Microfluidic Digital PCR with Degenerate Primers: Multiplex Molecular Community Analysis of Acetogenic Bacteria in the Termite Hindgut

Abstract

PCR-based molecular profiling techniques allow in-depth analysis of uncultured environmental microorganisms, but are limited by their single-gene nature. Here, we present a method for multiplex PCR interrogation of uncultured environmental bacteria using degenerate primers that target protein coding genes and 16S rRNA. The use of microfluidic digital PCR to generate environmental gene inventories from parallel analysis of separated bacterial cells should minimize the effect of PCR bias on library composition and eliminate chimeic sequence artifacts. The ability to perform multiplex gene inventories allows the discovery of 16S rRNA gene sequences of organisms carrying a genetic marker of interest, or the association of two or more metabolic markers to single strains of uncultured bacteria. We used this technique to discover the 16S rRNA-based species identities of termite gut bacteria carrying the gene for formyl-tetrahydrofolate synthetase, a key enzyme for CO_2-reductive acetogenesis.

Introduction

The use of molecular community profiling techniques has transformed the field of microbial ecology (16, 27). Assays that target ribosomal RNA genes are routinely used to characterize the species composition of complex environments (9, 20, 36), while assays targeting metabolic genes are used to evaluate the diversity of organisms carrying

a genetic capacity of interest (3, 18, 47). Molecular profiling experiments use PCR to amplify a subset of related sequences from an environmental DNA pool. This generates a PCR product pool sequences that reflects the diversity of organisms within a sample that carry the genetic marker of interest; the diversity of sequences in this pool can be measured indirectly by techniques such as terminal restriction length polymorphism or denaturing gradient gel electrophoresis, or directly through generation and sequencing of clone libraries (25).

Large-scale environmental sequencing projects have proven valuable sources of novel genes and insights into environmental processes (32, 38). Similarly, techniques for single cell genome sequencing show promise for the metabolic characterization of uncultured organisms (24, 37, 48). However, these analyses cannot match the target specificity (and therefore the potential survey depth) of PCR-based techniques. As an example, a recent survey used 16S rRNA primers in conjunction with high-density 454 pyrosequencing to generate 8 sequence libraries from deep sea DNA samples that contained 6,505–22,994 total sequences, 2,656–8,699 of them unique (36). In contrast, the initial Sargasso Sea metagenomic analysis comprised over 1.88 million sequence reads, yet yielded only 1,412 rRNA genes (40).

However, the targeted nature of PCR-based techniques can also represent a drawback. As genes are studied in isolation, it is difficult to establish relationships between phylotypes identified in inventories of diverse genes from a single sample. When metabolic genes are examined, the species identity of the source organisms can only be

hypothesized according to their similarity to gene sequences from previously cultivated and characterized organisms. This type of analysis can result in large clusters of unassigned sequences when cultured representatives are rare (7, 13). We have developed a microfluidics-based technique that allows the association of such sequences with the 16S rRNA gene sequences of the uncultured environmental bacteria that encode them.

In 2006, we demonstrated a technique for multiplex, microfluidic digital PCR-based interrogation of hundreds of bacterial cells in parallel (26). We used a microfluidic device to partition an environmental sample into hundreds of independent reaction chambers. At low sample dilutions, reaction chambers contained no more than one bacterial cell. Each of those cells was then used as template for a multiplex PCR reaction targeting bacterial 16S rRNA genes and a specific functional gene sequence. PCR products from chambers showing amplification of both genes were retrieved and sequenced.

This technique allowed 16S rRNA gene-based identification of uncultured bacteria carrying a genetic marker of interest. However, the utility of the technique as described was limited by the nature of the PCR chemistry utilized. PCR amplification within the microfluidic device was detected using amplicon-specific Taqman probes. While the diversity of sequences amplified in PCR assays can be expanded through the use of degenerate primers, Taqman probes are highly sequence specific. The strong sequence conservation of the 16S rRNA gene allowed the design of an “all bacterial” probe with low degeneracy, yet broad specificity (39). However, protein-encoding genes are

generally less well conserved, and a nondegenerate probe will only detect closely related sequence subgroups.

Here, we present an approach that allows multiplex digital PCR with degenerate primers that target genes encoding clusters of orthologous proteins. We used a universal-template probe strategy developed by Zhang et al. (49), in which a probe-binding sequence is attached to the 5′ end of a real-time PCR primer. This sequence is incorporated into the amplicon during the first round of amplification, allowing the Taqman probe to detect amplification of that product. Zhang et al. proposed this approach as a method to reduce the costs associated with real-time PCR. However, we have adapted this strategy to pair nondegenerate Taqman probes with degenerate primers for multiplex PCR.

We used this technique in the context of our on-going efforts to characterize the acetogenic community of the termite hindgut. In wood-feeding termites, CO_2-reducing acetogens are the primary consumers of H_2 generated during the fermentation of wood polysaccharides; the acetate produced by these bacteria powers 22%–26% of the insect's energy metabolism (4, 5, 30). The gene for formyl-tetrahydrofolate synthetase (FTHFS), a key enzyme in the acetyl-CoA pathway, can be used as a genetic marker of acetogenic capability (19, 22). FTHFS-based molecular community analyses have been carried out on a number of termite species, which are dominated by a sequence cluster that includes FTHFS genes from two acetogenic spirochetes, *Treponema primitia* ZAS-1 and ZAS-2 (17, 29, 34). However, many of the recovered sequences are only distantly related to these two isolates, and the termite *Treponeme* cluster as a whole affiliates with FTHFS

sequences from *Firmicute* acetogens. As a result, even phylum-level classification of the bacteria that encode key FTHFS types is ambiguous.

We have designed a multiplex, microfluidic digital PCR assay that allows the use of all-bacterial 16S rRNA gene primers and FTHFS primers that target the Lovell cluster of acetogenic FTHFS types. Using this assay, we have identified FTHFS-bearing organisms from the termite hindgut, and used 16S rRNA gene phylogeny to discover their species identity. To demonstrate the general applicability of our strategy, we also designed and implemented an assay to discover the gene sequences for the ATPase subunit of the Clp protease (ClpX) of uncultured termite *Treponemes*. ClpX was chosen because it a potential target for design of species-specific internal controls for environmental expression analyses.

Materials and Methods

Laboratory Maintenance of Termites and Bacterial Strains

Zootermopsis nevadensis specimens were collected from fallen Ponderosa Pine (*Pinus ponderosa*) at the Chilao Campground in the Angeles National Forest. Colonies were maintained in the laboratory on Ponderosa at 23 ºC and at a constant humidity of 96%, achieved via incubation over saturated solutions of KH_2PO_4 within 10-gallon aquaria (43). *Treponema primitia* ZAS-1 was maintained in the laboratory as described in (17).

