C h a p t e r 3

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

Abstract

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

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

Introduction

Hydrogen plays a prominent role in the digestion of wood by termites (1, 6, 13, 39, 40, 44). Hydrogen concentrations in the guts of some termites can reach concentrations exceeding those measured for any other biological system (13, 16, 44, 46, 48, 49, 51). Turnover rates have been measured in some species at fluxes as high as 33 m^3/m^3 gut volume (44). The environment is also spatially complex, comprising a matrix of microenvironments characterized by different hydrogen concentrations (8, 9, 13, 24, 25, 44).

This hydrogen is produced during the fermentation of lignocellulosic polysaccharides by the symbiotic microbial community residing in the termite gut, particularly the protozoa (15, 17, 18, 40, 53, 57, 58). The termites are dependent upon this complex symbiosis for the degradation of wood (2-4, 7, 10, 11, 39). The primary product of this symbiosis is acetate, which the termites use as their primary carbon and energy source (41). Most of the hydrogen produced in the gut is used by CO_2-reducing bacteria to produce up to 1/3 of this acetate in reductive acetogenesis (1, 6, 26, 41, 44). Methanoarchaea consume only a small portion of this hydrogen (1, 24).

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

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

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

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

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

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

Methods

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

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

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

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

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

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

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

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

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

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

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

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

The products of each of these reactions were then subjected to digestion with HinPI1 and the resulting restriction fragment length polymorphism (RFLP) patterns were analyzed by agarose gel electrophoresis.

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

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

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

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

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

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

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

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

Results

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

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

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

Table 3-1. Quantifying hydrogenase clone library diversity.

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

[a]Number of unique restriction fragment polymorphism patterns (RFLPs) observed.

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

[c]Chao1 species-richness index calculated using the classic method in EstimateS. OTUs representing Family 3 [FeFe] hydrogenases with their respective abundances were used as program inputs.

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

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

Figure 3-1.

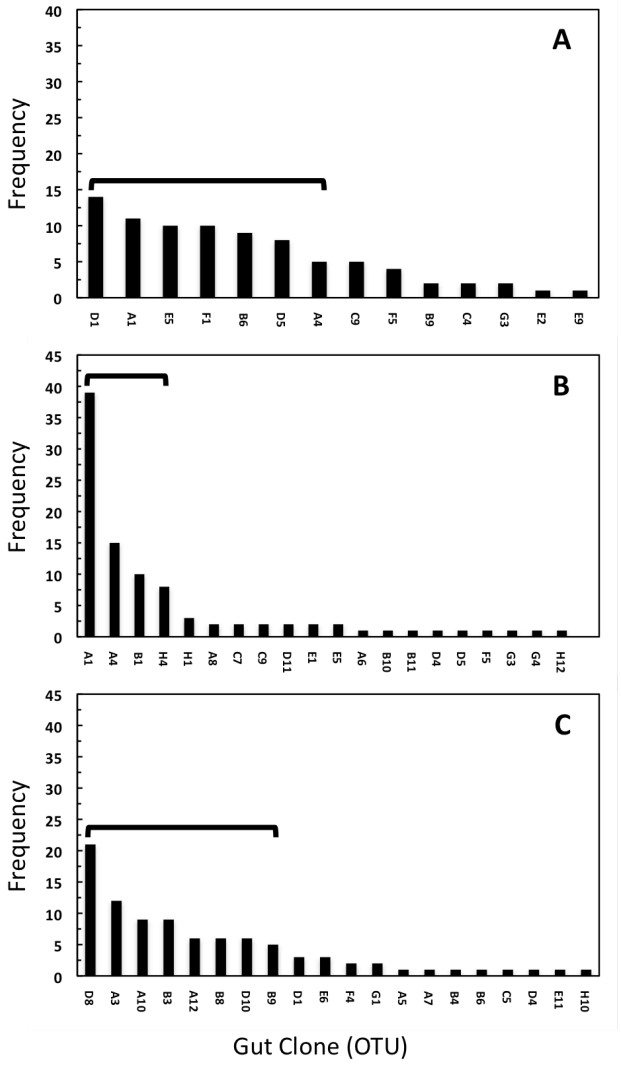

Figure 3-1. Collector's curves for lower termite samples. The horizontal brackets in each figure indicate the number of OTUs comprising 75% of all sequences cloned. Each bin represents an OTU calculated using the furthest-neighbor method in DOTUR (47) with a minimum of 97% amino-acid similarity used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

Figure 3-2.

