

Analysis of genes of tetrahydrofolate-dependent metabolism from cultivated spirochaetes and the gut community of the termite *Zootermopsis angusticollis*

Tina M. Salmassit and Jared R. Leadbetter

Correspondence

Jared R. Leadbetter
jleadbetter@caltech.edu

Environmental Science and Engineering, M/C 138-78, California Institute of Technology, Pasadena, CA 91125, USA

The hindguts of wood-feeding termites are the sites of intense, CO₂-reductive acetogenesis. This activity profoundly influences host nutrition and methane emissions. Homoacetogens previously isolated from diverse termites comprised novel taxa belonging to two distinct bacterial phyla, *Firmicutes* and *Spirochaetes*. Little else is known about either the diversity or abundance of homoacetogenic species present in any given termite or the genetic details underlying CO₂-reductive acetogenesis by *Spirochaetes*. A key enzyme of CO₂-reductive acetogenesis is formyltetrahydrofolate synthetase (FTHFS). A previously designed primer set was used to amplify FTHFS genes from three isolated termite-gut spirochaetes. Sequencing DNA flanking the FTHFS gene of *Treponema* strain ZAS-2 revealed genes encoding two acetogenesis-related enzymes, methenyltetrahydrofolate cyclohydrolase and methylenetetrahydrofolate dehydrogenase. Although termite-gut spirochaetes are only distantly related to clostridia at the ribosomal level, their tetrahydrofolate-dependent enzymes appear to be closely related. In contrast, homologous proteins identified in the non-homoacetogenic oral spirochaete *Treponema denticola* were only distantly related to those from clostridia and the termite-gut treponemes. Having demonstrated their utility with spirochaete pure cultures, the FTHFS primers were used to construct a 91-clone library from the termite-gut community DNA. From this, 19 DNA and eight amino acid FTHFS types were identified. Over 75 % of the retrieved clones formed a novel, coherent cluster with the FTHFS homologues obtained from the termite-gut treponemes. Thus, FTHFS gene diversity in the gut of the termite *Zootermopsis angusticollis* appears to be dominated by spirochaetes. The homoacetogenic capacity of termite-gut spirochaetes may have been acquired via lateral gene transfer from clostridia.

Received 14 March 2003

Revised 13 May 2003

Accepted 28 May 2003

INTRODUCTION

The most important bacterially mediated, hindgut activity known to have an impact on the energy metabolism of wood-feeding termites is CO₂-reductive acetogenesis: $4 [2H] + 2 CO_2 \rightarrow CH_3COOH + 2 H_2O$ (Breznak, 1994). Key reductants identified in this process include H₂ and lactate (Tholen & Brune, 2000). Gut production rates as

high as 2–5 μmol CO₂-derived acetate (g fresh weight termite)⁻¹ h⁻¹ have been observed for at least a half dozen wood-feeding termite species (Brauman *et al.*, 1992). No other environment has been demonstrated to catalyse CO₂-reductive acetogenesis at such high rates. Gut-derived acetate serves to fuel up to 100 % of the energy metabolism of termites, and it has been estimated that CO₂ reduction can generate up to one-third of this key nutrient for the insect (Breznak & Switzer, 1986; Odelson & Breznak, 1983). Thus, homoacetogens play an important role in the nutrition of their insect hosts.

In many termite species, bacterial homoacetogens out-process methanoarchaea for reductant, most notably H₂ (Brauman *et al.*, 1992). As a result, termites emit less CH₄ than might be expected if their fermentation pattern were similar to that observed in ruminants, and the basis for this is not well understood. Termites are as intriguing for

[†]Present address: Faculty for Biological Sciences, BS 344, California State University, Los Angeles, Los Angeles, CA 90032-4226, USA.

Abbreviations: FTHFS, formyltetrahydrofolate synthetase; MTHFC, methenyltetrahydrofolate cyclohydrolase; MTHFD, methylenetetrahydrofolate dehydrogenase; THF, tetrahydrofolate.

The nucleotide and deduced amino acid sequences for the FTHFS genes recovered from pure cultures and environmental samples during the course of this study can be found in GenBank under accession numbers AY162294–AY162314, AY162316 and AY254548.

the CH₄ that they do not emit, as much as for the CH₄ that they do.

Because of the influence of acetogenesis on termite nutrition and the production of greenhouse gases, there has been considerable interest in learning more about the gut microbiota accounting for this activity (Breznak, 1994). Four homoacetogenic, spore-forming *Firmicutes* have previously been isolated from four different termite species and described as novel taxa (Boga *et al.*, 2003; Breznak *et al.*, 1988; Kane & Breznak, 1991; Kane *et al.*, 1991). Two homoacetogenic spirochaetes have been isolated from a fifth termite species (Leadbetter *et al.*, 1999). The latter strains, *Treponema* ZAS-1 and *Treponema* ZAS-2, represented the first examples of spirochaetes isolated from any termite-gut environment, as well as the first demonstration of lithotrophic metabolism (much less CO₂-reductive acetogenesis) in this phylum-level group of bacteria. The isolates exhibited enzymic activities associated with the Wood–Ljungdahl pathway used by homoacetogenic *Firmicutes* (Leadbetter *et al.*, 1999), a pathway also used in the oxidative energy metabolism and autotrophic anabolism of diverse archaea and sulfate-reducing bacteria (Drake *et al.*, 1997; Ragsdale, 1997; Wood, 1991). However, the evolutionary history of this pathway in spirochaetes, and its relationships to those from better-studied organisms, is not currently known. The actual contribution to gut acetogenesis by any of the aforementioned isolates and the full diversity of homoacetogens present and active in any single termite species are also poorly understood. It has proven tempting to speculate that spirochaete species may dominate acetogenesis in the guts of termites (Leadbetter *et al.*, 1999), especially since spirochaetes are among the most numerous microbiota present in the guts of virtually all termite species, representing as many as 50% of all prokaryotic cells (Breznak & Leadbetter, 2002; Lilburn *et al.*, 1999; Paster *et al.*, 1996). However, at least 50% of the prokaryotic cells found in termite guts are not spirochaetes and not all *Zootermopsis*-gut spirochaetes are CO₂-reducing acetogens. One strain isolated from this environment, *Treponema* ZAS-9, does not display this mode of metabolism (Breznak & Leadbetter, 2002; Lilburn *et al.*, 2001).