PCR Primer Design

Degenerate primers were designed using the CODEHOP program (33). FTHFS sequences from acetogenic bacteria and partial termite *Treponeme* FTHFS genes from the *Nasutitermes* metagenome (42) were used to design primers. A 57-60 ºC consensus region was found to be optimal for the 60 ºC extension/annealing temperature used in real-time PCR experiments. Consensus regions suggested by the CODEHOP program were adjusted to match codon preferences observed in termite *Treponeme* FTHFS sequences.

Table 5.1. Primers Used in Microfluidic Digital PCR

Name	Sequence	Target
357F	CTCCTACGGGAGGCAGCAG	All bacterial 16S rRNA
1492RL2D	TACGGYTACCTTGTTACGACTT	All bacterial 16S rRNA
1409Ra	GGGTACCTCCAACTCGGATGGTG	Termite *Treponeme* 16S rRNA
1409Rb	CGGGTACCCTCTACTCGGATGGTG	Termite *Treponeme* 16S rRNA
533F	AGGGTTGCGCTCGTTG	16S rRNA sequencing
1100R	GTGCCAGCMGCCGCGGTAA	16S rRNA sequencing
FTHFS-Fa	GGICCIGTITTYGGIGTIAARGG	FTHFS, unprobed
FTHFS-Ra	CCIGGCATIGTCATIATITCICCIGT	FTHFS, unprobed
FTHFS-Fb	ACCTGCACTTCACCGGAGAYTTYCAYGCIAT	FTHFS, probed
UP149-FTHFS-Fb	GGCGGCGAACCTGCACTTCACCGGAGAYTTYCAYGCIAT	FTHFS, probed
FTHFS-Rb	ACGCCTTCGCCACCCTTIKCCCAIAC	FTHFS, probed
ClpX_F	CGAAGCGGGCTATGTCGGIGARGAYGT	ClpX, probed
ClpX_R	GATGGGAAGCCTGCCGATGAAYTCIGGDAT	ClpX, probed
UP149-ClpX-R	GGCGGCGAGATGGGAAGCCTGCCGATGAAYTCIGGDAT	ClpX, probed

ClpX protease primers were designed using a similar strategy. ClpX protease sequences were downloaded from the *Nasutitermes* metagenome data set (42) and aligned with sequences from published microbial genomes. A putative termite *Treponeme* cluster of ClpX sequences was identified based on phylogenetic similarity to ClpX sequences from published *Treponeme* genomes (*T. denticola* (35) and *T. pallidum* (11)) and unpublished

ZAS-2 ClpX sequence (Eric Matson, personal communication). Primer sequences are presented in Table 5.1. Inosine base analogues (denoted as "I" in the primer sequence) were used in the place of N to reduce the degeneracy of the primers. Other than the Roche universal probe, all primers and probes were purchased from Integrated DNA Technologies.

Template Preparation

DNA was purified from *T. primitia* ZAS-1 using the Qiagen DNeasy Tissue kit, with the protocol described for extraction of DNA from gram negative bacteria cells (DNeasy Tissue Handbook, July 2003 version). Template concentrations were measured using the Hoefer DyNAQuant 200 fluorometer and DNA quantification system (Amersham Pharmacia Biotech) using reagents and procedures directed in the manual (DQ200-IM, Rev C1, 5-98).

A "synthetic gut fluid" (SGF) salt solution was used for suspension and dilution of bacterial cells prior to addition to the PCR reaction. This solution contained 29.4 mM, K_2HPO_4, 11.6 mM KH_2PO_4, 5.6 mM KCl, and 30 mM NaCl. DNase-free RNase (Roche) was added just prior to cell dilution at 0.5 μg/mL to prevent PCR inhibition by ribosomal RNA. *T. primitia* ZAS-1 cells were collected from late exponential phase cultures and diluted in sterile SGF. Single *Z. nevadensis* hindguts were extracted from worker larvae, suspended in sterile SGF, and physically disrupted by crushing the gut with a sterile pipette tip followed by brief (2–3 s) pulses of vortexing. Suspensions were allowed to stand briefly to sediment large particles, then diluted to working concentrations in SGF

and mixed 1 to 15 v/v with the PCR reaction mixture for immediate loading onto microfluidic chips.

PCR on Microfluidic Chips

Microfluidic devices were purchased from Fluidigm Corporation. On-chip multiplex PCR reactions contained iQ Multiplex Powermix (BioRad, 170-8848), 0.1% Tween-20, and 150 nM ROX standard. 16S rRNA amplifications used primers and probes described in (26): 357F and 1492RL2D at 400nM for all bacterial 16S rRNA, and 357F, 1409Ra, and 1409Rb at 400nM each for “spirochete-specific” 16S rRNA amplification; all 16S rRNA reactions used the 1389 probe (HEX-CTTGTACACACCGCCCGTC-BHQ1) at 267 nM. Unprobed FTHFS reactions used FTHFS-Fa and FTHFS-Ra at 400 nM each. For universal template probe reactions, Roche Universal Probe #149 was included at 267 nM, the unlabeled primer (FTHFS-Rb or ClpX-F) added at 400 nM. The best signal intensity for universal template probe reactions was obtained when the primer with attached binding site was mixed 50:50 with the same primer without the binding site. Reactions contained 200 nM probe-binding primer (149-FTHFS-Fb or 149-ClpX-R) and 200 nM conventional primer (FTHFS-Fb or ClpX-R).

Chips were loaded and PCR performed using the BioMark system as recommended by Fluidigm. The cycling protocol was 95 ºC 3 min, (95 ºC 15 s, 60 ºC 90 s) x 45. When chips were intended for product retrieval, a final extension step of 10min at 60 ºC was added. Amplification curves and reaction results were evaluated using BioMark Digital PCR analysis software (Fluidigm, v.2.0.6).

Sample Retrieval and Analysis

Single cell PCR products were retrieved from amplification-positive chambers. Pressure was released from the accumulators, and chips were then peeled from the carrier and silicon heat sink. Target chambers were located using a dissecting microscope, and the tip of a 26 gauge syringe needle was inserted into each chamber through the bottom surface of the chip. Needle tips were then swirled briefly in 10 μL of TE to release the PCR product.

Retrieved samples were evaluated for the presence of target genes via simplex PCR at a benchtop scale. For functional gene analysis, primers without the probe binding site were used at 400nM, and the probe-binding primer omitted. In some samples, re-amplification with 357 and 1492RL2D resulted in secondary bands. Utilization of 533F in place of 357 eliminated these artifacts. The cycling protocol for conventional PCR was 95 ºC 3 min, (95 ºC 15 s, 60 ºC 60 s, 72 ºC 60 s) x 35. The presence or absence of product was evaluated using agarose gel electrophoresis.

PCR products from successful retrievals were purified using the Qiagen PCR purification kit, and sequenced using the FTHFS PCR primers and 16S rRNA gene internal primers 1100R and 533F. Cycle sequencing reactions were carried out by Laragen (Los Angeles, CA) Sequences were assembled and edited using the Lasergene software package (DNASTAR, version 7.2.1). Sequence alignment and phylogenetic analysis were carried out using the ARB software package (23).

