

Published in final edited form as:

*Environ Microbiol.* 2008 October ; 10(10): 2550–2573. doi:10.1111/j.1462-2920.2008.01679.x.

## The Complete Genome Sequence of *Moorella thermoacetica* (f. *Clostridium thermoaceticum*)

Elizabeth Pierce<sup>1</sup>, Gary Xie<sup>2,3</sup>, Ravi D. Barabote<sup>2,3</sup>, Elizabeth Saunders<sup>2,3</sup>, Cliff S. Han<sup>2,3</sup>, John C. Detter<sup>2,3</sup>, Paul Richardson<sup>2,4</sup>, Thomas S. Brettin<sup>2,3</sup>, Amaresh Das<sup>5</sup>, Lars G. Ljungdahl<sup>5</sup>, and Stephen W. Ragsdale<sup>1</sup>

<sup>1</sup> Department of Biological Chemistry, University of Michigan, Ann Arbor, Michigan

<sup>2</sup> Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico

<sup>3</sup> Department of Energy Joint Genome Institute, Walnut Creek, CA

<sup>4</sup> Lawrence Berkeley National Laboratory, Berkeley, CA

<sup>5</sup> Department of Biochemistry and Molecular Biology, University of Georgia

### Summary

This paper describes the genome sequence of *M. thermoacetica* (f. *Clostridium thermoaceticum*), which is the model acetogenic bacterium that has been widely used for elucidating the Wood-Ljungdahl pathway of CO and CO<sub>2</sub> fixation. This pathway, which is also known as the reductive acetyl-CoA pathway, allows acetogenic (often called homoacetogenic) bacteria to convert glucose stoichiometrically into three mol of acetate and to grow autotrophically using H<sub>2</sub> and CO as electron donors and CO<sub>2</sub> as an electron acceptor. Methanogenic archaea use this pathway in reverse to grow by converting acetate into methane and CO<sub>2</sub>. Acetogenic bacteria also couple the Wood-Ljungdahl pathway to a variety of other pathways to allow the metabolism of a wide variety of carbon sources and electron donors (sugars, carboxylic acids, alcohols, and aromatic compounds) and electron acceptors (CO<sub>2</sub>, nitrate, nitrite, thiosulfate, dimethylsulfoxide, and aromatic carboxyl groups). The genome consists of a single circular 2628784 bp chromosome encoding 2615 open reading frames, which includes 2523 predicted protein-encoding genes. Of these, 1834 genes (70.13%) have been assigned tentative functions, 665 (25.43%) matched genes of unknown function, and the remaining 24 (0.92%) had no database match. Two thousand three hundred eighty-four (91.17%) of the ORFs in the *M. thermoacetica* genome can be grouped in ortholog clusters. This first genome sequence of an acetogenic bacterium provides important information related to how acetogens engage their extreme metabolic diversity by switching among different carbon substrates and electron donors/acceptors and how they conserve energy by anaerobic respiration. Our genome analysis indicates that the key genetic trait for homoacetogenesis is the core *acs* gene cluster of the Wood-Ljungdahl pathway.

### Introduction

Here we report the first complete genome sequence of an acetogenic bacterium. Acetogens are obligately anaerobic bacteria that use a mechanism of anaerobic CO and CO<sub>2</sub> fixation, called the reductive acetyl-CoA or the Wood-Ljungdahl pathway as their main mechanism for energy conservation and for synthesis of acetyl-CoA and cell carbon from CO<sub>2</sub> (Drake et al., 1994). Other terms to describe an organism using this type of metabolism are

Please address correspondence to: Dr. Stephen W. Ragsdale, Department of Biological Chemistry, University of Michigan Medical School, 1150 W. Medical Center Dr., 5301 MSRB III, Ann Arbor, MI 48109-0606, Tel: (734) 615-4621, Fax: 734-763-4581, Email: sragdsal@umich.edu.

“homoacetogen” and “CO<sub>2</sub>-reducing acetogen”. We determined the genome sequence of *Moorella thermoacetica*, originally named *Clostridium thermoaceticum*, which has been the model acetogen since its isolation in 1942 (Fontaine et al., 1942) for elucidating the Wood-Ljungdahl pathway. As discussed below, this pathway is found in a broad range of phylogenetic classes, and is used in both the oxidative and reductive directions. Besides being used for energy production and autotrophic carbon assimilation in acetogens (Ljungdahl, 1994; Ragsdale, 1997; Drake et al., 2002), the Wood-Ljungdahl pathway is used for acetate catabolism by sulfate reducing bacteria (Schauder et al., 1988; Spormann and Thauer, 1988) and aceticlastic methanogens (Ferry, 1999) and for CO<sub>2</sub> fixation by hydrogenotrophic methanogens (Stupperich et al., 1983; Ladapo and Whitman, 1990).