We have adopted the use of a functional-gene approach employing degenerate, oligonucleotide primers developed to amplify portions of genes encoding a key enzyme of the CO₂-reductive acetogenesis pathway: formyltetrahydrofolate synthetase (FTHFS) (Leaphart & Lovell, 2001; Leaphart *et al.*, 2003). These primers were successfully applied to three termite-gut spirochaete isolates. The sequences for FTHFS partial genes for two of these spirochaetes have also been recently reported by others (Leaphart *et al.*, 2003). Here, we detail our analysis of a sizeable collection of FTHFS genes derived from spirochaetal isolates and the gut community of the termite *Zootermopsis*, as well as two putative acetogenesis-related genes flanking the FTHFS gene in *Treponema* strain ZAS-2.

METHODS

Termites and strains. Specimens of *Zootermopsis angusticollis*, the Pacific dampwood termite, were collected from a rotted *Pinus ponderosa* (Pacific ponderosa pine) stump located on the north face of the summit of Mount Wilson in the San Gabriel Mountains of Southern California. Specimens were healthily maintained in the laboratory in polypropylene trays containing pieces of the stump material and were used within a month of their collection. Microbial isolates examined in the study [*Treponema* strains ZAS-1 (DSM 12426), ZAS-2 (DSM 12427) and ZAS-9 (DSM 13862)] and media for their cultivation have been described previously (Breznak & Leadbetter, 2002; Leadbetter *et al.*, 1999; Lilburn *et al.*, 2001).

DNA extraction. *Z. angusticollis*-gut community DNA was purified using a modification of a previously described technique (Purdy *et al.*, 1996). DNA was extracted from the entire, intact hindgut of nine uniformly sized worker larvae. Three termite hindguts were combined into each of three 2 ml microcentrifuge tubes containing 0.2 g of 0.1 mm Biospec zirconium/silica beads, 700 µl of 1% (w/v) polyvinylpyrrolidone in 120 mM sodium phosphate, pH 8.0, 50 µl of 20% (w/v) SDS and 500 µl TRIS-equilibrated phenol, pH 8.0. The samples were disrupted using a Biospec MiniBeadbeater-8 as described previously (Purdy *et al.*, 1996). The reaction mixtures were pooled and applied to hydroxyapatite-containing microcentrifuge spin columns (Bio-Rad). Sample application, washing and elution steps were performed by gently spinning the column for <10 s in a fixed-angle bench-top microcentrifuge. DNA was extracted from the treponemal isolates using the DNeasy extraction kit (Qiagen). DNA samples were stored at -20 °C until used.

PCR amplification, cloning and RFLP analysis. FTHFS homologues were amplified from DNA templates prepared from cultivated spirochaete strains and termite-gut contents using the primers and PCR conditions exactly as described by others previously (Leaphart & Lovell, 2001), except that Roche *Taq* polymerase and reagents were used. Reactions yielded the expected ~1.1 kb gene fragments as products. When community DNA was used as template, reaction products of various other sizes were also observed. PCR products generated from genomic templates from isolated strains were cloned directly using the Topo TA cloning kit (Invitrogen); products amplified from gut template were similarly cloned after purification of the desired-size band using the QIAEX II Gel Extraction Kit (Qiagen). Using an adaptation of a previously described method (Ng *et al.*, 1996), the community-derived clone library was characterized via RFLP analyses after double digestion with *Msp*I and *Hin*P1I following the manufacturer's recommendations (New England Biolabs). RFLPs were visualized via gel electrophoresis using 2.5% (w/v) low-melting-point agarose and analysed manually.

Inverse PCR generation of *Treponema* ZAS-2 genomic fragment. A synopsis of this technique has been described previously (Raponi *et al.*, 2000). Genomic DNA was digested with *Nco*I according to the manufacturer's guidelines (New England Biolabs). The digested-DNA fragments were circularized using T4 DNA ligase. The ligation mixture was purified using the QIAEX II Gel Extraction Kit (Qiagen). Five microlitres of the purified ligation mixture was used as template in 20 µl reactions using the Failsafe PCR System and PCR PreMix Selection Kit (Epicenter). The internal FTHFS 'out' primers employed were 5'-GGCCGAAGAAGTCGTAAAC-3', designed to amplify the downstream gene sequence, and 5'-CCAGGTTCGGTTTCAGGGC-3', designed to amplify the upstream gene sequence. The amplification conditions used were initial denaturation at 94 °C for 3 min, followed by 30 cycles of denaturation at 94 °C for 30 s, annealing at 55 °C for 1 min and extension at

72 °C for 6 min, with a final extension at 72 °C for 10 min. Products were RFLP-sorted and cloned as outlined above.

Nucleotide sequencing and analysis. Sequencing was performed at the DNA Sequencing Core Facility at the Beckman Institute of Caltech using the dideoxy chain termination method, Sequenase (United States Biochemical) and a Perkin-Elmer ABI 373A automated sequencer. Sequence reads were assembled and edited using SEQUENCHER software for Windows (Genecodes). Multiple sequence alignments, translations and phylogenetic analyses were performed using the ARB freeware package (<http://www.arb-home.de/>) running within the Linux environment. FTHFS trees were constructed using PUZZLE-MAP 5.0 maximum-likelihood analyses (Schmidt *et al.*, 2002). For analyses of aligned protein sequences, 10 000 puzzling-steps were employed using the WAG model for substitution-rate frequencies and uniform models for rate heterogeneity to construct the input matrices (Goldman & Whelan, 2000). For analysis of FTHFS clones at the DNA sequence level, the Schöniger & von Haeseler substitution and uniform rate heterogeneity models were employed (Schöniger & von Haeseler, 1999). Transversion parameters and doublet frequencies were estimated from the dataset by the software. Phylogenetic-tree layout-editing was performed using TREEVIEW 1.6.6 for Windows (Page, 1996). Collector's curve analyses were used to estimate the completeness of the clone inventory and were performed using ANALYTIC RAREFACTION 1.3, written and made freely available by Steven M. Holland (<http://www.uga.edu/~strata/software/>), and ESTIMATES 6.01b1, written and made freely available by Robert K. Colwell (<http://viceroy.eeb.uconn.edu/EstimateS>).

