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

Abstract

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

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

Introduction

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

Investigations on the nature of termite-microbe nutritional mutualisms indicate lignocellulose degradation by gut microbes is stepwise and results in the production of substantial levels of acetate, the main carbon and energy source of the insect host (11, 39, 43). Polysaccharides are first hydrolyzed from wood and fermented to acetate, H_2, and CO_2. CO_2 reductive bacteria, using the Wood-Ljungdahl pathway for acetogenesis, then consume the great majority of fermentation-derived H_2 and CO_2 (i.e., 82–100% in lower termites) and produce additional acetate for the insect host (11, 39, 43). Acetate generated from CO_2 reductive acetogenesis may account for up to 30% of gut acetate (11,

39). The remaining H_2 from fermentation does not benefit the host, but is instead consumed by methanogenic *Archaea* and emitted as methane.

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

Investigations aimed at elucidating the processes involved in digestion of non-woody substrates have also been undertaken, but interpretations have been challenged by the complex nature of food substrates like soil. Nevertheless, several important observations have been made. Radiotracer studies comparing carbon and reductant flows in higher termites with different feeding habits revealed rates of acetogenesis and methanogenesis could vary by an order of magnitude (8, 9, 11). In these experiments, CO_2 reduction to acetate was the dominant terminal electron accepting process in grass- and wood-feeding termites, but methanogenesis outcompeted acetogenesis for H_2 in fungus- and soil-feeding termites (9, 50). Efforts aimed at understanding the organisms responsible for such differences have been largely focused on ribosome-based identifications (1, 8, 35,

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

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

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

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

Materials and Methods

Insect collection and identification

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

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

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

DNA extraction

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

***fdhF* amplification and cloning**

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

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

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

RFLP analysis, sequencing, diversity assessment

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

COII amplification for termite identification

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

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

Primer design and PCR for a major clade of lower termite and wood roach Cys FDH$_H$ alleles

Degenerate primers (Cys499F1b, 1045R) for a major clade of selenium independent (Cys) FDH$_H$ alleles present in lower termites and the wood roach *C. punctulatus* were designed manually using all sequences recovered from these insects (Chapter 2). Forward primer Cys499F1b (5'– ATG TCS CTK TCS ATI CCG GAA A –3') specificity is as follows: 38.9% of the sequences are perfectly matched, 22.2% have 1 mismatch, 27.8% have 2 mismatches, and 8.3% have 3 mismatches. No mismatches are in located in the terminal 3' position. The reverse primer 1045R (5'– CIC CCA TRT CGC AGG YIC CCT G –3') was designed based on 154 sequences from higher termites, lower termites and *C. punctulatus*. The primer targets both Sec and Cys *fdhF* variants; 60.3% of the sequences have 0 primer mismatches, 32.4% have 1, 5.8% have 2, and 1.3% have 3 mismatches. All sequences are perfectly matched at the terminal 3' position. PCR reactions contained 0.4 ng · μl^{-1} of DNA template, 200 nM of Cys4991F1b, 200 nM 1045R, 1X FAILSAFE Premix D (EPICENTRE), and 0.05 U · μl^{-1} Taq polymerase (New England Biolabs). Thermocycling conditions were 2 min at 95°C, 30 cycles of (95°C for 30 sec, 60°C for 30 sec, 72°C for 45 sec), followed by 10 min at 72°C.

Primer design and PCR for a novel group of FDH$_H$ alleles identified in subterranean and litter feeding termites.

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

Phylogenetic and Principle Component Analysis

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

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

Results

Termite classification

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

Table 3.1. Characteristics of insects examined in this study.

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

[1] Termite family classifications were based on Kambhampati and Eggleton (26) and Grimaldi and Engel (22).

[2] Nest type (Arboreal versus subterranean) and collection location, CR = Costa Rica, JT= Joshua Tree, CA.

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

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

[5] Predicted level of soil exposure based on nest location (subterranean or above ground), food substrate, and foraging style.

[6] It is unclear whether *Amitermes* sp. affiliate within the subfamily *Termitinae* or rather constitute their own subfamily (26).

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

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

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

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

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

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

Phylotype abundance related to lifestyle and insect phylogeny

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

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

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

Subfamily Termitinae
- *Microcerotermes arboreus*
- **Microcerotermes sp. Cost006**
- *Microcerotermes parvus*
- **Microcerotermes sp. Cost008**

Subfamily Nasutitermitinae
- *Cornitermes pugnax*
- *Cornitermes walkeri*
- *Labiotermes labralis*
- **Rhynchotermes sp. Cost004**
- *Nasutitermes sp.* warnecke-2007
- **Nasutitermes sp. Cost003**
- *Nasutitermes ephratae*
- *Nasutitermes corniger*
- **Nasutitermes corniger cost007**
- *Nasutitermes nigriceps*

Subfamily Termitinae
- **Gnathamitermes sp. JT5**
- **Amitermes sp. JT2**
- **Amitermes sp. Cost010**
- *Amitermes evuncifer*
- *Amitermes dentatus*

Higher termites: Termitidae

- **Reticulitermes tibialis JT1**
- *Reticulitermes tibialis*
- **Reticulitermes hesperus ChiA2**
- *Reticulitermes speratus*
- *Reticulitermes santonensis*
- *Coptotermes testaceus*
- **Coptotermes sp. Cost009**
- *Coptotermes niger*
- **Incisitermes minor Pas1**
- *Incisitermes immigrans*
- *Cryptotermes domesticus*
- *Cryptotermes secundus*
- *Kalotermes flavicollis*
- **Zootermopsis nevadensis ChiA1**
- *Zootermopsis angusticollis*
- *Archotermopsis wroughtoni*
- *Mastotermes darwiniensis*
- *Hodotermopsis japonica*
- *Hodotermopsis sp.*

Lower termites: Rhinotermitidae, Kalotermitidae, Termopsidae, Hodotermitidae, Mastotermitidae

- *Cryptocercus punctulatus*
- **Cryptocercus punctulatus nymph**
- *Cryptocercus clevelandi*
- *Cryptocercus primarius*

Wood-feeding roaches

- *Deropeltis erythrocephala*
- *Periplaneta australasiae*
- *Periplaneta americana*

Omnivorous roaches

0.10

Table 3.2. FDH inventories constructed in this study.

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

[1] Two libraries were constructed for each of the following templates: Cost003, Cost004, and Cost010. These differ most significantly in PCR annealing temperature (details in Table 3.4). PCR was performed at 55°C for libraries 3L1 and 4L1, 53.6 °C for 10L1, and 51°C for 3L2, 4L2, and 10L2.