As early as 1932, organisms were discovered that could convert H<sub>2</sub> and CO<sub>2</sub> into acetic acid (Fischer et al., 1932). In 1936, Wieringa reported the first acetogenic bacterium, *Clostridium aceticum* (Wieringa, 1936, 1940), which was reisolated in 1981 (Braun et al., 1981). *M. thermoacetica* (Collins et al., 1994), a clostridium in the Thermoanaerobacteriaceae family, attracted wide interest when it was isolated because of its unusual ability to convert glucose almost stoichiometrically to three moles of acetic acid (Fontaine et al., 1942). It took many years to uncover the biochemical reactions responsible for this homoacetogenic pathway. The novelty of this pathway was recognized with the discovery that growth of *M. thermoacetica* on glucose involves incorporation of two molecules of CO<sub>2</sub> directly into acetic acid (Barker and Kamen, 1945; Wood, 1952). The isolation of *Acetobacterium woodii* in 1977 and the description of methods to isolate organisms that can grow on H<sub>2</sub> and CO<sub>2</sub> (Balch et al., 1977) led to the isolation of many acetogens that can grow autotrophically, and in 1983 *M. thermoacetica* was shown to grow on H<sub>2</sub>/CO<sub>2</sub> (Kerby and Zeikus, 1983). The identification of a hydrogenase in 1982 provided a biochemical basis for autotrophic growth of acetogens on hydrogen (Drake, 1982). Subsequently, acetogens have been shown to use a wide variety of carbon sources and electron donors and acceptors. *M. thermoacetica* is a versatile heterotroph, growing on sugars, two-carbon compounds (glyoxylate, glycolate, and oxalate), lactate, pyruvate, short-chain fatty acids, and methoxylated aromatic compounds. Besides CO<sub>2</sub>, electron acceptors include nitrate, nitrite, thiosulfate, and dimethylsulfoxide.

Acetogenic bacteria play an important role in the global carbon cycle, producing over 10<sup>12</sup> kg (10 billion US tons) of acetic acid on a huge scale annually (Wood and Ljungdahl, 1991; Drake et al., 1994), which dwarfs the total output of acetate (an important chemical feedstock) by the world's chemical industry (~10<sup>10</sup> kg/yr). While most acetogens like *M. thermoacetica* are in the phylum Firmicutes, acetogens include Spirochaetes (Leadbetter et al., 1999; Salmassi and Leadbetter, 2003), δ-proteobacteria like *Desulfotignum phosphitoxidans* (Schink et al., 2002), and Acidobacteria like *Holophaga foetida* (Liesack et al., 1994). Important in the biology of the soil, lakes, and oceans, acetogens have been isolated from diverse environments, including the GI tract of animals and termites (Brezna, 1994; Mackie and Bryant, 1994), rice paddy soils (Rosencrantz et al., 1999), hypersaline waters (Ollivier et al., 1994), surface soils (Peters and Conrad, 1995; Kusel et al., 1999), and deep subsurface sediments (Liu and Suflita, 1993). Like methanogens, acetogens act as a H<sub>2</sub> sink, depleting H<sub>2</sub> that is generated in anaerobic environments during the natural biodegradation of organic compounds. Since the build-up of H<sub>2</sub> inhibits biodegradation by creating an unfavorable thermodynamic equilibrium, acetogens enhance biodegradative capacity by coupling the oxidation of hydrogen gas to the reduction of CO<sub>2</sub> to acetate. Acetogens also have been found in a methanogenic mixed population from an army ammunition manufacturing plant waste water treatment facility (Adrian and Arnett, 2004) and a dechlorinating community that has been enriched for bioremediation (Macbeth et al., 2004).

For organisms that house acetogens in their digestive systems, like humans, termites, and ruminants (Greening and Leedle, 1989; Tholen and Brune, 1999; Chassard and Bernalier-Donadille, 2006), the acetate generated by microbial metabolism is a beneficial nutrient for the host and for other microbes within the community. In these ecosystems, acetogens can compete directly with hydrogenotrophic methanogenic archaea, or interact syntrophically with acetotrophic methanogens that use H<sub>2</sub> and CO<sub>2</sub> to produce methane (Chassard and Bernalier-Donadille, 2006). Organisms containing the *acs* gene cluster are also enriched in the marine and terrestrial vertebrate and invertebrate gut biomes (Gill et al., 2006; Warnecke et al., 2007). In the termite gut, acetogens are the dominant hydrogen sinks (Pester and Brune, 2007), and it has been proposed that acetate is the major energy source for the termite (Odelson and Breznak, 1983). However, methanogens are the dominant hydrogenotrophs in many environments since methanogens have a lower threshold for H<sub>2</sub> than acetogens (Le Van et al., 1998) and since the energy yield from the conversion of CO<sub>2</sub> and H<sub>2</sub> to methane is greater than that for conversion to acetate (Schink, 1997). Under such conditions, acetogens must often resort to other metabolic pathways for growth and, thus, have a highly diverse metabolic menu that comprises the biodegradation products of most natural polymers like cellulose and lignin, including sugars, alcohols, organic acids and aldehydes, aromatic compounds, and inorganic gases like CO, H<sub>2</sub>, and CO<sub>2</sub>. They also can use a variety of electron acceptors, e.g., CO<sub>2</sub>, nitrate, dimethylsulfoxide, fumarate, and protons. The *M. thermoacetica* genome sequence described here provides important information related to how acetogens engage this extreme metabolic diversity and how they switch with such facility among different carbon substrates and electron donors and acceptors.