Nucleotide accession numbers. The GenBank accession numbers for the sequences generated in this study and other FTHFS sequences used in the construction of Fig. 2 are: *Moorella thermoacetica*, J02911; Horse Manure Clone H 1, AF295711; Horse Manure Clone H 2, AF295715; Horse Manure Clone H 4, AF295714; Horse Manure Clone H 5, AF295713; *Spartina* Clone SR 10, AF295723; *Acetobacterium woodii*, AF295701; *Clostridium acetivum*, AF295705; *Clostridium magnum*, AF295703; *Eubacterium limosum*, AF295706; *Proteus vulgaris*, AF295710; *Ruminococcus productus*, AF295707; *Sporomusa ovata*, AF295708; *Sporomusa termitida*, AF295709; *Thermoanaerobacter kivui*, AF295704; *Za*-gut Clone A, AY162294; *Za*-gut Clone C, AY162295; *Za*-gut Clone E, AY162296; *Za*-gut Clone E2, AY162297; *Za*-gut Clone F, AY162298; *Za*-gut Clone F2, AY162299; *Za*-gut Clone G, AY162300; *Za*-gut Clone G2, AY162301; *Za*-gut Clone H, AY162302; *Za*-gut Clone I, AY162303; *Za*-gut Clone L, AY162304; *Za*-gut Clone M, AY162305; *Za*-gut Clone N, AY162306; *Za*-gut Clone P, AY162307; *Za*-gut Clone R, AY162308; *Za*-gut Clone T, AY162309; *Za*-gut Clone U, AY162310; *Za*-gut Clone Y, AY162311; *Za*-gut Clone Z, AY162312; *Treponema* strain ZAS-1A, AY162313; *Treponema* strain ZAS-1B, AY162314; *Treponema* strain ZAS-2, AY254548; *Treponema* strain ZAS-9, AY162316. An FTHFS gene was identified in the unpublished genome of *Treponema denticola* (<http://www.hgsc.bcm.tmc.edu/microbial/Tdenticola/>).

The 2371 bp genome contig containing three tetrahydrofolate (THF)-dependent enzyme-encoding genes from *Treponema* strain ZAS-2 has been deposited in GenBank under accession number AY254548. Sequences used in the construction of Fig. 4 were obtained from the NCBI BLAST with microbial genomes (http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi?) or DOE Joint Genome Institute (http://www.jgi.doe.gov/JGI_microbial/html/index.html) from both published and unpublished genome projects (Bao *et al.*, 2002; Casjens *et al.*, 2000; Eisen *et al.*, 2002; Ferretti *et al.*, 2001; Fraser *et al.*, 1998; Goodner *et al.*, 2001; Kapatral *et al.*, 2002; Kuroda *et al.*, 2001; Nolling *et al.*, 2001; Ruepp *et al.*, 2000; Shimizu *et al.*, 2002; White *et al.*, 1999; Wood *et al.*, 2001).

RESULTS

Amplification and cloning of FTHFS homologues from three cultivated termite-gut treponemes

Two nearly identical FTHFS homologues were cloned from *Treponema* strain ZAS-1, and one each was cloned from *Treponema* strain ZAS-2 and the termite-gut diazotroph *Treponema* strain ZAS-9. The latter strain is not known to be a homoacetogen (Lilburn *et al.*, 2001). When aligned, the FTHFS homologues from the three termite-gut treponemes shared 82.2–88.8 % identity with each other at the amino acid level and 77–78 % identity with each other at the DNA level. They shared 74–78 % amino acid identity with the FTHFS homologue identified from the bona fide homoacetogen *Clostridium magnum*, but were completely dissimilar in DNA sequence, even at the level of G+C content (53.3–56.4 mol% for the spirochaetes, in comparison to 36.6 mol% for *Clostridium magnum*). The FTHFS homologues from the three termite-gut treponemes shared less than 50 % amino acid identity with the FTHFS homologue identified via BLAST homology-searching of the unpublished genome sequence of the non-homoacetogenic, human oral treponeme, *T. denticola*. The published genomes of two other spirochaetes, *Treponema pallidum* and *Borrelia burgdorferi*, did not contain any obvious FTHFS homologues.

Amplification and cloning of diverse FTHFS homologues from the termite-gut community

DNA was purified from the gut contents of robust specimens. Amplification reactions using the FTHFS primer set and gut DNA as template successfully yielded products of the anticipated ~1.1 kb size. These were ligated into a cloning vehicle and transformed into *Escherichia coli*. From this, a library of 91 insert-containing plasmid clones was collected. After sorting of the library by comparing RFLP patterns, 31 clones were targeted for sequencing and further analyses. BLAST analyses confirmed that all of the clones encoded FTHFS homologues. When the 31 sequences were aligned and compared with each other, several were found to share near identity (greater than 99.5 % identity at the DNA level) and were thereafter considered as being representative of the same genotype. From these data, 19 unique clone types were established. Each was assigned a *Za*-gut epithet (for *Zootermopsis angusticollis*). After *in silico* translation of FTHFS genes, the deduced amino acid sequences for the 19 *Za*-gut clone types were aligned and compared. Despite clear differences at the DNA level, many of these were nearly identical in their amino acid sequences. Using 98 % amino acid identity as a cut-off, the clone inventory was collapsed into eight distinct amino acid groups. The rank abundances of the DNA and amino acid types recovered in the 91-clone collection are presented in Fig. 1. At both levels, the recovery of unique clones in the collection appeared to be close to saturation: rarefaction and collector's curve analyses indicated that the

