The ISME Journal (2011) 5, 1133–1142

www.nature.com/ismej

ORIGINAL ARTICLE

RNA-seq reveals cooperative metabolic interactions between two termite-gut spirochete species in co-culture

Adam Z Rosenthal[1], Eric G Matson[1], Avigdor Eldar[2] and Jared R Leadbetter[1]
[1]*Ronald and Maxine Linde Center for Global Environmental Science, California Institute of Technology, Mailcode 138-78, Pasadena, CA, USA and* [2]*Howard Hughes Medical Institute and Division of Biology and Department of Applied Physics, California Institute of Technology, Pasadena, CA, USA*

The hindguts of wood-feeding termites typically contain hundreds of microbial species. Together with their insect host, these gut microbes degrade lignocellulose into usable catabolites. Although past research revealed many facets of the stepwise flow of metabolites in this scheme, not much is known about the breadth of interactions occurring between termite-gut microbes. Most of these microbes are thought to depend on, and to have co-speciated with, their host and each other for millions of years. In this study, we explored the interactions of two spirochetes previously isolated from the very same termite species. As hydrogen (H_2) is the central free intermediate in termite-gut lignocellulose digestion, we focused on interactions between two closely related termite-gut spirochetes possessing complementary H_2 physiologies: one produces H_2, while the other consumes it. *In vitro*, these two *Treponema* species markedly enhanced each other's growth. RNA sequencing resolved the transcriptomes of these two closely related organisms, revealing that co-cultivation causes comprehensive changes in global gene expression. The expression of well over a 100 genes in each species was changed >twofold, with over a dozen changed >10-fold. Several changes implicating synergistic cross-feeding of known metabolites were validated *in vitro*. Additionally, certain activities beneficial to the host were preferentially expressed during consortial growth. However, the majority of changes in gene expression are not yet understandable, but indicate a broad, comprehensive and mutualistic interaction between these closely related, co-resident gut symbionts. The results suggest that staggeringly intricate networks of metabolic and gene interactions drive lignocellulose degradation and co-evolution of termite gut microbiota.
The ISME Journal (2011) **5,** 1133–1142; doi:10.1038/ISMEJ.2011.3; published online 17 February 2011
Subject Category: microbe–microbe and microbe–host interactions
Keywords: co-culture; RNA Seq; symbiosis; termite-gut

Introduction

A bacterium's capacity to perform physiological tasks in isolation is not necessarily predictive of how it performs in the natural environment. Take for example the classic model bacterium, *Escherichia coli*, a versatile organism that degrades many substrates under many growth conditions in pure culture (Boone *et al.*, 2001). In environmental context, much less is known about its physiology (Chang *et al.*, 2004). *E. coli* probably degrades only a few substrates *in situ*, because it is outnumbered and likely outprocessed by a diversity of other specialists. Likewise, some species do not express certain traits when grown alone, and what appears to be a cryptic activity only blooms when growth occurs in the presence of other species (Straight *et al.*, 2007). Lastly, host-associated organisms may respond to cues from their host, and in response catalyze activities that impact the host (Palmer *et al.*, 2007). Neither of these activities are typically recapitulated during standard cultivation regimes.

Physiological analyses of model organisms, like *E. coli* above, have greatly benefited from high-throughput tools, especially microarrays (Chang *et al.*, 2004). However, arrays require extensive design and ground-truthing before even the first experiment can be performed. These constraints are significant, considering that the number of candidate organisms for expression studies has exploded in this era of genome sequencing. Moreover, any successful interpretation of array data becomes complicated when investigating global gene expression of closely related species (sharing high nucleic acid similarities) grown in either defined co-cultures or their natural habitats. The probe-binding nature of the microarray approach is not well suited to distinguish between conserved

Correspondence: JR Leadbetter, Ronald and Maxine Linde Center for Global Environmental Science, California Institute of Technology, Mailcode 138-78, Pasadena, CA 91125, USA.
E-mail: jleadbetter@caltech.edu
Received 26 August 2010; revised 16 November 2010; accepted 22 December 2010; published online 17 February 2011

nucleotide sequences. Thus, array-based studies on co-cultures and communities are essentially limited to examples involving species having very different genomes (Saleh-Lakha *et al.*, 2005).

Breakthroughs in next-generation sequencing not only allow the comparison of expression from highly similar sequences, but also dramatically expedite the pace at which data can be gathered from less-well-studied microbes (Holt *et al.*, 2008; Yoder-Himes *et al.*, 2009) and complex biological environments (Gilbert *et al.*, 2008; Urich *et al.*, 2008). Although transcriptomic calculations are now possible, studies of co-cultures and intact environments are challenging, as they entail inherent difficulties when making comparisons and contrasts. For one, it is very difficult to directly compare the gene expression patterns of cultures when the growth conditions are very dissimilar, and this is the case when environmentally reliant symbionts are removed from their environment—or even one from another in defined co-culture. Furthermore, common RNA-seq techniques do not differentiate the direction in which a transcript is oriented, and there appears to be times when anti-coding RNAs are present, which may not be fully accounted for by many studies (Dornenburg *et al.*, 2010). However, studies of transcriptional output in different conditions do allow one to analyze the impact of individual microbes at a particular point of time, and to produce testable hypotheses about the functions of different members of a microbial community, issues at the crux of molecular microbial ecology.

In this study, we combined the use of classic microbiological cultivation techniques with next-generation sequencing to explore gene expression patterns underlying the metabolic interactions between closely related bacteria that had been isolated from the same microliter-in-scale termite hindgut environment.

The gut communities of wood-feeding termites ferment the polysaccharide fractions of lignocellulose, ultimately providing the termite with compounds such as carbon and energy substrate acetate, amino acids and other essential growth factors (Odelson and Breznak, 1983; Brauman *et al.*, 1992; Abe *et al.*, 2000). During the course of lignocellulose fermentation, the free intermediate hydrogen (H_2) is produced and accumulates to concentrations near saturation (Pester and Brune, 2007), before being consumed by homoacetogenic gut bacteria and methanogenic archaea (Breznak and Switzer, 1986; Brauman *et al.*, 1992). It has been estimated that $H_2 + CO_2$ acetogenic bacteria supply their termite host with up to a third of its acetate supply (Breznak and Switzer, 1986). Together, the fermentation and CO_2-reductive acetogenesis-derived acetate can contribute up to 70–100% of the termite host's energy requirements (Drake, 1994).