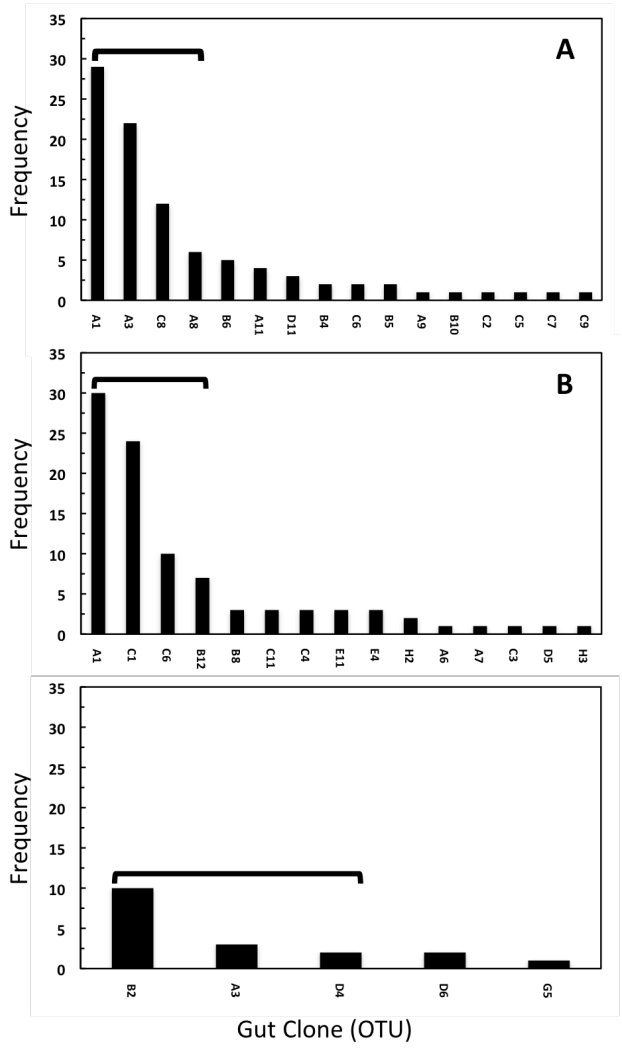

Figure 3-2. Collector's curves for roach samples. The horizontal brackets in each figure indicate the number of OTUs comprising 97% of all sequences cloned. Each bin represents an OTU calculated using the furthest-neighbor method in DOTUR (47) with a 97% amino-acid similarity cut-off. A) *Cryptocercus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

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

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

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

Discussion

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

Figure 3-3.