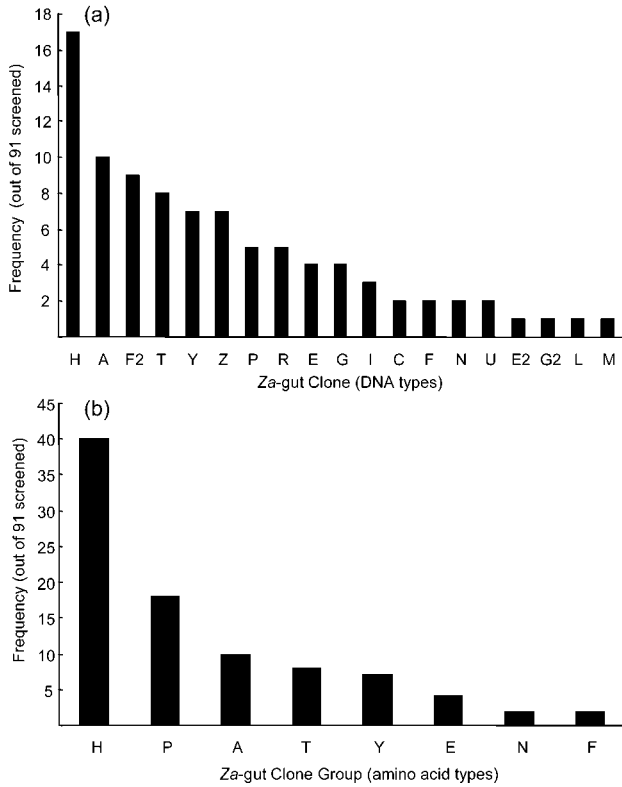


Fig. 1. Rank abundances of recovered FTHFS Za-gut clone types. (a) Frequency at which 19 distinct DNA types were encountered in a 91-clone inventory of FTHFS partial genes amplified from the gut contents of *Z. angusticollis*. Upon translation, the 19 DNA types collapsed into eight distinct amino acid clone groups. (b) Frequency at which the eight distinct amino acid clone groups were encountered in the inventory.

FTHFS inventory was essentially complete using this particular template DNA and primer set. With ‘infinite’ sampling, rarefaction analysis suggested that no more than 23 DNA and nine amino acid types would have been recovered. Chou and ACE estimators (available within the ESTIMATES program) predicted that no more than 22 DNA and the observed eight amino acid types would have been recovered.

Phylogenetic analysis

Using distance (not shown), maximum-parsimony (not shown) and maximum-likelihood methods (Fig. 2a), the deduced FTHFS peptides from *Treponema* strains ZAS-1, ZAS-2 and ZAS-9 clustered within the clostridial-homoacetogen FTHFS homology group previously identified by Lovell and co-workers. Within this cluster of known homoacetogens (Fig. 2a), strain ZAS-9 is the only cultivated organism *not* known to have this physiology. With the exception of the spirochaetes, which belong to their own unique phylum, all other sequences comprising the Lovell cluster were derived from clostridia, i.e. belonging

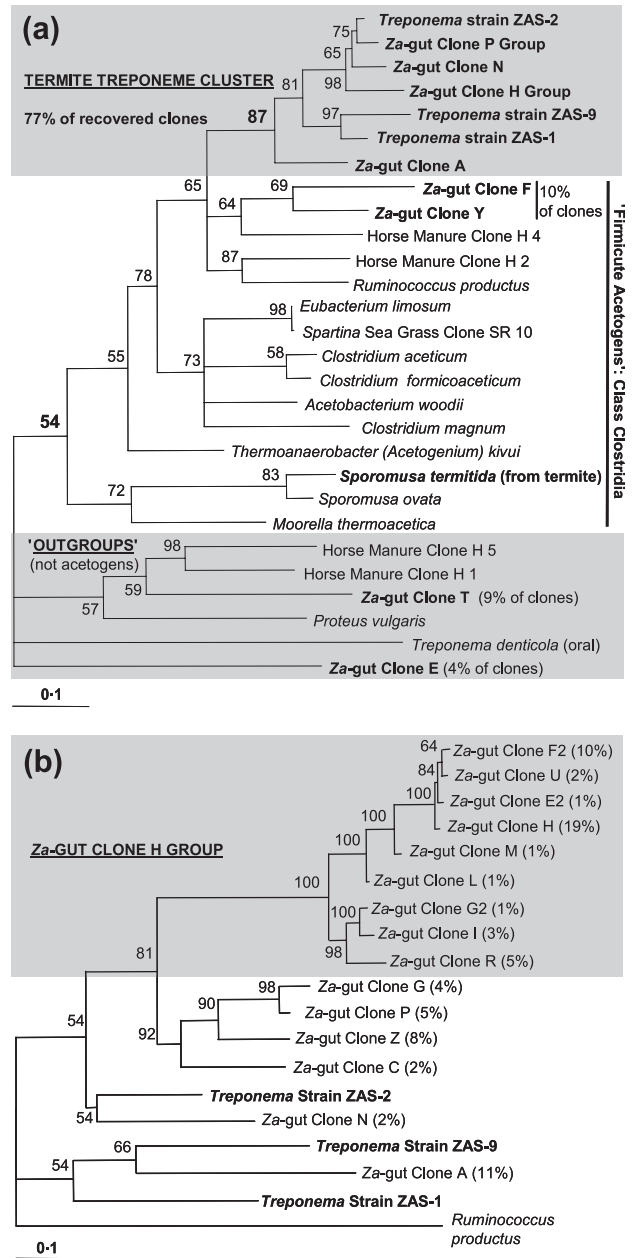


Fig. 2. Phylogenetic position of FTHFS homologues cloned from termite-gut contents, gut isolates or other isolates and environments. Percentages denote frequency of the encounter of a particular clone type or group within the inventory. Numbers are bootstrap values. Bars represent evolutionary distance as 0.1 changes per amino acid or nucleotide position. See Methods for GenBank accession numbers. (a) Maximum-likelihood phylogeny of the deduced FTHFS protein sequences, generated using 323 unambiguously aligned amino acid residue positions. Shaded boxes highlight the termite treponeme cluster (upper) and presumptive or known non-homoacetogens (lower) used to outgroup the unrooted tree. (b) Maximum-likelihood phylogeny of FTHFS genes comprising the termite treponeme cluster, generated using 1062 unambiguously aligned nucleotide positions. The FTHFS gene from *R. productus* was used to outgroup this unrooted tree. Shaded box highlights clones comprising the Za-gut Clone H group.

to the phylum *Firmicutes*, or were derived from not-yet-cultivated organisms (Leaphart & Lovell, 2001).

Over 85% of the FTHFS homologues recovered from termite-gut contents grouped within the Lovell cluster, and most of these comprised a unique, distinct and well-supported branch together with the homologues obtained from the treponemal isolates (Fig. 2a). All clone types comprising this 'termite treponeme' cluster shared greater than 80% amino acid identity with those derived from the treponemal isolates. Of particular note, members of *Za*-gut Clone P group and *Za*-gut Clone N shared 96.3 and 94.3% sequence identity, respectively, with the FTHFS homologue identified from *Treponema* strain ZAS-2. The nine clone types comprising the H group accounted for 43% of the clones recovered from gut-community DNA, and shared ~91% amino acid identity with the FTHFS homologue from strain ZAS-2. Despite close similarities with each other at the amino acid level (>98% identity), clone types comprising the *Za*-gut Clone H group were far from identical at the DNA level, thus they appear to represent different microbial strains or species (Fig. 2b). For example, *Za*-gut Clone F2 and *Za*-gut Clone R shared only 88% DNA sequence identity despite sharing 99.3% amino acid identity.