In this study, we have sought to reveal possible interactions occurring between two previously isolated termite-gut treponemes: *Treponema primitia* ZAS-2 and *Treponema azotonutricium* ZAS-9. These two species reside in close proximity in the hindgut of the dampwood termite *Zootermopsis angusticolis* (Graber *et al.*, 2004), wherein (as in most termites) a diversity of spirochetes constitute a large portion of the total gut microbiota (Breznak, 2002). Previous studies suggest that these bacteria have separate physiological roles in the complex mutualism (Leadbetter *et al.*, 1999; Lilburn *et al.*, 2001; Graber and Breznak, 2004; Graber *et al.*, 2004). However, to what extent they interact with and impact each other and other gut species is less well examined. Certainly, *T. primitia* is known to depend on an essential growth factor, folate, derived from other gut microbiota (Graber and Breznak, 2005). *T. azotonutricium* fixes nitrogen (Lilburn *et al.*, 2001) and produces H_2 as a fermentation product (Graber *et al.*, 2004), whereas *T. primitia* is an H_2-consuming CO_2-reducing homoacetogen (Leadbetter *et al.*, 1999). Their complementary H_2 physiologies and their close proximity to each other in the small termite-gut environment begs the question of whether they might engage in other interactions. The genomes of *T. azotonutricium* and *T. primitia* are relatively large (3.86 Mb and 4.06 Mb, respectively (Graber *et al.*, 2004) compared with other sequenced treponemes (*T. denticola* 2.84 Mb (Seshadri *et al.*, 2004); *T. pallidum* 1.14 Mb (Fraser *et al.*, 1998)). We sought to identify genes that may be relevant to interspecies interactions by analyzing global gene expression during consortial growth of the two, via Illumina transcript-sequencing technology (Illumina, San Diego, CA, USA).

Materials and methods

Bacterial strains and culture media

T. primitia (ZAS-2) and *T. azotonutricium* (ZAS-9) cultures were grown in 4-YACo media (4% yeast autolysate) supplemented with 20 mM maltose and 80% H_2 and 20% CO_2 in the headspace. Cultures were grown in 5 ml volumes in 25-ml Balch tubes, with crimp top stoppers in the dark at room temperature. Vitamin, amino-acid and cofactor preparations used to supplement media were as follows:

Vitamin B7: 0.3–0.5 μg biotin (Sigma Aldrich, St Louis, MO, USA) was added per culture tube.

Vitamin B6: 120–200 μg pyridoxal-HCl and pyridoxal-phosphate (Sigma Aldrich) were added per culture tube.

B12 and corrinoids: 30–50 μg of one of the following: (1) hydroxocobalamin acetate salt, (2) hydroxocobalamin hydrochloride, (3) methylcobalamin and (4) cyanocobalamin (Sigma Aldrich), was added per culture tube.

Tryptophan: 60–100 μg tryptophan (Sigma Aldrich) was added per culture tube.

RNA isolation and processing
Total RNA was isolated from two biological replicates per growth condition. Procedures for sample preparation followed from the standard Illumina protocol for RNA-Seq sample preparation available from the manufacturer (Illumina). In short, RNA was isolated using RNeasy kit (Qiagen, Valencia, CA, USA) as per the manufacturer's protocol. Samples were run through the RNeasy procedure twice, with the optional DNase treatment performed in both instances. Total RNA was fragmented using the Ambion RNA fragmentation kit (Ambion, Austin, TX, USA) and protocol. First-strand complementary DNA was prepared following the SuperScript II method (Invitrogen, Carlsbad, CA, USA), using the Invitrogen hexamer random primers to ensure low primer bias. Second-strand complementary DNA was subsequently synthesized by adding to the first-strand reaction second-strand buffer (500 mM Tris-HCl pH 7.8, 50 mM $MgCl_2$, 10 mM dithiothreitol), deoxyribonucleotide triphosphate (0.3 mm), RNaseH (2 U μl^{-1}; Invitrogen #18021-014), Invitrogen hexamer primers and DNA polymerase I (Invitrogen). The final reaction volume for second-strand synthesis was 100μl, and reactions were carried out at 16 °C for 2.5 h.

Complementary DNA sequencing
Fragmented second-strand complementary DNA samples were submitted to the Caltech Sequencing Core facility (Pasadena, CA, USA). Libraries were sequenced as 37-mers using the standard Solexa (Illumina) protocol and pipeline. Sequencing depth information is summarized in Supplementary Table 3.

RNA-Seq data analysis
Illumina raw data provided by the GERALD (Illumina) software package were aligned to a FASTA file containing both *T. azotonutricium* (ZAS-9) and *T. primitia* (ZAS-2) genomes using the Maq short read aligning program (Wellcome Trust Sanger Institute, Hinxton, UK). For initial analysis of mapping quality (see Table 1), zero or one misses were allowed per read. All subsequent samples were analyzed with a maximum of one mismatch allowed.

Reads from biological replicates were first compared with each other (a) graphically after mapping onto the two genomes, and then (b) by looking for differences in fold regulation when compared in all pairwise combinations of other replicates of interest. Biological replicates broadly were far more similar to each other than to other samples. Reads from biological replicates were merged and averaged for all further analysis.