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

[3] Mean of the diversity estimator Chao1 (SD, standard deviation) calculated using EstimateS.

[4] Lower (LCI) and higher (HCI) 95% confidence interval for mean Chao1.

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

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

Higher termite sequences affiliate with four major FDH$_H$ clades

Higher termite FDH$_H$ sequences phylogenetically cluster into four major clades (Figure 3.2, clades A-D) within the FDH$_H$ family of enzymes, composed of sequences from enteric γ-*Proteobacteria*, *Spirochaetes*, *Firmicutes,* and uncultured organisms from lower termite and wood-roach hindguts.   The relative abundances of the different sequence types in each inventory are listed in Table 3.3.
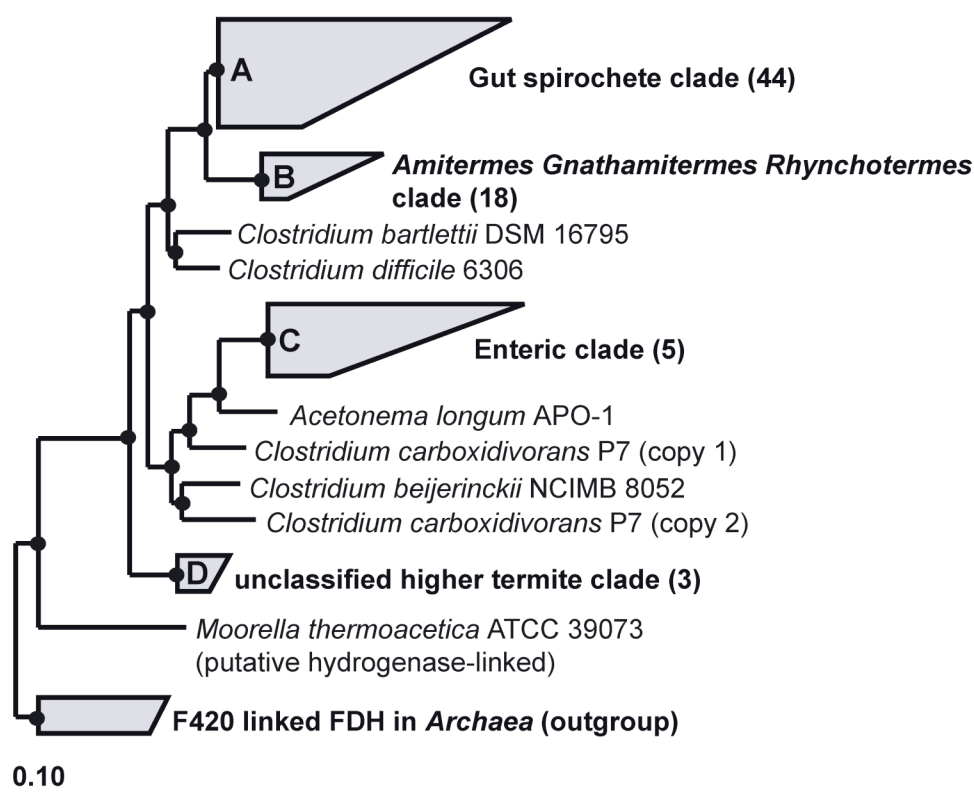
Figure 3.2. Higher termite, lower termite, wood-roach, and pure culture sequences form four major FDH$_H$ clades (A, B, C, D). The numbers of sequences within grouped clades are indicated in parentheses. Tree was constructed with 542 aligned amino acids with the maximum likelihood phylogenetic algorithm Phylip PROTML. A metagenomic FDH$_H$ sequence fragment (tgut2b_BHZN47861_b2) from the gut of *Nasutitermes* sp. Warnecke-2007 (54) was added in by parsimony and falls within the Gut spirochete clade (clade A). Filled circles indicate nodes supported by PROTML and parsimony (Phylip PROPARS, 100 bootstraps) methods. The tree was outgrouped with F420-linked FDH from methanogenic *Archaea*. Scale bar indicates 0.1 units of amino acid change per alignment position.

Table 3.3. Distribution of clones in each major FDH$_H$ clade.

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

[1] Libraries L1 and L2 were combined for abundance calculation.

Sweeping loss of 'Cys clade' alleles from higher termite gut communities

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

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

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

We then hypothesized that the absence of Cys clade alleles in higher termites may be related to insect habitat. To explore the relationship between habitat and the presence of Cys clade genes, we performed PCR screens of two lower termite species collected in the same habitats as certain higher termites (Costa Rican lower termite *Coptotermes* sp. Cost009 collected near Cost006 and Cost008; desert-adapted lower termite *R. tibialis* sp. JT1 collected near JT1 and JT5). PCR amplicons were observed for each lower termite sample (Figure 3.7, Appendix). Dilution-to-extinction PCRs suggest that Cys clade alleles are at least 1000-fold more abundant in lower termites than higher termites (calculations in the legend of Figure 3.7, Appendix). Taken together, the results of targeted PCR assays are consistent with inventory findings and the hypothesis that sweeping gene loss has occurred in the FDH$_H$ gene pool of higher termites.
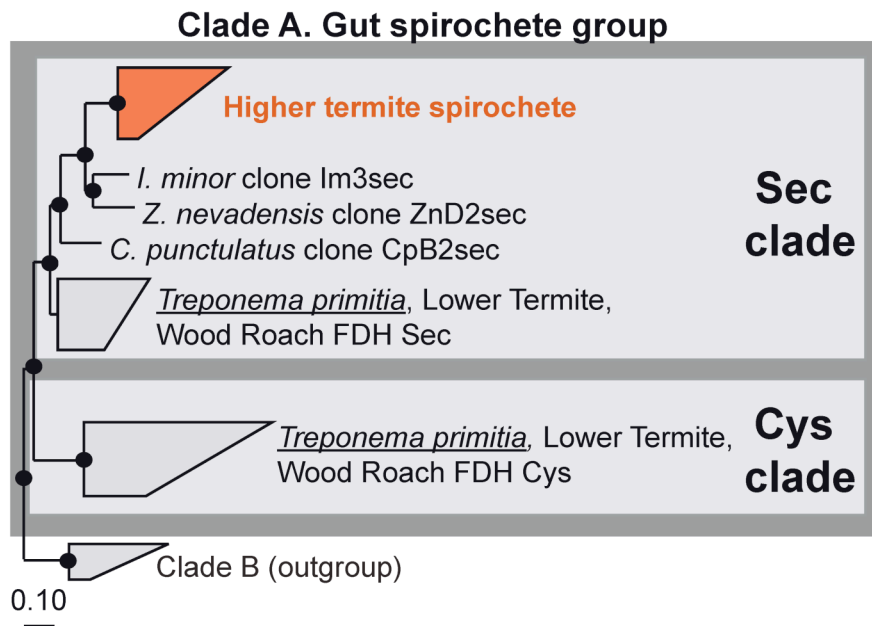
Figure 3.3. Sec and Cys clades within the "Gut spirochete clade" (Clade A, Figure 3.2). Higher termite sequences, marked in red, form the "higher termite spirochete group". Tree was constructed using the methods described in the legend of Figure 3.2. Filled circles indicate nodes were supported by PROTML and parsimony methods of analyses. Scale bar indicates 0.1 amino acid changes per alignment position.

