

# RNA-Seq and microfluidic digital PCR identification of transcriptionally active spirochetes in termite gut microbial communities

## Abstract

CO<sub>2</sub>-reductive acetogenesis in termite hindguts is a bacterial process with significant impact on the nutrition of wood-feeding termites. Acetogenic spirochetes have been identified as key mediators of acetogenesis. Here, we use high-throughput, short transcript sequencing (RNA-Seq) and microfluidic, multiplex digital PCR to identify uncultured termite gut spirochetes transcribing genes for hydrogenase-linked formate dehydrogenase (FDH<sub>H</sub>) enzymes, which are required for acetogenic metabolism in the spirochete, *Treponema primitia*. To assess FDH<sub>H</sub> gene (*fdhF*) transcription within the gut community of a wood-feeding termite, we sequenced ca. 28,000,000 short transcript reads of gut microbial community RNA using Illumina Solexa technology. RNA-Seq results indicate that *fdhF* transcription in the gut is dominated by two *fdhF* genotypes: ZnD2sec and Zn2cys. This finding was independently corroborated with cDNA inventory and qRT-PCR transcription measurements. We, therefore, propose that RNA-Seq mapping of microbial community transcripts is specific and quantitative. Following transcriptional assessments, we performed microfluidic, multiplex digital PCR on single termite gut bacterial cells to discover the identity of uncultured bacteria encoding *fdhF*

genotypes ZnD2sec and Zn2cys. We identified the specific 16S rRNA gene ribotype of the bacterium encoding Zn2cys *fdhF* and report that the bacterium is a spirochete. Phylogenetic analysis reveals that this uncultured spirochete, like *T. primitia*, possesses genes for acetogenic metabolism – formyl-tetra-hydrofolate synthetase and both selenocysteine and cysteine variants of formate dehydrogenase. Microfluidic results also imply a spirochetal origin for ZnD2sec *fdhF*, but further gene pair associations are required for verification. Taken together, the results (i) show novel transcriptomic and single cell approaches can be successfully combined to study active microbes in natural microbial communities, (ii) underscore the continued relevance of leveraging investigations of uncultured bacteria with the results from pure culture studies, and (iii) imply that termite gut acetogenesis is largely mediated by spirochetes which represent only a small portion of total acetogenic spirochete diversity.

## Introduction

The structure and function of natural microbial communities are primary targets of study for microbial ecologists. Molecular profiling using 16S rRNA gene inventory (14, 27, 40) and metagenomic (2, 12, 40) techniques have proved incredibly useful for elucidating community structure, particularly in environments that have yielded few culturable microbes. Similar methods have been utilized to outline a suite of potential functions encoded in community DNA (2, 40, 41). These efforts have led to surveys of actual community function at the level of transcription [e.g., (9, 16, 21, 31, 39, 46)]. Quantitative reverse transcriptase PCR (qRT-PCR), microarray, cDNA inventory, and mRNA-based terminal-restriction fragment length polymorphism (T-RFLP) techniques have been commonly employed to monitor community transcription (7, 28, 29, 38). However, all these environmental transcriptomic methods suffer drawbacks related to primer/probe binding specificity and/or PCR biases (1, 8, 20, 30, 43). Furthermore, the extent of primer/probe cross-binding can not be easily assessed, as data sets yield accurate information on either transcript abundance or transcript sequence.

The recent advent of high-throughput Illumina/Solexa Genome Analyzer sequencing technology (26) has enabled researchers to obtain massive amounts of short DNA sequences (37-75 base pairs) from their sample quickly, with no primer or cloning bias. Unlike previous transcriptomic methods, this technology and similar high-throughput sequencing methods (e.g., Roche 454 GS20 pyrosequencing) yield *both* transcript sequence (i.e., verification) and transcript abundance in a single data set. However, the difficulty of sequence fragment mapping makes data interpretation a major challenge.

Previous knowledge of gene sequence, which can serve as a “scaffold” for fragment mapping, is usually required. Recent studies have utilized Solexa technology to deep sequence transcriptomes (RNA-Seq) for eukaryotes (13, 23, 44) or defined cultures of prokaryotes (33, 45), but studies of natural microbial community transcription have so far only utilized 454 pyrosequencing technology (9, 42).

Here, we demonstrate that high-throughput sequencing of transcripts via Illumina-Solexa RNA-Seq can be leveraged by traditional DNA and cDNA library data (used as scaffolding for fragment assembly and interpretation) to rapidly assess environmental functional gene transcription in microbial communities. This approach differs from mRNA-T-RLFP, as the entire length of scaffold sequence is informative and, more importantly, the sequence fragment serves as both signal (abundance of particular fragment) and verification (sequences can be mapped to library scaffolds). While this approach is still scaffold-limited, we expect that a combination of RNA-Seq and inventory data can serve as a tool for microbial ecologists interested in assessing transcription in environments with high allelic diversity.

With some knowledge of community structure and function in hand, microbial ecologists then face the challenge of linking community members (structure) with the respective activities they carry out (function). This is straight-forward when pure culture isolates representing different functional groups are available, but in the majority of cases, researchers find themselves confronted with a diversity of 16S rRNA and functional gene sequences from uncultured organisms which can not be related to one another based on

phylogenetic inference. Ottesen *et al.* (26) have recently shown that microfluidic multiplex digital PCR assays on single cells can resolve such relationships in natural microbial communities.

In this study, we combine RNA-Seq and single cell techniques to investigate functionally important uncultured bacteria in the symbiotic microbial community of a wood-feeding termite. All phylogenetically “lower” wood-feeding termites harbor a species-rich hindgut community of symbiotic protozoa and bacteria that efficiently degrades lignocellulose into acetate, the major carbon and energy source of their insect host (3, 4, 6). CO<sub>2</sub>-reducing acetogens play an important role in this nutritional mutualism: these bacteria consume the majority of the lignocellulose fermentation byproducts H<sub>2</sub> and CO<sub>2</sub>, generating up to a third of gut acetate (5, 6). Inventory surveys of key acetogenesis genes (28, 35) and pure culture studies (17) imply that spirochetes of the bacterial phylum *Treponema* are responsible for acetogenesis in wood-feeding termites. Moreover, phylotype abundance for the functional gene encoding the hydrogenase-linked acetogenesis enzyme formate dehydrogenase (FDH<sub>H</sub>, *fdhF*) in the wood-feeding termite *Zootermopsis nevadensis* suggest the acetogenic spirochete population comprises as many as 7–15 different types of Treponemes (Chapter 2). This estimate is roughly consistent with phylotype abundance (3–11) observed for another key acetogenesis enzyme (formyl-tetrahydrofolate synthetase, FTHFS) in *Zootermopsis* and other wood-feeding termites (28, 35).

Sequence and phylogenetic analyses revealed *fdhF* phlotypes in *Z. nevadensis* could be classified into two clades: one comprised of sequences encoding selenocysteine (Sec) at the FDH<sub>H</sub> active site (*fdhF*<sub>Sec</sub>) and the other of sequences that encode cysteine (Cys) at the homologous position (*fdhF*<sub>Cys</sub>). Studies with the pure culture acetogenic spirochete *Treponema primitia* str. ZAS-2 indicated *fdhF*<sub>Sec</sub> and *fdhF*<sub>Cys</sub> could be present in the same organism and that *both* are transcriptionally controlled by the trace element selenium (22). It is unknown whether various *fdhF*<sub>Sec</sub> and *fdhF*<sub>Cys</sub> variants in *Z. nevadensis* belong to the same spirochete or are differentially transcribed. Here, we use novel sequencing and single-cell techniques (26) to (i) assess transcription of hydrogenase-linked FDH genes within the species-rich symbiotic gut microbial community of *Z. nevadensis* and (ii) determine the 16S rRNA sequence identity of uncultured termite gut bacteria encoding transcriptionally active FDH<sub>H</sub> genes.

## Materials and Methods

### Termite collection

Worker specimens of the dampwood termite *Zootermopsis nevadensis* were collected in the San Gabriel Mountains of California. Some were maintained in plastic boxes at 95% humidity in foil-covered glass aquaria in the laboratory. The entire gut tracts of ~5 worker termites were preserved in 50 – 200 µl of RNA stabilization buffer (RNA Protect Bacteria Reagent, QIAGEN, Valencia, CA) at -80°C until nucleic acid extraction for RNA-Seq and inventory experiments.

**Termite gut nucleic acid extraction**

100  $\mu$ l of TE buffer (1 mM Tris-HCl, 0.1 mM EDTA, pH 8.0) was added to an ice-thawed tube containing worker guts. Guts were then homogenized (3 x 30 sec) by bead beating with sterile zirconia/silica beads (0.1 mm) using a MiniBeadbeater-8 (BioSpec Products, Inc., Bartlesville, OK). Lysozyme (Sigma, St. Louis, MO) was added to the homogenate (1 mg); this mixture was incubated at room temperature for 15 min. DNA and total RNA were extracted from 150  $\mu$ l aliquots of gut homogenate using a DNeasy Tissue Kit (QIAGEN) and RNeasy Kit (QIAGEN), respectively. Purification details for total RNA can be found in Chapter 2. Total RNA was used for Illumina RNA-Seq and cDNA library experiments.

**RNA-Seq: Processing and sequencing**

Samples were prepared using the Illumina protocol for RNA-Seq sample preparation V2 (<https://icom.illumina.com>). Briefly, total RNA (at least 5  $\mu$ g) was fragmented using an Ambion RNA fragmentation kit and then converted to single-strand cDNA using an Invitrogen SuperScript II kit (Invitrogen, Carlsbad, CA). Second Strand Buffer (500 mM Tris-HCl, pH 7.8, 50 mM MgCl<sub>2</sub>, 10mM DTT), dNTP (0.3 mM), RNaseH (2 U  $\cdot$   $\mu$ l<sup>-1</sup>, Invitrogen) and DNA polymerase I (Invitrogen) were then added to the first-strand reaction to synthesize second strand cDNA (16°C, 2.5 hours). Fragmented second strand cDNA samples were sequenced as 37-mers using the standard Solexa (Illumina) protocol and pipeline at Caltech's Sequencing Core Facility (Pasadena, CA).

### **RNA-Seq Data Analysis**

Illumina raw data, obtained using GERALD (a software package within the Illumina pipeline), was aligned to a FASTA file containing FDH gene sequences (Table 4.4, Appendix) with the Maq short read aligning program (18). Samples were analyzed for perfect matches only. Signal intensities were visualized graphically by converting Maq aligned reads into a .BAR file using the Cisgenome software (15) and viewed on the Cisgenome browser and on the IGB genomic browser (<http://www.affymetrix.com>).

### **cDNA inventories**

Separate cDNA libraries for *fdhF<sub>Sec</sub>* and *fdhF<sub>Cys</sub>* gene variants were generated from gut cDNA. A forward primer for *fdhF<sub>Sec</sub>* (Sec427F, Table 4.1) that targets the selenocysteine FDH<sub>H</sub> active site was designed manually. Sec427F was used with 1045R (Chapter 3, Table 4.1) to amplify *fdhF<sub>Sec</sub>* from gut cDNA. The *fdhF<sub>Cys</sub>* cDNA library was constructed with primers Cys499F1b and 1045R (Chapter 3, Table 4.1). PCR reactions contained 200 nM forward primer (Sec427F or Cys499F1b), 200 nM 1045R, 1X FAILSAFE Premix D (EPICENTRE, Madison, WI), 0.07 U ·  $\mu\text{l}^{-1}$  of EXPAND High Fidelity polymerase (Roche Applied Science, Indianapolis, IN), and 0.5 ng ·  $\mu\text{l}^{-1}$  gut cDNA. Thermocycling conditions on a Mastercycler Model 5331 thermocycler (Eppendorf, Westbury, NY) were 2 min at 95°C, 30 cycles of (95°C for 30 sec, 60°C for 1 min, 72°C for 1 min), followed by 10 min at 72°C. Amplicon size was checked on 1.5% agarose gels (Invitrogen) and the products were TOPO-TA cloned (Invitrogen). Plasmids were extracted (QIAprep Spin Miniprep Kit, QIAGEN) from 48 randomly chosen clones and sequenced (Laragen Inc., Los Angeles, CA).