## Results

### General genome features

The major features of the *M. thermoacetica* genome are listed in Table 1 and shown in Supplementary Figure S1. The genome consists of a single circular 2,628,784 bp chromosome that has a GC content of 56 % (Supplementary Figure S1). Base pair one of the chromosome was assigned within the putative origin of replication. The genome contains 51 tRNAs, 3 rRNAs (5S, 16S, and 23S), and 38 other RNAs predicted by the Rfam. Analysis of 16S rRNA sequences confirmed that *M. thermoacetica* belongs to the phylum Clostridia (Figure 1) and to the class *Thermoanaerobacteriaceae*. Figure 1 also highlights the known homoacetogenic microbes, revealing the wide distribution of phylogenetic classes in which this trait appears, clearly indicating that acetogenesis is a metabolic, not a phylogenetic trait. Besides the bacteria shown in Fig. 1, which are related to *M. thermoacetica*, the Wood-Ljungdahl pathway is also found in archaea. On the other hand, homoacetogenesis is a rare occurrence within a particular phylum. For example, two relatives of *M. thermoacetica* in the *Thermoanaerobacteriaceae* family whose genomes have been sequenced, *Thermoanaerobacter tengcongensis* (Bao et al., 2002), and *Thermoanaerobacter ethanolicus*, (draft sequence available at JGI), can also grow anaerobically on several mono- and disaccharides, and starch (Wiegel and Ljungdahl, 1981); however, among these three strains, only *M. thermoacetica* is homoacetogenic, and only *M. thermoacetica* has acetyl coenzyme A synthase (ACS), a key enzyme in the Wood-Ljungdahl pathway. Fermentation of glucose by *T. tengcongensis* or *T. ethanolicus* yields ethanol and/or acetate, 1.5–2 moles of CO<sub>2</sub>, and small amounts of H<sub>2</sub>. Thus, the genomic results clearly indicate that the trait of acetogenesis was distributed by horizontal transfer of the core genes of the Wood-Ljungdahl pathway.

The *M. thermoacetica* genome contains 2523 predicted protein-encoding genes, of which 1834 (70.13%) have been assigned tentative functions. Six hundred sixty-five predicted proteins (25.43%) matched proteins of unknown function, and the remaining 24 (0.92%) had

no database match. A total of 429 paralogous protein families were identified in *M. thermoacetica* strain ATCC 39073, containing 1307 predicted proteins (49.98% of the total). Two thousand three hundred eighty-four (91.17%) of the ORFs in the *M. thermoacetica* genome can be grouped in ortholog clusters. Fifty-eight pseudogenes were identified in the *M. thermoacetica* genome by the IMG annotation. A complete list of pseudogenes and gene remnants is provided in Supplementary Table S1. Analysis of these genes in *M. thermoacetica* indicates that 41 out of 58 are redundant genes, with paralogs in the *M. thermoacetica* genome. The seventeen that do not have paralogs encode either putative or hypothetical proteins. These results suggest that elimination of the functions of these redundant genes may be tenuously ascribed to selective pressures.

### The Wood-Ljungdahl pathway genes

Figure 2 shows the genes encoding the key enzymes in the Wood-Ljungdahl pathway of CO<sub>2</sub> fixation in which two molecules of CO<sub>2</sub> are reduced to a methyl and a carbonyl group, and are joined with coenzyme A (CoA) to make acetyl-CoA. Acetyl-CoA is then incorporated into cell carbon or converted to acetyl phosphate, whose phosphoryl group is transferred to ADP to generate ATP and acetate, the main growth product of acetogenic bacteria (for reviews see (Ljungdahl, 1986; Drake and Daniel, 1997; Ragsdale, 1997)). This pathway has been described to consist of an Eastern (blue coloring scheme) and a Western (red) branch (Figure 2) (Ragsdale, 1997).

**Genes encoding the Eastern branch of the Wood-Ljungdahl pathway**—The genes encoding enzymes in the Eastern branch of the pathway that are involved in the conversion of CO<sub>2</sub> to methyl-H<sub>4</sub>folate (Fig. 2) are common to most organisms, though they are expressed at much higher levels in acetogens. As shown in Figure 3, many of the genes encoding functions of the Eastern branch of the pathway are dispersed on the genome at locations distant from the core genes of the Wood-Ljungdahl pathway, which are involved in converting methyl-H<sub>4</sub>folate, CoA, and CO<sub>2</sub> to acetyl-CoA. The functional assignments of all these genes (except phosphotransacetylase, below) are backed by biochemical evidence based on protein sequence analysis of well-characterized enzymes.