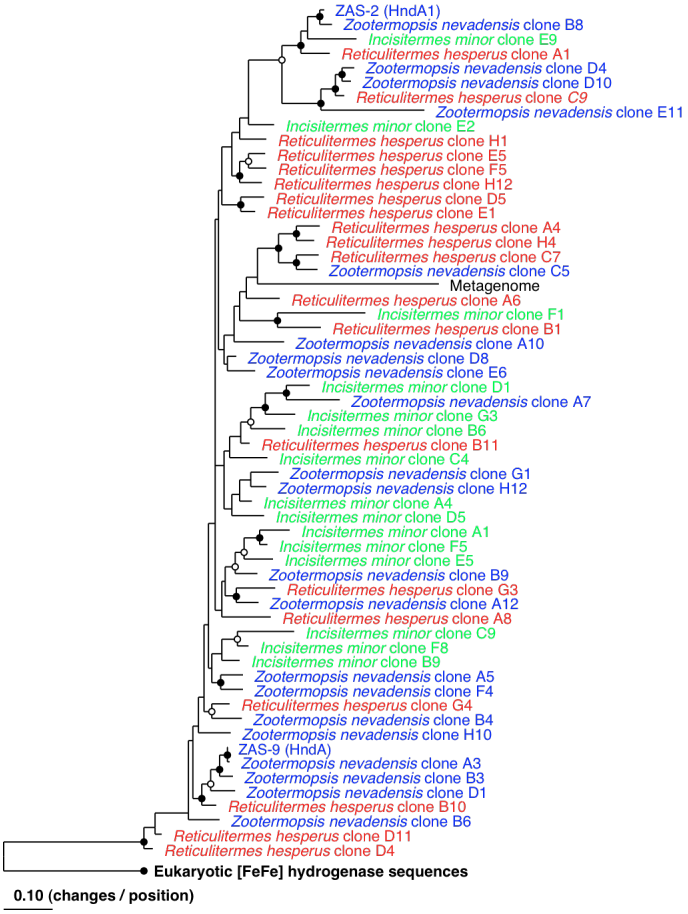

Figure 3-3. Phylogram of Family 3 [FeFe] hydrogenases cloned from the guts of lower termites. The tree was calculated using a maximum likelihood (Phylip ProML) method with 173 unambiguously aligned amino acid positions. Open circles designate groupings also supported by either parsimony (Phylip PROTPARS,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

Figure 3-4.

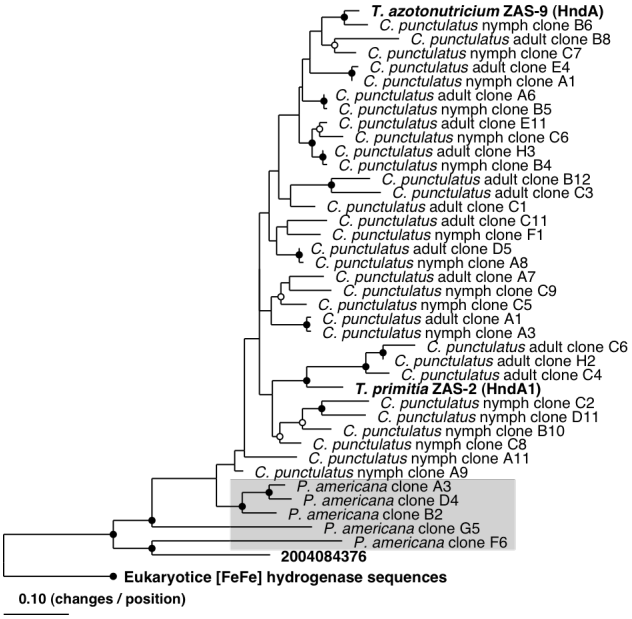

Figure 3-4. Phylogram of Family 3 [FeFe] hydrogenases cloned from the guts of an Adult and Nymph *C. punctulatus* samples. See Figure 3-3 caption for description of open and closed black circles and tree construction methods. All sequences cloned from *P. americana* are highlighted by a grey box.

Figure 3-5.