Two Lovell-cluster clone types, *Za*-gut Clone F and *Za*-gut Clone Y, showed no affiliation with the termite treponemes (Fig. 2a). These, together with a clone recovered from horse manure DNA, formed a unique radiation and were not highly similar to homologues identified in known homoacetogens. They shared less than 75% amino acid identity with the FTHFS from *R. productus*. Thus they may represent a novel, not-yet-cultivated microbial lineage. None of the recovered clones evidenced any affiliation with the FTHFS homologue identified in *S. termitida*, the spore-forming homoacetogen isolated from the gut of the Caribbean wood-feeding, 'higher' termite, *Nasutitermes nigriceps*. Approximately 15% of the recovered clones (i.e. *Za*-gut Clones T and E) did not affiliate with the Lovell cluster and thus may not represent homoacetogens. When our analyses of termite-gut-derived sequences were expanded to include a recently reported, larger collection of FTHFS homologues (Leaphart *et al.*, 2003), none of these aforementioned relationships were meaningfully changed or further resolved. *Za*-gut Clone E, for example, remained quite remote from any available FTHFS sequences.

Character analysis of FTHFS sequences

With a few minor exceptions, all FTHFS genes retrieved during the course of this study contained all of the conserved FTHFS polypeptide hallmarks that have previously been noted (Leaphart & Lovell, 2001; Leaphart *et al.*, 2003). The majority of the FTHFS genes comprising the 'termite treponeme cluster' contained a distinguishing polypeptide character not observed in other FTHFS genes analysed. As many as eight amino acid residues, instead of the more typically encountered two, were found inserted into a region

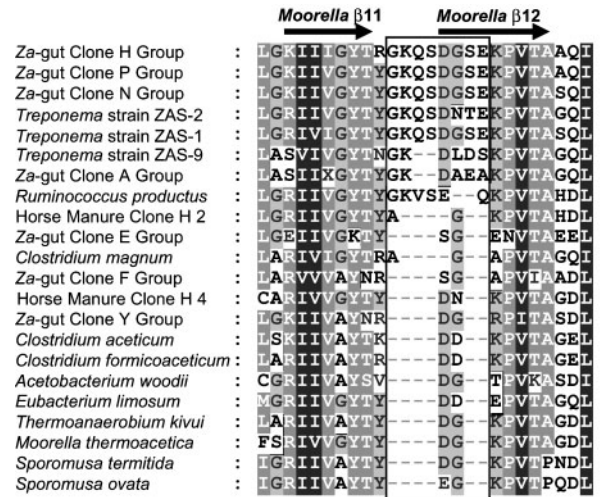


Fig. 3. Highly variable region found within diverse FTHFS homologues. Shown is a partial amino acid alignment corresponding to a stretch of residues corresponding to Phe222 to Leu241 in the FTHFS enzyme from *M. thermoacetica* (Lovell *et al.*, 1988; Radfar *et al.*, 2000a, b). Arrows denote β -sheets evidenced in the crystal structure of the FTHFS from *M. thermoacetica*. The box highlights an eight-residue stretch that is essentially unique to FTHFS homologues cloned from *Zootermopsis*-gut (*Za*-gut) *Treponema* isolates and community DNA. With the exception of those cloned from environmental sources and from *Treponema* strain ZAS-9, all the listed homologues are from organisms shown to be homoacetogens.

corresponding to *Moorella* residues Tyr231 and Lys234 (Fig. 3). Of all other FTHFS homologues analysed, only *R. productus* contained additional residues within this region. This unique region was filtered from alignments used for making phylogenetic comparisons, thus serves as an independent character or 'signpost' providing additional support for the coherency of the termite treponeme radiation.

Inverse PCR-mediated cloning of FTHFS-flanking DNA

Using sequence from the partial gene, outward-oriented primers were designed to amplify DNA flanking the FTHFS locus in *Treponema* strain ZAS-2 (see Methods). The primer set was used in an inverse PCR (iPCR) amplification (see Methods). From this, an 1880 bp genome fragment was amplified, cloned and sequenced. The resultant data, when assembled together with those for the FTHFS partial gene, resulted in 2371 bp of contiguous sequence. This included sufficient data to evaluate the site targeted by Lovell's forward primer, i.e. used to amplify the FTHFS partial gene. The site was identical to a single oligonucleotide (5'-TTCCTGGTGATTTCCATGCC-3') in the 24-fold degenerate 'FTHFS forward primer' mixture. No

other upstream data were revealed, as the majority of the iPCR-generated sequence corresponded to the region immediately downstream from the FTHFS partial gene. This included an unambiguous sequence that contained a single nucleotide mismatch in the binding site (5'-TGCATGGCCAAGACCCAATACAGC-3'; with C in place of the target G, previously considered to be invariable) for an oligonucleotide in the 12-fold degenerate 'FTHFS reverse primer' mixture. The contig also included a sequence encoding the remainder of the carboxyl-terminus of the FTHFS homologue.

Downstream from the FTHFS-encoding gene was an obvious ORF. The deduced 209 residue protein shared 50% amino acid sequence identity with a putative methenyltetrahydrofolate cyclohydrolase (MTHFC) from the unpublished genome of *Desulfitobacterium hafniense*, a spore-forming member of the class *Clostridia*. MTHFCs catalyse the enzymic step in acetogenesis following that exerted by FTHFS and are also involved in other THF-dependent metabolisms. Because the gene encoding this enzyme was not available from any known homoacetogen, the homologue from strain ZAS-2 was compared with those identified homologues from the *D. hafniense* and 20 other, mostly unpublished, genome sequences. It clustered within a well-supported branch along with the homologues from *D. hafniense*, *Clostridium botulinum*, *Clostridium difficile* and *Carboxydotherrnus hydrogenoformans*, all of which are *Firmicutes* belonging to the class *Clostridia* (Fig. 4a). This radiation was distinct from the branch which contained a MTHFC homologue identified from the unpublished genome of the non-acetogenic, human oral treponeme, *T. denticola* (Fig. 4a). As was the case for FTHFS-encoding genes, the published genomes of two other spirochaetes, *T. pallidum* and *B. burgdorferi*, did not contain any obvious MTHFC homologues.