Conversion of CO<sub>2</sub> to formate is catalyzed by the well characterized selenocysteine-containing formate dehydrogenase (Li et al., 1966; Andreesen and Ljungdahl, 1973; Yamamoto et al., 1983). Moth\_2312 and Moth\_2313, which are actually one gene, match the sequence deposited by Li, Ljungdahl, and Gollin in GenBank (gi 1658401) as the formate dehydrogenase a subunit. The two ORFs were separated in the original genome annotation because they contain the codon for selenocysteine, a stop codon in most contexts. Moth\_2314 matches the sequence of the β subunit. In acetogens, formate dehydrogenase catalyzes an especially interesting reaction since the reduction of CO<sub>2</sub> to formate ( $E_0 = -420$ ) is coupled to the oxidation of NADPH to NADP<sup>+</sup> ( $E_0 = -340$  mV). The *M. thermoacetica* formate dehydrogenase was the first enzyme shown to contain tungsten (Ljungdahl and Andreesen, 1975). Another selenocysteine-containing formate dehydrogenase (Moth\_2193), which appears to be part of a formate hydrogen lyase system (Moth\_2184-2193), is described below.

The genes encoding 10-formyl-H<sub>4</sub>folate synthetase (Moth\_0109), and the bifunctional 5,10-methenyl-H<sub>4</sub>folate cyclohydrolase/5,10-methylene-H<sub>4</sub>folate dehydrogenase (Moth\_1516) are dispersed on the genome and distant from the *acs* gene cluster by at least 300 genes (Fig 3). In the reaction catalyzed by 10-formyl-H<sub>4</sub>folate synthetase, formate undergoes an ATP-dependent condensation with H<sub>4</sub>folate, forming 10-formyl-H<sub>4</sub>folate. This enzyme has been purified, characterized, and sequenced from a number of sources including *M. thermoacetica* (Himes and Wilder, 1968; O'Brien et al., 1975; McGuire and Rabinowitz,

1978; Lovell et al., 1988; Lovell et al., 1990). In higher organisms, this enzyme exists as one of the activities of a trifunctional C1-H<sub>4</sub>folate synthase (also containing cyclohydrolase and dehydrogenase) (Staben and Rabinowitz, 1986; Shannon and Rabinowitz, 1988), whereas the bacterial formyl-H<sub>4</sub>folate synthetases are monofunctional and share similar properties (Lovell et al., 1990). While the cyclohydrolase and dehydrogenase are part of a bifunctional protein in *M. thermoacetica*, they are monofunctional proteins in other acetogens, e.g., *Clostridium formicoaceticum* and *Acetobacterium woodii* (Moore et al., 1974; Clark and Ljungdahl, 1982; Ragsdale and Ljungdahl, 1984).

The enzyme that catalyzes the last step in the eastern branch is encoded by the 5,10-methylene-H<sub>4</sub>folate reductase gene (Moth\_1191). In *M. thermoacetica*, this is an oxygen-sensitive enzyme containing an iron-sulfur cluster, zinc, and FAD, which utilizes reduced ferredoxin as an electron donor (Clark and Ljungdahl, 1984; Park et al., 1991). Thauer et al. (Thauer et al., 1977) suggested that the 5,10-methylene-H<sub>4</sub>folate reductase may link to an electron transport chain that begins with an electron donor (e.g., hydrogenase, CODH, or NADH dehydrogenase) through membrane-bound electron carriers, finally to methylene-H<sub>4</sub>folate. In this scenario, electron transport could result in the generation of a transmembrane proton potential coupled to ATP synthesis. In *A. woodii*, one of the steps involving methylene-H<sub>4</sub>folate reductase, the methyltransferase or the CFeSP (see Figure 2) was shown to generate a sodium motive force across the membrane (Heise et al., 1989). However, since the reductase appears to be located in the cytoplasm in *M. thermoacetica* (Clark and Ljungdahl, 1984) and in other organisms, it appeared that this is not the energy-conserving step in the Wood-Ljungdahl pathway (Muller, 2003). On the other hand, the location of the 5,10-methylene-H<sub>4</sub>folate reductase gene very near genes encoding a hydrogenase (Moth\_1193) and heterodisulfide reductase (Moth\_1194-1196), which are directly downstream from the *acs* gene cluster (Moth\_1197-1204), suggests a possible mechanism of energy conservation involving the reductase. Thus, perhaps methylene-H<sub>4</sub>folate reductase links to the membrane-associated hydrogenase and/or heterodisulfide reductase to generate the proton motive force.

**Genes encoding the Western branch of the Wood-Ljungdahl pathway**—The core genes of the Western branch of the Wood-Ljungdahl pathway were first identified in *M. thermoacetica* by a reverse genetics approach to be located in the *acs* gene cluster (Figure 3b), which contains genes encoding carbon monoxide dehydrogenase (CODH), acetyl-CoA synthase (ACS), methyltransferase (MeTr), and both subunits of the corrinoid iron-sulfur protein (CFeSP) (Roberts et al., 1989). The *acs* gene cluster also includes a gene encoding an iron-sulfur protein of unknown function (*orf7*), and two genes, *cooC* and *acsF*, that encode proteins with sequences that are 44 and 27 percent identical to that of *Rhodospirillum rubrum cooC*, which is required for nickel insertion into carbon monoxide dehydrogenase (Jeon et al., 2001). Thus, the ubiquitous genes in the Eastern branch of the Wood-Ljungdahl pathway are scattered around the genome, while what can be considered to be the “core set of Wood-Ljungdahl genes” in the Western part of the pathway are co-localized. The genome sequence reveals an additional carbon monoxide dehydrogenase (Moth\_1972), which is similar to an uncharacterized carbon monoxide dehydrogenase from *Clostridium cellulolyticum* and to CODH IV from *Carboxydotherrmus hydrogenoformans* (Wu et al., 2005), and is unlinked to the *acs* gene cluster.