Results

We successfully designed primer/probe sets that allowed Taqman-based detection of FTHFS sequences in multiplex reactions with "all-bacterial" 16S rRNA primers on microfluidic chips (Figure 5.1). Low levels of amplification (5-15 chambers per panel) were detected for the FTHFS channel in "no template added" reactions, most likely as a result of primer dimer formation with 16S rRNA primers. In simplex FTHFS reactions, no amplification was observed in "no template added" reactions. Minor amplification (5-10 positives per panel) was also detected for 16S rRNA in template-free controls; this is likely due to low levels of bacterial DNA contamination (common in commercial PCR reagents (6)).

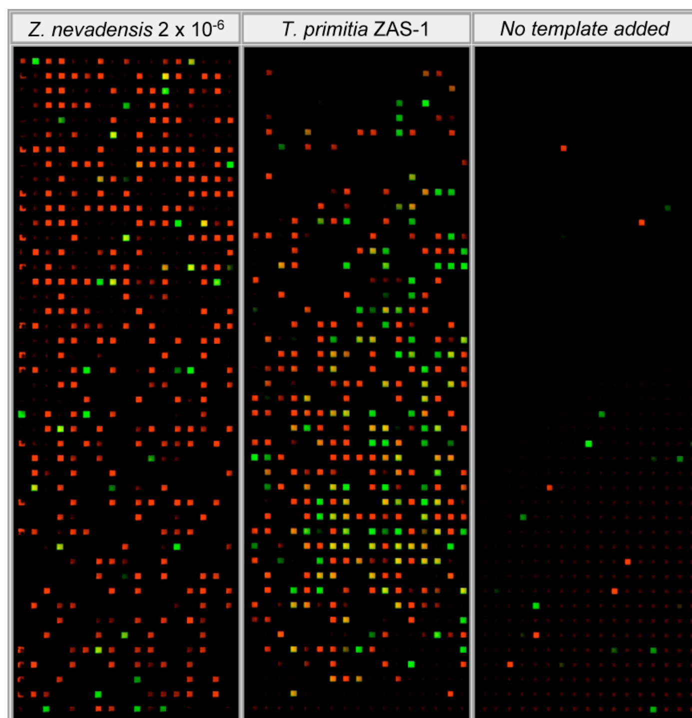

Figure 5.1. Microfluidic Digital PCR for "Lovell cluster" FTHFS and all-bacterial 16S rRNA genes. Three sample panels from a representative chip are shown at amplification cycle 45. FTHFS signal shown in green, 16S rRNA gene signal shown in red. The template source for panel 1 was *Z. nevadensis* hindgut contents, for panel 2 was a late exponential phase *T. primitia* ZAS-1 culture. No template was added to the PCR mixture for panel 3.

Treponema primitia ZAS-1 was chosen for use as a positive control template because it carries two genomic copies each of FTHFS and 16S rRNA (12, 34), allowing direct comparison of amplification efficiency between these two genes. The observed amplification success rate for on-chip, multiplex PCR from purified genomic DNA was 42% for FTHFS and 73% for 16S rRNA (calculated as measured number of copies per μL divided by expected copes/μL according to DNA concentration). In multiplex PCR reactions from cultured *T. primitia* ZAS-1 cells, approximately 25% of chambers with either FTHFS or 16S rRNA amplification were positive for both genes. This is more than twice the number of colocalizations expected if FTHFS and 16S rRNA were assorting independently. The presence of chambers in which either FTHFS or 16S rRNA amplified alone may due either to multiplexing failure (where amplification of one gene outcompetes amplification of the other) or to lysis of ZAS-1 cells followed by genome fragmentation.

FTHFS and 16S rRNA genes were successfully amplified in multiplex PCR from hindgut luminal contents from the lower termite *Zootermopsis nevadensis*. *Z. nevadensis* hindgut contents were diluted in SGF salt solution and added to PCR reactions immediately prior to chip loading. The standard for single cell separation was 33% occupancy or less. PCR products were retrieved, reamplified, and sequenced from chambers in which both FTHFS and 16S rRNA genes had amplified. The resultant FTHFS and 16S rRNA sequences were binned with a similarity cutoff of 99.5% and characterized by phylogenetic analysis (Figure 5.2, Table 5.2).

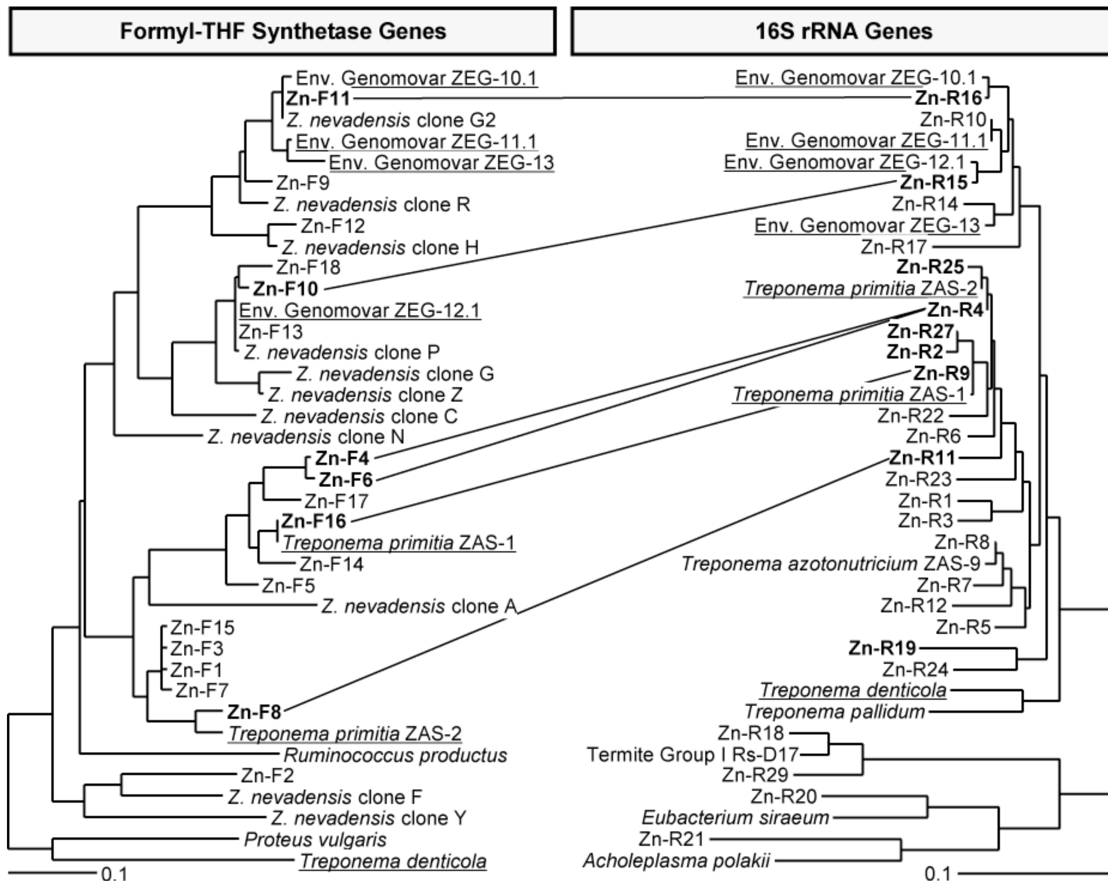

Figure 5.2. Phylogenetic analysis of FTHFS and 16S rRNA gene sequences amplified using microfluidic digital PCR. Trees calculated using Phylip distance Fitch algorithm. **Left,** An FTHFS gene tree constructed using 726 unambiguous, aligned base pairs; short sequences (ZEG sequences and Zn-F18) were added to the finished tree using 192 alignment positions. **Right,** A 16S rRNA gene tree calculated using 722 unambiguous, aligned base pairs. Scale bars represent 0.1 changes per alignment position. Lines identify FTHFS-16S rRNA gene pairs supported by repeated colocalizations or similarity to established associations (Table 5.2). Sequences assigned to an environmental genomovar (Table 5.5) marked in bold.