### Quantitative PCR

Quantitative RT-PCR for select FDH genotypes (ZnD2sec, ZnB5sec, *T. primitia fdhF<sub>Sec</sub>*) was performed on termite gut cDNA and DNA. Quantitative PCR primers for these genotypes were: ZnD2sec (ZnO-1636F, 5'– ACT ATG ACC GGC AAT TGT CGC CTG TT –3'; ZnO-1729R, 5'– TCA GAC CCA TAT CAC GGC AAA GTT –3'), ZnB5sec (ZnB5-1636F, 5'– ACG ATG ACG GGC AAC TGC CGG ATG TT –3'; ZnB5-1729R, 5'– TAT GCC GAG AGC ATT GGC ATC TT –3'), and *T. primitia fdhF<sub>Sec</sub>* (ZAS-1636F, 5'– ACC ATG ACC GGT AAC TGC CGG ACC CT –3'; ZAS-1729R, 5'– TTA TAC CGA GCT TTT CCG CAT CCC –3'). Primers were designed with Primer3 software (34) and amplify the same region in *fdhF* genes to avoid primer site biases. Standard curves (10-fold dilutions ranging from  $10^9$  –  $10^6$  copies/reaction) were generated from TOPO-TA plasmid templates containing the relevant inserts. QPCR reactions (20  $\mu$ l) contained iQ SYBR Green Supermix (Bio-Rad laboratories, Irvine, CA), 500 nM forward primer, 500 nM reverse primer, 5 ng cDNA or 10 ng DNA. All reactions were run in duplicate. Thermocycling conditions on a Bio-Rad DNAEngine thermocycler (Chromo4 real time detector) were: 3 min at 95°C, followed by 44 cycles of 95°C for 15 sec, and 60°C for 30 sec.

### Microfluidic multiplex digital PCR

For each microfluidic chip experiment, the entire gut tract of one worker termite was extracted and suspended in 250  $\mu$ l Synthetic Gut Fluid solution (25) containing  $0.5 \mu\text{g} \cdot \text{mL}^{-1}$  Dnase-free RNase (Roche Applied Science). Cells were released from the gut tract by aspirating the sample 3 – 5 times with a sterile 200  $\mu$ l pipet tip. Large particles were

allowed to sediment for ~5 sec. Cell dilutions ( $10^{-5}$  –  $2.5 \times 10^{-6}$  range) were added (1:20 v/v) to PCR reactions.

FDH<sub>H</sub> genes in spirochetes were surveyed using multiplex digital PCR. PCR reactions (20  $\mu$ l) contained iQ Multiple Powermix (Bio-Rad Laboratories, Discontinued Cat. No. 170-8848), 0.1% Tween-20, and 75 – 175 nM ROX standard. Final reaction concentrations of primer and probes (Table 4.1) were 100 – 400 nM. Specific concentrations for each chip experiment are described in Table 4.5 (Appendix). 16S rRNA primers and a general bacterial 16S rRNA probe (1389Prb) were designed by Ottesen *et al.* (26). A new 16S rRNA probe specific for spirochetes (1409RaPrb) was designed based on 1409Ra, a spirochete-specific primer (26). Functional gene primers (Cys499F1b, 1045R) for formate dehydrogenase genes have been described in Chapter 3. Sec427F and Cys538F primers were designed to target *fdhF*<sub>Sec</sub> and *fdhF*<sub>Cys</sub> gene variants, respectively. Sec427F targets all *fdhF*<sub>Sec</sub> genes, whereas Cys538F has a more limited target range for *fdhF*<sub>Cys</sub>. PCR reactions were loaded on microfluidic chips (Biomark 12.765 Digital Array series) purchased from Fluidigm Corporation (San Francisco, CA). Microfluidic chip thermocycling conditions were: 2 min at 95°C, 45 cycles of (95°C for 15 sec, 60°C for 1 min, 72°C for 1 min), followed by 10 min at 72°C.

Samples were retrieved based on amplification of spirochete DNA, accomplished using spirochete specific-primers and a general bacterial 16S rRNA probe, or general bacterial 16S rRNA primers and a spirochete-specific 16S rRNA probe (Table 4.2). Fluorescence above background for amplification-positive wells was typically detected  $\leq$  cycle 35.

Total bacterial concentration in panels sampled for retrieval was inferred from the total number of positive 16S rRNA gene amplifications observed in a separate panel loaded with template at the same dilution and general bacterial 16S rRNA primers/probes. Only panels that corresponded to template dilutions resulting in < 250 all bacteria hits (~1/3 of all chambers) were sampled for retrieval. Samples were manually retrieved into 10 µl TE from chip chambers using a dissecting microscope and 30 gauge needles (Becton, Dickinson, and Company, Franklin Lakes, NJ) as described by Ottensen *et al.* (26).

Chip samples were screened for 16S rRNA and *fdhF* gene products via simplex PCR with microfluidic chip primers on a Mastercycler Model 5331 thermocycler (Eppendorf, Westbury, NY) and agarose gel electrophoresis (1.5%, Invitrogen). PCR reactions (50 µl) contained iQ Multiple Powermix (Bio-Rad Laboratories), 200 – 300 nM of each primer, and 2.5 µl of template. Benchtop thermocycling conditions were 2 min at 95°C, 30 or 35 cycles of (95°C for 15 sec, 60°C for 1 min, 72°C for 1 min), followed by 10 min at 72°C. Products from samples that yielded both 16S rRNA and *fdhF* amplicons were PCR purified (QIAquick PCR purification, QIAGEN). 16S rRNA PCR products were cloned in TOPO-TA vectors (TOPO-TA cloning kit, Invitrogen) for low-yield PCR purifications; plasmids from 8 randomly chosen clones were purified (QIAprep Spin Miniprep, QIAGEN). 16S rRNA PCR products and plasmids were sequenced with the internal primers 533F and 1100R (26); *fdhF* products were sequenced with microfluidic chip primers. All sequencing reactions were performed at Laragen, Inc. (Los Angeles, CA).

**Sequence Analysis**

Sequences were assembled and edited using DNA-Star Lasagene software (Madison, WI). The software DOTUR was used to group sequences into operational taxonomic units (OTU) based on 8% Phylip DNA distance between OTUs, a cutoff which corresponds to the definition (3% amino acid distance) used to distinguish protein phylotypes in Chapters 2 and 3 (36). Phylogenetic trees were constructed using algorithms implemented within the ARB software environment (19). Tree construction details can be found in figure legends. The accession numbers of sequences used for phylogenetic analysis appear in Table 4.4 (Appendix).

**Table 4.1.** Primer and probes used in this study.

| Primer    | Sequence   | Target <sup>1</sup>                                     | Experiments <sup>2</sup>   | Reference     |
|-----------|--|---|--|---------------|
| 357F      | 5' – CTC CTA CGG GAG GCA<br>GCA G – 3'                 | Gen Bac 16S rRNA  | chip 1-5   | (26)          |
| 1409Ra    | 5' – GGG TAC CTC CAA CTC<br>GGA TGG TG – 3'            | Spirochete 16S rRNA                                     | chip 1, 2  | (26)          |
| 1492RL2D  | 5' – TAC GGY TAC CTT GTT<br>ACG ACT T – 3'             | Gen Bac 16S rRNA  | chip 1-5   | (26)          |
| 1389Prb   | 5' – HEX-CTT GTA CAC ACC<br>GCC CGT C-3BHQ1 – 3'       | Gen Bac 16S rRNA<br>(probe)                             | chip 1-5   | (26)          |
| 1409RaPrb | 5' – HEX-CGG GTA CCT CCA<br>ACT CGG ATG GTG-3BHQ1 – 3' | Spirochete 16S rRNA<br>(probe)                          | chip 3-5   | this<br>study |
| Sec427F   | 5' – CGI ATA TGA CAC GCT CCT<br>TCT GTA GC – 3'        | <i>fdhF</i> <sub>Sec</sub>                              | chip 1-5,<br><i>fdhF</i> <sub>Sec</sub> lib.                                 | this<br>study |
| Cys538F   | 5' – TAY AAY GCG GCG GCI<br>TCC CAC – 3'               | <i>fdhF</i> <sub>Cys</sub>                              | chip 1, 2  | this<br>study |
| Cys499F1b | 5' – ATG TCS CTK TCS ATI CCG<br>GAA A – 3'             | <i>fdhF</i> <sub>Cys</sub>                              | chip 3-5,<br>Cys lib.  | Chap.<br>3    |
| 1045R     | 5' – CIC CCA TRT CGC AGG YIC<br>CCT G – 3'             | <i>fdhF</i> <sub>Sec</sub> + <i>fdhF</i> <sub>Cys</sub> | chip 1-5,<br><i>fdhF</i> <sub>Sec</sub> , <i>fdhF</i> <sub>Cys</sub><br>lib. | Chap.<br>3    |

<sup>1</sup> Gen Bac, general bacterial.

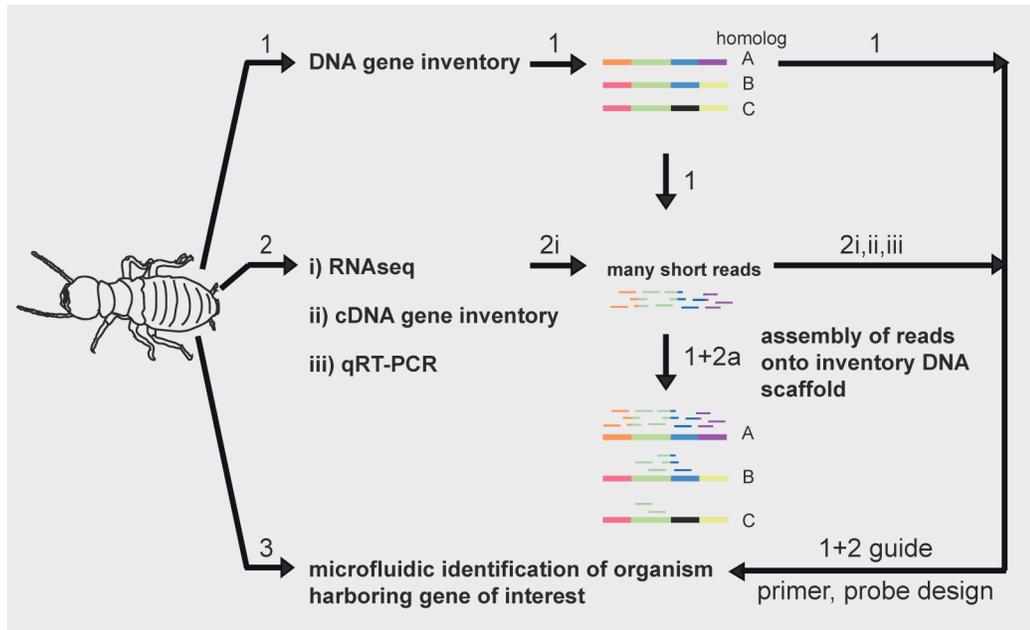
<sup>2</sup> chip, microfluidic chip experiment (Table 4.5, Appendix); lib. = cDNA library

## Results

In this study, we employ an approach that combines gene inventory, Illumina RNA-Seq, and microfluidic digital PCR techniques to assess transcription of a key acetogenesis gene (*fdhF*) in the gut community of a wood-feeding termite and identify bacteria encoding transcriptionally active *fdhF* genotypes. Figure 4.1 outlines the components of this approach. Briefly, we first mapped Illumina transcript reads of gut community RNA to gene inventory and pure culture sequence data to identify highly transcribed *fdhF* genotypes (arrows leading to 1+2a in Figure 4.1). We then corroborated the results using two independent methods. Finally, we performed microfluidics to discover the identity of organisms encoding transcribed *fdhF* genotypes (arrows leading to 3 in Figure 4.1).

### **RNA-Seq and other transcriptional assessments reveal two *fdhF* phylotypes dominate gut community *fdhF* transcription**

Total RNA was extracted from two collections of worker termites and sequenced by Illumina Solexa; one set was processed immediately after field collection, the other was maintained in the laboratory. RNA-Seq runs yielded 13,913,270 total 37-base pair reads (37-mers) for lab maintained termites and 14,043,698 reads for field-collected termites. Accounting for ribosomal RNA (~ 90% total) (24) and protozoa RNA [~ 90% of gut volume, (3)], we estimate bacterial functional gene transcripts only represent ~300,000 of total reads. We combined RNA-Seq reads from two Illumina runs into one large dataset (~28,000,000 reads) to increase bacterial functional gene read density.