Phosphotransacetylase has been characterized from *M. thermoacetica* (Drake et al., 1981), but the gene sequence has not been reported. *M. thermoacetica* lacks any gene that is homologous to the phosphotransacetylase genes whose products are used for phosphorylation of acetate from acetyl-CoA in related organisms, like *Clostridium acetobutyricum* (Boynton et al., 1996). However, Moth\_1181 and Moth\_0864 are homologous to the recently characterized *pduL*, which is used for 1,2-propanediol

degradation by *Salmonella enterica* (Liu et al., 2007). Therefore, Moth\_1181 and Moth\_0864 are likely to encode phosphotransacetylases in *M. thermoacetica*. The acetate kinase (Moth\_0940) gene is distant from and, thus, is not linked to Moth\_1181, Moth\_0864, or the *acs* gene cluster; however, Moth\_1181 is located only 16 genes downstream of the methyltransferase in the Wood-Ljungdahl pathway (or ten genes further downstream from that of the gene encoding methylene-H<sub>4</sub>folate reductase).

### Comparison of Wood-Ljungdahl pathway gene clusters

Acetyl-CoA synthase (COG1614, *acsB*) is found in five Clostridia, three Deltaproteobacteria, and three Dehalococcoidetes species among the completely sequenced bacteria, and in the almost complete genome sequence of the uncultured bacterium *Kuenenia stuttgartiensis* (Strous et al., 2006), and seems to be a marker for the presence of the Wood-Ljungdahl pathway in bacteria. COG1614 is also found in archaea and corresponds to the ~53 kDa beta subunit that harbors the active site dinickel-iron-sulfur cluster of acetyl-CoA decarbonylase synthase in methanogens and archaeoglobus (Pfam 03598, *cdhC*) (Maupin-Furlow and Ferry, 1996; Gencic and Grahame, 2003). From these genomes with acetyl-CoA synthase genes, only *Dehalococcoides ethenogenes* (Seshadri et al., 2005), and *Dehalococcoides* sp. BAV1 and CBDB1 do not have the complete Wood-Ljungdahl pathway. Though these genomes encode many enzymes that are part of the pathway, they lack homologs of known methylene-tetrahydrofolate reductase, methyltransferase (*acsE*), CO dehydrogenase (*acsA*), phosphotransacetylase (*pta* or *pduL*), and acetate kinase sequences. We have been unable to locate any gene in the *Dehalococcoides* genome that is homologous to any of the CODH (*acsA*) genes from *M. thermoacetica* or *C. hydrogenoformans*. Given the small genome size and metabolic specialization of *D. ethenogenes*, one can speculate that the complete pathway was present in an ancestor, and has partly been lost as the organism's metabolism became limited to its current capabilities. *Alkaliphilus metalliredigens* (Ye et al., 2004), Carboxydotherrmus hydrogenoformans (Wu et al., 2005), two *Clostridium difficile* strains, two *Desulfitobacterium hafniense* strains (Nonaka et al., 2006), *Syntrophobacter fumaroxidans* (Harmsen et al., 1998), Deltaproteobacterium MLMS-1, *K. stuttgartiensis* (Strous et al., 2006), and *Candidatus Desulfococcus oleovorans* Hxd3, formerly Deltaproteobacterium Hxd3 (So et al., 2003), appear to have complete Wood-Ljungdahl pathways, although none of these organisms produces acetate as their main growth product.

Since acetogenesis is a physiological trait that is not phylogenetically conserved (see Figure 1), we reasoned that co-occurrence of acetyl-CoA synthase with a set of genes only in genomes of acetogenic bacteria could provide clues about which genes are required to confer the ability to grow as an acetogen. Thus, the *M. thermoacetica* genome was compared to those of all Firmicute, Chlorflexi, and Deltaproteobacterial sequences available at IMG, using a Blast analysis of each *M. thermoacetica* genes against these other genomes, to find *M. thermoacetica* genes with homologs in the other bacterial genomes that also have acetyl-CoA synthase, but without homologs in any of the sequences without acetyl-CoA synthase. The only genes that show this pattern of co-occurrence and co-omission are ACS and both subunits of the corrinoid iron-sulfur protein (CFeSP). The CFeSP must have a highly conserved function as part of the Wood-Ljungdahl pathway, since homologs of the CFeSP genes are not found in any genome that does not also have acetyl-CoA synthase and other Wood-Ljungdahl pathway genes.