To exclude reads that may align ambiguously (that is, to either genome), we assembled hypothetical databases of all possible 37′-mer sequences for each genome, and mapped them to the opposing genome with one mismatch allowed. Additionally, a database of pooled Illumina sequencing data from each pure culture was also aligned to the opposing genome. Reads that ambiguously mapped were excluded from transcriptional analysis. Gene expression values were determined by normalizing the number of reads mapped to a particular gene (excluding ambiguous regions) divided by the size of the gene (also excluding ambiguous regions). The resulting value is the normalized reads per kilobase, in a manner consistent with the gene expression index calculations of previously published reports (Yoder-Himes *et al.*, 2009). In order to adjust for intensity between samples, the ribosomal signal from each sample was used as a standard, and each sample's intensity was multiplied by a factor that would yield an equal ribosomal RNA (rRNA) signal.

In considering up- or downregulated genes, a cutoff of twofold increase in transcription was used.

Table 1 A hypothetical short sequence data set and experimental RNA-sequencing data preferentially aligned to the cognate genome

	Total DB size	*Total hits (% rRNA)*	*Non-rRNA hits*	*Unique loci*[a]	*Genes with a hit (%)*
Hypothetical					
Exact	3 855 671	6518 (45%)	2923	2923	109 (2.8%)
1 miss	3 855 671	14 086 (27%)	10 283	10 283	290 (7.6%)
Actual[b]					
Before mask	10 943 994	1 936 998 (99.9%)	1525	575	161 (4.1%)
After mask[c]	542 600	646 (0%)	646	340	139 (3.6%)
Actual[d]					
Before mask	15 151 014	2 400 846 (99.9%)	1280	409	194 (4.8%)
After mask[c]	427 444	380 (0%)	380	204	151 (3.8%)

Abbreviation: rRNA, ribosomal RNA.
[a]Unique loci refer to the number of distinct non-ribosomal sequences of *T. primitia* that had at least one hit.
An *in-silico*-generated data set of all possible 37 base-pair sequences in the genome of *T. azotonutricium* (hypothetical) and the RNA-seq data from a sample of *T. azotonutricium* (denoted as 'Actual[b]' in the table) were mapped to the genome of *T. primitia*. Database size describes the number of short sequences in the data sets. The total hits column displays the number of short sequences that mapped to the *T. primitia* genome, and the percentage of hits that align to ribosomal 16S or 23S.
[c]After mask sequences are those remaining after the most similar sequences between the two genomes were removed from sampling.
[d]Refers to the RNA-Seq data from a sample of *T. primitia*, when mapped to the genome of *T. azotonutricium*.

Additionally, except in the case of functional groups and gene clusters used to generate Supplementary Table 1, only genes with greater than 50 adjusted hits per kb of coding DNA were considered in analyses.

Signal intensities were visualized graphically by converting Maq-aligned reads into a .BAR file using the Cisgenome software, and viewed on the Cisgenome browser (Stanford University, Stanford, CA, USA) (Ji *et al.*, 2008).

Quantitative RT-PCR of several *T. azotonutricium* genes:

clpX (TREAZ# 737).

Endo-1,4-beta-xylanase A precursor (TREAZ# 2717).

Endo-1,4-beta-xylanase A precursor (TREAZ# 2718).

Glycoprotein gp2, Endo-1,4-beta-xylanase A precursor (TREAZ#2716).

These qRT measurements reveal expression ratios similar to those reported by RNA-Seq.

Quantitative PCR

In all, 100 μl samples from bacterial cultures were taken daily. Samples were centrifuged at 13 000 *g* for 5 min, aspirated, resuspended in an equal volume of dH_2O, and frozen for later use. Quantitative PCR on bacterial samples followed, using primers specific for the *clpX* gene of either *T. primitia* or *T. azotonutricium*. Quantitative PCR primers of 20 bp length were designed, with the following sequences:

T. azotonutricium clpX (fwd): 5′-GGAACTTTTCG ATGCTCTGC-3′

T. azotonutricium clpX (rev): 5′-GCGCTTAAGGT CTTCCCTCT-3′

T. primitia clpX (fwd): 5′-CTCCCGTTTCATTTCT TCCA-3′

T. primitia clpX (rev): 5′-GAAATGTTAGACGCCC TCCA-3′

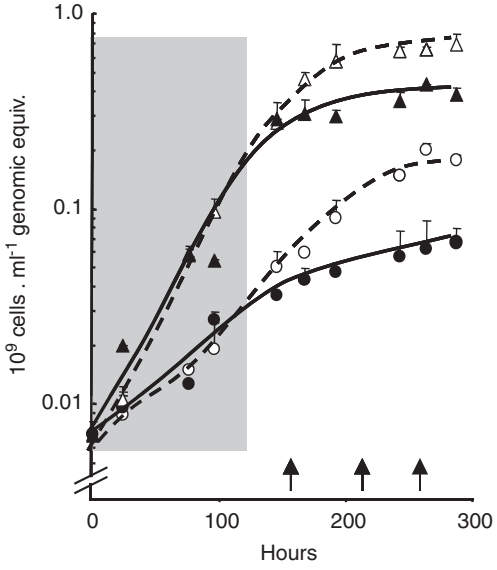
Primers were designed to have a roughly equal amplification product length to avoid large differences in fluorescence intensities. Each pair of primers was tested for melting temperature, specificity and amplification efficiency, and was found to be specific and suitable for q-PCR (each having an *E* factor greater than 1.75). Quantitative PCR reactions were performed with the iTaq SYBR Green PCR kit and a DNA engine chromo-4 qPCR instrument (Bio-Rad, Hercules, CA, USA). A standard curve of known dilutions of *T. primitia* and *T. azotonutricium* genomic DNA was performed with each set of experiments. Triplicate q-PCR reactions were used for each data point. To obtain genomic cell equivalent units, the ribosomal copy number was obtained from the standard curve calculations, and divided by 2 to account for the number of 16S copies in each genome.