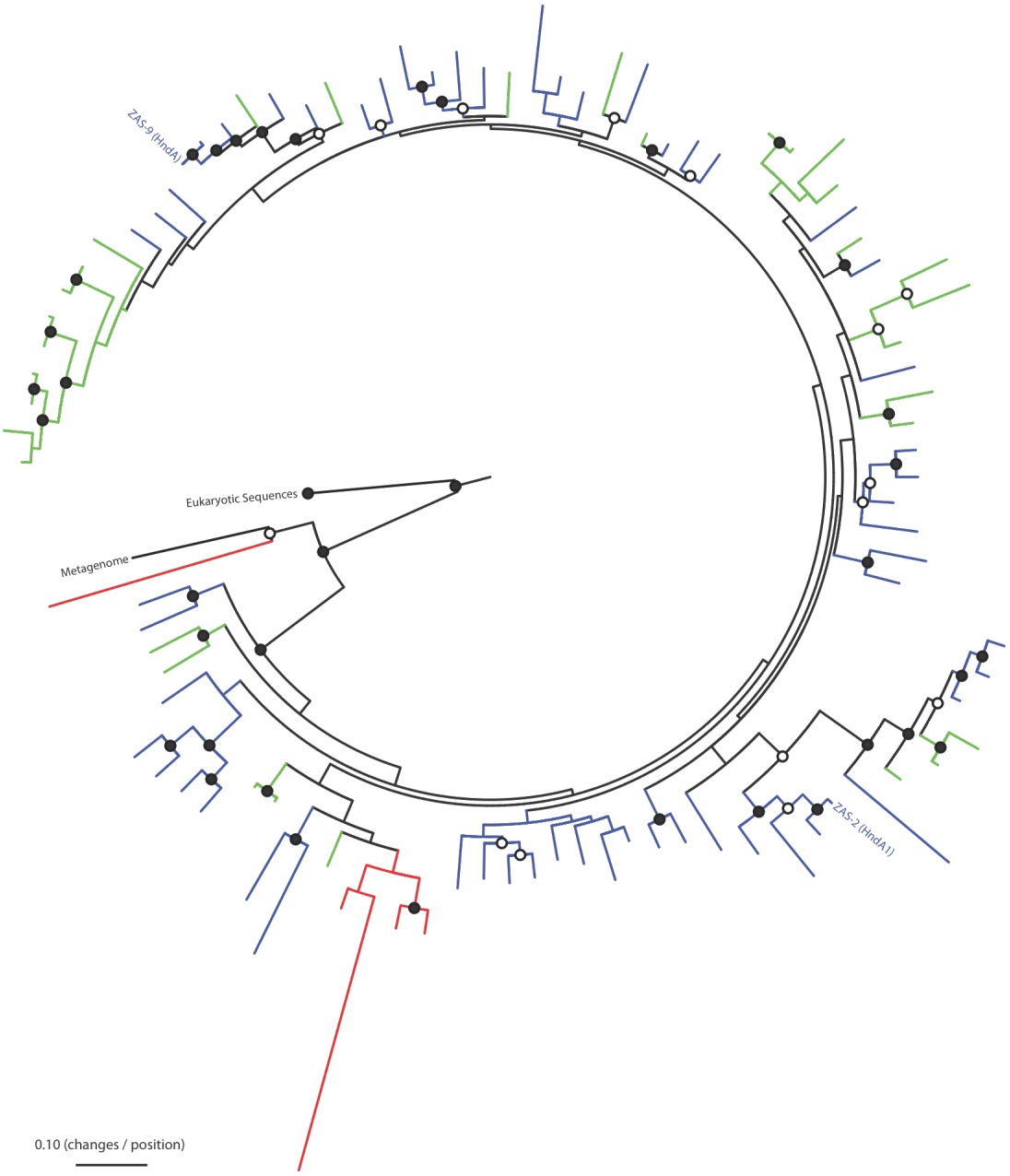

Figure 3-5. Maximum likelihood tree of all cloned Family 3 [FeFe] hydrogenase sequences. See Figure 3-3 caption for description of open and closed black circles and tree construction methods. Each leaf represents an OTU. Leaves and branches representing OTUs cloned from lower termites are in blue, from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

Figure 3-6.

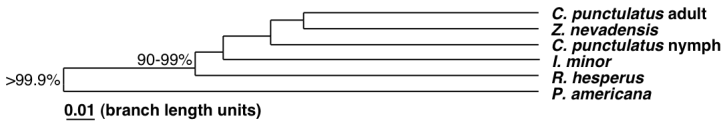

Figure 3-6. Unifrac jackknife clustering of all cloned Family 3 [FeFe] hydrogenase sequences. The maximum-likelihood tree shown in Figure 3-5 and the OTUs with their respective abundance weights given in Table 3-S1 were used as inputs to Unifrac. The analysis was completed using normalized abundance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

Figure 3-7.