Downstream from the putative MTHFC homologue from *Treponema* strain ZAS-2 was a partial gene interrupted by the terminus of the iPCR-generated amplification product. The first 102 amino acids of the deduced polypeptide were homologous to the catalytic domains of FoLD-type methylenetetrahydrofolate dehydrogenases (MTHFDs), and shared 40% amino acid identity with that from the unpublished genome of *D. hafniense*. MTHFD catalyses the enzymic step in acetogenesis that follows those exerted by FTHFS and MTHFC and is also involved in other THF-dependent metabolisms in other organisms. As was the case with MTHFC, the gene encoding a MTHFD was not yet available from any known homoacetogen. Accordingly, the ZAS-2 homologue was used in a phylogenetic comparison with those identified in the *D. hafniense* and 10 other, mostly unpublished, genome sequences. The homologue from strain ZAS-2 was the deepest branch on a well-supported cluster comprising the enzyme homologues from *D. hafniense*, *Clostridium botulinum*, *Clostridium difficile* and *Carboxydotherrnus hydrogenoformans* (Fig. 4b). FoLD homologues identified in the

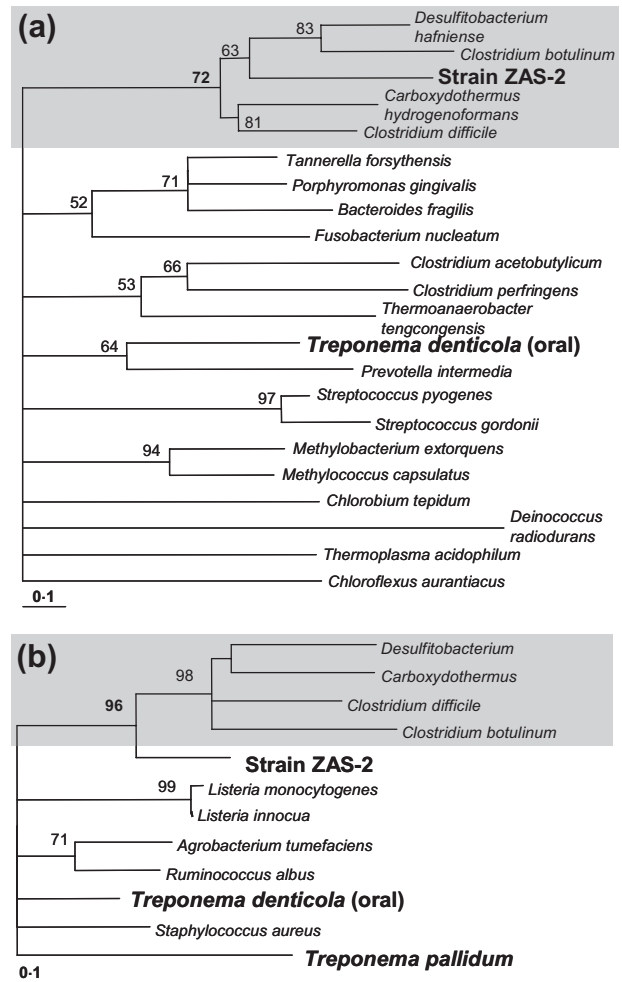


Fig. 4. Phylogenetic position of THF-dependent enzyme homologues cloned from *Treponema* strain ZAS-2. Homologues of these genes have not yet been cloned from any other homoacetogen. The majority of the homologues used in the analysis were obtained from unpublished genome sequences. Numbers are bootstrap values. Bars represent evolutionary distance as 0.1 changes per amino acid or nucleotide position. (a) Phylogeny of 'MTHFC' homologues. Maximum-likelihood analysis used 174 unambiguously aligned amino acid residue positions of the deduced protein sequences. Shaded box highlights a cluster comprising homologues from *Treponema* strain ZAS-2 and from four anaerobes belonging to the class *Clostridia*. (b) Phylogeny of the catalytic region of bifunctional 'FoLD-type' and other 'MTHFD' homologues. Maximum-likelihood analysis used 95 unambiguously aligned amino acid residue positions of the deduced protein sequences. Shaded box highlights the cluster formed by the homologues from the same four clostridia comprising the box shaded in (a) (above).

unpublished genome sequences of the non-homoacetogenic oral treponeme *T. denticola*, and published genome of the syphilis spirochaete *T. pallidum*, did not cluster phylogenetically with *Treponema* strain ZAS-2 or the four clostridia (Fig. 4b). As was the case for genes encoding

FTHFS and MTHFC, the published genome of the Lyme disease spirochaete *B. burgdorferi* did not contain an obvious MTHFD homologue. The branching topology did not change significantly with the addition of more than 50 other demonstrated or deduced MTHFDs to the analysis (not shown).

DISCUSSION

We have successfully used FTHFS primers to amplify genes likely to be involved in the CO₂-reductive metabolism of termite-gut spirochaetes and the gut community of *Z. angusticollis*. We have cloned and analysed (1) FTHFS partial genes encoded by *Treponema* strains ZAS-1, ZAS-2 and ZAS-9, including extended sequence for strain ZAS-2, (2) a substantial collection of diverse FTHFS partial genes obtained from the gut community of *Z. angusticollis*, and (3) loci encoding two putative acetogenesis pathway enzymes from ZAS-2. To our knowledge, this represents the first appraisal of homoacetogen diversity in any known 'hot spot' for this activity, and also the first report on two key acetogenesis-related genes from any of the ~100 established species of homoacetogens.