Results and discussion

Full and accurate analysis of transcript pools from two closely related organisms grown together relies on the ability to distinguish between highly homologous sequences. Therefore, before analyzing the experimental data, we determined the frequency of cross-identification between the two treponemes used in this study. To do this, a data set of all possible short nucleotide sequences (37 bp, the exact length of the reads) from each genome was generated *in silico* and aligned to the genome of the other species (Materials and methods). The results (Table 1) demonstrate that only a small number of all possible 37-bp sequences (6518, ca. 0.15%) are identical in both bacteria. A large percentage of these, ca. 45%, are r16S and r23S RNAs, whereas greater than 97% of all non-rRNA genes are completely unique (that is, do not contain a single sequence read having an exact match to the other genome). Relaxing the stringency to allow one mismatch per 37-bp sequence does not yield dramatically less favorable results, with over 92% of the genes being completely unique. Moreover, in no case was a single gene in either genome fully masked; all genes in both organisms had at least one unique 37-bp identifier. This resolution allows complete transcript discrimination and genome-wide analysis in co-culture experiments.

Experimental considerations compelled us to analyze the background noise expected in actual data. To this end, a database totaling nearly 11 million RNA-Seq reads from *T. azotonutricium* was mapped to the combined genomes of both bacteria. As predicted, the majority of reads (9 million; 82.5%) aligned to *T. azotonutricium*, and practically all (over 99.9%) of the reads that did map to *T. primitia* were rRNA (Table 1) This result was expected, as ribosomal transcription typically comprises over 90% of the total transcript pool (Neidhardt and Curtiss, 1996), and rRNAs are typically the most conserved sequences between the genomes (the small subunit rRNA has 93% nucleotide similarity; (Graber *et al.*, 2004)). When all rRNA sequences, except highly variable regions and sequences with up to one mismatch similarity, were masked from the analysis, only 646 reads from one species mapped to the other genome. This led us to conclude that this method was effective at discriminating the signals from two similar genomes. The reciprocal analyses yielded similar trends, and are presented in Table 1.

For gene expression profiling, bacterial cultures of *T. primitia*, *T. azotonutricium* and their co-culture were grown on 4YACo medium supplemented with H_2 and maltose (Graber *et al.*, 2004). The growth kinetics of each species was followed as a function of genome copy number, that is, using qPCR of unique chromosomal DNA markers from each genome. These measurements reveal that initially each species grew at similar rates, whether in pure or co-culture (shaded region, Figure 1). However, after this initial growth phase, the growth of each species was markedly stimulated in both rate and yield in co-cultures, relative to growth in



	T_d 1	T_d 2	Yield
▲ *T. primitia*	32.2	179	0.43
△ *T. prim co-culture*	30.9	78.5	0.69
● *T. azotonutricium*	64	158.8	0.2
○ *T. azoto co-culture*	49	59.5	0.07

Figure 1 Synergistic growth effects in co-culture of termite-gut treponemes. Quantitative PCR measurements are given for *T. primitia* and *T. azotonutricium* abundance in either pure culture (solid lines) or co-culture (dashed lines). The growth phase in which bacteria in both pure cultures and co-cultures grow at roughly equal rates is marked by a shaded box. RNA-sequencing samples were taken at approximately 160 hours (co-culture), 210 h (*T. primitia*) and 260 h (*T. azotonutricium*). These collection times are in the later growth phase and are marked by arrows on the *X* axis. Samples were run in triplicates, and the s.d. is denoted by error bars. Growth rates and yield of each sample are summarized below the curve. Yields are in number of genomic equivalents per ml. Doubling times were calculated from data points either within the first 117 h of growth (Td1) or within the productive stage of growth that proceeds (Td2).

monoculture. The greater effect was observed for *T. azotonutricium* (Figure 1). Importantly, although these two species might compete for anabolic nutrients in the medium's yeast autolysate base, at no point were the two species observed to have any negative impact on each other's growth. All samples for gene expression analysis were collected during a period within the later, synergism-dominated growth phase ($OD_{600} \approx 0.35$) (Figure 1). Overall, a large set of genes, consisting of approximately 7% of all *T. azotonutricium* and 4.5% of all *T. primitia*-coding regions, were differentially regulated as a function of co-cultivation (Supplementary Table 2). In agreement with growth measurements, several genes related to cell division and other aspects of growth are upregulated in co-cultures (Supplementary Table 2). The observation that in the later growth phase co-cultures maintain near-log phase growth rates (while the pure cultures are nearing or entering stationary phase) can also be used to explain the differences between transcript pools in the co-cultures. This is especially true for *T. azotonutricium*, for which samples were taken in a late stage to allow for sufficient cell growth.

The positive co-culture growth effect on *T. primitia* might be anticipated when considering the H_2 physiologies of the two species. Fermenting gut microbes, including *T. azotonutricium*, provide homoacetogens like *T. primitia* a source of H_2 (Pester and Brune, 2007). In the culture media, an initial amount of H_2 gas (80%), which roughly mimics physiological conditions within the gut, was supplied in the headspace but was not replenished during growth. Previously, the growth of *T. primitia* has been observed to be noticeably restricted if H_2 is not initially supplied in the medium (Graber and Breznak, 2004). Thus, one predicted aspect of synergy between the two strains is the continued production of H_2 by *T. azotonutricium,* as this gas is consumed by *T. primitia*. Indeed, *T. primitia* genes associated with H_2 utilization were differentially regulated in co-cultivation: among the most highly upregulated *T. primitia* genes were those that encoded several hydrogenase-like proteins; in turn, the genes for several other hydrogenase-like proteins are markedly downregulated during co-cultivation (Table 2, Supplementary Table 1). These findings are consistent with previously published reports of hydrogenase expression in methanogenic H_2 consumers (*Desulfovibrio vulgaris*), when cultured together with a hydrogen producer (Walker *et al.*, 2009).

When the two treponemes were co-cultured, an additional effect on *T. primitia* genes associated with CO_2-reductive acetogenesis was observed. In pure culture, both a selenocysteine-containing ($fdhF_{sec}$) and a non-selenocysteine ($fdhF_{cys}$) variant of the enzyme formate dehydrogenase, which catalyzes an early step of the Wood–Ljungdahl pathway, were transcribed (Matson *et al.*, 2010). During growth in co-cultures, the selenocysteine form of the enzyme was upregulated, whereas the non-selenium form was downregulated (Table 2, Supplementary Tables 1 and 4).