Higher termite FDH$_H$ sequences form a single clade within the 'Sec clade' of the Gut spirochete group

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

Besides being the largest, the "higher termite spirochete" clade is also the most diverse group with regard to termite species representation (Table 3.3). The relative abundances of higher termite spirochete clade sequences in the inventories (74-100%) indicate spirochete-like FDH$_H$ types dominate FDH$_H$ diversity in all higher termite species but the litter-feeding termite Cost004, in which they are the second most abundant FDH$_H$ type (47%).
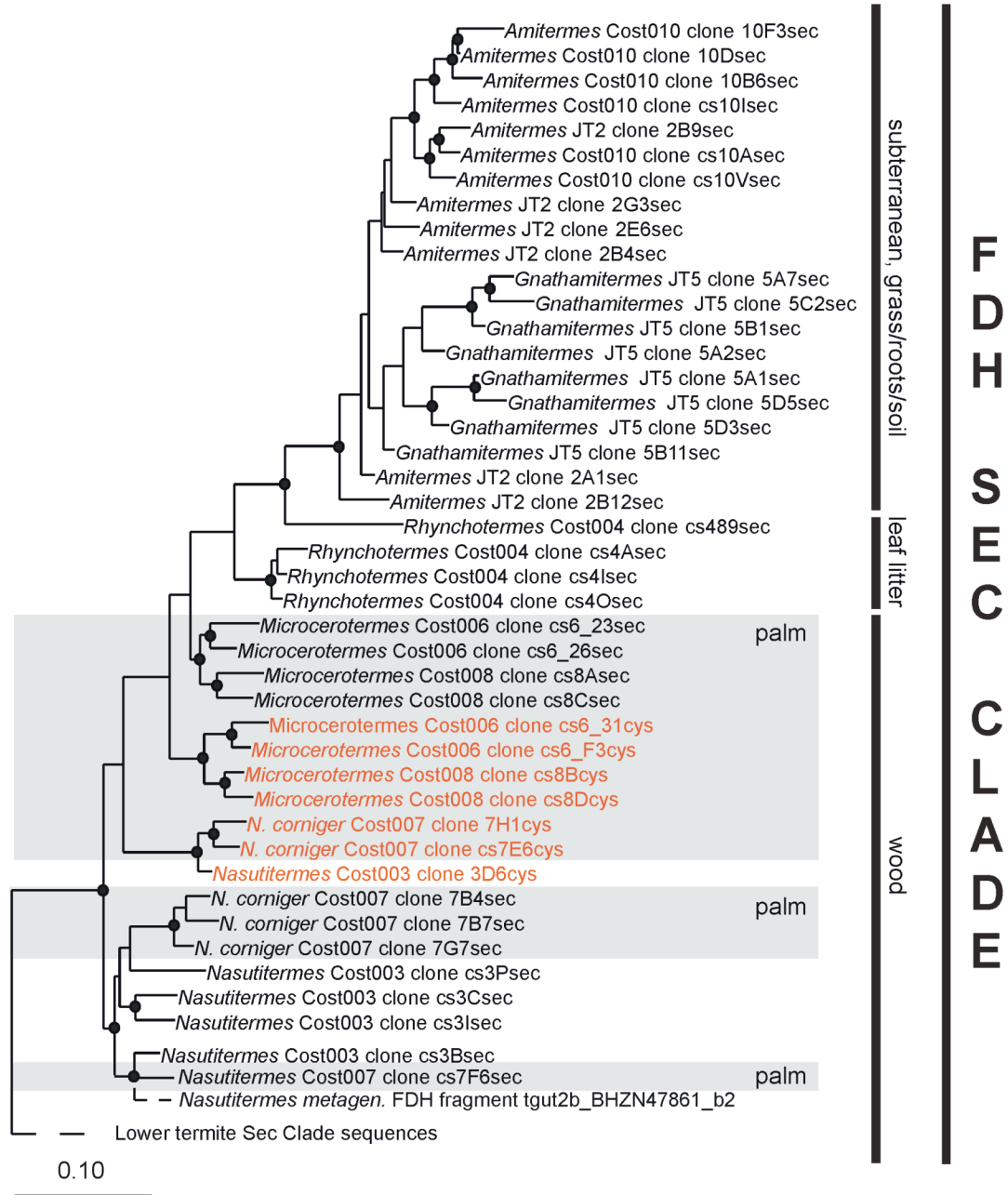
The broad distribution of higher termite spirochete clade sequences among termites from different subfamilies, coupled with the finding that they are the *only* spirochete-like FDH$_H$ types in the Gut spirochete group, indicates the higher termite spirochete clade represents an important evolutionary radiation within the FDH$_H$ gene pool of acetogenic spirochetes. We hypothesize this radiation is associated with the loss of most Sec clade

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

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

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

Amitermes Cost010 clone 10F3sec
Amitermes Cost010 clone 10Dsec
Amitermes Cost010 clone 10B6sec
Amitermes Cost010 clone cs10Isec
Amitermes JT2 clone 2B9sec
Amitermes Cost010 clone cs10Asec
Amitermes Cost010 clone cs10Vsec
Amitermes JT2 clone 2G3sec
Amitermes JT2 clone 2E6sec
Amitermes JT2 clone 2B4sec
Gnathamitermes JT5 clone 5A7sec
Gnathamitermes JT5 clone 5C2sec
Gnathamitermes JT5 clone 5B1sec
Gnathamitermes JT5 clone 5A2sec
Gnathamitermes JT5 clone 5A1sec
Gnathamitermes JT5 clone 5D5sec
Gnathamitermes JT5 clone 5D3sec
Gnathamitermes JT5 clone 5B11sec
Amitermes JT2 clone 2A1sec
Amitermes JT2 clone 2B12sec
Rhynchotermes Cost004 clone cs489sec
Rhynchotermes Cost004 clone cs4Asec
Rhynchotermes Cost004 clone cs4Isec
Rhynchotermes Cost004 clone cs4Osec
Microcerotermes Cost006 clone cs6_23sec
Microcerotermes Cost006 clone cs6_26sec
Microcerotermes Cost008 clone cs8Asec
Microcerotermes Cost008 clone cs8Csec
Microcerotermes Cost006 clone cs6_31cys
Microcerotermes Cost006 clone cs6_F3cys
Microcerotermes Cost008 clone cs8Bcys
Microcerotermes Cost008 clone cs8Dcys
N. corniger Cost007 clone 7H1cys
N. corniger Cost007 clone cs7E6cys
Nasutitermes Cost003 clone 3D6cys
N. corniger Cost007 clone 7B4sec
N. corniger Cost007 clone 7B7sec
N. corniger Cost007 clone 7G7sec
Nasutitermes Cost003 clone cs3Psec
Nasutitermes Cost003 clone cs3Csec
Nasutitermes Cost003 clone cs3Isec
Nasutitermes Cost003 clone cs3Bsec
Nasutitermes Cost007 clone cs7F6sec
Nasutitermes metagen. FDH fragment tgut2b_BHZN47861_b2
Lower termite Sec Clade sequences