Table 5.2 lists FTHFS and 16S rRNA gene pairs colocalized using microfluidic digital PCR. In our initial microfluidic digital PCR experiments (26), we found that apparent single cell dilutions sometimes contained multiple 16S rRNA types or a single 16S rRNA type that did not match those found in other experiments. This was attributed to the nature of the dilution method used for cell separation; physically associated cell aggregates sort as one particle. Sorting of cell aggregates followed by a skewed PCR

balance result in false associations. As a result, a single colocalization event was considered insufficient evidence of association; FTHFS-16S rRNA pairs identified in at least two independent colocalizations or supported by similarity to cultured strains or prior-colocalizations are marked in Figure 5.2.

Table 5.2. FTHFS/16S rRNA gene pairs[a]

Experiment	FTHFS	16S rRNA	Comment
1	Zn-F1	Zn-R1	
2	Zn-F2	Zn-R2	
3	Zn-F3	Zn-R3	
4	Zn-F4	Zn-R4	Similar to 15; ZEG 14.1
5	Zn-F5	Zn-R5	
6	Zn-F5	Zn-R6	
7	Zn-F5	Zn-R7	
8	Zn-F4	Zn-R8	
9	Zn-F6	Zn-R9	
10	Zn-F6	Zn-R10	
11	Zn-F7	Zn-R7	
12	Zn-F7	Zn-R11	
13	Zn-F8	Zn-R12	
14	Zn-F4	Zn-R4	
15	Zn-F6	Zn-R4	Similar to 4; ZEG 14.2
16	Zn-F8	Zn-R14	
17	Zn-F8	Zn-R10	
18	Zn-F8	Zn-R11	Repeated in 20, 24; ZEG 16.1
19	Zn-F9	Zn-R14	
20	Zn-F8	Zn-R11	Repeated in 20, 24; ZEG 16.1
21	Zn-F10	Zn-R15	Similar to ZEG 12; ZEG 12.5
22	Zn-F11	Zn-R16	Similar to ZEG 10; ZEG 10.5
23	Zn-F12	Zn-R17	
24	Zn-F8	Zn-R11	Repeated in 20, 24; ZEG 16.1
25	Zn-F13	Zn-R18	
26	Zn-F14	Zn-R9	
27	Zn-F15	Zn-R2	
28	Zn-F16	Zn-R19	
29	Zn-F16	Zn-R20	
30	Zn-F16	Zn-R9	Similar to ZAS-1; ZEG 15.1
31	Zn-F17	Zn-R21	

[a] Colocalizations 1-17 were collected in experiments using an unprobed FTHFS primer set and termite *Treponeme* specific 16S rRNA primers. Termite gut dilutions were analyzed in side-by-side reactions with all-bacterial 16S rRNA primers to confirm single cell separation. PCR products were retrieved from chambers positive for *Treponeme* 16S rRNA and screened off-chip for FTHFS amplification.

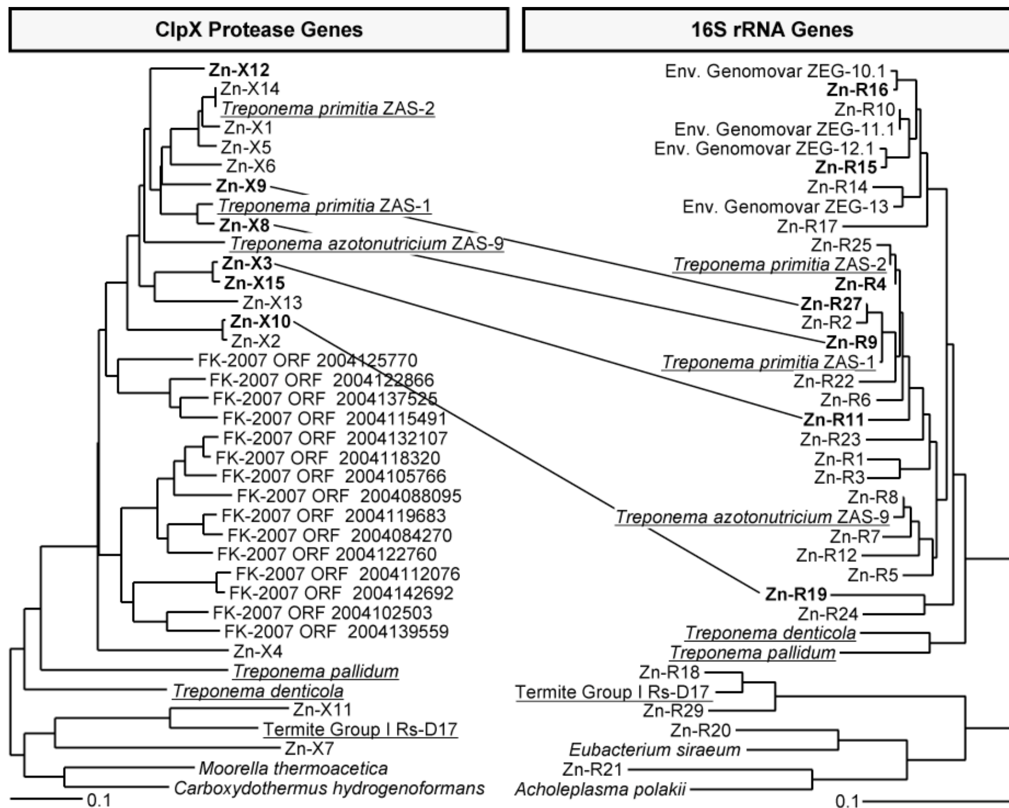

Figure 5.3. Phylogenetic analysis of ClpX and 16S rRNA gene sequences amplified using microfluidic digital PCR. Trees calculated using Phylip distance Fitch algorithm. **Left,** A ClpX gene tree constructed using 397 unambiguous, aligned base pairs. Sequences from the *Nasutitermes* metagenome named as FK-2007 ORF [JGI database GOI] **Right,** 16S rRNA gene tree (calculation described in Figure 5.2). Scale bars represent 0.1 changes per alignment position. Lines identify ClpX-16S rRNA gene pairs supported by repeated colocalizations or similarity to established associations (Table 5.3). Sequences assigned to an environmental genomovar (Table 5.5) marked in bold.

