQLTRVPEFV IHEKDP AKRIHAYYITGEDPAQSDPD
CpF9cys	LDNKVGIQLTRVPEFV IHDKDP AKRIHAYYITGEDPAQSDPD
CpClcys	LDNKVGIQLTRVPEFV VHEKDP AKRIHAYYITGEDPAQSDPD
CpH1cys	LDNKVGIQLTRVPEFV IHEKDP AKRIHAYYITGEDPAQSDPD
CpD1cys	LDNQVGLQLTRVPEFV LHEKDP KKRIHAYYITGEDPAQFDPD
CpF1cys	LDNKVGIQLTRVPEFV IHEKDP AKRIHAYYITGEDPAQSDPD
Im11cys	LDDKVGIQLTRVPEFV QHMSDP AKRLHAYYITGEDPCQSDPD
Im42cys	LDDKAGIQLTRVPEFV QHESDP AKRIHAYYITGEDPCQSDPD
<i>Treponema primitia</i> str. ZAS-2 fdhFcys	LSNKAGIQLTRVPEFV IHEKDP AKRIHAYYITGEDPAQSDPD
ZnH6cys	LSNKVGLQLTRVPEFV IHEKDP AKRIHAYYITGEDPAQSDPD
Zn9cys	LSNKAGIQLTRVPEFV IHEKDP AKRIHAYYITGEDPAQSDPD
Zn9cysRT	LSNKAGIQLTRVPEFV IHEKDP AKRIHAYYITGEDPAQSDPD
Zn67cysRT	LSNKAGIQLTRVPEFV IHEKDP AKRIHAYYITGEDPAQSDPD
ZnB9cys	LSNKAGIQLTRVPEFV IHEKDP AKRIHAYYITGEDPAQSDPD
<i>Treponema primitia</i> str. ZAS-1 fdhFcys	LSNKAGIQLTRVPEFV IHEKDP AKRIHAYYITGEDPAQSDPD
Zn2cys	LSNKVGIQLTRVPEFV IHEKDP AKRIHAYYITGEDPAQSDPD
Zn2cysRT	LSNKVGIQLTRVPEFV IHEKDP AKRIHAYYITGEDPAQSDPD

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

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

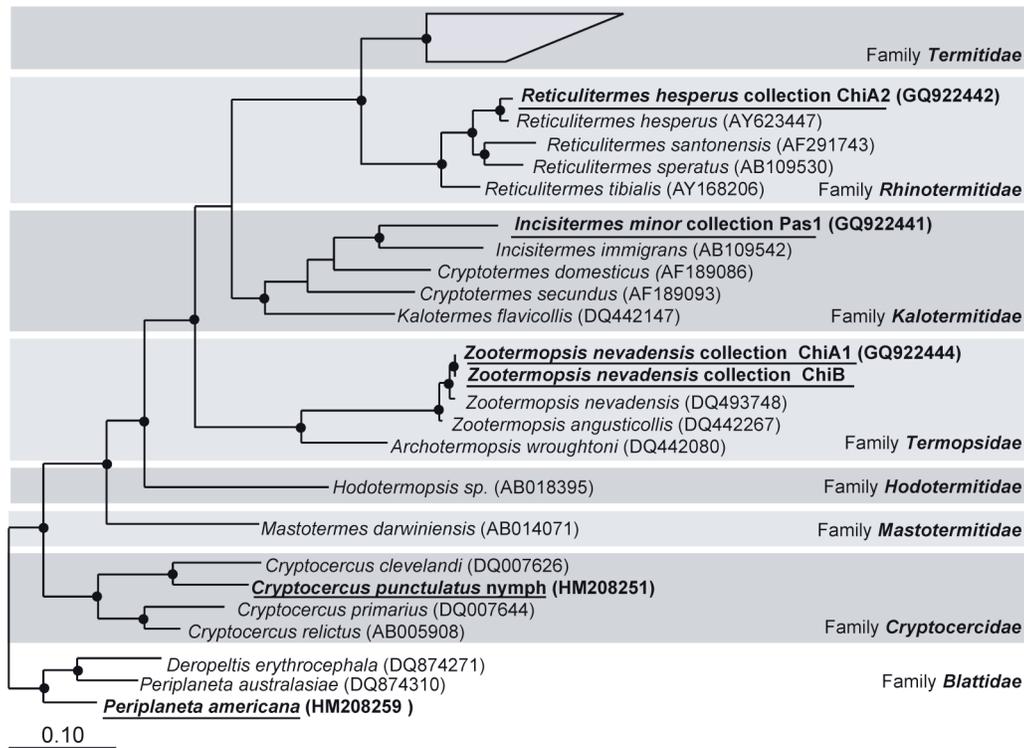
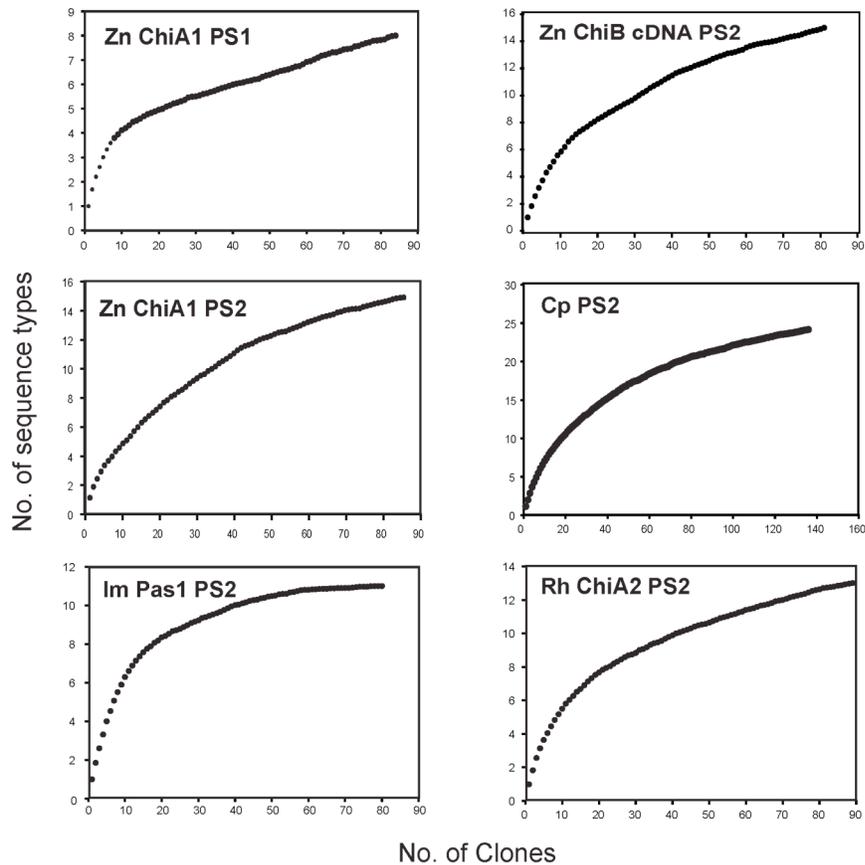