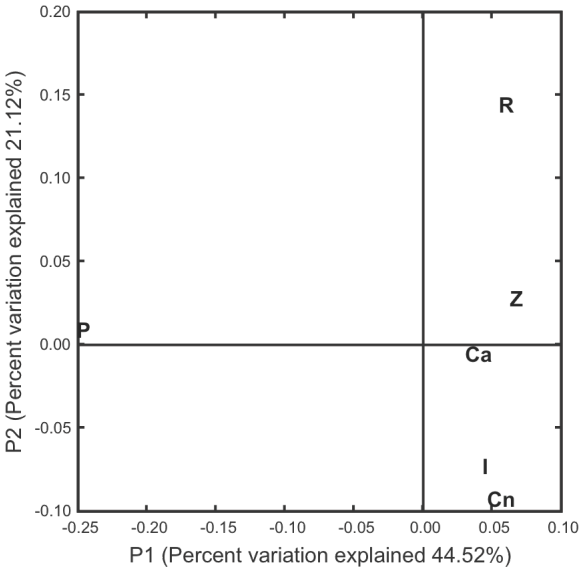

Figure 3-7. Unifrac principle components analysis of all cloned Family 3 [FeFe] hydrogenase sequences. The maximum-likelihood tree shown in Figure 3-5 and the OTUs with their respective abundance weights given in Table 3-S1 were used as inputs to Unifrac. The analysis was completed using normalized abundance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

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

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

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

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

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

References

1. **Brauman, A., M. D. Kane, M. Labat, and J. A. Breznak.** 1992. Genesis of Acetate and Methane by Gut Bacteria of Nutritionally Diverse Termites. Science **257:**1384-1387.
2. **Breznak, J. A.** 2000. Ecology of prokaryotic microbes in the guts of wood- and litter-feeding termites, p. 209-231. *In* T. Abe, D. E. Bignell, and M. Higashi (ed.), Termites: evolution, sociality, symbioses, ecology. Kluwer Academic Publishers, Boston.
3. **Breznak, J. A.** 1982. Intestinal microbiota of termites and other xylophagous insects. Annu. Rev. Microbiol. **36:**323-343.
4. **Breznak, J. A.** 1975. Symbiotic Relationships Between Termites and Their Intestinal Microbiota. Symp. Soc. Exp. Biol. **:**559-580.
5. **Breznak, J. A.** 2006. Termite Gut Spirochetes, p. 421-444. *In* J. Radolf and S. Lukehart (ed.), Pathogenic treponema: molecular and cellular biology. Caister Academic Press.
6. **Breznak, J. A., and J. M. Switzer.** 1986. Acetate Synthesis from H2 plus CO2 by Termite Gut Microbes. Appl. Environ. Microbiol. **52:**623-630.
7. **Brune, A.** 2006. Symbiotic Associations Between Termites and Prokaryotes. Prokaryotes **1:**439-474.
8. **Brune, A.** 1998. Termite guts: the world's smallest bioreactors. Trends in Biotechnol. **16:**16-21.
9. **Brune, A., and M. Friedrich.** 2000. Microecology of the termite gut: structure and function on a microscale. Curr. Opin. Microbiol. **3:**263-269.

10. **Cleveland, L. R.** 1923. Correlation Between the Food and Morphology of Termites and the Presence of Intestinal Protozoa. Am. J. Hyg. **3:**444-461.

11. **Cleveland, L. R.** 1923. Symbiosis between Termites and Their Intestinal Protozoa. Proc. Natl. Acad. Sci. U.S.A. **9:**424-428.

12. **Combet, C., C. Blanchet, C. Geourjon, and G. Deléage.** 2000. NPS@: network protein sequence analysis. Trends Biochem. Sci. **25:**147-50.

13. **Ebert, A., and A. Brune.** 1997. Hydrogen Concentration Profiles at the Oxic-Anoxic Interface: a Microsensor Study of the Hindgut of the Wood-Feeding Lower Termite *Reticulitermes flavipes* (Kollar). Appl. Environ. Microbiol. **63:**4039-4046.

14. **Felsenstein, J.** 1989. PHYLIP - Phylogeny Inference Package (Version 3.2). Cladistics **5:**164-166.