subterranean, grass/roots/soil
leaf litter
palm
wood
palm
palm

F D H   S E C   C L A D E

0.10

**Multiple reinventions of selenium-independent FDH$_H$ alleles in arboreal termites**

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

Sequences from subterranean and litter-feeding termites form a novel FDH$_H$ clade

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

and 8% of JT5 inventories (Table 3.3; Figure 3.5, left panel). We were not able to infer the identity of the uncultured organisms encoding sequences in this clade, as the clade has no pure culture representatives and falls outside the Gut spirochete group. However, the presence of an amino acid indel (Figure 3.5, right panel) characteristic of the Gut spirochete group may indicate a spirochetal origin. If this origin is confirmed, these FDH_H types would function in the direction of CO_2 fixation within the context Wood-Ljungdahl pathway. No additional information could be extracted from the genomic context of FDH_H in *Clostridium difficile,* which falls basal to the *Amitermes-Gnathamitermes-Rhynchotermes* clade. However, we note that *C. difficile* possesses several Wood-Ljungdahl pathway genes.

Based on our inventory findings, we hypothesized that *Amitermes-Gnathamitermes-Rhynchotermes* FDH_H sequences represent a group of gut symbionts present only in subterranean and litter-feeding termites (Table 3.3). To identify *Amitermes-Gnathamitermes-Rhynchotermes* FDH_H sequences in other termites, we designed clade-specific primers (193F, 1045R) and screened lower and higher termite gut DNA using nested PCR methods (Figure 3.9, Appendix). Robust amplicons were detected in every subterranean and litter-feeding termite, but in no other termite species. This result implies *Amitermes-Gnathamitermes-Rhynchotermes* clade alleles are only present in subterranean and litter-feeding termites gut communities.
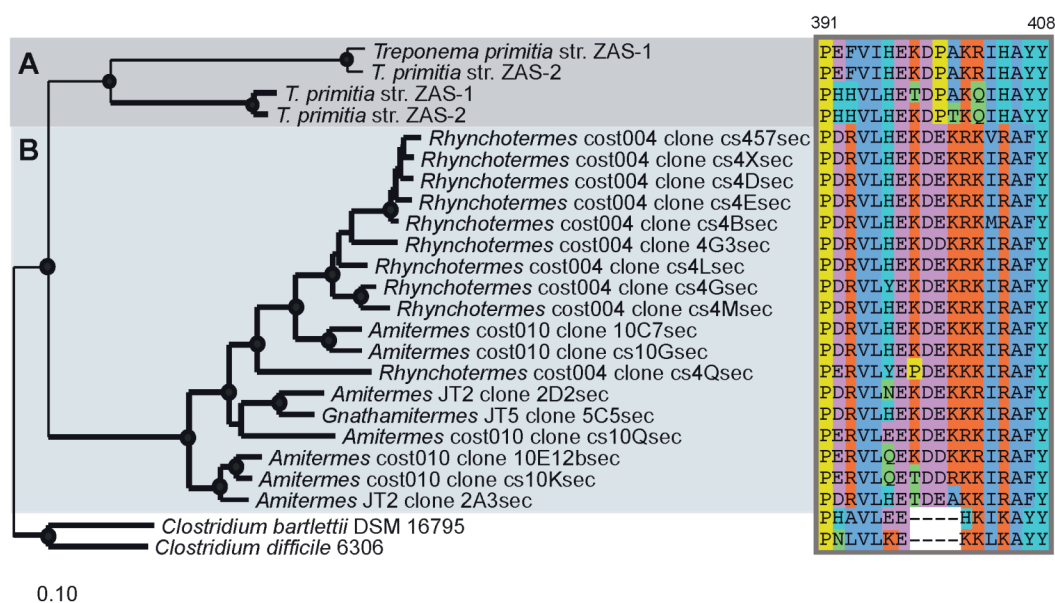
Figure 3.5. *Amitermes-Gnathamitermes-Rhynchotermes* FDH$_H$ clade (Clade B, Figure 3.2) detailed phylogeny (left panel, grey box B) and amino acid character analysis (right panel). Dark grey box A highlights the Gut spirochete group (Clade A, Figure 3.2), which is represented by *T. primitia.* Clone names containing 'sec' and branches in bold correspond to selenocysteine encoding FDH$_H$; those with 'cys' correspond to cysteine FDH$_H$. Tree was constructed with 595 aligned amino acids using the maximum likelihood algorithm Phylip PROTML. Filled circles indicate nodes were supported by PROTML, parsimony, and distance methods of analyses. Scale bar indicates 0.1 amino acid changes per alignment position. Numbers above the alignment refer to amino acid positions in the Sec FDH$_H$ of *T. primitia* str. ZAS-2.