Co-cultivation with *T. primitia* significantly enhanced the growth rate and yield of *T. azotonutricium* (Figure 1). Examination of the gene expression data and the genome sequences of the two treponemes suggested several potential growth factors that might be involved in the interaction. The differential regulation of many genes involved in corrinoid production or transport, vitamin B_7 (biotin), vitamin B_6 (pyridoxal phosphate) and coenzyme-A (Table 2, Supplementary Table 1), was observed. In the case of vitamin B_7, several relevant regulatory or transport genes are downregulated when *T. azotonutricium* is co-cultivated with *T. primitia* (Table 2, Supplementary Table 1), but only the latter has a complement of the genes to

Table 2 Major groups of transcriptionally up- and downregulated genes in co-culture

Process	*Function*	*Up in co-culture*	*Down in co-culture*
Treponema primitia			
Metabolism	Hydrogen and C1 metabolism	5	7
Vitamins and cofactors	B_{12} and corrinoid related	2	23
	Tryptophan/phenylalanine/tyrosine biosynthesis	10	
Amino acids	Methionine synthesis and transport	3	
	Isoleucine/leucine/valine transport	2	
Treponema azotonutricium			
Vitamins and cofactors	B_{12} and corrinoid-related		8
	Biotin transport, regulation, metabolism		3
	Vitamin B_6 precursor synthesis		1
	Enzymes requiring B6 for activity		3
Amino acids	Isoleucine/leucine/valine biosynthesis	4	
	Serine[a]		3
	Cysteine[a]		2

Regulated genes and gene clusters of *T. primitia* and *T. azotonutricium* are listed by major cellular pathways. Included are genes with clear annotation and fold change that is above background (see Materials and methods), and which are discussed in the manuscript. The values in the up and down columns describe the number of genes associated with a specific process.
[a]Some genes involved in the serine and cysteine biosynthesis pathways require vitamin B_6, and appear in both categories.

synthesize B_7 (Supplementary Figure 1). Similarly, an important gene for the synthesis of vitamin B_6 precursors (for example, phosphoserine aminotransferase) (Lam and Winkler, 1990) and several genes that require vitamin B_6 for full activity were also downregulated during consortial growth of *T. azotonutricium* (Table 2, Supplementary Table 1). The enzyme responsible for producing the active form of vitamin B_6 (pyridoxal kinase) is exclusive to the genome of *T. primitia* (Supplementary Figure 1). Finally, key genes for coenzyme-A biosynthesis are absent in *T. azotonutricium* but present in *T. primitia* (Supplementary Figure 1).

One explanation for the upregulation of genes that transport vitamin B_7 or require vitamin B_6 for activity is that as the bacterial population size increases, the amount of available vitamins is increasingly scarce, and thus bacteria are more reliant on importing and using what little vitamins are present.

To examine if any of these candidate growth factors might be relevant to the observed synergistic growth of the species during their co-culture, the growth medium of *T. azotonutricium* pure cultures was supplemented. The results (Figure 2) demonstrate that vitamin B_6, and to a lesser extent B_7 and coenzyme-A, have a positive impact on growth. None of the corrinoid or B_{12} preparations tested improved the growth of *T. azotonutricium*, nor were the additions of any of these supplements observed to have any beneficial effect on the growth of *T. primitia*.

In addition to the candidate cofactors and vitamins that were identified and explored above, we note that a number of hypothetical genes are also regulated differently in co-culture (Figure 3, Supplementary Table 2). The large number of these genes suggests that additional processes may be involved in complex mutualistic symbiosis.

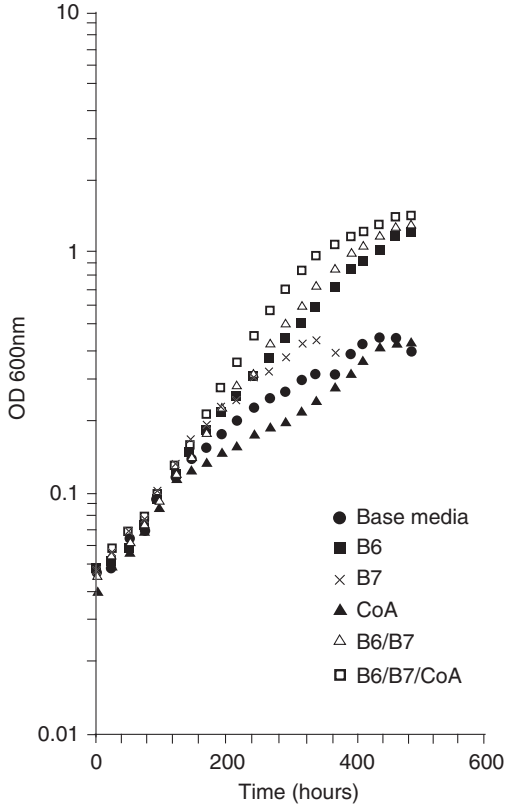


Figure 2 Positive growth effects of the vitamins B_6, B_7 and coenzyme-A on *T. azotonutricium*. *T. azotonutricium* growth rate and yield benefit from the addition of vitamin B_6, coenzyme-A and vitamin B_7 work in consortia with B_6 to further increase the yield and rate of the culture media. Data from duplicates were averaged and plotted.