**Figure 4.1.** Schematic of gene inventory, RNA-Seq, microfluidic PCR work-flow.  $FDH_H$  gene inventories and NCBI database sequences serve as scaffolds for RNA-Seq read mapping and data analysis. RNA-Seq based identifications of candidate genotypes belonging to transcriptionally important organisms can be corroborated using independent transcriptomic methods (cDNA gene inventory, qRT-PCR). Microfluidic, multiplex digital PCR on single cells can then be employed to obtain more genetic information on these important organisms.

To identify *fdhF* genotypes transcribed in the gut community, we analyzed the combined 28,000,000-read dataset using three nucleotide scaffold data sets, which contained: (i) 44 *fdhF* genotypes representing 23 phylotypes (8% DNA distance between OTUs) from *Z. nevadensis*; (ii) 92 *fdhF* genotypes representing 71 phylotypes from *Z. nevadensis*, two other phylogenetically “lower” termites, and a wood-feeding roach; (iii) 224 FDH genotypes representing 167 phylotypes for hydrogenase (*fdhF*), *Archaeal* F420, NAD(P)H, and respiratory chain-linked FDH enzymes from insects and the NCBI public database (Table 4.4, Appendix). Only reads that were perfectly matched to scaffold data set sequences were counted as hits. Reads were considered “unique” when they could only be mapped to genotypes within one FDH phylotype.

A total of 69 unique reads mapped onto the *Z. nevadensis* data set; these reads were distributed amongst 10 phylotypes (Table 4.2). Nearly half of all hits (30 reads, 43.4%) mapped to a single *Treponeme*-like *fdhF*<sub>sec</sub> phylotype (ZnD2sec). Almost all hits (27 out of 30) were distributed at unique positions along the entire length of the scaffold sequence, consistent with our inference that ZnD2sec is highly transcribed within the gut community. ZnD2sec also represented the majority of hits when laboratory maintained and field collected RNA-Seq reads were considered separately. The remaining 39 hits mapped to ZnHcys, Zn2cys, *fdhF*<sub>sec</sub> in *T. primitia* str. ZAS-2, and other *Treponeme*-like phylotypes (Table 4.2). All ZnHcys hits were derived from the field termite RNA-Seq dataset. Hits for other phylotypes were approximately evenly split between lab and field termites. The number of unique hits (69 reads) did not increase when reads were mapped to the 92-genotype data set, which contained sequences from four insect species.

Mapping onto the largest data set (224 genotypes representing 167 phylotypes for functionally diverse FDH enzymes) yielded only 1 more unique hit. This read mapped to an *Escherichia coli* hydrogenase-linked FDH<sub>H</sub> gene (Table 4.2). Although the total number of hits (70 reads) is small based on the abundance of ribosome and eukaryote transcripts, our results indicate RNA-Seq reads can be mapped to specific genotypes and phylotypes within an inventory containing several different homologs of a functional gene. We expect that increased sequencing, combined with effective rRNA depletion methods, will yield a more finely-resolved assessment of transcription.

To verify RNA-Seq results, we constructed separate cDNA libraries for *fdhF<sub>Sec</sub>* and *fdhF<sub>Cys</sub>* genes and performed SYBR-green qPCR assays using genotype specific primers (Table 4.3). In particular, we sought to determine whether ZnD2sec transcription was dominant relative to other *fdhF* sequences as RNA-Seq results indicated. Analysis of the *fdhF<sub>Sec</sub>* cDNA inventory from lab-maintained termite guts indicated the ZnD2sec phylotype accounts for 67% of all clone sequences. Comparison of ZnD2sec transcription with that of ZnB5sec and *T. primitia* Sec *fdhF<sub>Sec</sub>* using SYBR green qPCR assays yielded further confirmation of cDNA and RNA-Seq transcriptional patterns. Transcription of ZnD2cys was highest, followed by Zn70sec transcription; ZnB5sec transcription was not consistently detected. This order is consistent with the order of transcriptional abundance observed in RNA-Seq and cDNA library data. Both RNA-Seq and *fdhF<sub>Cys</sub>* cDNA libraries also identified the Zn2cys phylotype of *fdhF<sub>Cys</sub>* variants as relatively transcriptionally active. The absence of ZnHcys sequences from the cDNA dataset can be explained by samples differences, as all RNA-Seq reads mapping to this phylotype were

from field collected termite gut cDNA sample, which we did not analyze using cDNA inventory techniques. While we believe more RNA-Seq reads and qPCR assays are needed for an accurate picture of *fdhF* transcription in the gut, at least two independent methods indicate *ZnD2sec* and *Zn2cys* transcripts are relatively abundant in the *fdhF* transcript pool.

**Table 4.2.** *Z. nevadensis* gut community FDH gene transcription: RNA-Seq, *fdhF<sub>Sec</sub>* and *fdhF<sub>Cys</sub>* cDNA libraries, qRT-PCR.

| Phylotype<br>(FDH category) <sup>1</sup>              | RNA-Seq<br>Unique hits<br>(% total) <sup>2</sup> | <i>fdhF<sub>Sec</sub></i><br>cDNA<br>library<br>(%) <sup>3</sup> | <i>fdhF<sub>Cys</sub></i><br>cDNA<br>library<br>(%) <sup>3</sup> | qPCR<br>(copies/ng gut<br>cDNA) <sup>4</sup> |
|---|--|--|--|--|
| Gut clone ZnD2sec ( <i>fdhF<sub>Sec</sub></i> )       | 30 (42.8)  | 67   | —  | 131  |
| Gut clone ZnHcys ( <i>fdhF<sub>Cys</sub></i> )        | 11 (15.7)  | —  | —  | —  |
| Gut clone Zn2cys ( <i>fdhF<sub>Cys</sub></i> )        | 10 (14.3)  | —  | 54   | —  |
| <i>T. primitia</i> ( <i>fdhF<sub>Sec</sub></i> )      | 7 (10.0)   | —  | —  | 44   |
| <i>T. primitia</i> ( <i>fdhF<sub>Cys</sub></i> )      | 3 (4.3)  | —  | 2  | —  |
| Gut clone ZnB5sec ( <i>fdhF<sub>Sec</sub></i> )       | 3 (4.3)  | 2  | —  | NCD  |
| Gut clone Zn61sec ( <i>fdhF<sub>Sec</sub></i> )       | 2 (2.9)  | —  | —  | —  |
| Gut clone ZnF7sec ( <i>fdhF<sub>Sec</sub></i> )       | 1 (1.4)  | —  | —  | —  |
| Gut clone ZnB8sec ( <i>fdhF<sub>Sec</sub></i> )       | 1 (1.4)  | —  | —  | —  |
| Gut clone Zn72secRT ( <i>fdhF<sub>Sec</sub></i> )     | 1 (1.4)  | —  | —  | —  |
| <i>Escherichia coli</i> ( <i>fdhF<sub>Sec</sub></i> ) | 1 (1.4)  | —  | —  | —  |

<sup>1</sup> RNA-Seq reads were mapped to a dataset containing genes for hydrogenase-, NADPH-, F420-, respiratory chain-linked FDH enzymes ('FDH category'). All FDH reads mapped to hydrogenase-linked FDH genes (*fdhF*). Selenocysteine *fdhF* variants are denoted as *fdhF<sub>Sec</sub>*; cysteine variants are denoted as *fdhF<sub>Cys</sub>*. *Zn* gut clone phylotypes recovered from *Z. nevadensis* inventories are likely encoded by uncultured acetogenic spirochetes, as they phylogenetically group with *T. primitia* sequences (see Chapter 2).

<sup>2</sup> Reads were drawn from the combined 28 million read RNA-Seq dataset. Only reads that perfectly matched scaffold sequences within the same phylotype were considered "unique hits."

<sup>3</sup> Percentage of clones from *fdhF<sub>Sec</sub>* or *fdhF<sub>Cys</sub>* inventories constructed from laboratory maintained termite gut cDNA.

<sup>4</sup> Copies/ng lab maintained termite gut cDNA. NCD = not consistently detected.  $1\sigma < 2$  copies/ng.

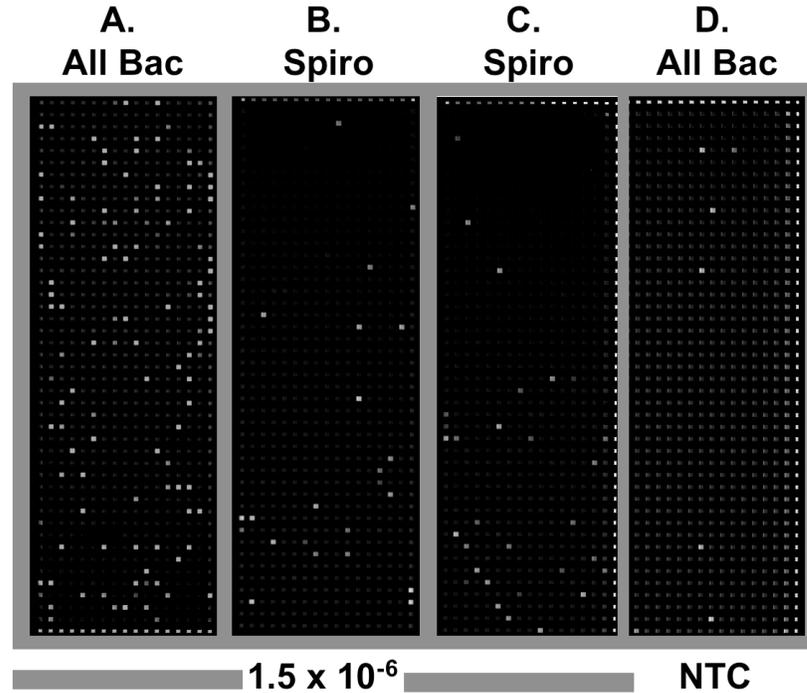
**Microfluidic digital PCR identification of two important *fdhF*-bearing spirochetes**

Previously in Chapter 2, we hypothesized that *fdhF<sub>Sec</sub>* and *fdhF<sub>Cys</sub>* phylotypes ZnD2sec and Zn2cys belong to spirochetes and that each of these spirochetes harbors both *fdhF<sub>Sec</sub>* and *fdhF<sub>Cys</sub>* gene variants. This was based on the phylogenetic clustering of ZnD2sec and Zn2cys with the dual *fdhF<sub>Sec</sub>* and *fdhF<sub>Cys</sub>* genes in *T. primitia* str. ZAS-1 and ZAS-2 [Chapter 2, (22)]. However, closely related treponemes like *T. azotonutricium* str. ZAS-9 are not acetogenic (10), nor encode FDH genes of any type (unpublished closed genome). Here, we performed microfluidic, multiplex digital PCR with 16S rRNA and *fdhF* primers on single termite gut bacterial cells to determine (i) whether uncultured spirochetes in *Z. nevadensis* guts encode *fdhF* and (ii) whether these spirochetes possess dual *fdhF<sub>Sec</sub>* and *fdhF<sub>Cys</sub>* as observed in *T. primitia*.

We utilized a broad approach, which differs from that of Ottesen *et al.* (26), as the design of functional gene probes with broad target ranges for *fdhF* was highly problematic. Microfluidic chip PCR reactions contained 16S rRNA primer and probes sets targeting spirochetes, and *fdhF* primers targeting Sec and/or Cys gene variants, but no functional gene probe. Most microfluidic PCR reactions were constructed to amplify 16S rRNA and *fdhF<sub>Sec</sub>* genes (i.e., duplex PCR); a few reactions targeted 16S rRNA, *fdhF<sub>Sec</sub>*, and *fdhF<sub>Cys</sub>* genes (i.e., triplex PCR). Samples were retrieved based on 16S rRNA probe fluorescence for spirochete 16S rRNA ribotypes rather than probe fluorescence for the functional gene; retrieved samples were then screened off-chip in simplex PCR reactions for the presence of 16S rRNA and *fdhF* gene products.