The Wood-Ljungdahl pathway gene clusters in these twelve bacteria can be divided into five groups within which the order of genes is absolutely conserved (Figure 4). These groupings mostly follow the currently accepted taxonomy found at NCBI. The *M. thermoacetica*, *C. hydrogenoformans*, *D. hafniense*, and *K. stuttgartiensis* (group A) genes are arranged in the same way, and, like *M. thermoacetica*, *C. hydrogenoformans* has genes encoding

heterodisulfide reductase directly following *acsE*. The *K. stuttgartiensis* gene cluster is identical in arrangement to the *D. hafniense* gene cluster, except for the insertion of a gene with sequence similar to clp protease sequences and one small gene of unknown function after *acsB*. The *S. fumaroxidans* and *Candidatus Desulfococcus oleovorans* (group B) gene clusters are alike, and minimal, with only the *acsABCDE* genes. *C. D. oleovorans* has four proteins in the same COG (3640) as the *M. thermoacetica* accessory proteins, and *S. fumaroxidans* has two proteins from this COG, but none of the genes encoding these putative accessory proteins are near the *acs* gene cluster in either species. *A. metalliredigenes* and *C. difficile* (group C) have much larger Wood-Ljungdahl pathway gene clusters, which, besides the genes present in the *M. thermoacetica* *acs* gene cluster, include genes encoding formyl-tetrahydrofolate synthetase (*fts*), the bifunctional formyl-tetrahydrofolate cyclohydrolase/methylene-tetrahydrofolate dehydrogenase (*folD*), and methylene-tetrahydrofolate reductase, which function in the eastern branch of the pathway (Figure 2). The three *Dehalococcoides* species (group D) have gene clusters that include *acsBCD*, *fts*, and *folD*; however, they surprisingly lack *acsA*. The deltaproteobacterium MLMS-1 gene cluster (group E) is similar to the *Dehalococcoides* Wood-Ljungdahl gene clusters, but includes *acsA*, genes which belong in COG0493 and COG3383 (see the anaerobic respiration and electron acceptors discussions), one heterodisulfide reductase subunit A gene, and two genes encoding homologs of the archaeal hydrogenase delta subunit that is found in the HDR gene clusters in *M. thermoacetica*. Based on this analysis, it appears that the core genes of the Wood-Ljungdahl pathway are located in the *acs* gene cluster, which includes *acsA-acsE*, and acquisition of this gene cluster might be sufficient for gaining the ability to grow as an acetogen.

### Genes involved in anaerobic respiration

In the Wood-Ljungdahl pathway, one ATP is consumed in the formation of formyl-tetrahydrofolate from tetrahydrofolate and formate in the formyl-tetrahydrofolate synthase reaction (Sun et al., 1969), and one is generated by phosphoryl transfer to ADP in the acetate kinase reaction (Figure 2) (Schaupp and Ljungdahl, 1974). Therefore, there is no net gain of ATP by substrate level phosphorylation during autotrophic growth. Thus, it is clear that ATP is generated by anaerobic respiration. During growth using the Wood-Ljungdahl pathway, energy should be conserved by transfer of electrons along a membrane-bound electron transfer chain, coincident with generation of a transmembrane proton gradient (Figure 5). *M. thermoacetica* has an  $F_1F_0$  ATP synthase, which could use this proton gradient to drive ATP synthesis (Das and Ljungdahl, 1997). The genes encoding the eight subunits of the ATP synthase are Moth\_2377-84. *M. thermoacetica* has integral membrane-bound electron carriers including *b* cytochromes (see Table 2) and a menaquinone (Gottwald et al., 1975; Hugenholz et al., 1987; Das et al., 1989) that could be part of this system. There is evidence in the *M. thermoacetica* genome for genes that encode systems that could be associated with energy generation by transmembrane proton gradient formation (Table S2), including NADH dehydrogenase, hydrogenase or formate hydrogen lyase, and heterodisulfide reductase.

There is considerable homology of both subunit compositions and sequences among NADH dehydrogenases and some energy conserving NiFe hydrogenases (Deppenmeier, 2004; Hedderich, 2004). Examples of the former are mitochondrial complex I and *E. coli* NADH dehydrogenase I. The latter include *E. coli* hydrogenases 3 and 4, which are parts of formate hydrogen lyase complexes (Andrews et al., 1997), the CO-inducible hydrogenase from *Rhodospirillum rubrum* (Fox et al., 1996), a CO dehydrogenase-associated hydrogenase from *Carboxydotherrmus hydrogenoformans* (Soboh et al., 2002), and the Ech hydrogenase of *Methanosarcina barkeri* (Meuer et al., 1999). Two sets of *M. thermoacetica* genes, Moth\_0977-87 and Moth\_2184-92, encode proteins homologous to these complexes. A

BLAST search of *C. hydrogenoformans*, *E. coli*, *M. barkeri*, and *R. rubrum* genomes with genes from Moth\_0977-87 (Table 3) shows that these genes are most closely related to NADH dehydrogenase I in bacteria, and to the F<sub>420</sub>-dependent hydrogenase of *M. barkeri*. In some bacteria, as shown in Table 3 for *R. rubrum* and *E. coli*, the NADH dehydrogenase operon contains genes encoding subunits E, F, and G of the NADH dehydrogenase complex. In *M. thermoacetica*, homologs to these subunits are encoded separately, on two different gene clusters, Moth\_1717-9 and Moth\_1886-8. Moth\_1717 and Moth\_1888 align only with the first 240 and 280 amino acids of nuoG sequences from *E. coli* and *Thermus thermophilus*. In *T. thermophilus*, this N-terminal portion of the protein has the residues that ligate the three iron-sulfur clusters that are involved in electron transfer through the NADH dehydrogenase complex (Sazanov and Hinchliffe, 2006).