Table 5.3. ClpX/16S rRNA gene pairs

Experiment	ClpX	16S rRNA	Comment
1	Zn-X1	Zn-R22	
2	Zn-X2	Zn-R23	
3	Zn-X3	Zn-R11	Triplex Zn-F8 to Zn-R11 to Zn-X3, X15; ZEG 16.2
4	Zn-X4	Zn-R24	
5	Zn-X5	Zn-R25	
6	Zn-X6	Zn-R18	
7	Zn-X7	Zn-R11	
8	Zn-X8	Zn-R9	Similar to ZAS-1; ZEG 15.2
9	Zn-X9	Zn-R27	Repeated in 10; ZEG 17.1
10	Zn-X9	Zn-R27	Repeated in 9; ZEG 17.1
11	Zn-X10	Zn-R19	Repeated in 13, 14; ZEG 18.1
12	Zn-X11	Zn-R29	Termite Group I
13	Zn-X10	Zn-R19	Repeated in 11, 14; ZEG 18.1
14	Zn-X10	Zn-R19	Repeated in 11, 13; ZEG 18.1

To demonstrate the general applicability of the universal-template probe strategy to degenerate primer based PCR, we designed a primer set to amplify the gene for the ATP-binding subunit of the Clp protease complex (ClpX) from termite gut *Treponemes*. ClpX was chosen as a target because it is highly conserved, a *Treponeme* sequence cluster for this gene was clearly identifiable in the *Nasutitermes* sp. metagenomic dataset, and prior experiments in this laboratory (Matson and Leadbetter, manuscript in preparation) have demonstrated this gene's utility as an internal standard for quantitative PCR-based transcriptional analyses. Multiplex PCR reactions containing *Z. nevadensis* hindgut contents were carried out as described for FTHFS and 16S rRNA. ClpX-16S rRNA associations are presented in Figure 5.3 and Table 5.3. A limited number of FTHFS-ClpX association experiments were also carried out (Figure 5.4, Table 5.4).

Table 5.4. FTHFS/ClpX gene pairs[a]

Experiment	FTHFS	ClpX	Comment
1	Zn-F10	Zn-X12	Repeated in 2; ZEG 12.6
2	Zn-F10	Zn-X12	Repeated in 1; ZEG 12.6
3	Zn-F18	Zn-X2	
4	Zn-F13	Zn-X13	
5	Zn-F10	Zn-X6	
6	Zn-F10	Zn-X14	
7	Zn-F8	Zn-X14	
8	Zn-F8	Zn-X15	Triplex Zn-F8 to Zn-R11 to Zn-X3, X15; ZEG 16.3

[a] Colocalizations 1-4 were retrieved from reactions using ClpX and the H-group specific FTHFS primer/probe set described in (26). All other colocalizations used the universal template probe FTHFS primer set.

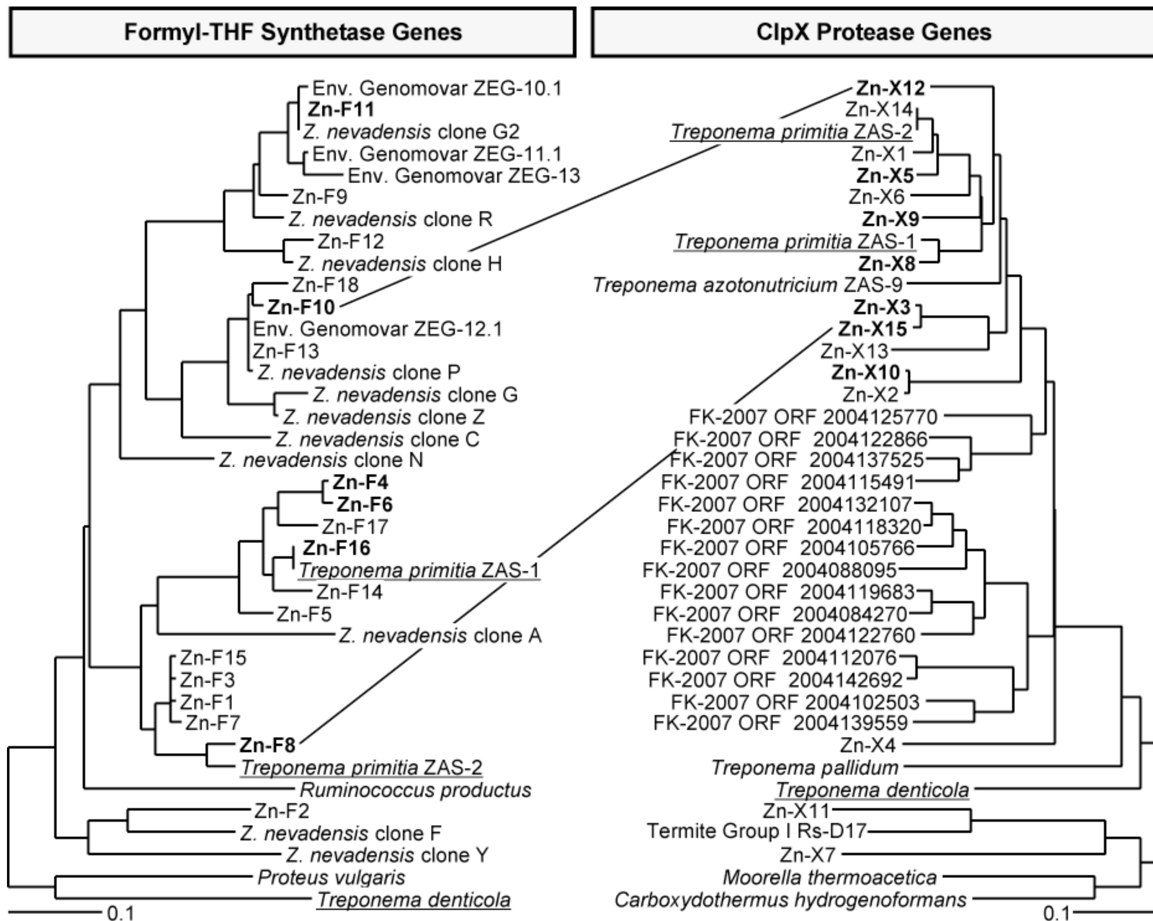

Figure 5.4. Phylogenetic analysis of FTHFS and ClpX gene sequences amplified using microfluidic digital PCR. Trees calculated as described in Figures 5.2 and 5.3. Scale bars represent 0.1 changes per alignment position. Lines identify FTHFS-ClpX gene pairs supported by repeated colocalizations or similarity to established associations (Table 5.4). Sequences assigned to an environmental genomovar (Table 5.5) marked in bold.