Spirochetes 16S rRNA genes were initially targeted with spirochete specific primers (357F, 1409Ra) and a general bacterial 16S rRNA probe (1389Prb) (microfluidic chip experiments 1, 2, Appendix 4, Table 4.5). However, the presence of non-spirochete 16S rRNA sequences within the same chamber as the target sequence could not be ruled out, since these sequences would not be amplified by 16S rRNA spirochete primers. Therefore, chip experiments 3 – 5 (Appendix 4, Table 4.5) were run with general bacterial 16S rRNA primers (357F, 1492RL2D), a spirochete specific 16S rRNA probe (1409RaPrb), and un-probed *fdhF* primers. Despite the increased sampling and screening steps associated this approach, we can, nevertheless, identify organisms encoding vastly different *fdhF* types (including those that are transcribed) as well as detect multiple 16S rRNA ribotypes in a sample to verify single cell amplification.

Microfluidic chip panels loaded with  $\sim 1 - 2 \times 10^{-6}$  dilutions of *Z. nevadensis* gut contents were sampled for retrieval. Panel A of Figure 4.2 shows end-point amplification from a typical gut dilution that yields  $< 150$  positive amplifications when general 16S rRNA bacteria gene primers are used. Assuming the distribution of cells on-chip follows a Poisson distribution, we estimate  $\sim 2.6\%$  of chambers contain more than a single cell. Replicate panels B and C (Figure 4.2) show the same gut dilution run with general bacterial 16S rRNA gene primers and a spirochete 16S rRNA gene probe. Well-separated amplification positive wells in spirochete specific panels were sampled for retrieval. Spirochetes accounted for  $12.5 \pm 6.5\%$  ( $1\sigma$ ) of all bacteria amplified on chip, consistent with previous observations in *Zootermopsis* (26). No template controls for PCR reactions targeting all bacteria (Panel D, Figure 4.2) typically yielded  $< 15$  positive amplifications.



**Figure 4.2.** Microfluidic digital PCR using an all-bacterial 16S rRNA gene primers-probe sets (panel A, D) or a spirochete specific 16S rRNA primer-probe set (panel B, C) with *fdhF* primers. Panels A, B, C contain *Z. nevadensis* gut contents diluted by  $1.5 \times 10^6$ . Panel D shows the no template control panel. Samples are retrieved from panels like B and C, which contain gut dilutions resulting in  $> 70\%$  chambers empty of any bacteria (Panel A).

On average,  $9.5 \pm 0.6 \%$  of spirochetes retrieved from triplex (16S rRNA – *fdhF*<sub>Sec</sub> – *fdhF*<sub>Cys</sub>) microfluidic chip PCR reactions run with Sec427F and Cys538F primers (Table 4.5, Appendix, experiments 1 and 2) were positive for *fdhF* amplification upon simplex screening. Screens of duplex (16S rRNA – *fdhF*<sub>Sec</sub>) chip reactions indicated  $10.1 \pm 5.9 \%$  of spirochetes had a gene for *fdhF*<sub>Sec</sub> (Table 4.5, Appendix, experiments 3a, 4, 5). Microfluidic chip triplex PCR with Sec427F and Cys519Fb primers yielded a much higher *fdhF* amplification rate from spirochete samples (40%) but this is likely due to the low sampling effort (only 5 wells were sampled in Experiment 3b).

We estimate 10% of all spirochetes and ~ 1% of all bacteria in *Z. nevadensis* carry *fdhF* genes. The latter result is consistent with Ottesen *et al.* (26), which estimates that 1% of all bacteria harbor the acetogenesis marker gene FTHFS, and with our previous findings that the majority of *fdhF* recovered from *Z. nevadensis* guts phylogenetically group with acetogenic spirochete sequences (Chapter 2). However, we note that *fdhF* genes were not recovered from 90% of spirochete on-chip retrievals. The effect of primer efficiency needs to be determined before a better estimate can be made of the true percentage of spirochetes harboring *fdhF*.

**Table 4.3.** Microfluidic chip retrieval of *fdhF* from spirochetes.

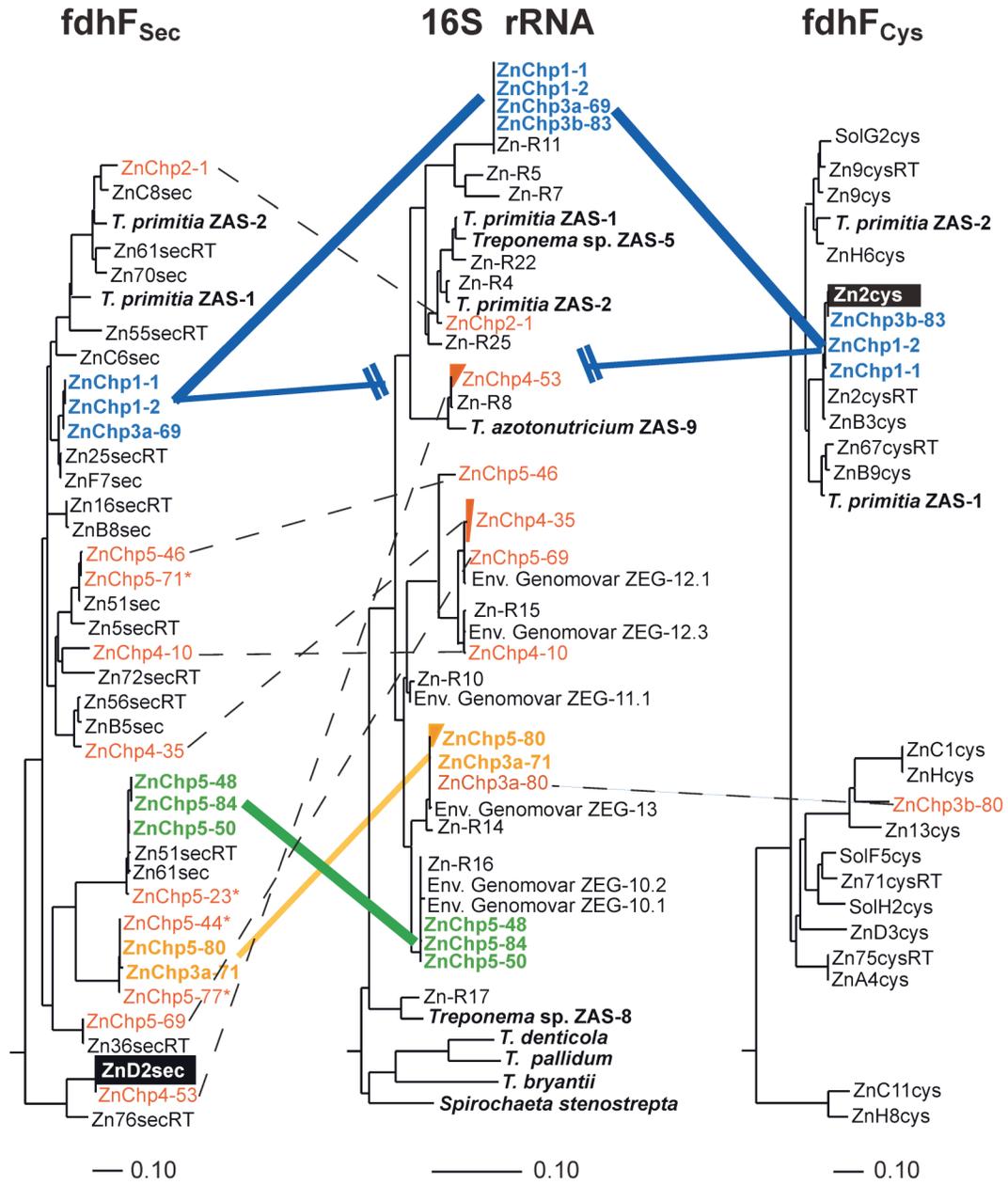
| Chip Experiment | Targets  | Spirochete with <i>fdhF</i> (%) |
|-----------------|--|---------------------------------|
| 1               | 16S rRNA – <i>fdhF</i> <sub>Sec</sub> – <i>fdhF</i> <sub>Cys</sub> | 9.1                             |
| 2               | 16S rRNA – <i>fdhF</i> <sub>Sec</sub> – <i>fdhF</i> <sub>Cys</sub> | 10.0                            |
| 3a              | 16S rRNA – <i>fdhF</i> <sub>Sec</sub>                              | 16.7                            |
| 3b              | 16S rRNA – <i>fdhF</i> <sub>Sec</sub> – <i>fdhF</i> <sub>Cys</sub> | 40.0                            |
| 4               | 16S rRNA – <i>fdhF</i> <sub>Sec</sub>                              | 5.3                             |
| 5               | 16S rRNA – <i>fdhF</i> <sub>Sec</sub>                              | 8.3                             |

Figure 4.3 shows microfluidic chip sequence phylogeny. The results support the hypothesis that *fdhF* are encoded by uncultured spirochetes (Figure 4.3, gene pairs marked in orange, blue, green). The results also indicate at least one uncultured spirochete [Zn-R11 ribotype, (25)] possesses both *fdhF*<sub>Sec</sub> and *fdhF*<sub>Cys</sub>, like *T. primitia* (ZnF7sec and Zn2cys phylotypes). We believe the 16S rRNA – *fdhF* gene pairs in Figure 4.3 are colocalized to single bacterial cells as mixed templates were not apparent in 16S

rRNA sequence traces. Multiple identifications of the same gene pair serve as further evidence of single cell amplification.

We found that the most commonly retrieved gene pairs were associated with the Zn-R11 16S rRNA ribotype, suggesting it is an abundant member of the gut spirochete population. RNA-Seq and cDNA analyses, which indicate this organism is responsible for a significant proportion of the *fdhF* transcript pool, are consistent with our inference that the Zn-R11 ribotype represents an important acetogen in the gut community. Moreover, Ottesen *et al.* (25) have previously co-localized FTHFS and Clp protease (ClpX) genes to the Zn-R11 16S rRNA ribotype (FTHFS, Zn-F8; ClpX, Zn-X3). Our results extend the genetic inventory of this uncultured spirochete (*Zootermopsis* environmental genomovar, ZEG 16) to include two more functional genes associated with acetogenesis. More experiments are needed to confirm the 16S rRNA identity (ZnR8 ribotype) of the organism encoding ZnD2cys (the most highly transcribed *fdhF* within the gut community) as well as other single 16S – *fdhF* colocalizations (Figure 4.2, gene pairs connected by dashed lines).

**Figure 4.3.** 16S rRNA (middle panel) and *fdhF* phylogeny (left panel, *fdhF<sub>Sec</sub>*; right panel, *fdhF<sub>Cys</sub>*) of microfluidic chip sequences. Chip samples are labeled “ZnChp(Chip number)-sample” and highlighted in color (orange, red, green, blue). Pure culture sequences are highlighted in bold. Environmental genomovars are uncultured spirochetes from *Z. nevadensis* that encode the canonical acetogenesis marker gene FTHFS (26). *fdhF* sequences outlined by black boxes were highly transcribed in RNA-Seq and cDNA datasets. Lines connecting sequences highlight 16S rRNA - *fdhF* colocalizations (duplex gene pairs for all but ZnChp1-1, ZnChp1-2 samples, which contained 16S rRNA, *fdhF<sub>Sec</sub>* and *fdhF<sub>Cys</sub>* gene products). Line thickness corresponds to the number of repeated colocalizations and indicates our confidence in the observed associations. Dotted lines denote only one instance of colocalization. Blue line with hatch marks connects *fdhF<sub>Sec</sub>* and *fdhF<sub>Cys</sub>*. Grouped clades are composed of chip 16S rRNA sequences that were re-amplified and cloned into plasmids prior to sequencing; all other sequences were from PCR products. 16S rRNA tree was constructed using the neighbor joining algorithm implemented in ARB (19) based on 705 SINA (SILVA Incremental aligner) aligned nucleotides (32). ZnChp5-84, ZnChp2-1, and ZnChp4-10 sequences were added in by parsimony using 600 aligned nucleotides. A PhyML-maximum likelihood (11) *fdhF* tree was constructed using 1818 aligned nucleotides from *fdhF<sub>Sec</sub>* and *fdhF<sub>Cys</sub>* genes. Chip *fdhF* sequences were added in by parsimony using 380 aligned nucleotides using ARB. The tree was then split into *fdhF<sub>Sec</sub>* and *fdhF<sub>Cys</sub>* clades for ease of viewing. Scale bars denote 0.1 base pair changes per alignment position.