Proteins encoded by Moth\_2184-92 (Table 4) are most closely related to *E. coli* hydrogenases 4 and 3 among experimentally characterized proteins. In *C. hydrogenoformans*, Moth\_2184-88 and Moth\_2190-92 are most closely related to CO-associated hydrogenase, although there is lower sequence homology between these *M. thermoacetica* genes and CO-associated hydrogenase than between Moth\_2184-92 and *E. coli* hydrogenases 3 and 4. Moth\_2184-90 are more similar to subunits of hydrogenase 4 than subunits of hydrogenase 3, and Moth\_2187-89 encode homologs of *E. coli* hydrogenase 4 genes *hyfDEF*, which do not have homologs in hydrogenase 3. Hydrogenase 3 does not generate a proton gradient, and probably is used by *E. coli* to metabolize formate produced during fermentative growth. From their analysis of the *hyf* gene cluster, Andrews et al. concluded that hydrogenase 4 is likely to have proton-translocating ability (although this has not been verified experimentally), because, like the NADH dehydrogenase I complex and the *R. rubrum* CO-inducible hydrogenase, and unlike hydrogenase 3 from *E. coli*, *hyfBEF* are homologous to ND2, ND4, and ND5 components of the quinone-reducing and proton translocating machinery in NADH dehydrogenase I (Andrews et al., 1997). *M. thermoacetica* genes Moth\_2176-83, which are immediately downstream from this second gene cluster, are homologs of the *hypA-F*, *hycI*, and *hycH* genes of *E. coli*, which make up an operon that is required for maturation of the Ni-Fe hydrogenase 3 (Lutz et al., 1991; Blokesch et al., 2004). Immediately upstream of the hydrogenase gene cluster is a gene encoding a selenocysteine-containing formate dehydrogenase (Moth\_2193) that is separate from the formate dehydrogenase that feeds formate into the Wood-Ljungdahl pathway. It is likely that Moth\_2184-93 encodes a formate hydrogenlyase system, which by homology to hydrogenase 4, could generate a proton gradient in the presence of formate.

In methanogenic archaea, hydrogenases distinct from the Ech hydrogenase discussed above are also involved in energy generation as parts of membrane-bound electron transport systems, along with heterodisulfide reductase. In these systems, electrons are transferred from hydrogen to the disulfide formed between coenzyme M and coenzyme B during the last step in methanogenesis. Both two-subunit heme *b*-containing and three-subunit (non-heme) heterodisulfide reductases have been characterized (Setzke et al., 1994; Deppenmeier, 2004). Although it is not an archaeon, *M. thermoacetica* has three heterodisulfide reductase gene clusters that are homologous to the three-subunit heterodisulfide reductase from *Methanothermobacter marburgensis* (Table 5), one of which (Moth\_1196) is directly after the CODH/ACS gene cluster. Moth\_0441 (a heterodisulfide reductase A subunit) includes about 200 amino acids (from 400–600) that do not align with HdrA domains from *Methanosarcina* or most bacterial hits. These amino acids align with lower E-values with sequences from *Syntrophobacter fumaroxidans*, *Methanosarcina mazei*, an uncultured archaeon, *Desulfotomaculum reducens*, two *Geobacter* species, and *Candidatus Desulfococcus oleovorans*. Moth\_1194 aligns with HdrA, except between amino acids 150 and 550. This region aligns with COG0493, which is made of the NADPH-dependent glutamate synthase beta chain, and related oxidoreductases. Moth\_0811 and Moth\_0812



make up one selenocysteine-containing gene, which aligns with other HdrA sequences along its full length.

### Genes involved in autotrophic growth

One-carbon compounds that *M. thermoacetica* can use for growth include H<sub>2</sub>/CO<sub>2</sub>, CO, formate, methanol, and methyl groups from many methoxylated aromatic compounds. The enzymes that catalyze transfer of the methyl group of methoxylated aromatic compounds to tetrahydrofolate are encoded by *mtvABC* (Moth\_0385-7 and Moth\_1316-8), and have been purified and identified biochemically by enzymatic function and protein sequence analysis (Naidu and Ragsdale, 2001). This pathway is shown in Figure 6, along with that of a methanol methyltransferase system (MtaABC, encoded by (Moth\_1208-9 and Moth\_2346) was recently characterized from *M. thermoacetica* (Das et al., 2007). *M. thermoacetica* has two nearly identical *mtvA* genes, each of which is next to a gene that we assign as *mtvB*, and a predicted cobalamin-binding *mtvC* gene. The putative *mtvB* genes are eighty percent identical to each other, and the *mtvC* genes differ by one amino acid. Both *mtvA* and *mtvC* genes are similar to cobalamin-dependent methionine synthase (5-methyltetrahydrofolate:homocysteine S-methyltransferase, encoded by *metH* in *E. coli*).