In our previous microfluidic experiments, we proposed the term *"environmental genomovar"* to describe uncultured organisms that have been shown to encode particular gene combinations. New environmental genomovars proposed based on the results of these experiments have been assigned monikers ZEG-14 through 18, as listed in Table 5.4. These include the first triplex association, as the FTHFS and 16S rRNA genes from ZEG 16 have been independently associated with highly similar (98%) ClpX sequences.

Table 5.5. Proposed Environmental Genomovars

Name	16S rRNA	FTHFS	ClpX
ZEG 10.5	Zn-R16	Zn-F11	
ZEG 12.6	Zn-R15	Zn-F10	
ZEG 12.7		Zn-F10	Zn-X12
ZEG 14.1	Zn-R4	Zn-F4	
ZEG 14.2	Zn-R4	Zn-F6	
ZEG 15.1	Zn-R9	Zn-F16	
ZEG 15.2	Zn-R9		Zn-X8
ZEG 16.1	Zn-R11	Zn-F8	
ZEG 16.2	Zn-R11		Zn-X3
ZEG 16.3		Zn-F8	Zn-X15
ZEG 17.1	Zn-R27		Zn-X9
ZEG 18.1	Zn-R19		Zn-X10

Discussion

We have developed a strategy for multiplex, microfluidic digital PCR that allows simultaneous amplification and detection of "Lovell cluster" FTHFS genes and bacterial 16S rRNA genes. These primers were used in combination with "all-bacterial" 16S rRNA gene primers to discover the species identity of uncultured, FTHFS-bearing bacteria in the termite gut. An assay targeting the ATP-binding subunit of the Clp protease complex (ClpX) from termite gut *Treponemes* was also developed. This primer/probe set was used to associate ClpX genes with both 16S rRNA gene sequences and FTHFS genes.

The last few decades have seen widespread use of degenerate primers to build environmental inventories of genes encoding enzymes involved in key environmental processes such as nitrogen fixation (*nifH* gene) (45, 46), methanotrophy (*pmoA, mmoX, mxaF*) (8), and sufate reduction (*dsrAB*) (41). The ability to carry out such analyses at the level of single environmental cells, however, should greatly expand the information

derived from such techniques. With the use of universal template probes and degenerate primers, microfluidic digital PCR now allows the construction of such inventories in parallel with 16S rRNA analysis, placing the targeted genetic capability within the context of the phylogenetic species identity of the host and the environmental species assemblage. Additionally, the "one template, one reaction" nature of this technique circumvents some of the technical caveats of environmental inventories (1), eliminating chimeric product formation and minimizing the role of PCR bias in determining library composition.

The utility of this approach is not limited to assignment of 16S rRNA species identities to organisms bearing a genetic capacity of interest. In these experiments, we also built associations between FTHFS and ClpX genes. The gene for ClpX protease exhibits steady-state expression during growth of *Treponema primitia* strain ZAS-2, and might therefore be useful as an internal standard for quantitative expression analyses (Matson and Leadbetter, manuscript in preparation). Cross-sample comparison of expression levels among uncultured environmental microbes is currently based on normalization of total RNA concentrations (45), a metric that is highly sensitive to RNA sample quality (10). The incorporation of invariant control transcripts should greatly enhance the resolution of environmental expression analyses.

In conclusion, we have developed a microfluidic digital PCR technique that allows the highly parallel interrogation of individual environmental cells using multiplex, degenerate primers. We have used this technique to simultaneously inventory and

identify (based on rRNA species phylogeny) acetogenic bacteria in termite hindgut samples. The ability to build metabolic gene inventories from environmental samples while simultaneously identifying the ribosomal phylotype of the organisms that carry these genes will greatly enhance the utility of PCR-based molecular community profiling. The ability to carry out in-depth analyses targeting major metabolic guilds is highly complementary to environmental genomic and metagenomic analyses, and will continue as an important element of the microbial ecologist's arsenal.

Chapter Five Appendix

Table 5.6. Sequences used in phylogenetic analysis

Source/Sequence Type	Designation	Accession	Gene	Reference
Z. nevadensis Gut Clone	A	FTHFS	AY162294	(34)
Z. nevadensis Gut Clone	C	FTHFS	AY162295	(34)
Z. nevadensis Gut Clone	F	FTHFS	AY162298	(34)
Z. nevadensis Gut Clone	G	FTHFS	AY162300	(34)
Z. nevadensis Gut Clone	G2	FTHFS	AY162301	(34)
Z. nevadensis Gut Clone	H	FTHFS	AY162302	(34)
Z. nevadensis Gut Clone	N	FTHFS	AY162306	(34)
Z. nevadensis Gut Clone	P	FTHFS	AY162307	(34)
Z. nevadensis Gut Clone	R	FTHFS	AY162308	(34)
Z. nevadensis Gut Clone	Y	FTHFS	AY162311	(34)
Z. nevadensis Gut Clone	Z	FTHFS	AY162312	(34)
Z. nevadensis Genomovar	ZEG 10.1	FTHFS	DQ420342	(26)
Z. nevadensis Genomovar	ZEG 11.1	FTHFS	DQ420346	(26)
Z. nevadensis Genomovar	ZEG 12.1	FTHFS	DQ420353	(26)
Z. nevadensis Genomovar	ZEG 13.1	FTHFS	DQ420358	(26)
Proteus vulgaris		FTHFS	AF295710	(19)
Ruminococcus productus		FTHFS	AF295707	(19)
Treponema denticola		FTHFS	NC_002967	(35)
Treponema primitia ZAS-1	ZAS-1a	FTHFS	AY162313	(34)
Treponema primitia ZAS-2	ZAS-2	FTHFS	AY162315	(34)
Acholeplasma polakii		16S	AF031479	(2)
Eubacterium siraeum	ATCC 29066	16S	L34625	
Termite Group I bacterium	Rs-D17	16S	AB089048	(14)
Treponema azotonutricium ZAS-9	ZAS-9	16S	AF320287	(21)
Treponema denticola	II:11:33520	16S	M71236	(28)
Treponema pallidum	Nichols	16S	M88726	(28)
Treponema primitia ZAS-1	ZAS-1	16S	AF093251	(17)
Treponema primitia ZAS-2	ZAS-2	16S	AF093252	(17)
Z. nevadensis Genomovar	ZEG 10.1	16S	DQ420325	(26)
Z. nevadensis Genomovar	ZEG 11.1	16S	DQ420329	(26)
Z. nevadensis Genomovar	ZEG 12.1	16S	DQ420336	(26)
Z. nevadensis Genomovar	ZEG 13.1	16S	DQ420341	(26)
Carboxydothermus hydrogenoformans	Z-2901	ClpX	NC_007503.1	(44)
Moorella thermoacetica	ATCC 39703	ClpX	NC_007644	(31)
Nasutitermes sp. FK-2007 metagenome	2004084270	ClpX	JGI GOI 2004084270	(42)
Nasutitermes sp. FK-2007 metagenome	2004088095	ClpX	JGI GOI 2004088095	(42)
Nasutitermes sp. FK-2007 metagenome	2004102503	ClpX	JGI GOI 2004102503	(42)
Nasutitermes sp. FK-2007 metagenome	2004105766	ClpX	JGI GOI 2004105766	(42)
Nasutitermes sp. FK-2007 metagenome	2004112076	ClpX	JGI GOI 2004112076	(42)
Nasutitermes sp. FK-2007 metagenome	2004115491	ClpX	JGI GOI 2004115491	(42)
Nasutitermes sp. FK-2007 metagenome	2004118320	ClpX	JGI GOI 2004118320	(42)
Nasutitermes sp. FK-2007 metagenome	2004119683	ClpX	JGI GOI 2004119683	(42)
Nasutitermes sp. FK-2007 metagenome	2004122760	ClpX	JGI GOI 2004122760	(42)
Nasutitermes sp. FK-2007 metagenome	2004122866	ClpX	JGI GOI 2004122866	(42)
Nasutitermes sp. FK-2007 metagenome	2004125770	ClpX	JGI GOI 2004125770	(42)
Nasutitermes sp. FK-2007 metagenome	2004132107	ClpX	JGI GOI 2004132107	(42)