## Discussion

In this study, we used gene inventories as guides for transcriptome and single cell experiments on uncultured bacteria from termite hindgut communities. In particular, we interpreted microbial community RNA-Seq data based on the results of previous gene inventory and pure culture studies. We then corroborated RNA-Seq results with traditional transcription assays. Lastly, we employed transcript and gene inventory information as a guide for microfluidic experiments, which aimed at identifying uncultured organisms possessing target genes of interest.

The analysis of *fdhF* transcription within the symbiotic gut microbial community of a wood-feeding termite using new RNA-Seq, cDNA inventory, and qRT-PCR techniques revealed that two *fdhF* phylotypes (ZnD2sec and Zn2cys) account for a significant proportion of all *fdhF* gene transcripts. These results indicate that RNA-Seq reads can be mapped to specific *fdhF* genotypes/phylotypes to obtain a snapshot of transcription in a species-rich community. However, it is obvious that RNA-Seq read depth for bacterial functional genes needs to be increased. Read depth can be enhanced with mRNA enrichment techniques, which can increase mRNA sample content by 5-50% (37). Although less significant, total *fdhF* read density may also be improved by extending functional gene library coverage. We did not construct a scaffold library with DNA from termites used for RNA-Seq, thus any genotypes (and transcripts) unique to these samples would not be detected.

Following transcriptome analyses, we performed microfluidic multiplex digital PCR on single termite gut bacterial cells to learn more about the uncultured bacteria encoding highly transcribed *fdhF* genotypes. Microfluidic chip experiments indicated that these *Treponeme*-like *fdhF* genotypes are encoded by uncultured spirochetes (Zn2cys encoded by spirochete with the Zn-R11 ribotype; ZnD2sec encoded by spirochete with the Zn-R8 ribotype). These results not only provide additional support to the concept that spirochetes dominate acetogenesis in termite guts, they also suggest that the bulk of acetogenesis in the gut may be due to relatively few spirochete species.

We also identified an uncultured spirochete that possesses a repertoire of acetogenesis genes similar to *T. primitia*. The spirochete defined by the Zn-R11 ribotype encodes the highly transcribed Zn2Cys *fdhF* and possesses genes for FTHFS and a second *fdhF* allele (ZnF7sec) for selenocysteine FDH<sub>H</sub>. This finding underscores the relevance of *T. primitia* to understanding carbon and energy flows mediated by uncultured acetogenic bacteria in the gut community. Additionally, the genomic context provided by multiplex digital PCR enables environmental transcription studies of this specific uncultured organisms (Zn-R11), wherein the organism's ClpX gene (Zn-X3) can be used as a quantitative internal transcription standard rather than total RNA content, which does not indicate whether transcriptional changes are due to variations in organism abundance or transcriptional upregulation.

The results presented herein provide a framework for future studies of transcriptionally active spirochetes. We employed a degenerate primer approach to identify spirochetes

that highly transcribe *fdhF*. However, a targeted primer and probe strategy may prove more time efficient. After 16S rRNA identification, other techniques like fluorescent in-situ hybridisation (FISH) can be used to study the environment niches uncultured organisms occupy. Targeted approaches are especially appropriate for confirming the 16S rRNA ribotype (Zn-R8) associated with the ZnD2sec *fdhF* phylotype. Phylogenetic analysis (Chapter 3) indicates the ZnD2sec phylotype is basal to a lineage of spirochete-like *fdhF* genes that have likely persisted through a sweeping gene loss within gut communities during termite evolution to subsequently diversify in the guts of ecologically successful phylogenetically “higher” termites. The fact that ZnD2sec is highly transcribed underscores the importance of obtaining more genetic information on the uncultured bacterium bearing this apparently successful functional gene allele. We envision that 16S rRNA identification will enable future studies (e.g., FISH, single cell whole genome amplification) that enhance our understanding of this organism’s ecological role in termite gut communities.

## **Acknowledgements**

Much of this work would not have been possible without Adam Rosenthal. He performed the RNA-Seq and qRT-PCR portions of the study and should be considered an equal contributor to the work presented herein.

## Appendix

**Table 4.4.** RNA-Seq scaffold data set.

**Table 4.5.** Microfluidic chip experiment details.

**Table 4.4.** RNA-Seq scaffold data set. FDH nucleotide sequences are categorized into Sec and Cys enzyme variants and major FDH types (FDH-H, hydrogen-linked; FDH-NADH, NADH-linked; FDH-N and FDH-O, respiratory chain linked).

| Sequence Source                                   | Variant | FDH type | Nucleotide Accession <sup>1</sup> |
|---|---------|----------|-----------------------------------|
| <i>Cryptocercus puntulatus</i> gut clone Cp10sec  | Sec     | FDH-H    | GU563433                          |
| <i>Cryptocercus puntulatus</i> gut clone Cp14sec  | Sec     | FDH-H    | GU563436                          |
| <i>Cryptocercus puntulatus</i> gut clone Cp16sec  | Sec     | FDH-H    | GU563432                          |
| <i>Cryptocercus puntulatus</i> gut clone Cp24sec  | Sec     | FDH-H    | GU563451                          |
| <i>Cryptocercus puntulatus</i> gut clone Cp28sec  | Sec     | FDH-H    | GU563450                          |
| <i>Cryptocercus puntulatus</i> gut clone Cp34sec  | Sec     | FDH-H    | GU563452                          |
| <i>Cryptocercus puntulatus</i> gut clone Cp3sec   | Sec     | FDH-H    | GU563434                          |
| <i>Cryptocercus puntulatus</i> gut clone Cp72cys  | Cys     | FDH-H    | GU563437                          |
| <i>Cryptocercus puntulatus</i> gut clone Cp78sec  | Sec     | FDH-H    | GU563453                          |
| <i>Cryptocercus puntulatus</i> gut clone Cp82sec  | Sec     | FDH-H    | GU563454                          |
| <i>Cryptocercus puntulatus</i> gut clone Cp94sec  | Sec     | FDH-H    | GU563455                          |
| <i>Cryptocercus puntulatus</i> gut clone Cp9cys   | Cys     | FDH-H    | GU563441                          |
| <i>Cryptocercus puntulatus</i> gut clone CpB10sec | Sec     | FDH-H    | GU563442                          |
| <i>Cryptocercus puntulatus</i> gut clone CpB2sec  | Sec     | FDH-H    | GU563446                          |
| <i>Cryptocercus puntulatus</i> gut clone CpB3sec  | Sec     | FDH-H    | GU563440                          |
| <i>Cryptocercus puntulatus</i> gut clone CpC1cys  | Cys     | FDH-H    | GU563444                          |
| <i>Cryptocercus puntulatus</i> gut clone CpC3sec  | Sec     | FDH-H    | GU563443                          |
| <i>Cryptocercus puntulatus</i> gut clone CpD1cys  | Cys     | FDH-H    | GU563445                          |
| <i>Cryptocercus puntulatus</i> gut clone CpD8sec  | Sec     | FDH-H    | GU563439                          |
| <i>Cryptocercus puntulatus</i> gut clone CpE8cys  | Cys     | FDH-H    | GU563447                          |
| <i>Cryptocercus puntulatus</i> gut clone CpF1cys  | Cys     | FDH-H    | GU563435                          |
| <i>Cryptocercus puntulatus</i> gut clone CpF8cys  | Cys     | FDH-H    | GU563449                          |
| <i>Cryptocercus puntulatus</i> gut clone CpF9cys  | Cys     | FDH-H    | GU563448                          |
| <i>Cryptocercus puntulatus</i> gut clone CpH1cys  | Cys     | FDH-H    | GU563438                          |

---

|  |     |       |          |
|--|-----|-------|----------|
| <i>Incisitermes minor</i> gut clone Im10sec      | Sec | FDH-H | GQ922349 |
| <i>Incisitermes minor</i> gut clone Im11cys      | Cys | FDH-H | GQ922364 |
| <i>Incisitermes minor</i> gut clone Im15sec      | Sec | FDH-H | GQ922351 |
| <i>Incisitermes minor</i> gut clone Im22sec      | Sec | FDH-H | GQ922353 |
| <i>Incisitermes minor</i> gut clone Im24cys      | Cys | FDH-H | GQ922369 |
| <i>Incisitermes minor</i> gut clone Im26sec      | Sec | FDH-H | GQ922354 |
| <i>Incisitermes minor</i> gut clone Im27sec      | Sec | FDH-H | GQ922355 |
| <i>Incisitermes minor</i> gut clone Im3sec       | Sec | FDH-H | GQ922356 |
| <i>Incisitermes minor</i> gut clone Im42cys      | Cys | FDH-H | GQ922371 |
| <i>Incisitermes minor</i> gut clone Im5cys       | Cys | FDH-H | GQ922373 |
| <i>Incisitermes minor</i> gut clone Im63sec      | Sec | FDH-H | GQ922361 |
| <i>Reticulitermes hesperus</i> gut clone Rh15cys | Cys | FDH-H | GQ922398 |
| <i>Reticulitermes hesperus</i> gut clone Rh24sec | Sec | FDH-H | GQ922383 |
| <i>Reticulitermes hesperus</i> gut clone Rh2sec  | Sec | FDH-H | GQ922381 |
| <i>Reticulitermes hesperus</i> gut clone Rh35sec | Sec | FDH-H | GQ922385 |
| <i>Reticulitermes hesperus</i> gut clone Rh36cys | Cys | FDH-H | GQ922410 |
| <i>Reticulitermes hesperus</i> gut clone Rh41sec | Sec | FDH-H | GQ922386 |
| <i>Reticulitermes hesperus</i> gut clone Rh47cys | Cys | FDH-H | GQ922402 |
| <i>Reticulitermes hesperus</i> gut clone Rh53sec | Sec | FDH-H | GQ922389 |
| <i>Reticulitermes hesperus</i> gut clone Rh54cys | Cys | FDH-H | GQ922404 |
| <i>Reticulitermes hesperus</i> gut clone Rh65cys | Cys | FDH-H | GQ922406 |
| <i>Reticulitermes hesperus</i> gut clone Rh71sec | Sec | FDH-H | GQ922391 |
| <i>Reticulitermes hesperus</i> gut clone Rh93cys | Cys | FDH-H | GQ922409 |
| <i>Reticulitermes hesperus</i> gut clone Rh9sec  | Sec | FDH-H | GQ922397 |
| <i>Treponema primitia</i> str. ZAS-1             | Cys | FDH-H | GQ922450 |
| <i>Treponema primitia</i> str. ZAS-1             | Sec | FDH-H | GQ922449 |
| <i>Treponema primitia</i> str. ZAS-2             | Cys | FDH-H | FJ479767 |
| <i>Treponema primitia</i> str. ZAS-2             | Sec | FDH-H | FJ479767 |
| <i>Zootermopsis nevadensis</i> gut clone Zn13cys | Cys | FDH-H | GQ922430 |