Enteric *Proteobacteria* and unclassified FDH$_H$ alleles

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

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

Figure 3.6. Detailed phylogenies of enteric *Proteobacteria* (Clade C, Figure 3.2) and unclassified FDH$_H$ sequences (Clade D, Figure 3.2). Sequences from higher termites are highlighted in red. Clone names containing 'sec' correspond to selenocysteine encoding FDHs; those with 'cys' correspond to cysteine FDHs. Tree was constructed using methods and setting specified in the legend of Figure 3.2. Filled circles indicate nodes were supported by PROTML and parsimony methods of analyses. Scale bar indicates 0.1 amino acid changes per alignment position.
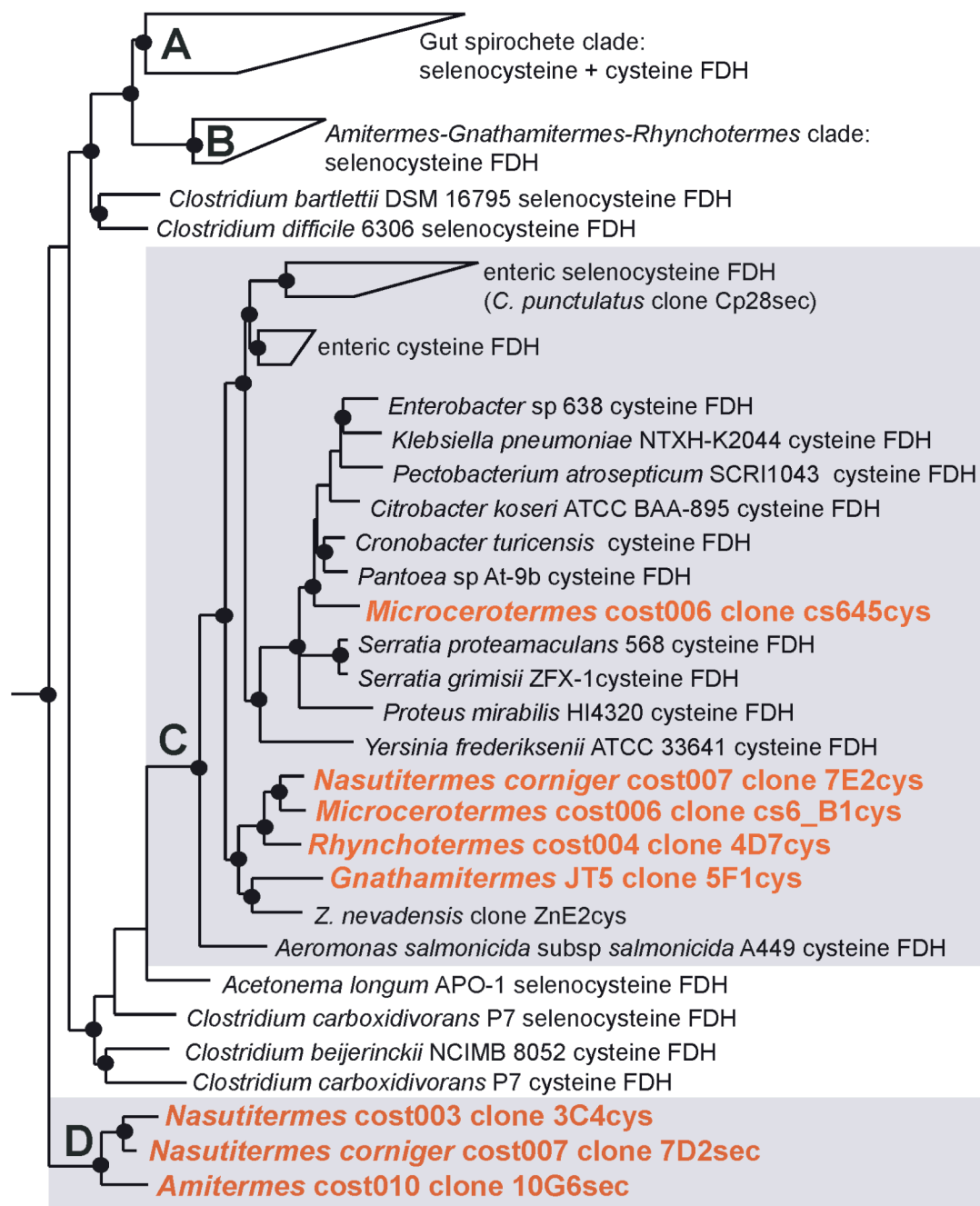
Principal component analysis reveals data cluster by higher order insect taxonomy, presence of enteric-like FDH$_H$, and degree of soil exposure.

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

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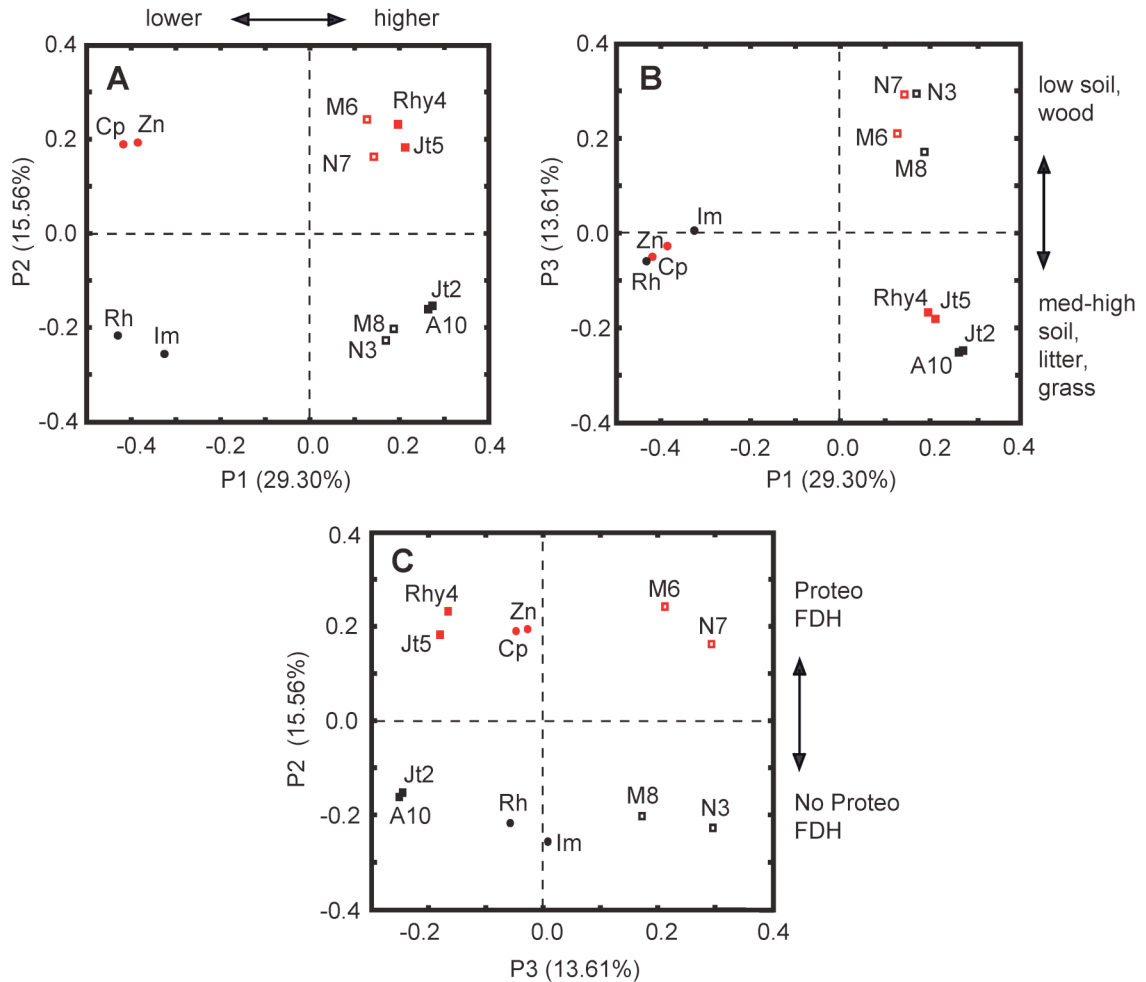
Discussion