The abundance of genes regulated by these conditions also serves to underscore the large effects that growth in the presence of other microbes has on the physiology and behavior of another bacterial

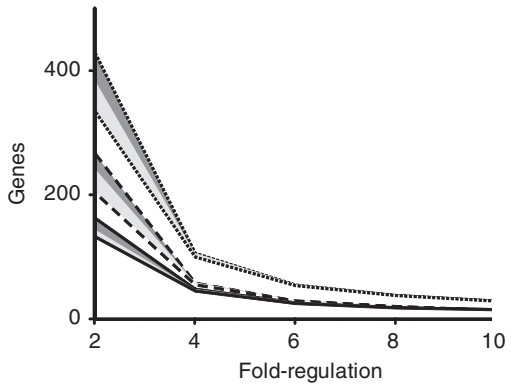


Figure 3 The expression of a large number of genes in both bacterial species is effected by co-culturing. The number of genes with expression levels regulated by twofold, fourfold, sixfold, eightfold and 10-fold (*X* axis) plotted for *T. primitia* (solid lines), *T. azotonutricium* (dashed line) and both species (dotted line). Best-fit lines were added to aid in data visualization. The area below the dark-shaded box shows the number of genes that are on average twofold regulated, and also have three of the four possible biological replicate combinations above 1.7-fold regulated. The values below the light gray shade are genes that are on average twofold regulated, and in which all four biological replicate combinations are above 1.7-fold regulated.

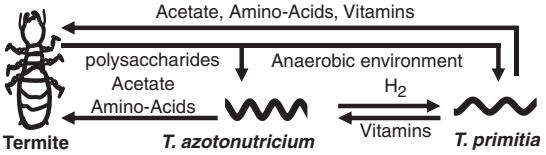


Figure 4 Schematic representation of the symbiosis between *T. primitia* and *T. azotonutricium* and the termite host.

species. Looking ahead, it is possible that the accumulation of expression studies of microbial communities can be used to highlight uncharacterized genes that act as putative symbiosis determinants.

A hallmark of several insect–microbe symbioses is the bacterial production of factors vital to the host, including vitamins, cofactors and essential amino acids (reviewed in Moran, 2006). Vitamins B_6 and B_7, the production of which appears to aid in the growth of *T. azotonutricium* when it is grown in co-culture with *T. primitia*, have a role in the symbiosis between tse-tse flies and their endosymbiont *Wigglesworthia glossinidia brevipalpis* (Akman *et al*., 2002). Genes for several amino-acid biosynthesis pathways were upregulated during the consortial growth of the two treponemes, including those for aromatic amino acids and methionine biosyntheses in *T. primitia*, and for branched-chain amino acids in *T. azotonutricium* (Table 2). Together these represent six out of ten amino acids known to be essential in insects (Gil *et al*., 2003). Curiously, the tryptophan synthesis pathway appears to be absent in the diazotroph *T. azotonutricium*, but the expression of these genes was highly induced in *T. primitia* when the two were grown together. Past studies had revealed that *T. azotonutricium* is capable of growth as a *bona-fide* N_2-fixing bacterium (Lilburn *et al*., 2001), yet required that the medium be supplemented with yeast autolysate while doing so, presumably at least in part to complement its need for tryptophan. Thus, our results suggest that the fixation of N_2, ultimately into a full suite of essential amino acids, is a consortial activity in the N-limited lignocellulose-degrading environment of the termite gut. In the case of the branched-chain amino acids (for which both strains appear to have all required biosynthetic components), co-cultivation stimulates expression of key synthesis genes in *T. azotonutricium* and key transport genes in *T. primitia* (Table 2, Supplementary Table 1). Taken together, the gene expression patterns observed during the consortial growth of the two species suggest a streamlined division of biosynthetic labor and shared production of key nutrients.

The upregulation of essential amino-acid biosynthesis genes in co-culture suggests that their production may also benefit the host and other community members, as these essential amino acids are often implicated in insect/bacteria symbioses (Shigenobu *et al*., 2000; Gil *et al*., 2003; Nakabachi *et al*., 2006). The recently published genomes of two endosymbionts of termite-gut protozoa have also suggested the importance of other aspects of amino-acid and vitamin production in the symbiosis (Hongoh *et al*., 2008a, b).

The genomes of intracellular endosymbionts of insects are often highly reduced (Shigenobu *et al*., 2000; Akman *et al*., 2002; Gil *et al*., 2003; Nakabachi *et al*., 2006; Hongoh *et al*., 2008a, b) and underlie the obligate reliance of these bacteria on nutrients and factors supplied by the host. In return, many such endosymbionts act almost exclusively as essential amino acids and vitamin factories, capable of little else (Shigenobu *et al*., 2000; Akman *et al*., 2002; Gil *et al*., 2003; Nakabachi *et al*., 2006). In contrast, these termite-gut spirochetes are not intracellular residents of their host's tissues, and have comparatively large genomes, suggesting that they are capable of performing many more tasks in their species-rich symbiotic environment than their endosymbiontic counterparts. Several of these tasks, as implicated by expression and activity data from this and previous studies, are presented in Figure 4. They include acetate production during sugar fermentation and CO_2-reductive acetogenesis, H_2 cycling, nitrogen fixation, amino-acid biosynthesis and vitamin and co-factor production (Leadbetter *et al*., 1999; Lilburn *et al*., 2001; Graber *et al*., 2004). In return, the termite provides its microbiota with finely ground particles of recalcitrat carbon and energy source, lignocellulose and a controlled environment.

In addition to providing insights into the symbiotic interactions of these two treponemes in the context of the termite gut, the RNA-Seq data that

were mapped onto the two genomes also provided information on genes and intergenic transcripts that had not yet been predicted. For example, on scanning the two genomes for intergenic loci with high gene expression, we were able to identify seven highly transcribed intergenic regions in the two genomes. Five of these are between previously annotated genes of *T. primitia* (between gene pairs 175:176, 2323:2324, 2456:2457, 2870:2871 and 3885:3886). The other two intergenic transcripts were in *T. azotonutricium* (between gene pairs 486:487 and 953:954). The high transcript levels of these genes are displayed in Supplementary Table 4, and are generally higher than those of mRNAs from the predicted, annotated genes. Upon further analysis we were able to predict a putative role for three of these intergenics. All three of these newly annotated genes appear to be ribozymes, with RnaseP genes being found in both genomes and what appears to be tmRNA in the genome of *T. primitia*. There remains an inherent difficulty in the *in-silico* identification of genes based on predictions of RNA folding (as opposed to proposing canonical open reading frames after scanning all possible translation reading-frames). Thus, RNA-Seq provides a wet-bench experimental method to aid in the identification of non-canonical open reading frames. Data on the expression of one of these genes, the tmRNA from the genome of *T. primitia*, are provided in Figure 5a.