Members of the genus *Treponema*, a genus within the phylum-level assemblage *Spirochaetes*, are only distantly related to the classical homoacetogens, most of which belong to the phylum *Firmicutes*. Treponemes are also only distantly related to other homoacetogens, such as *Desulfotignum phosphitoxidans*, a δ -proteobacterium, and *Holophaga foetida*, one of the few cultivated representatives of the phylum *Acidobacteria* (Liesack *et al.*, 1994; Schink *et al.*, 2002). It was somewhat surprising that primers designed to amplify FTHFS genes from clostridia functioned in doing so from termite-gut spirochaetes, an observation that has also been recently reported by others (Leaphart *et al.*, 2003). More surprising, FTHFS homologues from termite-gut *Treponema* strains clustered phylogenetically with those from homoacetogenic clostridia and not, for example, with that from the non-homoacetogenic oral spirochaete *T. denticola* (Fig. 2a). This pattern was re-encountered during the analysis of each of two other genes likely to encode THF-dependent enzymes identified from *Treponema* ZAS-2 (Fig. 4a, b). This suggests that THF-dependent enzymes from closely related termite-gut and human-associated treponemes may have distinctly different evolutionary origins. Thus, it can be postulated that the termite-gut spirochaete pathway may have been acquired via lateral gene transfer from a species closely related to *R. productus*. If so, was the event before or after spirochaetes had first become established within the symbiotic, termite-gut communities? The identification of a 'homoacetogen-like' FTHFS gene from the H₂-producing isolate *Treponema* ZAS-9 raises similar intrigue, as this strain is not a homoacetogen (Breznak & Leadbetter, 2002; Lilburn *et al.*, 2001). Of all the FTHFS genes that comprise Lovell's homoacetogen cluster, to date this is the only example representing a known non-homoacetogen. Might

this strain represent a 'snapshot' of a species in evolutionary transition either into or from having CO₂-reductive capacities?

Our initial concerns that the FTHFS primer set might be biased towards the overamplification of genes from the homoacetogenic clostridia (for which they had been designed) did not play out. Three-quarters of the clones recovered clustered closely with the spirochaete-derived sequences, several sharing remarkable amino acid similarities with the isolates, for example, 96.3% similarity shared between *Za*-gut Clone P and *Treponema* ZAS-2 (Fig. 2a). The results of this study establish that there could be as many as 20 strains representing at least two homoacetogenic phyla present in *Zootermopsis* guts. We are now interested in examining the nature and physical location of the cells encoding these genes, and in determining which types actually dominate FTHFS expression *in situ*. Are the organisms representing Clone P and Clone H truly spirochaetes, are they as abundant and active as the inventory results might suggest, and are they actually homoacetogens? Are the cells representing Clone F and Clone Y actually spore-forming clostridia? The termite-gut community is now recognized as being highly structured (Brune & Friedrich, 2000). Are all homoacetogen species found localized to the same temporal and spatial niches within the gut? The results of the current study should help facilitate the design of quantitative mRNA, cell-localization, community fingerprinting, and other approaches to address all such questions.

ACKNOWLEDGEMENTS

This research was supported by infrastructure funds from the National Science Foundation (DBI-0107908), and a gift from the Schlumberger Foundation. We thank Rick Lovell for helpful advice and sharing data on an extensive FTHFS inventory in advance of its publication, and E. R. Leadbetter and our laboratory co-workers for their helpful comments.

REFERENCES

- Bao, Q., Tian, Y., Li, W. & 18 other authors (2002). A complete sequence of the *T. tengcongensis* genome. *Genome Res* **12**, 689–700.
- Boga, H. I., Ludwig, W. & Brune, A. (2003). *Sporomusa aerivorans* sp. nov., an oxygen-reducing homoacetogenic bacterium from the gut of a soil-feeding termite. *Int J Syst Evol Microbiol* **53**, 1397–1404.
- Brauman, A., Kane, M. D., Labat, M. & Breznak, J. A. (1992). Genesis of acetate and methane by gut bacteria of nutritionally diverse termites. *Science* **257**, 1384–1387.
- Breznak, J. A. (1994). Acetogenesis from carbon dioxide in termite guts. In *Acetogenesis*, pp. 303–330. Edited by H. L. Drake. New York: Chapman & Hall.
- Breznak, J. A. & Leadbetter, J. R. (2002). Termite gut spirochetes. In *The Prokaryotes – an Evolving Electronic Resource for the Microbiological Community*. Edited by M. Dworkin, S. Falkow, E. Rosenberg, K. H. Schleifer & E. Stackebrandt. New York: Springer.