Figure 2.8. Rarefaction curves calculated using EstimateS for each insect gut *fdhF* DNA or cDNA inventory. Sequences were first binned into operational taxonomic units at a cutoff of 97% amino acid similarity (Jones-Thorton-Taylor corrected 3% amino acid difference) using DOTUR. Inventory templates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



References

- Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389-3402.
- Axley, M.J., Böck, A., and Stadtman, T.C. (1991) Catalytic properties of an *Escherichia coli* formate dehydrogenase mutant in which sulfur replaces selenium. *Proc Natl Acad Sci U S A* **88**: 8450-8454.
- Berghöfer, Y., Agha-Amiri, K., and Klein, A. (1994) Selenium is involved in the negative regulation of the expression of selenium-free [NiFe] hydrogenases in *Methanococcus voltae*. *Mol Gen Genet* **242**: 369-373.
- Berry, M.J., Maia, L., Kieffer, J.D., Harney, J.W., and Larsen, P.R. (1992) Substitution of cysteine for selenocysteine in type I iodothyronine deiodinase reduces the catalytic efficiency of the protein but enhances its translation. *Endocrinology* **131**: 1848-1852.
- Böck, A. (2000) Biosynthesis of selenoproteins: an overview. *BioFactors* **11**: 77.
- Böhm, R., Sauter, M., and Böck, A. (1990) Nucleotide sequence and expression of an operon in *Escherichia coli* coding for formate hydrogenylase components. *Mol Microbiol* **4**: 231-243.
- Boyington, J.C., Gladyshev, V., Khangulov, S.V., Stadtman, T., and Sun, P.D. (1997) Crystal structure of formate dehydrogenase H: catalysis Involving Mo, molybdopterin, selenocysteine, and an Fe₄S₄ cluster. *Science* **275**: 1305-1308.
- Brauman, A., Kane, M.D., Labat, M., and Breznak, J.A. (1992) Genesis of acetate and methane by gut bacteria of nutritionally diverse termites. *Science* **257**: 1384-1387.
- Breznak, J.A. (1982) Intestinal microbiota of termites and other xylophagous insects. *Annu Rev Microbiol* **36**: 323-343.
- Breznak, J.A., and Switzer, J.M. (1986) Acetate synthesis from H₂ plus CO₂ by termite gut microbes. *Applied and Environment Microbiology* **52**: 623-630.
- Breznak, J.A., and Brune, A. (1994) Role of microorganisms in the digestion of lignocellulose by termites. *Ann Rev Entomol* **39**: 453-487.
- Breznak, J.A., and Warnecke, F. (2008) *Spirochaeta cellobiosiphila* sp. nov., a facultatively anaerobic, marine spirochaete. *Int J Syst Evol Microbiol* **58**: 2762-2768.
- Chao, A. (1987) Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* **43**: 783-791.
- Colwell, R.K. (2009). EstimateS: statistical estimation of species richness and shared species from samples. version 8.2.0. URL <http://viceroy.eeb.uconn.edu/EstimateS>