15. **Graber, J. R., J. R. Leadbetter, and J. A. Breznak.** 2004. Description of *Treponema azotonutricium* sp. nov. and *Treponema primitia* sp. nov., the First Spirochetes Isolated from Termite Guts. Appl. Environ. Microbiol. **70:**1315-1320.

16. **Hoehler, T. M., B. M. Bebout, and D. J. Des Marais.** 2001. The role of microbial mats in the production of reduced gases on the early Earth. Nature **412:**324-327.

17. **Hungate, R. E.** 1939. Experiments on the Nutrition of Zootermopsis. III. The anaerobic carbohydrate dissimilation by the intestinal protozoa. Ecology **20:**230-245.

18. **Hungate, R. E.** 1943. Quantitative Analysis on the Cellulose Fermentation by Termite Protozoa. Ann. Entomol. Soc. Am. **36:**730-9.

19. **Inoue, J.-I., K. Saita, T. Kudo, S. Ui, and M. Ohkuma.** 2007. Hydrogen Production by Termite Gut Protists: Characterization of Iron Hydrogenases of Parabasalian Symbionts of the Termite *Coptotermes formosanus*. Eukaryotic Cell **6:**1925-1932.

20. **Inward, D., G. Beccaloni, and P. Eggleton.** 2007. Death of an order: a comprehensive molecular phylogenetic study confirms that termites are eusocial cockroaches. Biol. Lett. **3:**331-335.

21. **Inward, D. J. G., A. P. Volger, and P. Eggleton.** 2007. A comprehensive phylogenetic analysis of termites (Isoptera) illuminates key aspects of their evolutionary biology. Mol. Phylogenet. Evol. **44:**953-967.

22. **Kambhampati, S., and P. Eggleton.** 2000. Taxonomy and Phylogeny of Termites, p. 1-23. *In* T. Abe, D. E. Bignell, and M. Higashi (ed.), Termites: evolution, sociality, symbioses, ecology. Kluwer Academic Publishers, Boston.

23. **Klass, K.-D., C. Nalepa, and N. Lo.** 2008. Wood-feeding cockroaches as models for termite evolution (Insecta: Dictyoptera): *Cryptocercus* vs. *Parasphaeria boleiriana*. Mol. Phylogenet. Evol. **46:**809-817.

24. **Leadbetter, J. R.** 1996. Physiological ecology of *Methanobrevibacter cuticularis* sp. nov. and *Methanobrevibacter curvatus* sp. nov., isolated from the hindgut of the termite *Reticulitermes flavipes*. Appl. Environ. Microbiol. **62:**3620-31.

25. **Leadbetter, J. R., L. D. Crosby, and J. A. Breznak.** 1998. *Methanobrevibacter filiformis* sp. nov., a filamentous methanogen from termite hindguts. Arch. Microbiol. **169:**287-92.

26. **Leadbetter, J. R., T. M. Schmidt, J. R. Graber, and J. A. Breznak.** 1999. Acetogenesis from H2 Plus CO2 by Spirochetes from Termite Guts. Science **283:**686-689.

27. **Legendre, F., M. F. Whiting, C. Bordereau, E. M. Cancello, T. A. Evans, and P. Grandcolas.** 2008. The phylogeny of termites (Dictyoptera: Isoptera) based on mitochondrial and nuclear markers: Implications for evolution of the worker and pseudergate castes, and foraging behaviors. Mol. Phylogenet. Evol. **48:**615-627.

28. **Lemke, T., T. v. Alen, J. H. P. Hackstein, and A. Brune.** 2001. Cross-Epithelial Hydrogen Transfer from the Midgut Compartment Drives Methanogenesis in the Hindgut of Cockroaches. Appl. Environ. Microbiol. **67:**4657-4661.

29. **Lilburn, T. G., K. S. Kim, N. E. Ostrom, K. R. Byzek, J. R. Leadbetter, and J. A. Breznak.** 2001. Nitrogen Fixation by Symbiotic and Free-Living Spirochetes. Science **292:**2495-2498.