---

---

|   |     |       |          |
|---|-----|-------|----------|
| <i>Zootermopsis nevadensis</i> gut clone<br>Zn16secRT | Sec | FDH-H | GU563476 |
| <i>Zootermopsis nevadensis</i> gut clone<br>Zn25secRT | Sec | FDH-H | GU563475 |
| <i>Zootermopsis nevadensis</i> gut clone<br>Zn2cys    | Cys | FDH-H | GQ922431 |
| <i>Zootermopsis nevadensis</i> gut clone<br>Zn2cysRT  | Cys | FDH-H | GU563472 |
| <i>Zootermopsis nevadensis</i> gut clone<br>Zn36secRT | Sec | FDH-H | GU563477 |
| <i>Zootermopsis nevadensis</i> gut clone<br>Zn51sec   | Sec | FDH-H | GQ922423 |
| <i>Zootermopsis nevadensis</i> gut clone<br>Zn51secRT | Sec | FDH-H | GU563478 |
| <i>Zootermopsis nevadensis</i> gut clone<br>Zn55secRT | Sec | FDH-H | GU563479 |
| <i>Zootermopsis nevadensis</i> gut clone<br>Zn56secRT | Sec | FDH-H | GU563473 |
| <i>Zootermopsis nevadensis</i> gut clone<br>Zn5secRT  | Sec | FDH-H | GU563471 |
| <i>Zootermopsis nevadensis</i> gut clone<br>Zn61sec   | Sec | FDH-H | GQ922426 |
| <i>Zootermopsis nevadensis</i> gut clone<br>Zn61secRT | Sec | FDH-H | GU563480 |
| <i>Zootermopsis nevadensis</i> gut clone<br>Zn62sec   | Sec | FDH-H | GQ922427 |
| <i>Zootermopsis nevadensis</i> gut clone<br>Zn67cysRT | Cys | FDH-H | GU563482 |
| <i>Zootermopsis nevadensis</i> gut clone<br>Zn70sec   | Sec | FDH-H | GQ922428 |
| <i>Zootermopsis nevadensis</i> gut clone<br>Zn71cysRT | Cys | FDH-H | GU563483 |
| <i>Zootermopsis nevadensis</i> gut clone<br>Zn72secRT | Sec | FDH-H | GU563484 |
| <i>Zootermopsis nevadensis</i> gut clone<br>Zn75cysRT | Cys | FDH-H | GU563481 |
| <i>Zootermopsis nevadensis</i> gut clone<br>Zn76secRT | Sec | FDH-H | GU563485 |
| <i>Zootermopsis nevadensis</i> gut clone<br>Zn9cys    | Cys | FDH-H | GQ922435 |
| <i>Zootermopsis nevadensis</i> gut clone<br>Zn9cysRT  | Cys | FDH-H | GU563474 |
| <i>Zootermopsis nevadensis</i> gut clone<br>ZnA4cys   | Cys | FDH-H | GU563456 |
| <i>Zootermopsis nevadensis</i> gut clone<br>ZnB3cys   | Cys | FDH-H | GU563459 |
| <i>Zootermopsis nevadensis</i> gut clone<br>ZnB5sec   | Sec | FDH-H | GU563460 |
| <i>Zootermopsis nevadensis</i> gut clone<br>ZnB8sec   | Sec | FDH-H | GU563461 |
| <i>Zootermopsis nevadensis</i> gut clone<br>ZnB9cys   | Cys | FDH-H | GU563462 |
| <i>Zootermopsis nevadensis</i> gut clone<br>ZnC11cys  | Cys | FDH-H | GU563466 |
| <i>Zootermopsis nevadensis</i> gut clone<br>ZnC1cys   | Cys | FDH-H | GU563463 |

---

---

|   |     |         |                                |
|---|-----|---------|--------------------------------|
| <i>Zootermopsis nevadensis</i> gut clone ZnC6sec            | Cys | FDH-H   | GU563464                       |
| <i>Zootermopsis nevadensis</i> gut clone ZnC8sec            | Sec | FDH-H   | GU563465                       |
| <i>Zootermopsis nevadensis</i> gut clone ZnD2sec            | Sec | FDH-H   | GU563467                       |
| <i>Zootermopsis nevadensis</i> gut clone ZnD3cys            | Cys | FDH-H   | GU563468                       |
| <i>Zootermopsis nevadensis</i> gut clone ZnE2cys            | Cys | FDH-H   | GU563469                       |
| <i>Zootermopsis nevadensis</i> gut clone ZnF7sec            | Sec | FDH-H   | GU563458                       |
| <i>Zootermopsis nevadensis</i> gut clone ZnH6cys            | Cys | FDH-H   | GU563457                       |
| <i>Zootermopsis nevadensis</i> gut clone ZnH8cys            | Cys | FDH-H   | GU563470                       |
| <i>Zootermopsis nevadensis</i> gut clone ZnHcys             | Cys | FDH-H   | GQ922420                       |
| <i>Zootermopsis nevadensis</i> gut clone ZnJcys             | Cys | FDH-H   | GQ922417                       |
| <i>Zootermopsis nevadensis</i> gut clone ZnKcys             | Cys | FDH-H   | GQ922418                       |
| <i>Zootermopsis nevadensis</i> gut clone ZnLsec             | Sec | FDH-H   | GQ922412                       |
| <i>Zootermopsis nevadensis</i> gut clone ZnMsec             | Sec | FDH-H   | GQ922413                       |
| <i>Zootermopsis nevadensis</i> gut clone ZnOsec             | Sec | FDH-H   | GQ922415                       |
| <i>Zootermopsis nevadensis</i> gut clone ZnPcys             | Cys | FDH-H   | GQ922419                       |
| <i>Aeromonas salmonicida</i> subsp. <i>salmonicida</i> A449 | Sec | FDH-H   | NC_009348.1:1906100-1908244    |
| <i>Aggregatibacter aphrophilus</i> NJ8700                   | Sec | FDH-H   | NC_012913.1: c1159571-1157412  |
| <i>Acetoneama longum</i> APO-1                              | Sec | FDH-H   | GQ922445                       |
| <i>Buttiauxiella</i> SN1                                    | Sec | FDH-H   | GQ922446                       |
| <i>Carboxydotherrmus hydrogenoformans</i> Z-2901            | Sec | FDH-NAD | NC_007503.1:646163-648844      |
| <i>Carboxydotherrmus hydrogenoformans</i> Z-2901            | Sec | FDH-O   | NC_007503.1:702113-705121      |
| <i>Citrobacter koseri</i> ATCC BAA-895                      | Cys | FDH-H   | NC_009792.1:1727418-1729565    |
| <i>Citrobacter koseri</i> ATCC BAA-895                      | Sec | FDH-H   | NC_009792.1:3531364-3533511    |
| <i>Citrobacter rodentium</i> ICC168 fdhFsec                 | Sec | FDH-H   | NC_013716.1:c3662542-3660395   |
| <i>Citrobacter rodentium</i> ICC168 fdhFsec                 | Sec | FDH-H   | NC_013716.1:c3568359-3566212   |
| <i>Citrobacter</i> str. TSA-1                               | Sec | FDH-H   | GQ922447                       |
| <i>Citrobacter</i> sp. 30_2                                 | Cys | FDH-H   | NZ_GG657366.1:c1094197-1096347 |
| <i>Citrobacter</i> sp. 30_2                                 | Sec | FDH-H   | NZ_GG657366.1:c93031-90884     |

---

|   |     |         |                                 |
|---|-----|---------|---------------------------------|
| <i>Citrobacter</i> sp. 30_2                   | Sec | FDH-N   | NZ_GG657366.1:c1468196-1465035  |
| <i>Citrobacter</i> sp. 30_2                   | Sec | FDH-O   | NZ_GG657366.1:c37521-34471      |
| <i>Citrobacter youngae</i> ATCC 29220         | Sec | FDH-H   | NZ_ABWL01000021.1::c93031-90884 |
| <i>Citrobacter youngae</i> ATCC 29220         | Cys | FDH-H   | NZ_ABWL01000021.1:c24883-27030  |
| <i>Citrobacter youngae</i> ATCC 29220         | Sec | FDH-O   | NZ_ABWL01000021.1:c43554-40504  |
| <i>Clostridium bartlettii</i> DSM 16795       | Cys | FDH-NAD | NZ_ABEZ02000007.1:22423-25119   |
| <i>Clostridium bartlettii</i> DSM 16795       | Sec | FDH-H   | NZ_ABEZ02000007.1:c36324-34174  |
| <i>Clostridium beijerinckii</i> NCIMB 8052    | Cys | FDH-H   | NC_009617.1:c4364248-4366389    |
| <i>Clostridium bolteae</i> ATCC BAA-613       | Cys | FDH-H   | NZ_ABCC02000017.1:93731-95716   |
| <i>Clostridium carboxidivorans</i> P7         | Sec | FDH-H   | NZ_ACVI01000105.1:231-2378      |
| <i>Clostridium carboxidivorans</i> P7         | Cys | FDH-H   | NZ_ACVI01000010.1:36001-38157   |
| <i>Clostridium difficile</i> 630              | Sec | FDH-H   | NC_009089.1:c3884230-3882086    |
| <i>Cronobacter turicensis</i>                 | Cys | FDH-H   | NC_013282.1:1996635-1998845     |
| <i>Cronobacter turicensis</i>                 | Sec | FDH-H   | NC_013282.1:2002311-2004458     |
| <i>Cronobacter turicensis</i>                 | Cys | FDH-NAD | NC_013282.1:c1009687-1006715    |
| <i>Desulfotobacterium hafniense</i> DCB-2     | Sec | FDH-NAD | NC_011830.1:1504497-1507178     |
| <i>Dickeya dadantii</i> Ech586                | Cys | FDH-N   | NC_013592.1:c3063358-3066408    |
| <i>Dickeya dadantii</i> Ech586                | Cys | FDH-H   | NC_013592.1:2958853-2961003     |
| <i>Dickeya dadantii</i> Ech703                | Cys | FDH-H   | NC_012880.1:c1450903-1453053    |
| <i>Dickeya dadantii</i> Ech703                | Cys | FDH-N   | NC_012880.1:c2955857-2958907    |
| <i>Dickeya dadantii</i> Ech703                | Cys | FDH-O   | NC_012880.1:c1523376-1526423    |
| <i>Dickeya zaeae</i> Ech1591                  | Cys | FDH-H   | NC_012912.1:3084906-3087056     |
| <i>Desulfatibacillum alkenivorans</i> AK-01   | Sec | FDH-NAD | NC_011768.1:5447766-5450528     |
| <i>Desulfobacterium autotrophicum</i> HRM2    | Cys | FDH-NAD | NC_012108.1:1930486-1933251     |
| <i>Desulfotomaculum acetoxidans</i> 5575      | Sec | FDH-NAD | NC_013216.1:c3713225-3715906    |
| <i>Escherichia coli</i> O157:H7 str. FRIK2000 | Sec | FDH-H   | NZ_ACXO01000060.1:c38313-36585  |
| <i>Escherichia coli</i> O157:H7 str. FRIK966  | Sec | FDH-H   | NZ_ACXN01000050.1:79269-81416   |
| <i>Escherichia coli</i> 83972                 | Sec | FDH-H   | NZ_ACGN01000114.1:89871-92018   |
| <i>Escherichia coli</i> APEC O1               | Sec | FDH-H   | NC_008563.1:c4646031-4643884    |