Source/Sequence Type	Designation	Accession	Gene	Reference
Nasutitermes sp. FK-2007 metagenome	2004137525	ClpX	JGI GOI 2004137525	(42)
Nasutitermes sp. FK-2007 metagenome	2004139559	ClpX	JGI GOI 2004139559	(42)
Nasutitermes sp. FK-2007 metagenome	2004142692	ClpX	JGI GOI 2004142692	(42)
Termite Group I bacterium	Rs-D17	ClpX	AP009510	(15)
Treponema denticola		ClpX	NC_002967	(35)
Treponema pallidum		ClpX	NC_000919	(11)

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. **Angulo, A. F., R. Reijgers, J. Brugman, I. Kroesen, F. E. Hekkens, P. Carle, J. M. Bove, J. G. Tully, A. C. Hill, L. M. Schouls, C. S. Schot, P. J. Roholl, and A. A. Polak-Vogelzang.** 2000. *Acholeplasma vituli* sp. nov., from bovine serum and cell cultures. Int J Syst Evol Microbiol **50 Pt 3:**1125-1131.
3. **Béjà, O., E. N. Spudich, J. L. Spudich, M. Leclerc, and E. F. DeLong.** 2001. Proteorhodopsin phototrophy in the ocean. Nature **411:**786-789.
4. **Breznak, J. A.** 1994. Acetogenesis from carbon dioxide in termite guts, p. 303-330. *In* H. A. Drake (ed.), Acetogenesis. Chapman and Hall, New York.
5. **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, Third ed, vol. I. Springer.
6. **Corless, C. E., M. Guiver, R. Borrow, V. Edwards-Jones, E. B. Kaczmarski, and A. J. Fox.** 2000. Contamination and sensitivity issues with a real-time universal 16S rRNA PCR. J Clin Microbiol **38:**1747-1752.
7. **Dhillon, A., A. Teske, J. Dillon, D. A. Stahl, and M. L. Sogin.** 2003. Molecular characterization of sulfate-reducing bacteria in the Guaymas Basin. Appl Environ Microbiol **69:**2765-2772.

8. **Dumont, M. G., and J. C. Murrell.** 2005. Community-level analysis: key genes of aerobic methane oxidation. Methods Enzymol **397:**413-427.

9. **Elshahed, M. S., N. H. Youssef, A. M. Spain, C. Sheik, F. Z. Najar, L. O. Sukharnikov, B. A. Roe, J. P. Davis, P. D. Schloss, V. L. Bailey, and L. R. Krumholz.** 2008. Novelty and uniqueness patterns of rare members of the soil biosphere. Appl Environ Microbiol **74:**5422-5428.

10. **Fleige, S., and M. W. Pfaffl.** 2006. RNA integrity and the effect on the real-time qRT-PCR performance. Mol Aspects Med **27:**126-139.

11. **Fraser, C. M., S. J. Norris, G. M. Weinstock, O. White, G. G. Sutton, R. Dodson, M. Gwinn, E. K. Hickey, R. Clayton, K. A. Ketchum, E. Sodergren, J. M. Hardham, M. P. McLeod, S. Salzberg, J. Peterson, H. Khalak, D. Richardson, J. K. Howell, M. Chidambaram, T. Utterback, L. McDonald, P. Artiach, C. Bowman, M. D. Cotton, C. Fujii, S. Garland, B. Hatch, K. Horst, K. Roberts, M. Sandusky, J. Weidman, H. O. Smith, and J. C. Venter.** 1998. Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. Science **281:**375-388.

12. **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.

13. **Henckel, T., U. Jackel, S. Schnell, and R. Conrad.** 2000. Molecular analyses of novel methanotrophic communities in forest soil that oxidize atmospheric methane. Appl Environ Microbiol **66:**1801-1808.

14. **Hongoh, Y., M. Ohkuma, and T. Kudo.** 2003. Molecular analysis of bacterial microbiota in the gut of the termite *Reticulitermes speratus* (Isoptera; Rhinotermitidae). FEMS Microbiol Ecol **44:**231-242.

15. **Hongoh, Y., V. K. Sharma, T. Prakash, S. Noda, T. D. Taylor, T. Kudo, Y. Sakaki, A. Toyoda, M. Hattori, and M. Ohkuma.** 2008. Complete genome of the uncultured Termite Group 1 bacteria in a single host protist cell. Proc Natl Acad Sci USA **105:**5555-5560.

16. **Hugenholtz, P., B. M. Goebel, and N. R. Pace.** 1998. Impact of culture-independent studies on the emerging phylogenetic view of bacterial diversity. J Bacteriol **180:**4765-4774.

17. **Leadbetter, J. R., T. M. Schmidt, J. R. Graber, and J. A. Breznak.** 1999. Acetogenesis from H_2 plus CO_2 by spirochetes from termite guts. Science **283:**686-689.

18. **Leaphart, A. B., M. J. Friez, and C. R. Lovell.** 2003. Formyltetrahydrofolate synthetase sequences from salt marsh plant roots reveal a diversity of acetogenic bacteria and other bacterial functional groups. Appl Environ Microbiol **69:**693-696.

19. **Leaphart, A. B., and C. R. Lovell.** 2001. Recovery and analysis of formyltetrahydrofolate synthetase gene sequences from natural populations of acetogenic bacteria. Appl Environ Microbiol **67:**1392-1395.

20. **Ley, R. E., M. Hamady, C. Lozupone, P. J. Turnbaugh, R. R. Ramey, J. S. Bircher, M. L. Schlegel, T. A. Tucker, M. D. Schrenzel, R. Knight, and J. I.**

Gordon. 2008. Evolution of mammals and their gut microbes. Science **320:**1647-1651.

21. **Lilburn, T. G., K. S. Kim, N. E. Ostrom, K. R. Byzek, J. R. Leadbetter, and J. A. Breznak.** 2001. Nitrogen fixation by symbiotic and free-living spirochetes. Science **292:**2495-2498.