Apart from intergenic sequences, we have scrutinized graphical displays of transcripts to identify atypical transcription patterns in the two genomes. An interesting example of an atypical transcript pattern is a gene that had overly abundant transcripts towards the gene center, as opposed to the two ends (Figure 5b). In order to determine which transcripts for this region of the gene were in the coding direction or noncoding (antisense) direction, we used qRT-PCR primers specific for either direction, targeting different areas of the gene. As shown in Figure 5b (and inset), whereas the coding transcripts are relatively constant throughout the gene, a large amount of transcripts in the antisense direction are present toward the gene center. This highly transcribed anti-sense region contains six evenly spaced direct repeats of unknown function, but which may help to explain the periodicity of transcript depth observed in the central portion of the gene. This finding is similar to those detailed in a recent publication that uses a variation of the RNA-seq protocol to compose a genome-wide map of both coding and antisense transcripts (Dornenburg *et al.*, 2010).

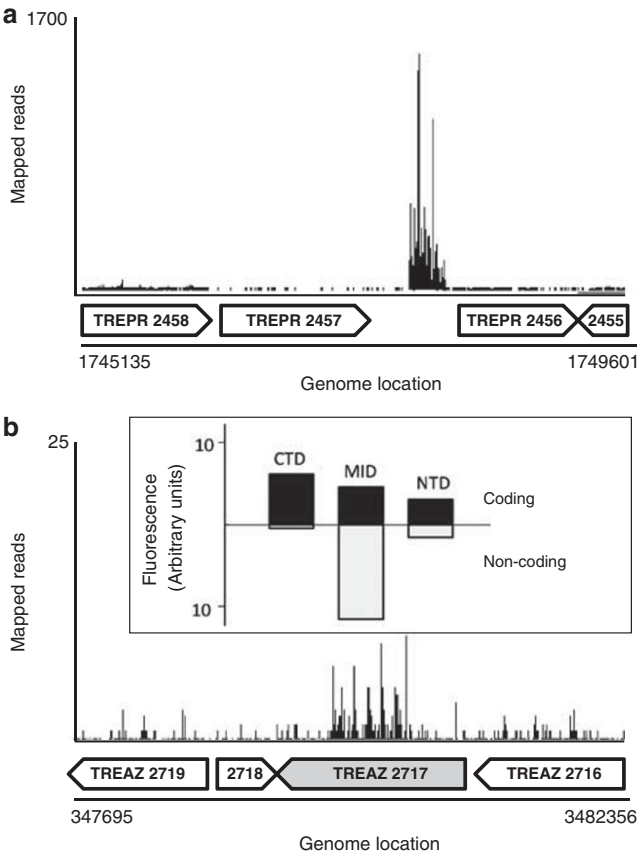


Figure 5 Transcription mapping of intergenic and atypical transcripts. (**a**) A graphical representation of transcripts mapped onto an annotated portion of the *T. primitia* genome containing one of several identified intergenic transcripts in this study. The sequence is similar to that of a tmRNA. Anotated genes are displayed as block arrows below the mapped reads. (**b**) An atypical transcript pattern in a putative glycohydrolase gene of *T. azotonutricium* (TREAZ 2717, shaded block arrow) shows a high transcript density in the gene center with repetitive pattern. Data from qRT-PCR experiments differentiating coding (black bars) from antisense (white bars) are shown in the inset box, and correspond to the section of the gene analyzed. qRT-PCR was performed with direction-specific primers in the amino-terminal domain (NTD), central portion (MID), and carboxy-terminal domain (CTD) of the gene.

Conclusions

This work reveals that, when grown together, two termite-gut *Treponema* species influence each other's gene expression in a far more comprehensive and nuanced manner than might have been predicted based on the results of previous studies on the respective pure cultures. Although H_2-based interactions are predicted by known physiologies, using new techniques we find a division of labor between the bacteria, and uncover unforeseen symbiotic interactions. On the road to understanding the even more complex interactions that occur among the hundreds of microbial species residing in termite and other gut environments, we believe that studies of defined microbial consortia of representative strain isolates become a promising avenue toward gaining a better understanding of such systems. Despite the inherent difficulties in comparing transcript signals from different organisms and co-cultures (as in this study), we expect that deep RNA sequencing will enable and expedite such

studies, revealing genes and interactions underlying the associations, and aiding in the development of many new hypotheses and new directions of research.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

We thank our laboratory colleagues and participants in Caltech's GEC facility for their insights and advice. This research was supported by the DOE (DE-FG02-07ER64484) and the NSF (EF-0523267).
Genome sequences for the two bacterial species are deposited with Genbank under the following accession numbers: (CP001843) and (CP001841).

References

Abe T, Bignell DE, Higashi M. (2000). *Termites: Evolution, Sociality, Symbioses, Ecology*. Kluwer Academic Publishers: Dordrecht and Boston.

Akman L, Yamashita A, Watanabe H, Oshima K, Shiba T, Hattori M *et al.* (2002). Genome sequence of the endocellular obligate symbiont of tsetse flies, *Wigglesworthia glossinidia*. *Nat Genet* **32**: 402–407.

Boone DR, Castenholz RW, Garrity GM. (2001). *Bergey's Manual of Systematic Bacteriology*. Garrity GM (ed.), 2nd edn. Springer: New York.

Brauman A, Kane MD, Labat M, Breznak JA. (1992). Genesis of acetate and methane by gut bacteria of nutritionally diverse termites. *Science* **257**: 1384–1387.

Breznak JA, Switzer JM. (1986). Acetate synthesis from H(2) plus CO(2) by termite gut microbes. *Appl Environ Microbiol* **52**: 623–630.

Breznak JA. (2002). Phylogenetic diversity and physiology of termite gut spirochetes. *Integr Comp Biol* **42**: 313–318.