---

|   |     |         |                                   |
|---|-----|---------|-----------------------------------|
| <i>Escherichia coli</i> O157:H7 str. EC4024                     | Sec | FDH-H   | NZ_ABJT01000004.1:c104404-106551  |
| <i>Escherichia coli</i> O157:H7 str. TW14588                    | Sec | FDH-H   | NZ_ABKY02000001.1:1646350-1648497 |
| <i>Escherichia</i> sp. 4_1_40B                                  | Sec | FDH-H   | NZ_ACDM01000067.1:c85542-83814    |
| <i>Escherichia coli</i> BL21(DE3)                               | Sec | FDH-H   | NC_012947.1:4135920-4138067       |
| <i>Escherichia coli</i> SE11                                    | Sec | FDH-H   | NC_011415.1:c4568500-4570647      |
| <i>Escherichia coli</i> UMN026                                  | Sec | FDH-H   | NC_011751.1:c4792216-4790069      |
| <i>Edwardsiella ictaluri</i> 93-146                             | Sec | FDH-H   | NC_012779.1:3156478-3158622       |
| <i>Edwardsiella tarda</i> EIB202                                | Sec | FDH-H   | NC_013508.1:3053142-3055286       |
| <i>Eggerthella lenta</i> VPI 0255                               | Cys | FDH-H   | NC_013204.1:c3320160..3322586     |
| <i>Cronobacter (Enterobacter) sakazakii</i> ATCC BAA-894        | Cys | FDH-NAD | NC_009778.1:2900970-2903942       |
| <i>Cronobacter (Enterobacter) sakazakii</i> ATCC BAA-894        | Cys | FDH-H   | NC_009778.1:c1996280-1998430      |
| <i>Enterobacter</i> sp. 638                                     | Sec | FDH-H   | NC_009436.1:c 329787-331934       |
| <i>Enterobacter</i> sp. 638                                     | Cys | FDH-H   | NC_009436.1:c1907448-1909598      |
| <i>Enterobacter cancerogenus</i> ATCC 35316                     | Sec | FDH-H   | NZ_ABWM02000022.1:21042-23189     |
| <i>Enterococcus faecalis</i> V583                               | Cys | FDH-NAD | NC_004668.1:1367291-1370011       |
| <i>Escherichia fergusonii</i> ATCC 35469                        | Sec | FDH-H   | NC_011740.1:4397249-4399396       |
| <i>Escherichia fergusonii</i> ATCC 35469                        | Sec | FDH-N   | NC_011740.1:1525306..1528353      |
| <i>Escherichia fergusonii</i> ATCC 35469                        | Sec | FDH-O   | NC_011740.1:3984322..3987372      |
| <i>Escherichia coli</i> str. K-12 substr. MG1655                | Sec | FDH-N   | NC_000913.2:1545425..1548472      |
| <i>Escherichia coli</i> str. K-12 substr. MG1655                | Sec | FDH-O   | NC_000913.2:c4080795..4083845     |
| <i>Escherichia coli</i> str. K-12 substr. MG1655                | Sec | FDH-H   | NC_000913.2:c4295242..4297389     |
| <i>Eubacterium acidaminophilum</i>                              | Sec | FDH-NAD | AJ312124.1:11347..14028           |
| <i>Eubacterium acidaminophilum</i>                              | Sec | FDH-NAD | AJ312125.1:2250..4943             |
| <i>Heliobacterium modesticaldum</i> Icel: NC_010337             | Cys | FDH-NAD | NC_010337.2:1747735..1750623      |
| <i>Klebsiella pneumoniae</i> subsp. <i>pneumoniae</i> MGH 78578 | Cys | FDH-H   | NC_009648.1:2290424..2292574      |
| <i>Klebsiella pneumoniae</i> subsp. <i>pneumoniae</i> MGH 78578 | Sec | FDH-H   | NC_009648.1:c4907710-4905563      |
| <i>Klebsiella pneumoniae</i> 342                                | Cys | FDH-H   | NC_011283.1:c2310716-2308566      |
| <i>Klebsiella pneumoniae</i> 342                                | Sec | FDH-H   | NC_011283.1:5239144-5241291       |
| <i>Klebsiella pneumoniae</i> NTUH-K2044                         | Sec | FDH-O   | NC_012731.1:c46019..49069         |

---

|   |     |          |                                   |
|---|-----|----------|-----------------------------------|
| <i>Klebsiella pneumoniae</i> NTUH-K2044                       | Sec | FDH-H    | NC_012731.1:c358869-356722        |
| <i>Klebsiella pneumoniae</i> NTUH-K2044                       | Sec | FDH-N    | NC_012731.1:c2794353..2797400     |
| <i>Klebsiella pneumoniae</i> NTUH-K2044                       | Cys | FDH-H    | NC_012731.1:3017444..3019594      |
| <i>Klebsiella pneumoniae</i> 342                              | Sec | FDH-N    | NC_011283.1:2546701..2549748      |
| <i>Klebsiella pneumoniae</i> 342                              | Sec | FDH-O    | NC_011283.1:5557641..5560691      |
| <i>Mannheimia succiniciproducens</i> MBEL55E                  | Cys | FDH-NAD  | NC_006300.1:684085..686892        |
| <i>Moorella thermoacetica</i> ATCC 39073                      | Sec | FDH-NAD  | NC_007644.1:c2432486..2435188     |
| <i>Moorella thermoacetica</i> ATCC 39073                      | Sec | FDH-H    | NC_007644.1:c2292497..2294737     |
| <i>Methanococcus maripaludis</i> S2                           | Sec | FDH-F420 | BX950229.1:145038..147068         |
| <i>Methanococcus vannielii</i> SB                             | Sec | FDH-F420 | CP000742.1:c663600..665624        |
| <i>Natranaerobius thermophilus</i> JW/NM-WN-LF                | Sec | FDH-NAD  | NC_010718.1:115206..117887        |
| <i>Oxalobacter formigenes</i> HOxBLS                          | Cys | FDH-H    | NZ_GG658151.1:2458842..2460998    |
| <i>Pantoea</i> sp. At-9b                                      | Cys | FDH-H    | NZ_ACYJ01000001.1:122540..124690  |
| <i>Pantoea</i> sp. At-9b                                      | Sec | FDH-O    | NZ_ACYJ01000014.1:c128676..131723 |
| <i>Pectobacterium carotovorum</i> subsp. carotovorum WPP14    | Cys | FDH-H    | NZ_ABVY01000027.1:c9266..11416    |
| <i>Pectobacterium carotovorum</i> subsp. brasiliensis PBR1692 | Cys | FDH-H    | NZ_ABVX01000086.1:c2739..4889     |
| <i>Pectobacterium atrosepticum</i> SCRI1043                   | Cys | FDH-H    | NC_004547.2:1420602..1422752      |
| <i>Pectobacterium atrosepticum</i> SCRI1043                   | Cys | FDH-H    | NC_004547.2:c1752061..1754157     |
| <i>Pectobacterium atrosepticum</i> SCRI1043                   | Cys | FDH-H    | BX950851.1:1752061..175415        |
| <i>Pectobacterium wasabiae</i> WPP163                         | Cys | FDH-H    | NC_013421.1:c1930748..1932898     |
| <i>Photobacterium profundum</i> 3TCK                          | Sec | FDH-H    | NZ_AAPH01000003.1:97396-99486     |
| <i>Pelobacter propionicus</i> DSM 2379                        | Cys | FDH-H    | NZ_AAJH01000001.1:11892..14606    |
| <i>Proteus mirabilis</i> ATCC 29906                           | Cys | FDH-H    | NZ_ACLE01000010.1:50054..52222    |
| <i>Proteus mirabilis</i> ATCC 29906                           | Sec | FDH-H    | NZ_ACLE01000010.1:30536-32701     |
| <i>Providencia alcalifaciens</i> DSM 30120                    | Sec | FDH-H    | NZ_ABXW01000042.1:35044-37197     |
| <i>Providencia alcalifaciens</i> DSM 30120                    | Sec | FDH-NAD  | NZ_ABXW01000042.1:c37197-35044    |
| <i>Providencia alcalifaciens</i> DSM 30120                    | Sec | FDH-O    | NZ_ABXW01000042.1:c129523-126476  |
| <i>Providencia alcalifaciens</i> DSM 30120                    | Sec | FDH-N    | NZ_ABXW01000042.1:235693-238740   |

---

|   |     |         |                                     |
|---|-----|---------|-------------------------------------|
| <i>Proteus mirabilis</i> HI4320   | Cys | FDH-H   | NC_010554.1:c3265604..3267772       |
| <i>Proteus mirabilis</i> HI4320   | Sec | FDH-H   | NC_010554.1:3909884-3912028         |
| <i>Providencia rettgeri</i> DSM 1131  | Sec | FDH-N   | NZ_ACCI02000030:c33183-30136        |
| <i>Providencia rustigianii</i> DSM 4541   | Sec | FDH-H   | NZ_ABXV02000023.1:88004-90157       |
| <i>Providencia rustigianii</i> DSM 4541   | Sec | FDH-N   | NZ_ABXV02000023.1:70811-73858       |
| <i>Psychromonas</i> sp. CNPT3 fdhFsec   | Sec | FDH-H   | NZ_AAPG01000013.1:c5742-3595        |
| <i>Ruminococcus</i> sp. 5_1_39B_FAA   | Cys | FDH-NAD | NZ_GG696049.1:c238140..240848       |
| <i>Salmonella enterica</i> subsp. <i>enterica</i><br>serovar Typhi str. CT18      | Sec | FDH-H   | NC_003198.1:4370484..4372631        |
| <i>Salmonella enterica</i> subsp. <i>enterica</i><br>serovar Typhimurium str. LT2 | Sec | FDH-H   | AE006468.1:c4525350..4527497        |
| <i>Salmonella enterica</i> subsp. <i>enterica</i><br>serovar Typhi str. CT18      | Sec | FDH-O   | NC_003198.1:3697528..3700578        |
| <i>Salmonella typhimurium</i> LT2   | Sec | FDH-H   | NC_003197.1:c4525350..4527497       |
| <i>Salmonella typhimurium</i> LT2   | Sec | FDH-N   | NC_003197.1:c1650442..1653489       |
| <i>Salmonella typhimurium</i> LT2   | Sec | FDH-O   | NC_003197.1:c4244758..4247808       |
| <i>Serratia proteamaculans</i> 568  | Cys | FDH-H   | NC_009832.1:c2657681..2659837       |
| <i>Serratia proteamaculans</i> 568  | Sec | FDH-N   | NC_009832.1:87013..90060            |
| <i>Serratia grimesii</i> ZFX-1  | Cys | FDH-H   | GQ922448                            |
| <i>Shigella flexneri</i> 2a str. 301  | Sec | FDH-O   | NC_004337.1:c4098182..4101232       |
| <i>Shigella</i> sp. D9  | Sec | FDH-H   | NZ_ACDL01000041:c39372-37225        |
| <i>Shigella sonnei</i> Ss046  | Sec | FDH-O   | NC_007384.1:c4296262..4299312       |
| <i>Shigella sonnei</i> Ss046  | Sec | FDH-N   | NC_007384.1:c1741118..1744165       |
| <i>Vibrio angustum</i> S14  | Sec | FDH-H   | NZ_AAOJ01000001.1:c1074316..1076460 |
| <i>Yersinia aldovae</i> ATCC 35236  | Cys | FDH-H   | NZ_ACCB01000002.1:136225..138372    |
| <i>Yersinia aldovae</i> ATCC 35236  | Sec | FDH-O   | NZ_ACCB01000003.1:36348..39395      |
| <i>Yersinia bercovieri</i> ATCC 43970   | Cys | FDH-H   | NZ_AALC02000017.1:13658..15805      |
| <i>Yersinia bercovieri</i> ATCC 43970   | Sec | FDH-O   | NZ_AALC02000005.1:103163..106210    |
| <i>Yersinia enterocolitica</i> subsp. <i>enterocolitica</i> 8081                  | Cys | FDH-H   | NC_008800.1:3050211..3052358        |
| <i>Yersinia enterocolitica</i> subsp. <i>enterocolitica</i> 8081                  | Sec | FDH-O   | NC_008800.1:c4525888..4528935       |
| <i>Yersinia frederiksenii</i> ATCC 33641  | Cys | FDH-H   | NZ_AALE02000011.1:c133500..135647   |
| <i>Yersinia frederiksenii</i> ATCC 33641  | Cys | FDH-H   | NZ_AALE02000004.1:63404..6554       |

---

|   |     |       |                                   |
|---|-----|-------|-----------------------------------|
| <i>Yersinia frederiksenii</i> ATCC 33641    | Sec | FDH-O | NZ_AAAL02000005.1:c136955-133908  |
| <i>Yersinia intermedia</i> ATCC 29909       | Cys | FDH-H | NZ_AAALF02000015.1:c38542..40698  |
| <i>Yersinia intermedia</i> ATCC 29909       | Sec | FDH-N | NZ_AAALF02000012.1:109282-112284  |
| <i>Yersinia kristensenii</i> ATCC 33638     | Cys | FDH-H | NZ_ACCA01000001.1:c40178..42325   |
| <i>Yersinia kristensenii</i> ATCC 33638     | Cys | FDH-H | NZ_ACCA01000002.1:c40178..42325   |
| <i>Yersinia kristensenii</i> ATCC 33638     | Sec | FDH-O | NZ_ACCA01000015.1:8904-11951      |
| <i>Yersinia mollaretii</i> ATCC 43969       | Cys | FDH-H | NZ_AALD02000036.1:52..2196        |
| <i>Yersinia mollaretii</i> ATCC 43969       | Cys | FDH-H | NZ_AALD02000005.1:c25400..27571   |
| <i>Yersinia mollaretii</i> ATCC 43969       | Sec | FDH-O | NZ_AALD02000033.1:c13893..16940   |
| <i>Yersinia pestis</i> KIM                  | Cys | FDH-H | NC_004088.1:678737..680884        |
| <i>Yersinia pseudotuberculosis</i> IP 32953 | Cys | FDH-H | NC_006155.1:474164..476311        |
| <i>Yersinia pseudotuberculosis</i> IP 32953 | Cys | FDH-H | NC_009708.1:c4151279..4153426     |
| <i>Yersinia rohdei</i> ATCC 43380           | Cys | FDH-H | NZ_ACCD01000002.1:c116227..118374 |
| <i>Yersinia rohdei</i> ATCC 43380           | Sec | FDH-N | NZ_ACCD01000004.1:c74607-71605    |
| <i>Yersinia ruckeri</i> ATCC 29473          | Cys | FDH-H | NZ_ACCC01000020.1:c42838..4496    |
| <i>Yersinia ruckeri</i> ATCC 29473          | Sec | FDH-N | NZ_ACCC01000005.1:93044-96546     |