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

We suggest that compartment specific sampling efforts and/or methodological artifacts may be the cause of low FDH recovery in the metagenome. Both of these possibilities have merit. With respect to compartment specific sampling, studies of methanogenesis in the soil-feeding termite *Cubitermes* demonstrate cross-epithelial H_2 transfer between gut compartments and imply formate transfer between gut compartments is also possible (50). This suggests FDH genes may be present outside the gut chamber sampled for metagenomic analysis. If this is true, acetogens inhabiting the largest gut compartment would be heterotrophs, as they would rely on outside sources of formate for acetate production. Gut compartment specific inventories of FDH should clarify the issue. With respect to methodology, one study of spirochete acetogenesis genes suggests that spirochete DNA is difficult to clone (32). Thus DNA toxicity issues may also contribute to low metagenome FDH gene recovery.

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

FDH_H phylogeny and diet

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

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

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

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

FDH$_H$ phylogeny and acetogenic spirochete evolution

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

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

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

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

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

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

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

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

Appendix

Figure 3.8. Targeted PCR assays on termite and roach gut DNA using lower termite spirochete group Cys clade *fdhF* specific primers.
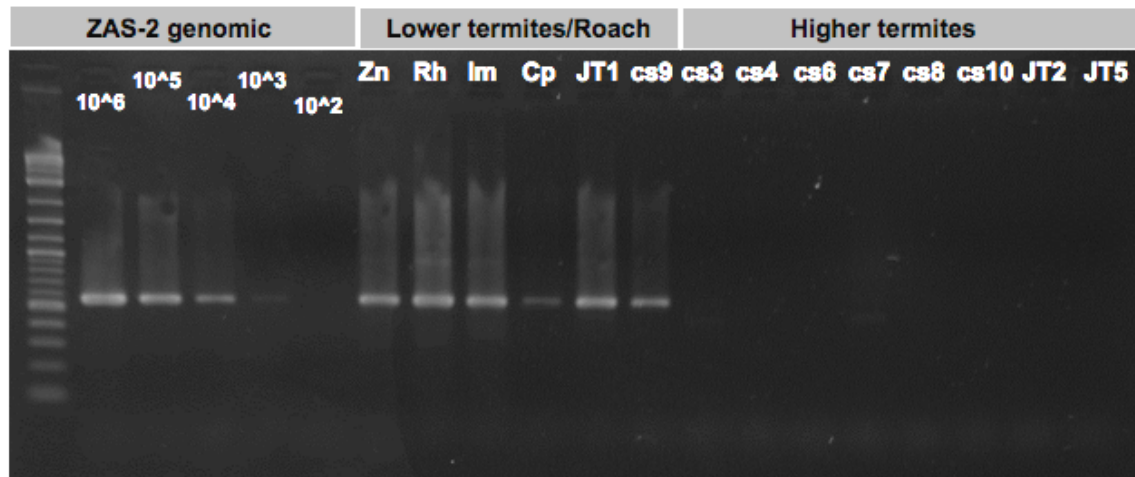
Figure 3.9. Targeted PCR assays using universal *fdhF* primers followed by *Amitermes-Gnathamitermes-Rhychotermes* clade specific primers on gut templates.

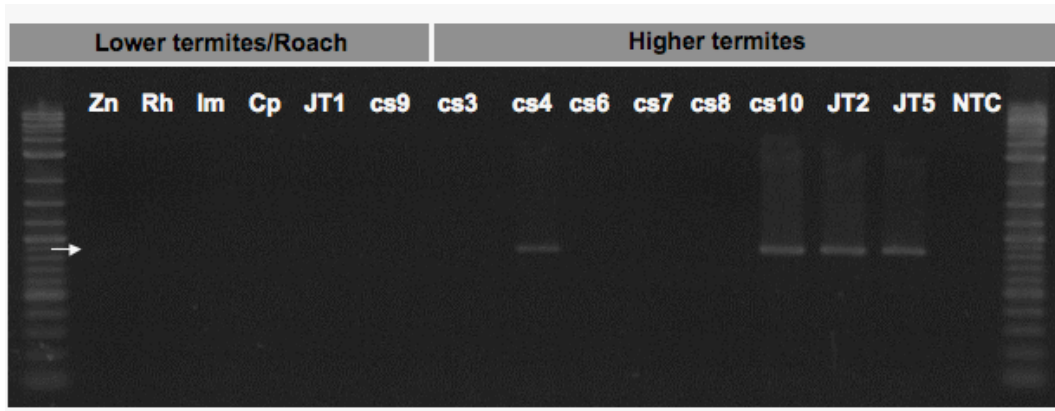**Table 3.4.** PCR conditions for clone library construction.

Table 3.5. Phylotype distribution in each library.

Table 3.6. Sequences used in phylogenetic anlaysis.

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

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

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

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

Table 3.5. Phylotype distribution in each library.

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

	7B7sec	13.3	
	7D2sec	6.7	
	7E2cys	6.7	
	7H1cys	6.7	
	7B4sec	3.3	
	7G7sec	3.3	
	total clones	30	

Microcerotermes sp. Cost008

		8L1	
	cs8Csec	42.9	
	cs8Asec	39.3	
	cs8Bcys	10.7	
	cs8Dcys	7.1	
	total clones	84	

Amitermes sp. Cost010

	10L1	**10L2**
cs10Dsec	62.8	30.4
cs10Asec	21.8	21.7
cs10Gsec	7.7	8.7
cs10Isec	2.6	4.3
cs10Ksec	2.6	0.0
cs10Qsec	1.3	0.0
cs10Vsec	1.3	0.0
10B6sec	0	13.0
10C7sec	0	4.3
10G6sec	0	8.7
10E12bsec	0	4.3
10F3sec	0	4.3
total clones	78	23

Amitermes sp. JT2

	Jt2L1
2A1sec	68.3
2D2sec	6.9
2B4sec	5.9
2G3sec	8.9
2B9sec	6.9
2E6sec	1.0
2B12sec	1.0
2A3sec	1.0
total clones	101

Gnathamitermes sp. JT5

	Jt5L1
5F1cys	17.9

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

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

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

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

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

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

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

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

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

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

References

1. **Bauer, S., A. Tholen, J. Overmann, and A. Brune.** 2000. Characterization of abundance and diversity of lactic acid bacteria in the hindgut of wood- and soil-feeding termites by molecular and culture-dependent techniques. Arch Microbiol **173:**126-37.