- Drake, H., Küsel, K., and Matthies, C. (2002) Ecological consequences of the phylogenetic and physiological diversities of acetogens. *Antonie van Leeuwenhoek* **81**: 203-213.
- Drake, H., Gossner, A., and Daniel, S. (2007) Old acetogens, new light. *Ann N Y Acad Sci* **1125**: 100-128.
- Drake, H.L. (1994) Introduction to acetogenesis. In *Acetogenesis*. Drake, H.L. (ed). New York: Chapman and Hall, pp. 3-60.
- Drake, H.L., Küsel, K., and Matthies, C. (2006) Acetogenic prokaryotes. In *The Prokaryotes*. Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.H., and Stackebrandt, E. (eds): Springer, pp. 354-420.
- Ebert, A., and Brune, A. (1997) Hydrogen concentration profiles at the oxic-anoxic interface: A microsensor study of the hindgut of the wood-feeding lower termite *Reticulitermes flavipes* (Kollar). *Appl Environ Microbiol* **63**: 4039-4046.
- Felsenstein, J. (1989) PHYLIP - Phylogeny inference package (version 3.2). *Cladistics* **5**: 164 - 166.
- Gladyshev, V.N., Khangulov, S.V., Milton, J.A., and Stadtman, T.C. (1994) Coordination of selenium to molybdenum in formate dehydrogenase H from *Escherichia coli*. *Proceedings of the National Academy of Sciences U S A* **91**: 7708-7711.
- Graber, J.R., and Breznak, J.A. (2004) Physiology and nutrition of *Treponema primitia*, an H₂/CO₂-acetogenic spirochete from termite hindguts. *Appl Environ Microbiol* **70**: 1307-1314.
- Graber, J.R., and Breznak, J.A. (2005) Folate cross-feeding supports symbiotic homoacetogenic spirochetes. *Appl Environ Microbiol* **71**: 1883-1889.
- Graber, J.R., Leadbetter, J.R., and Breznak, J.A. (2004) Description of *Treponema azotonutricium* sp. nov. and *Treponema primitia* sp. nov., the first spirochetes isolated from termite guts. *Appl Environ Microbiol* **70**: 1315-1320.
- Grimaldi, D., and Engel, M.S. (2005) *Evolution of the insects*. New York, NY: Cambridge University Press.
- Gromer, S., Johansson, L., Bauer, H., Arscott, L.D., Rauch, S., Ballou, D.P. *et al.* (2003) Active sites of thioredoxin reductases: why selenoproteins? *Proc Natl Acad Sci USA* **100**: 12618 - 12623.
- Hakobyan, M., Sargsyan, H., and Bagramyan, K. (2005) Proton translocation coupled to formate oxidation in anaerobically grown fermenting *Escherichia coli*. *Biophys Chem* **115**: 55-61.