- Breznak, J. A. & Switzer, J. M. (1986). Acetate synthesis from H₂ plus CO₂ by termite gut microbes. *Appl Environ Microbiol* **52**, 623–630.
- Breznak, J. A., Switzer, J. M. & Seitz, H. J. (1988). *Sporomusa termitida* sp. nov., an H₂/CO₂-utilizing acetogen isolated from termites. *Arch Microbiol* **150**, 282–288.
- Brune, A. & Friedrich, M. (2000). Microecology of the termite gut: structure and function on a microscale. *Curr Opin Microbiol* **3**, 263–269.
- Casjens, S., Palmer, N., van Vugt, R. & 12 other authors (2000). A bacterial genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete *Borrelia burgdorferi*. *Mol Microbiol* **35**, 490–516.
- Drake, H. L., Daniel, S. L., Küsel, K., Matthies, C., Kuhner, C. & Braus-Stromeier, S. (1997). Acetogenic bacteria: what are the in situ consequences of their diverse metabolic versatilities? *Biofactors* **6**, 13–24.
- Eisen, J. A., Nelson, K. E., Paulsen, I. T. & 32 other authors (2002). The complete genome sequence of *Chlorobium tepidum* TLS, a photosynthetic, anaerobic, green-sulfur bacterium. *Proc Natl Acad Sci U S A* **99**, 9509–9514.
- Ferretti, J. J., McShan, W. M., Ajdic, D. & 20 other authors (2001). Complete genome sequence of an M1 strain of *Streptococcus pyogenes*. *Proc Natl Acad Sci U S A* **98**, 4658–4663.
- Fraser, C. M., Norris, S. J., Weinstock, G. M. & 30 other authors (1998). Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science* **281**, 375–388.
- Goldman, N. & Whelan, S. (2000). Statistical tests of gamma-distributed rate heterogeneity in models of sequence evolution in phylogenetics. *Mol Biol Evol* **17**, 975–978.
- Goodner, B., Hinkle, G., Gattung, S. & 28 other authors (2001). Genome sequence of the plant pathogen and biotechnology agent *Agrobacterium tumefaciens* C58. *Science* **294**, 2323–2328.
- Kane, M. D. & Breznak, J. A. (1991). *Acetonema longum* gen. nov. sp. nov., an H₂/CO₂ acetogenic bacterium from the termite, *Pterotermes occidentis*. *Arch Microbiol* **156**, 91–98.
- Kane, M. D., Brauman, A. & Breznak, J. A. (1991). *Clostridium mayombe* sp. nov., an H₂/CO₂ acetogenic bacterium from the gut of the African soil-feeding termite, *Cubitermes speciosus*. *Arch Microbiol* **156**, 99–104.
- Kapatral, V., Anderson, I., Ivanova, N. & 22 other authors (2002). Genome sequence and analysis of the oral bacterium *Fusobacterium nucleatum* strain ATCC 25586. *J Bacteriol* **184**, 2005–2018.
- Kuroda, M., Ohta, T., Uchiyama, I. & 34 other authors (2001). Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*. *Lancet* **357**, 1225–1240.
- Leadbetter, J. R., Schmidt, T. M., Graber, J. R. & Breznak, J. A. (1999). Acetogenesis from H₂ plus CO₂ by spirochetes from termite guts. *Science* **283**, 686–689.
- Leaphart, A. B. & Lovell, C. R. (2001). Recovery and analysis of formyltetrahydrofolate synthetase gene sequences from natural populations of acetogenic bacteria. *Appl Environ Microbiol* **67**, 1392–1395.
- Leaphart, A. B., Friez, M. J. & Lovell, C. R. (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.
- Liesack, W., Bak, F., Kreft, J. U. & Stackebrandt, E. (1994). *Holophaga foetida* gen. nov., sp. nov., a new, homoacetogenic bacterium degrading methoxylated aromatic compounds. *Arch Microbiol* **162**, 85–90.
- Lilburn, T. G., Schmidt, T. M. & Breznak, J. A. (1999). Phylogenetic diversity of termite gut spirochaetes. *Environ Microbiol* **1**, 331–345.
- Lilburn, T. G., Kim, K. S., Ostrom, N. E., Byzek, K. R., Leadbetter, J. R. & Breznak, J. A. (2001). Nitrogen fixation by symbiotic and free-living spirochetes. *Science* **292**, 2495–2498.
- Lovell, C. R., Przybyla, A. & Ljungdahl, L. G. (1988). Cloning and expression in *Escherichia coli* of the *Clostridium thermoaceticum* gene encoding thermostable formyltetrahydrofolate synthetase. *Arch Microbiol* **149**, 280–285.
- Ng, W.-L., Schummer, M., Cirisano, F. D., Baldwin, R. L., Karlan, B. Y. & Hood, L. (1996). High-throughput plasmid mini preparations facilitated by micro-mixing. *Nucleic Acids Res* **24**, 5045–5047.
- Nolling, J., Breton, G., Omelchenko, M. V. & 16 other authors (2001). Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *J Bacteriol* **183**, 4823–4838.
- Odelson, D. A. & Breznak, J. A. (1983). Volatile fatty acid production by the hindgut microbiota of xylophagous termites. *Appl Environ Microbiol* **45**, 1602–1613.
- Page, R. D. M. (1996). TREEVIEW: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* **12**, 357–358.
- Paster, B. J., Dewhirst, F. E., Cooke, S. M., Fussing, V., Poulsen, L. K. & Breznak, J. A. (1996). Phylogeny of not-yet-cultured spirochetes from termite guts. *Appl Environ Microbiol* **62**, 347–352.
- Purdy, K., Embley, T., Takii, S. & Nedwell, D. (1996). Rapid extraction of DNA and rRNA from sediments by a novel hydroxyapatite spin-column method. *Appl Environ Microbiol* **62**, 3905–3907.
- Radfar, R., Shin, R., Sheldrick, G. M., Minor, W., Lovell, C. R., Odom, J. D., Dunlap, R. B. & Lebioda, L. (2000a). The crystal structure of N¹⁰-formyltetrahydrofolate synthetase from *Moorella thermoacetica*. *Biochemistry* **39**, 3920–3926.
- Radfar, R., Leaphart, A., Brewer, J. M., Minor, W., Odom, J. D., Dunlap, R. B., Lovell, C. R. & Lebioda, L. (2000b). Cation binding and thermostability of FTHFS monovalent cation binding sites and thermostability of N¹⁰-formyltetrahydrofolate synthetase from *Moorella thermoacetica*. *Biochemistry* **39**, 14481–14486.
- Ragsdale, S. W. (1997). The eastern and western branches of the Wood/Ljungdahl pathway: how the east and west were won. *Biofactors* **6**, 3–11.
- Raponi, M., Dawes, I. W. & Arndt, G. M. (2000). Characterization of flanking sequences using long inverse PCR. *Biotechniques* **28**, 838–844.
- Ruepp, A., Graml, W., Santos-Martinez, M. L. & 7 other authors (2000). The genome sequence of the thermoacidophilic scavenger *Thermoplasma acidophilum*. *Nature* **407**, 508–513.
- Schink, B., Thiemann, V., Laue, H. & Friedrich, M. W. (2002). *Desulfotignum phosphitoxidans* sp. nov., a new marine sulfate reducer that oxidizes phosphite to phosphate. *Arch Microbiol* **177**, 381–391.
- Schmidt, H. A., Strimmer, K., Vingron, M. & von Haeseler, A. (2002). TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* **18**, 502–504.
- Schöniger, M. & von Haeseler, A. (1999). Toward assigning helical regions in alignments of ribosomal RNA and testing the appropriateness of evolutionary models. *J Mol Evol* **49**, 691–698.
- Shimizu, T., Ohtani, K., Hirakawa, H. & 7 other authors (2002). Complete genome sequence of *Clostridium perfringens*, an anaerobic flesh-eater. *Proc Natl Acad Sci U S A* **99**, 996–1001.
- Tholen, A. & Brune, A. (2000). Impact of oxygen on metabolic fluxes and *in situ* rates of reductive acetogenesis in the hindgut of the

wood-feeding termite *Reticulitermes flavipes*. *Environ Microbiol* **2**, 436–449.

White, O., Eisen, J. A., Heidelberg, J. F. & 29 other authors (1999). Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* R1. *Science* **286**, 1571–1577.

Wood, H. G. (1991). Life with CO or CO₂ and H₂ as a source of carbon and energy. *FASEB J* **5**, 156–163.

Wood, D. W., Setubal, J. C., Kaul, R. & 48 other authors (2001). The genome of the natural genetic engineer *Agrobacterium tumefaciens* C58. *Science* **294**, 2317–2323.

Front cover illustration: A phase contrast image of diluted gut contents from the termite *Zootermopsis angusticollis* exhibiting a diversity of spirochaetal morphotypes. Micrograph courtesy Jared R. Leadbetter (Environmental Science & Engineering, California Institute of Technology, Pasadena, CA, USA) and John A. Breznak (Department of Microbiology, Michigan State University, East Lansing, MI, USA).