2. **Bignell, D. E.** 2000. Introduction to symbiosis, p. 189–208. *In* T. Abe, D. E. Bignell, and M. Higashi (ed.), Termites: Evolution, Sociality, Symbioses, Ecology. Kluwer Academic Publishers, Dordrecht, The Netherlands.

3. **Bignell, D. E.** 2006. Termites as soil engineers and soil processors, p. 183-220. *In* H. König and A. Varma (ed.), Intestinal microorganisms of termites and other invertebrates. Springer, Heidelberg, Germany.

4. **Bignell, D. E., and J. M. Anderson.** 1980. Determination of pH and oxygen status in the guts of lower and higher termites. J Insect Physiol **26:**183-188.

5. **Bignell, D. E., and P. Eggleton.** 1995. On the elevated intestinal pH of higher termites (*Isoptera, Termitidae*). Insectes Sociaux **42:**57-69.

6. **Bignell, D. E., and P. Eggleton.** 2000. Termites in Ecosystems, p. 363-387. *In* T. Abe, D. E. Bignell, and M. Higashi (ed.), Termites: Evolution, Sociality, Symbioses, Ecology. Springer, Dordrecht, The Netherlands.

7. **Brauman, A., D. E. Bignell, and I. Tayasu.** 2000. Soil-feeding termites: biology, microbial associations and digestive mechanisms, p. 233-259. *In* T. Abe, D. E. Bignell, and M. Higashi (ed.), Termites: Evolution, Sociality, Symbioses, Ecology. Kluwer Academic Publishers,, Dordrecht, The Netherlands.

8. **Brauman, A., J. Dore, P. Eggleton, D. Bignell, J. A. Breznak, and M. D. Kane.** 2001. Molecular phylogenetic profiling of prokaryotic communities in guts of termites with different feeding habits. FEMS Microbiol Ecol **35:**27-36.

9. **Brauman, A., M. D. Kane, M. Labat, and J. A. Breznak.** 1992. Genesis of acetate and methane by gut bacteria of nutritionally diverse termites. Science **257:**1384-1387.

10. **Breznak, J. A.** 2000. Ecology of prokaryotic microbes in the guts of wood-and litter-feeding termites, p. 209-231. *In* T. Abe, D. E. Bignell, and M. Higashi (ed.), Termites: Evolution, Sociality, Symbiosis, Ecology Kluwer Academic Publishers Dordrecht, The Netherlands.

11. **Breznak, J. A., and J. M. Switzer.** 1986. Acetate synthesis from H2 plus CO2 by termite gut microbes. Appl Environ Microbiol **52:**623–630.

12. **Brugerolle, G., and R. Radek.** 2006. Symbiotic Protozoa of Termites, p. 244-269. *In* H. König and A. Varma (ed.), Intestinal microorganisms of termites and other invertebrates. Springer, Heidelberg, Germany.

13. **Brune, A., and M. Kühl.** 1996. pH profiles of the extremely alkaline hindguts of soil-feeding termites (*Isoptera: Termitidae)* determined with microelectrodes. J Insect Physiol **42:**1121–1127.

14. **Colwell, R. K.** 2009, posting date. EstimateS: statistical estimation of species richness and shared species from samples. version 8.2.0. [Online.]

15. **Drake, H. L., K. Kü sel, and C. Matthies.** 2006. Acetogenic prokaryotes, p. 354-420. *In* M. Dworkin, S. Falkow, E. Rosenber, K. H. Schleifer, and E. Stackebrandt (ed.), The Prokaryotes, 3 ed. Springer.

16. **Eggleton, P.** 2006. The Termite Gut Habitat: Its Evolution and Co-Evolution, p. 373-404. *In* H. König and A. Varma (ed.), Intestinal microorganisms of termites and other invertebrates. Springer, Heidelberg, Germany.

17. **Eggleton, P., D. E. Bignell, W. A. Sands, N. A. Mawdsley, J. H. Lawton, T. G. Wood, and N. C. Bignell.** 1996. The diversity, abundance and biomass of termites under differing levels of disturbance in the Mbalmayo Forest Reserve, southern Cameroon. Philosophical Transactions - Royal Society Biological sciences **351:**51-68.

18. **Engel, M. S., D. A. Grimaldi, and K. Krishna.** 2009. Termites (*Isoptera*): Their Phylogeny, Classification, and Rise to Ecological Dominance. American Museum Novitates **3650:**1-27.

19. **Enquist, C. A. F.** 2002. Predicted regional impacts of climate change on the geographical distribution and diversity of tropical forests in Costa Rica. Journal of Biogeography **29:**519-534.

20. **Felsenstein, J.** 1989. PHYLIP - Phylogeny inference package (version 3.2). Cladistics **5:**164 - 166.

21. **Ferry, J. G.** 1990. Formate Dehydrogenase: Microbiology, Biochemistry and Genetics, p. 117-141. *In* G. A. Codd, L. Dijkhuizen, and F. R. Tabita (ed.), Autotrophic Microbiology and One Carbon Metabolism. Kluwer Academic Publishers, Dordrecht.

22. **Grimaldi, D., and M. S. Engel.** 2005. Evolution of the insects. Cambridge University Press, New York, NY.

23. **Hernes, P. J., and J. I. Hedges.** 2004. Tannin signatures of barks, needles, leaves, cones, and wood at the molecular level. Geochim Cosmochim Acta **68:**1293-1307.

24. **Holdridge, L. R., W. C. Grenke, W. H. Hatheway, and J. A. Tosi.** 1971. Forest environments in tropical life zones. A pilot study. Pergamon Press, New York.

25. **Inoue, T., O. Kitade, T. Yoshimura, and I. Yamaoka.** 2000. Symbiotic association with protists, p. 275-288. *In* T. Abe, D. E. Bignell, and M. Higashi (ed.), Termites: Evolution, Sociality, Symbioses. Kluwer Academic Publishers, Dordrecht, The Netherlands.