22. **Lovell, C. R., and A. B. Leaphart.** 2005. Community-level analysis: key genes of CO2-reductive acetogenesis. Methods Enzymol **397:**454-469.

23. **Ludwig, W., O. Strunk, R. Westram, L. Richter, H. Meier, Yadhukumar, A. Buchner, T. Lai, S. Steppi, G. Jobb, W. Förster, I. Brettske, S. Gerber, A. W. Ginhart, O. Gross, S. Grumann, S. Hermann, R. Jost, A. König, T. Liss, R. Lüßmann, M. May, B. Nonhoff, B. Reichel, R. Strehlow, A. Stamatakis, N. Stuckmann, A. Vilbig, M. Lenke, T. Ludwig, A. Bode, and K. H. Schleifer.** 2004. ARB: a software environment for sequence data. Nucl Acids Res **32:**1363-1371.

24. **Marcy, Y., C. Ouverney, E. M. Bik, T. Lösekann, N. Ivanova, H. G. Martin, E. Szeto, D. Platt, P. Hugenholtz, D. A. Relman, and S. R. Quake.** 2007. Dissecting biological "dark matter" with single-cell genetic analysis of rare and uncultivated TM7 microbes from the human mouth. Proc Natl Acad Sci USA **104:**11889-11894.

25. **Nocker, A., M. Burr, and A. K. Camper.** 2007. Genotypic microbial community profiling: a critical technical review. Microb Ecol **54:**276-289.

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. R.** 1997. A molecular view of microbial diversity and the biosphere. Science **276:**734-740.

28. **Paster, B. J., F. E. Dewhirst, W. G. Weisburg, L. A. Tordoff, G. J. Fraser, R. B. Hespell, T. B. Stanton, L. Zablen, L. Mandelco, and C. R. Woese.** 1991. Phylogenetic analysis of the spirochetes. J Bacteriol **173:**6101-6109.

29. **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.

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

31. **Pierce, E., G. Xie, R. D. Barabote, E. Saunders, C. S. Han, J. C. Detter, P. Richardson, T. S. Brettin, A. Das, L. G. Ljungdahl, and S. W. Ragsdale.** 2008. The complete genome sequence of *Moorella thermoacetica* (f. *Clostridium thermoaceticum*). Environ Microbiol **10:**2550-2573.

32. **Riesenfeld, C. S., P. D. Schloss, and J. Handelsman.** 2004. Metagenomics: Genomic analysis of microbial communities. Ann Rev Genet **38:**525-552.

33. **Rose, T. M., E. R. Schultz, J. G. Henikoff, S. Pietrokovski, C. M. McCallum, and S. Henikoff.** 1998. Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. Nucleic Acids Res **26:**1628-1635.

34. **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*. Microbiol **149:**2529-2537.

35. **Seshadri, R., G. S. Myers, H. Tettelin, J. A. Eisen, J. F. Heidelberg, R. J. Dodson, T. M. Davidsen, R. T. DeBoy, D. E. Fouts, D. H. Haft, J. Selengut, Q. Ren, L. M. Brinkac, R. Madupu, J. Kolonay, S. A. Durkin, S. C. Daugherty, J. Shetty, A. Shvartsbeyn, E. Gebregeorgis, K. Geer, G. Tsegaye, J. Malek, B. Ayodeji, S. Shatsman, M. P. McLeod, D. Smajs, J. K. Howell, S. Pal, A. Amin, P. Vashisth, T. Z. McNeill, Q. Xiang, E. Sodergren, E. Baca, G. M. Weinstock, S. J. Norris, C. M. Fraser, and I. T. Paulsen.** 2004. Comparison of the genome of the oral pathogen *Treponema denticola* with other spirochete genomes. Proc Natl Acad Sci USA **101:**5646-5651.

36. **Sogin, M. L., H. G. Morrison, J. A. Huber, D. Mark Welch, S. M. Huse, P. R. Neal, J. M. Arrieta, and G. J. Herndl.** 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". Proc Natl Acad Sci USA **103:**12115-12120.

37. **Stepanauskas, R., and M. E. Sieracki.** 2007. Matching phylogeny and metabolism in the uncultured marine bacteria, one cell at a time. Proc Natl Acad Sci USA **104:**9052-9057.

38. **Streit, W. R., and R. A. Schmitz.** 2004. Metagenomics--the key to the uncultured microbes. Curr Opin Microbiol **7:**492-498.

39. **Suzuki, M. T., L. T. Taylor, and E. F. DeLong.** 2000. Quantitative analysis of small-subunit rRNA genes in mixed microbial populations via 5'-nuclease assays. Appl Environ Microbiol **66:**4605-4014.

40. **Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y. H. Rogers, and H. O. Smith.** 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science **304:**66-74.

41. **Wagner, M., A. Loy, M. Klein, N. Lee, N. B. Ramsing, D. A. Stahl, and M. W. Friedrich.** 2005. Functional marker genes for identification of sulfate-reducing prokaryotes. Methods Enzymol **397:**469-489.

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

43. **Winston, P. W., Bates, Donald H.** 1960. Saturated solutions for the control of humidity in biological research. Ecology **41:**232-237.

44. **Wu, M., Q. H. Ren, A. S. Durkin, S. C. Daugherty, L. M. Brinkac, R. J. Dodson, R. Madupu, S. A. Sullivan, J. F. Kolonay, W. C. Nelson, L. J. Tallon, K. M. Jones, L. E. Ulrich, J. M. Gonzalez, I. B. Zhulin, F. T. Robb, and J. A. Eisen.** 2005. Life in hot carbon monoxide: The complete genome sequence of *Carboxydothermus hydrogenoformans* Z-2901. PLoS Gen **1:**563-574.

45. **Zehr, J. P., B. D. Jenkins, S. M. Short, and G. F. Steward.** 2003. Nitrogenase gene diversity and microbial community structure: a cross-system comparison. Environ Microbiol **5:**539-554.

46. **Zehr, J. P., and L. A. McReynolds.** 1989. Use of degenerate oligonucleotides for amplification of the nifH gene from the marine cyanobacterium *Trichodesmium thiebautii*. Appl Environ Microbiol **55:**2522-2526.

47. **Zeidner, G., C. M. Preston, E. F. Delong, R. Massana, A. F. Post, D. J. Scanlan, and O. Béjà.** 2003. Molecular diversity among marine picophytoplankton as revealed by psbA analyses. Environ Microbiol **5:**212-216.

48. **Zhang, K., A. C. Martiny, N. B. Reppas, K. W. Barry, J. Malek, S. W. Chisholm, and G. M. Church.** 2006. Sequencing genomes from single cells by polymerase cloning. Nat Biotechnol **24:**680-686.

49. **Zhang, Y., D. Zhang, W. Li, J. Chen, Y. Peng, and W. Cao.** 2003. A novel real-time quantitative PCR method using attached universal template probe. Nucleic Acids Res **31:**e123.