- Honigberg, B.M. (1970) Protozoa associated with termites and their role in digestion. In *Biology of Termites*. Krishna, K., and Weesner, F.M. (eds). New York: Academic Press, pp. 1-36.
- Inoue, T., Kitade, O., Yoshimura, T., and Yamaoka, I. (2000) Symbiotic association with protists. In *Termites: Evolution, Sociality, Symbioses*. Abe, T., Bignell, D.E., and Higashi, M. (eds). Dordrecht, The Netherlands: Kluwer Academic Publishers, pp. 275-288.
- Jones, J.B., and Stadtman, T.C. (1981) Selenium-dependent and selenium-independent formate dehydrogenases of *Methanococcus vannielii*. Separation of the two forms and characterization of the purified selenium-independent form. *J Biol Chem* **256**: 656-663.
- Jormakka, M., Byrne, B., and Iwata, S. (2003) Formate dehydrogenase-a versatile enzyme in changing environments. *Curr Opin Struct Biol* **13**: 418-423.
- Jormakka, M., Tornroth, S., Byrne, B., and Iwata, S. (2002) Molecular basis of proton motive force generation: structure of formate dehydrogenase-N. *Science* **295**: 1863-1868.
- Kambhampati, S., and Eggleton, P. (2000) Taxonomy and phylogeny of termites. In *Termites: Evolution, Sociality, Symbioses, Ecology*. Abe, T., Bignell, D.E., and Higashi, M. (eds). Dordrecht, The Netherlands: Kluwer Academic Publishers, pp. 1-24.
- Kane, M.D., and Breznak, J.A. (1991) *Acetonema longum* gen. nov. sp. nov., an H₂/CO₂ acetogenic bacterium from the termite, *Pterotermes occidentis*. *Arch Microbiol* **156**: 91-98.
- Kim, H.Y., and Gladyshev, V.N. (2005) Different catalytic mechanisms in mammalian selenocysteine- and cysteine-containing methionine-R-sulfoxide reductases. *PLoS Biol* **3**: e375.
- Krishna, K. (1970) Taxonomy, phylogeny, and distribution of termites. In *Biology of Termites*. Krishna, K., and Weesner, F.M. (eds). New York: Academic Press, pp. 127-152.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H. *et al.* (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947-2948.
- Leadbetter, J.R., Schmidt, T.M., Graber, J.R., and Breznak, J.A. (1999) Acetogenesis from H₂ plus CO₂ by spirochetes from termite guts. *Science* **283**: 686-689.
- Lee, S.R., Bar-Noy, S., Kwon, J., Levine, R.L., Stadtman, T.C., and Rhee, S.G. (2000) Mammalian thioredoxin reductase: Oxidation of the C-terminal cysteine/selenocysteine active site forms a thioselenide, and replacement of selenium with sulfur markedly reduces catalytic activity. *Proc Natl Acad Sci USA* **97**: 2521-2526.
- Lilburn, T.G., Kim, K.S., Ostrom, N.E., Byzek, K.R., Leadbetter, J.R., and Breznak, J.A. (2001) Nitrogen fixation by symbiotic and free-living spirochetes. *Science* **292**: 2495-2498.
- Liou, J.S.-C., Balkwill, D.L., Drake, G.R., and Tanner, R.S. (2005) *Clostridium carboxidivorans* sp. nov., a solvent-producing clostridium isolated from an agricultural

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

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

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

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

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

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

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

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

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

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

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

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

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

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

- Romão, M. (2009) Molybdenum and tungsten enzymes: a crystallographic and mechanistic overview. *Dalton Trans*, : 4053 - 4068.
- Salmassi, T.M., and Leadbetter, J.R. (2003) Analysis of genes of tetrahydrofolate-dependent metabolism from cultivated spirochaetes and the gut community of the termite *Zootermopsis angusticollis*. *Microbiology* **149**: 2529–2537.
- Sawers, G. (1994) The hydrogenases and formate dehydrogenases of *Escherichia coli*. *Antonie van Leeuwenhoek* **66**: 57-88.
- Schloss, P.D., and Handelsman, J. (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* **71**: 1501-1506.
- Schultz, J.E., and Breznak, J.A. (1978) Heterotrophic bacteria present in hindguts of wood-eating termites [*Reticulitermes flavipes* (Kollar)]. *Appl Environ Microbiol* **35**: 930-936.
- Stamatakis, A.P., Ludwig, T., and Meier, H. (2004) The AxML program family for maximum likelihood-based phylogenetic tree inference. *Concurrency and Computation: Practice and Experience* **16**: 975-988.
- Thauer, R.K. (1972) CO₂-reduction to formate by NADPH. The initial step in the total synthesis of acetate from CO₂ in *Clostridium thermoaceticum*. *FEBS Lett* **27**: 111-115.
- Valente, F.M., Almeida, C.C., Pacheco, I., Carita, J., Saraiva, L.M., and Pereira, I.A. (2006) Selenium is involved in regulation of periplasmic hydrogenase gene expression in *Desulfovibrio vulgaris* Hildenborough. *J Bacteriol* **188**: 3228-3235.
- Vorholt, J.A., and Thauer, R.K. (2002) Molybdenum and tungsten enzymes in C1 metabolism. *Met Ions Biol Syst* **39**: 571-619.
- Vorholt, J.A., Vaupel, M., and Thauer, R.K. (1997) A selenium-dependent and a selenium-independent formylmethanofuran dehydrogenase and their transcriptional regulation in the hyperthermophilic *Methanopyrus kandleri*. *Mol Microbiol* **23**: 1033-1042.
- Warnecke, F., Luginbühl, P., Ivanova, N., Ghassemian, M., Richardson, T., Stege, J. *et al.* (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**: 560-565.
- Wood, H.G., and Ljungdahl, L.G. (1991) Autotrophic character of the acetogenic bacteria. In *Variations in Autotrophic Life*. Shively, J.M., and Barton, L.L. (eds). New York: Academic Press, pp. 201-250.
- Yamamoto, I., Saiki, T., Liu, S.M., and Ljungdahl, L.G. (1983) Purification and properties of NADP-dependent formate dehydrogenase from *Clostridium thermoaceticum*, a tungsten-selenium-iron protein. *J Biol Chem* **258**: 1826-1832.

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

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

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