30. **Lo, N., G. Tokuda, H. Watanabe, H. Rose, M. Slaytor, K. Maekawa, C. Bandi, and H. Noda.** 2000. Evidence from multiple sequences indicates that termites evolved from wood-feeding cockroaches. Curr. Biol. **10:**801-804.

31. **Lozupone, C., M. Hamady, and R. Knight.** 2006. UniFrac -- An online tool for comparing microbial community diversity in a phylogenetic context. BMC Bioinformatics **7**.

32. **Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Förster, I. Brettske, S. Gerber, A. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. König, T. Liss, R. Lüssmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N.**

Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode, and K. Schleifer. 2004. ARB: a software environment for sequence data. Nucleic Acids Res. **32:**1363-1371.

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

34. **Meyer, J.** 2007. [FeFe] hydrogenases and their evolution: a genomic perspective. Cell. Mol. Life Sci. **64:**1063-1084.

35. **Miura, T., K. Maekawa, O. Kitade, T. Abe, and T. Matsumoto.** 1998. Phylogenetic relationships among subfamilies in higher termites (Isoptera: Termitidae) based on mitochondrial COII gene sequences. Ann. Entomol. Soc. Am. **91:**515-523.

36. **Néron, B., H. Ménager, C. Maufrais, N. Joly, T. Pierre, and C. Letondal.** 2008. Presented at the Bio Open Source Conference, Toronto.

37. **Nicolet, Y., C. Piras, P. Legrand, C. E. Hatchikian, and J. C. Fontecilla-Camps.** 1999. *Desulfovibrio desulfuricans* iron hydrogenase: the structure shows unusual coordination to an active site Fe binuclear center. Structure **7:**13-23.

38. **Noirot, C.** 1995. The Gut of Termites (Isoptera). Comparative Anatomy, Systematics, Phylogeny. I. Lower Termites. Ann. Soc. Entomol. Fr. **31:**197-226.

39. **Odelson, D. A., and J. A. Breznak.** 1985. Cellulase and Other Polymer-Hydrolyzing Activities of *Trichomitopsis termopsidis*, a Symbiotic Protozoan from Termites. Appl. Environ. Microbiol. **49:**622-626.

40. **Odelson, D. A., and J. A. Breznak.** 1985. Nutrition and Growth Characteristics of *Trichomitopsis termopsidis*, a Cellulolytic Protozoan from Termites. Appl. Environ. Microbiol. **49:**614-621.

41. **Odelson, D. A., and J. A. Breznak.** 1983. Volatile Fatty Acid Production by the Hindgut Microbiota of Xylophagous Termites. Appl. Environ. Microbiol. **45:**1602-1613.

42. **Ohkuma, M., H. Noda, Y. Hongoh, C. A. Nalepa, and T. Inoue.** 2009. Inheritance and diversification of symbiotic trichonymphid flagellates from a common anscestor of termites and the cockroach *Cryptocercus*. Proc. Biol. Sci. **276:**239-245.

43. **Ottesen, E. A.** 2009. The Biology and Community Structure of CO_2-Reducing Acetogens in the Termite Hindgut. California Institute of Technology, Pasadena.

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

45. **Peters, J. W., W. N. Lanzilotta, B. J. Lemon, and L. C. Seefeldt.** 1998. X-ray Crystal Structure of the Fe-Only Hydrogenase (CpI) from *Clostridium pasteurianum* to 1.8 Angstrom Resolution. Science **282:**1853-1858.

46. **Schink, B., F. S. Lupton, and J. G. Zeikus.** 1983. Radioassay for Hydrogenase Activity in Viable Cells and Documentation of Aerobic Hydrogen-Consuming Bacteria Living in Extreme Environments. Appl. Environ. Microbiol. **45:**1491-1500.

47. **Schloss, P. D., and J. Handelsman.** 2005. Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. Appl. Environ. Microbiol. **71:**1501-1506.

48. **Scranton, M. I., P. C. Novelli, and P. A. Loud.** 1984. The distribution and cycling of hydrogen gas in the waters of two anoxic marine environments. Limnol. Oceanogr. **29:**993-1003.