Chang DE, Smalley DJ, Tucker DL, Leatham MP, Norris WE, Stevenson SJ *et al.* (2004). Carbon nutrition of *Escherichia coli* in the mouse intestine. *Proc Natl Acad Sci USA* **101**: 7427–7432.

Dornenburg JE, Devita AM, Palumbo MJ, Wade JT. (2010). Widespread antisense transcription in Escherichia coli. *MBio* **1**: e00024-10.

Drake HL. (1994). *Acetogenesis*. Chapman & Hall: New York.

Fraser CM, Norris SJ, Weinstock GM, White O, Sutton GG, Dodson R *et al.* (1998). Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science* **281**: 375–388.

Gil R, Silva FJ, Zientz E, Delmotte F, González-Candelas F, Latorre A *et al.* (2003). The genome sequence of *Blochmannia floridanus*: comparative analysis of reduced genomes. *Proc Natl Acad Sci USA* **100**: 9388–9393.

Gilbert JA, Field D, Huang Y, Edwards R, Li W, Gilna P *et al.* (2008). Detection of large numbers of novel sequences in the metatranscriptomes of complex marine microbial communities. *PLoS One* **3**: e3042.

Graber JR, Breznak JA. (2004). Physiology and nutrition of *Treponema primitia*, an H2/CO2-acetogenic spirochete from termite hindguts. *Appl Environ Microbiol* **70**: 1307–1314.

Graber JR, Breznak JA. (2005). Folate cross-feeding supports symbiotic homoacetogenic spirochetes. *Appl Environ Microbiol* **71**: 1883–1889.

Graber JR, Leadbetter JR, Breznak JA. (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.

Holt KE, Parkhill J, Mazzoni CJ, Roumagnac P, Weill FX, Goodhead I *et al.* (2008). High-throughput sequencing provides insights into genome variation and evolution in *Salmonella Typhi*. *Nat Genet* **40**: 987–993.

Hongoh Y, Sharma VK, Prakash T, Noda S, Taylor TD, Kudo T *et al.* (2008a). Complete genome of the uncultured Termite Group 1 bacteria in a single host protist cell. *Proc Natl Acad Sci USA* **105**: 5555–5560.

Hongoh Y, Sharma VK, Prakash T, Noda S, Toh H, Taylor TD *et al.* (2008b). Genome of an endosymbiont coupling N2 fixation to cellulolysis within protist cells in termite gut. *Science* **322**: 1108–1109.

Ji H, Jiang H, Ma W, Johnson DS, Myers RM, Wong WH. (2008). An integrated software system for analyzing ChIP-chip and ChIP-seq data. *Nat Biotechnol* **26**: 1293–1300.

Lam HM, Winkler ME. (1990). Metabolic relationships between pyridoxine (vitamin B6) and serine biosynthesis in *Escherichia coli* K-12. *J Bacteriol* **172**: 6518–6528.

Leadbetter JR, Schmidt TM, Graber JR, Breznak JA. (1999). Acetogenesis from H2 plus CO2 by spirochetes from termite guts. *Science* **283**: 686–689.

Lilburn TG, Kim KS, Ostrom NE, Byzek KR, Leadbetter JR, Breznak JA. (2001). Nitrogen fixation by symbiotic and free-living spirochetes. *Science* **292**: 2495–2498.

Matson EG, Zhang X, Leadbetter JR. (2010). Selenium controls transcription of paralogous formate dehydrogenase genes in the termite gut acetogen, *Treponema primitia*. *Environ Microbiol* **12**: 2245–2258.

Moran NA. (2006). Symbiosis. *Curr Biol* **16**: R866–R871.

Nakabachi A, Yamashita A, Toh H, Ishikawa H, Dunbar HE, Moran NA *et al.* (2006). The 160-kilobase genome of the bacterial endosymbiont *Carsonella*. *Science* **314**: 267.

Neidhardt FC, Curtiss R. (1996). *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd edn. ASM Press: Washington, DC.

Odelson DA, Breznak JA. (1983). Volatile fatty acid production by the hindgut microbiota of xylophagous termites. *Appl Environ Microbiol* **45**: 1602–1613.

Palmer KL, Aye LM, Whiteley M. (2007). Nutritional cues control *Pseudomonas aeruginosa* multicellular behavior in cystic fibrosis sputum. *J Bacteriol* **189**: 8079–8087.

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

Saleh-Lakha S, Miller M, Campbell RG, Schneider K, Elahimanesh P, Hart MM *et al.* (2005). Microbial gene expression in soil: methods, applications and challenges. *J Microbiol Methods* **63**: 1–19.

Seshadri R, Myers GS, Tettelin H, Eisen JA, Heidelberg JF, Dodson RJ *et al.* (2004). Comparison of the genome of

the oral pathogen *Treponema denticola* with other spirochete genomes. *Proc Natl Acad Sci USA* **101**: 5646–5651.

Shigenobu S, Watanabe H, Hattori M, Sakaki Y, Ishikawa H. (2000). Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp APS. *Nature* **407**: 81–86.

Straight PD, Fischbach MA, Walsh CT, Rudner DZ, Kolter R. (2007). A singular enzymatic megacomplex from *Bacillus subtilis*. *Proc Natl Acad Sci USA* **104**: 305–310.

Urich T, Lanzén A, Qi J, Huson DH, Schleper C, Schuster SC. (2008). Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS ONE* **3**: e2527.

Walker CB, Stolyar S, Chivian D, Pinel N, Gabster JA, Dehal PS *et al.* (2009). Contribution of mobile genetic elements to Desulfovibrio vulgaris genome plasticity. *Environ Microbiol* **11**: 2244–2252.

Yoder-Himes DR, Chain PS, Zhu Y, Wurtzel O, Rubin EM, Tiedje JM *et al.* (2009). Mapping the *Burkholderia cenocepacia* niche response via high-throughput sequencing. *Proc Natl Acad Sci USA* **106**: 3976–3981.

Supplementary Information accompanies the paper on The ISME Journal website (http://www.nature.com/ismej)