26. **Kambhampati, S., and P. Eggleton.** 2000. Taxonomy and phylogeny of termites, p. 1-24. *In* T. Abe, D. E. Bignell, and M. Higashi (ed.), Termites: Evolution, Sociality, Symbioses, Ecology. Kluwer Academic Publishers, Dordrecht, The Netherlands.

27. **Leadbetter, J. R., T. M. Schmidt, J. R. Graber, and J. A. Breznak.** 1999. Acetogenesis from H2 plus CO2 by spirochetes from termite guts. Science **283:**686-689.

28. **Lozupone, C., and R. Knight.** 2005. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol **71:**8228-8235.

29. **Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Forster, I. Brettske, S. Gerber, A. W. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. Konig, T. Liss, R. Lussmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N. Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode, and K.-H. Schleifer.**

2004. ARB: a software environment for sequence data. Nucl Acids Res **32:**1363-1371.

30. **Lugo, A. E., S. L. Brown, R. Dodson, T. S. Smith, and H. H. Shugart.** 1999. The Holdridge life zones of the conterminous United States in relation to ecosystem mapping. Journal of Biogeography **26:**1025-1038.

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

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

33. **Miura, T., K. Maekawa, O. Kitade, T. Abe, and T. Matsumoto.** 1998. Phylogenetic relationships among subfamilies in higher termites (*Isoptera: Termitidae*) based on mitochondrial COII gene sequences. Annals of the Entomological Society of America **91:**515-521(

34. **Miura, T., Y. Roisin, and T. Matsumoto.** 2000. Molecular phylogeny and biogeography of the Nasute termite genus *Nasutitermes* (Isoptera: *Termitidae*) in the Pacific Tropics. Molecular Phylogenetics and Evolution **17:**1-10.

35. **Miyata, R., N. Noda, H. Tamaki, K. Kinjyo, H. Aoyagi, H. Uchiyama, and H. Tanaka.** 2007. Influence of feed components on symbiotic bacterial community structure in the gut of the wood-feeding higher termite *Nasutitermes takasagoensis*. Biosci Biotechnol Biochem **71:**1244-1251.

36. **Noda, S., M. Ohkuma, A. Yamada, Y. Hongoh, and T. Kudo.** 2003. Phylogenetic position and in situ identification of ectosymbiotic spirochetes on protists in the termite gut. Appl Environ Microbiol **69:**625–633

37. **Noirot, C.** 1992. From wood- to humus-feeding: an important trend in termite evolution. , p. 107–119. *In* J. Billen (ed.), Biology and Evolution of Social Insects. University Press, Leuven, Belgium.

38. **Nriagu, J. O.** 1989. Global cycling of selenium, p. 327-340. *In* M. Ihnat (ed.), Occurrence and Distribution of Selenium. CRC Press, Boca Baton, Florida.

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

40. **Ohkuma, M., Y. Hongoh, and T. Kudo.** 2006. Diversity and Molecular Analyses of Yet-Uncultivated Microorganisms. *In* H. König and A. Varma (ed.), Intestinal microorganisms of termites and other invertebrates. Springer, Heidelberg, Germany.

41. **Paster, B. J., F. E. Dewhirst, S. M. Cooke, V. Fussing, L. K. Poulsen, and J. A. Breznak.** 1996. Phylogeny of not-yet-cultured spirochetes from termite guts. Appl Environ Microbiol **62:**347-52.

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

43. **Pester, M., and A. Brune.** 2007. Hydrogen is the central free intermediate during lignocellulose degradation by termite gut symbionts. ISME J **1:**551-65.

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

45. **Schmitt-Wagner, D., and A. Brune.** 1999. Hydrogen profiles and localization of methanogenic activities in the highly compartmentalized hindgut of soil-feeding higher termites (*Cubitermes* spp.). Appl Environ Microbiol **65:**4490-4496.

46. **Schmitt-Wagner, D., M. W. Friedrich, B. Wagner, and A. Brune.** 2003. Axial dynamics, stability, and interspecies similarity of bacterial community structure in the highly compartmentalized gut of soil-feeding termites (*Cubitermes* spp.). Appl Environ Microbiol **69:**6018–6024

47. **Schmitt-Wagner, D., M. W. Friedrich, B. Wagner, and A. Brune.** 2003. Phylogenetic diversity, abundance, and axial distribution of bacteria in the intestinal tract of two soil-feeding termites (*Cubitermes* spp.). Appl Environ Microbiol **69:**6007-17.

48. **Stamatakis, A. P., T. Ludwig, and H. Meier.** 2004. The AxML program family for maximum likelihood-based phylogenetic tree inference. Concurrency and Computation: Practice and Experience **16:**975-988.

49. **Strassert, J. F., M. S. Desai, A. Brune, and R. Radek.** 2009. The true diversity of Devescovinid flagellates in the termite *Incisitermes marginipennis*. Protist.

50. **Tholen, A., and A. Brune.** 1999. Localization and *in situ* activities of homoacetogenic bacteria in the highly compartmentalized hindgut of soil-feeding higher termites (*Cubitermes* spp.). Appl Environ Microbiol **65:**4497-505.

51. **Thongaram, T., Y. Hongoh, S. Kosono, M. Ohkuma, S. Trakulnaleamsai, N. Noparatnaraporn, and T. Kudo.** 2005. Comparison of bacterial communities in the alkaline gut segment among various species of higher termites. Extremophiles **9:**229-238.

52. **Thorne, B. L., D. Grimaldi, and K. Krishna.** 2000. Early fossil history of termites, p. 77-93. *In* T. Abe, D. E. Bignell, and M. Higashi (ed.), Termites: Evolution, Sociality, Symbioses, Ecology. Springer, Dordrecht, The Netherlands.

53. **Tokuda, G. I. Y., and H. Noda.** 2000. Localization of symbiotic clostridia in the mixed segment of the termite *Nasutitermes takasagoensis* (Shiraki). Appl Environ Microbiol **66:**2199–2207.

54. **Warnecke, F., P. Luginbühl, N. Ivanova, M. Ghassemian, T. Richardson, J. Stege, M. Cayouette, A. Mchardy, G. Djordjevic, N. Aboushadi, R. Sorek, S. Tringe, M. Podar, H. Martin, V. Kunin, D. Dalevi, J. Madejska, E. Kirton, D. Platt, E. Szeto, A. Salamov, K. Barry, N. Mikhailova, N. Kyrpides, E. Matson, E. Ottesen, X. Zhang, M. Hernández, C. Murillo, L. Acosta, I. Rigoutsos, G. Tamayo, B. Green, C. Chang, E. Rubin, E. Mathur, D. Robertson, P. Hugenholtz, and J. Leadbetter.** 2007. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. Nature **450:**560-565.