49. **Smolenski, W. J., and J. A. Robinson.** 1988. In situ rumen hydrogen concentrations in steers fed eight times daily, measured using a mercury reduction detector. FEMS Microbiol. Ecol. **53:**95-100.

50. **Sugimoto, A., D. E. Bignell, and J. A. MacDonald.** 2000. Global Impact of Termites on the Carbon Cycle and Atmospheric Trace Gases, p. 409-435. *In* T. Abe, D. E. Bignell, and M. Higashi (ed.), Termites: evolution, sociality, symbioses, ecology. Kluwer Academic Publishers, Boston.

51. **Sugimoto, A., and N. Fujita.** 2006. Hydrogen concentrations and stable isotopic composition of methane in bubble gas observed in a natural wetland. Biogeochemistry **81:**33-44.

52. **Thorne, B. L.** 1990. A case for ancestral transfer of symbionts between cockroaches and termites. Proc. Biol. Sci. **241:**37-41.

53. **Trager, W.** 1934. The Cultivation of a Cellulose-Digesting Flagellate, *Trichomonas termopsidis*, and of Certain Other Termite Protozoa. Biol. Bull. **66:**182-190.

54. **Vignais, P. M., and B. Billoud.** 2007. Occurence, Classification, and Biological Function of Hydrogenases: An Overview. Chem. Rev. **107:**4206-4272.

55. **Vignais, P. M., B. Billoud, and J. Meyer.** 2001. Classification and phylogeny of hydrogenases. FEMS Microbiol. Rev. **25:**455-501.

56. **Warnecke, F., P. Luginbühl, N. Ivanova, M. Ghassemian, T. H. Richardson, J. T. Stege, M. Cayouette, A. C. McHardy, G. Djordjevic, N. Aboushadi, R. Sorek, S. G. Tringe, M. Podar, H. G. Martin, V. Kunin, D. Dalevi, J. Madejska, E. Kirton, D. Platt, E. Szeto, A. Salamov, K. Barry, N. Mikhailova, N. C. Kyrpides, E. G. Matson, E. A. Ottesen, X. Zhang, M. Hernández, C. Murillo, L. G. Acosta, I. Rigoutsos, G. Tamayo, B. D. Green, C. Chang, E. M. Rubin, E. J. Mathur, D. E. Robertson, P. Hugenholtz, and J. R. Leadbetter.** 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. Nature **450:**560-569.

57. **Yamin, M. A.** 1979. Cellulolytic activity of an axenically-cultivated termite flagellate, Trichomitopsis termopsidis. J. Gen. Microbiol. **113:**417-20.

58. **Yamin, M. A.** 1981. Cellulose Metabolism by the Flagellate *Trichonympha* from a Termite Is Independent of Endosymbiotic Bacteria. Science **211:**58-59.

59. **Yang, H., D. Schmitt-Wagner, U. Stingl, and A. Brune.** 2005. Niche heterogeneity determites bacterial community structure in the termite gut (*Reticulitermes santonensis*). Environ. Microbiol. **7:**916-932.

60. **Zhang, X.** 2010. Formate dehydrogenase gene diversity in lignocellulose-feeding insect gut microbial communities. California Institute of Technology, Pasadena.

61. **Zimmerman, P. R., J. P. Greenberg, S. O. Wandiga, and P. J. Crutzen.** 1982. Termites: A Potentially Large Source of Atmospheric Methane, Carbon Dioxide, and Molecular Hydrogen. Science **218:**563-565.

Appendix

Table 3-S1. Cloned sequences

Table 3-S2. Alignment used in primer design

Table 3-S1. Sequences cloned.

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

Continuing Table 3-S1.

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

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

[b]Number of cloned sequences grouped within each OTU.

[c]Percent of cloned sequences represented by each OTU.

[d]Sequences that could not be classified as belonging to any of the sequence families defined by Warnecke *et al*. (56).

[e]Sequence classified as belonging to the Family 7 hydrogenase sequence family defined by Warnecke *et al.* (56).

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