<sup>1</sup> 'c' in front of genome coordinates indicates complement sequence

**Table 4.5.** Microfluidic chip experiment details.

| Chip Exp. | Fluidigm Chip No. | Primers and Probes ( $\mu\text{M}$ )  | Targets   | Spirochete Specificity via                            |
|-----------|-------------------|---|---|---|
| 1         | 1151-005-038      | 357F (200 nM), 1409Ra (200 nM),<br>1389Prb (300 nM),<br>Sec427F (200 nM), Cys538F (200 nM),<br>1045R (200 nM)       | Spirochete<br>16S rRNA,<br><i>fdhF</i> <sub>Sec</sub> ,<br><i>fdhF</i> <sub>Cys</sub> | Spirochete specific<br>primers, All Bacteria<br>Probe |
| 2         | 1151-026-033      | 357F (200 nM), 1409Ra (200 nM),<br>1389Prb (300 nM),<br>Sec427F (200 nM), Cys538F (200 nM),<br>1045R (200 nM)       | Spirochete<br>16S rRNA,<br><i>fdhF</i> <sub>Sec</sub> ,<br><i>fdhF</i> <sub>Cys</sub> | Spirochete specific<br>primers, All Bacteria<br>Probe |
| 3a        | 1151-067-035      | 357F (200 nM), 1492RL2D (200 nM),<br>1409RaPrb (300 nM),<br>Sec427F (200 nM), 1045R (175 nM)                        | Spirochete<br>16S rRNA,<br><i>fdhF</i> <sub>Sec</sub> ,                               | All Bacteria primers,<br>Spirochete specific<br>Probe |
| 3b        | 1151-067-035      | 357F (200 nM), 1492RL2D (200 nM),<br>1409RaPrb (300 nM),<br>Sec427F (200 nM), Cys499F1b (125<br>nM), 1045R (175 nM) | Spirochete<br>16S rRNA,<br><i>fdhF</i> <sub>Sec</sub> ,<br><i>fdhF</i> <sub>Cys</sub> | All Bacteria primers,<br>Spirochete specific<br>Probe |
| 4         | 1151-067-038      | 357F (200 nM), 1492RL2D (200 nM),<br>1409RaPrb (300 nM),<br>Sec427F (200 nM), 1045R (200 nM)                        | Spirochete<br>16S rRNA,<br><i>fdhF</i> <sub>Sec</sub> ,                               | All Bacteria primers,<br>Spirochete specific<br>Probe |
| 5         | 1151-067-041      | 357F (200 nM), 1492RL2D (200 nM),<br>1409RaPrb (300 nM),<br>Sec427F (200 nM), 1045R (200 nM)                        | Spirochete<br>16S rRNA,<br><i>fdhF</i> <sub>Sec</sub> ,                               | All Bacteria primers,<br>Spirochete specific<br>Probe |

## References

1. **Acinas, S. G., R. Sarma-Rupavtarm, V. Klepac-Ceraj, and M. F. Polz.** 2005. PCR-Induced sequence artifacts and bias: Insights from comparison of two 16S rRNA clone libraries constructed from the same sample. *Appl Environ Microbiol* **71**:8966-8969.
2. **Allen, E. E., and J. F. Banfield.** 2005. Community genomics in microbial ecology and evolution. *Nature Rev Microbiol* **3**:489-498.
3. **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.
4. **Breznak, J. A., and A. Brune.** 1994. Role of microorganisms in the digestion of lignocellulose by termites. *Ann Rev Entomol* **39**:453-487.
5. **Breznak, J. A., and J. M. Switzer.** 1986. Acetate synthesis from H<sub>2</sub> plus CO<sub>2</sub> by termite gut microbes. *Appl Environ Microbiol* **52**:623-630.
6. **Brune, A.** 2006. Symbiotic associations between termites and prokaryotes, p. 439-474. *In* M. Dworkin, S. Falkow, E. Rosenber, K. H. Schleifer, and E. Stackebrandt (ed.), *The Prokaryotes*, 3 ed, vol. 1. Springer, New York.
7. **Dennis, P., E. A. Edwards, S. N. Liss, and R. Fulthorpe.** 2003. Monitoring gene expression in mixed microbial communities by using DNA microarrays. *Appl Environ Microbiol* **69**:769-778.
8. **Egert, M., and M. W. Friedrich.** 2003. Formation of pseudo-terminal restriction fragments, a PCR-related bias affecting terminal restriction fragment length polymorphism analysis of microbial community structure. *Appl Environ Microbiol* **69**:2555-2562.
9. **Frias-Lopez, J., Y. Shi, G. W. Tyson, M. L. Coleman, S. C. Schuster, S. W. Chisholm, and E. F. DeLong.** 2008. Microbial community gene expression in ocean surface waters. *Proc Natl Acad Sci USA* **105**:3805-3810.
10. **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.
11. **Guindon, S., F. Lethiec, P. Duroux, and O. Gascuel.** 2005. PHYML Online--a web server for fast maximum likelihood-based phylogenetic inference. *Nucl Acids Res* **33**:W557-559.
12. **Handelsman, J.** 2004. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol Mol Biol Rev* **68**:669-685.
13. **Hillier, L. W., V. Reinke, P. Green, M. Hirst, M. A. Marra, and R. H. Waterston.** 2009. Massively parallel sequencing of the polyadenylated transcriptome of *C. elegans*. *Genome Res* **19**:657-666.
14. **Hugenholtz, P., B. M. Goebel, and N. Pace.** 1998. Impact of culture-independent studies on the emerging phylogenetic view of Bacterial diversity. *J Bacteriol* **180**: 4765-4774.

15. **Ji, H., H. Jiang, W. Ma, D. S. Johnson, R. M. Myers, and W. H. Wong.** 2008. An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nat Biotechnol* **26**:1293-300.
16. **Kolb, S., C. Knief, S. Stubner, and R. Conrad.** 2003. Quantitative detection of methanotrophs in soil by novel pmoA-targeted real-time PCR assays. *Appl Environ Microbiol* **69**:2423–2429.
17. **Leadbetter, J. R., T. M. Schmidt, J. R. Graber, and J. A. Breznak.** 1999. Acetogenesis from H<sub>2</sub> plus CO<sub>2</sub> by spirochetes from termite guts. *Science* **283**:686-689.
18. **Li, H., J. Ruan, and R. Durbin.** 2008. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* **18**:1851-1858.
19. **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. König, 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.
20. **Lueders, T., and M. W. Friedrich.** 2003. Evaluation of PCR amplification bias by terminal restriction fragment length polymorphism analysis of small-subunit rRNA and mcrA genes by using defined template mixtures of methanogenic pure Cultures and soil DNA extracts. *Appl Environ Microbiol* **69**:320-326.
21. **Luo, Y. H., L. Steinberg, S. Suda, S. Kumazawa, and A. Mitsui.** 1991. Extremely low D/H ratios of photoproduced hydrogen by cyanobacteria. *Plant Cell Physiol* **32**:897-900.
22. **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*.
23. **Nagalakshmi, U., Z. Wang, K. Waern, C. Shou, D. Raha, M. Gerstein, and M. Snyder.** 2008. The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* **320**:1344-1349.
24. **Neidhardt, F. C., and R. Curtiss.** 1996. *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2 ed. ASM press, Washington D.C.
25. **Ottesen, E. A.** 2008. The biology and community structure of CO<sub>2</sub>-reducing acetogens in the termite hindgut. Ph.D. dissertation. California Institute of Technology, Pasadena.
26. **Ottesen, E. A., J. W. Hong, S. R. Quake, and J. R. Leadbetter.** 2006. Microfluidic digital PCR enables multigene analysis of individual environmental bacteria. *Science* **314**:1464-1467.
27. **Pace, N.** 1997. A molecular view of microbial diversity and the biosphere. *Science* **276**:734-740.
28. **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.

29. **Pester, M., M. W. Friedrich, B. Schink, and A. Brune.** 2004. *pmoA*-based analysis of methanotrophs in a littoral lake sediment reveals a diverse and stable community in a dynamic environment. *Appl Environ Microbiol* **70**:3138-42.
30. **Polz, M. F., and C. M. Cavanaugh.** 1998. Bias in template-to-product ratios in multitemplate PCR. *Appl Environ Microbiol* **64**:3724-3730.
31. **Poretzky, R. S., N. Bano, A. Buchan, G. LeClerc, J. Kleikemper, M. Pickering, W. M. Pate, M. A. Moran, and J. T. Hollibaugh.** 2005. Analysis of microbial gene transcripts in environmental samples. *Appl Environ Microbiol* **71**:4121-4126.
32. **Pruesse, E., C. Quast, K. Knittel, B. M. Fuchs, W. Ludwig, J. Peplies, and F. O. Glockner.** 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucl Acids Res* **35**:7188-7196.
33. **Rosenthal, A. Z., E. Matson, A. Eldar, and J. R. Leadbetter.** Deep-transcript sequencing reveals multifaceted interactions between two termite gut spirochetes in co-culture. *Unpublished*.
34. **Rozen, S., and H. J. Skaletsky.** 2000. Primer3 on the WWW for general users and for biologist programmers. , p. 365-386. *In* S. Krawetz and S. Misener (ed.), *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ.
35. **Salmassi, T. M., and J. R. Leadbetter.** 2003. Analysis of genes of tetrahydrofolate-dependent metabolism from cultivated spirochaetes and the gut community of the termite *Zootermopsis angusticollis*. *Microbiology* **149**:2529-2537.
36. **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.
37. **Sorek, R., and P. Cossart.** 2010. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nat Rev Genet* **11**:9-16.
38. **Steinberg, L. M., and J. M. Regan.** 2009. *mcrA*-targeted real-time quantitative PCR method to examine methanogen communities. *Appl Environ Microbiol* **75**:4435-4442.
39. **Todaka, N., T. Inoue, K. Saita, M. Ohkuma, C. A. Nalepa, M. Lenz, T. Kudo, and S. Moriya.** 2010. Phylogenetic analysis of cellulolytic enzyme genes from representative lineages of termites and a related cockroach. *PLoS One* **5**:e8636.
40. **Tringe, S. G., and E. M. Rubin.** 2005. Metagenomics: DNA sequencing of environmental samples. *Nature Rev Gen* **6**:805-814.
41. **Tringe, S. G., C. von Mering, A. Kobayashi, A. A. Salamov, K. Chen, H. W. Chang, M. Podar, J. M. Short, E. J. Mathur, J. C. Detter, P. Bork, P. Hugenholtz, and E. M. Rubin.** 2005. Comparative metagenomics of microbial communities. *Science* **308**:554-557.
42. **Urich, T., A. Lanzen, J. Qi, D. H. Huson, C. Schleper, and S. C. Schuster.** 2008. Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS One* **3**:e2527.
43. **Wang, Z., M. Gerstein, and M. Snyder.** 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**:57-63.

44. **Wilhelm, B. T., S. Marguerat, S. Watt, F. Schubert, V. Wood, I. Goodhead, C. J. Penkett, J. Rogers, and J. Bahler.** 2008. Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* **453**:1239-1243.
45. **Yoder-Himes, D. R., P. S. G. Chain, Y. Zhu, O. Wurtzel, E. M. Rubin, J. M. Tiedje, and R. Sorek.** 2009. Mapping the *Burkholderia cenocepacia* niche response via high-throughput sequencing. *Proc Natl Acad Sci USA* **106**:3976-3981.
46. **Zhang, L., T. Hurek, and B. Reinhold-Hurek.** 2007. A nifH-based oligonucleotide microarray for functional diagnostics of nitrogen-fixing microorganisms. *Microb Ecol* **53**:456-470.