The two *mtvB* genes are most similar to uroporphyrinogen decarboxylase sequences, among experimentally characterized proteins, although the function of MtvB is clearly associated with methyl transfer. In addition to these *mtv* genes, *M. thermoacetica* encodes several homologs of MtvA, MtvB and MtvC (Table 6). It is possible that some of these homologs are used for growth on methyl donors other than vanillate and methanol. Many homologs of the *M. thermoacetica* MtvA and C proteins are also found in *Desulfitobacterium hafniense*, which can also demethylate methoxylated aromatic compounds (Neumann et al., 2004). These are arranged in clusters with A and C genes and sequences aligning either with *M. thermoacetica* MtvB or sequences of similar function from *Acetobacterium dehalogenans*. One of the *M. thermoacetica* MtvB homologs (Moth\_1314) is fused to a domain that has not been functionally identified, and does not align with any COG identified at NCBI. Sequences that align with this whole fusion are found in *D. hafniense*, *Archaeoglobus fulgidus* and *Alkaliphilus metalliredigenes*.

### Genes involved in heterotrophic growth

*M. thermoacetica* can grow heterotrophically on fructose, glucose, xylose, ethanol, *n*-propanol, *n*-butanol (Beaty and Ljungdahl, 1991), oxalate and glyoxylate (Daniel and Drake, 1993), glycolate (Seifritz et al., 1999), pyruvate (Barker and Kamen, 1945), and lactate (Göbner et al., 1999). *M. thermoacetica* encodes a complete Embden-Meyerhof-Parnas pathway and all enzymes that are part of the oxidative branch of the pentose phosphate pathway (Supplementary Figure S2). Glucose and xylose are converted stoichiometrically to acetate by *M. thermoacetica* (Fontaine et al., 1942; Barker and Kamen, 1945). It seems likely that fructose is metabolized by the same pathway as glucose (Supplementary Figure S3), since growth yields of the two substrates are very similar (Andreesen et al., 1973).

Ethanol, propanol, and butanol are oxidized to acetate, propionate, and butyrate by *M. thermoacetica* (Beaty and Ljungdahl, 1991). Eight genes from four COGs containing alcohol dehydrogenases are present in the genome (Supplementary Table S3). Proteins encoded by three genes from COG2414, tungsten-containing aldehyde:ferredoxin oxidoreductases, might catalyze the second step in growth on these alcohols.

The pathways for growth of *M. thermoacetica* on the two-carbon substrates oxalate, glyoxylate, and glycolate are not known. The stoichiometries of acetate production suggest that they are oxidized by two, four, and six electrons, respectively (Seifritz et al., 1999).

Given differences in growth on these three substrates, it seems likely that *M. thermoacetica* uses different pathways for their oxidation. Only partial pathways known to be involved in glycolate metabolism in other organisms are present in *M. thermoacetica*. In *E. coli*, the glycolate metabolism genes in *E. coli* are located in an operon, which includes *glcA* (encodes a glycolate permease), *glcB* (encodes malate synthase), *glcG* (function is unknown), *glcF*, *E*, and *D* (encode the three subunits of glycolate oxidase), and *glcC* (encodes a transcriptional activator) (Pellicer et al., 1996), and the first enzymatic step is oxidation of glycolate to glyoxylate via an FMN-dependent oxidase (GlcDEF). The *M. thermoacetica* genes are arranged in two clusters, one containing *glcDEF* homologs (Moth\_0037-39), and the other containing homologs to *glcACF*, an unknown protein, another *glcF*, and a sequence that aligns with *glcD* and *E* (Moth\_2303-8). The glycolate permease gene (*glcA*) in *E. coli* and the proposed *M. thermoacetica glcA* are very similar to L-lactate permeases (Núñez et al., 2001). No homolog of malate synthase was found in *M. thermoacetica*, but *M. thermoacetica* does have a gene, Moth\_1954, with 51 percent sequence identity to a possible glyoxylate reductase that has been purified from *Thermococcus litoralis* (Ohshima et al., 2001). This enzyme catalyzes formation of malate from glyoxylate and acetyl-CoA and is homologous to D-lactate dehydrogenase.

*M. thermoacetica* is able to grow on lactate and pyruvate. Lactate is probably converted to pyruvate by Moth\_1826, which is similar to both *Clostridium acetobutylicum* and *Thermotoga maritima* L-lactate dehydrogenases. Pyruvate is metabolized using pyruvate:ferredoxin oxidoreductase (PFOR), which catalyzes the oxidative cleavage of pyruvate and attachment of the two carbon fragment to CoA, forming acetyl-CoA, CO<sub>2</sub>, and reduced ferredoxin (Supplementary Figure S4). PFOR is also essential in sugar metabolism, since it couples the Embden-Meyerhoff-Parnas pathway to the Wood-Ljungdahl pathway. In sugar metabolism, 4 additional electrons are available from the glyceraldehyde-phosphate dehydrogenase reaction, so 1 mol of glucose, i.e., 2 mol of pyruvate are required to supply the 8 electrons for reducing 2 mol of CO<sub>2</sub> to acetate. The PFOR that has been purified from *M. thermoacetica* is a homodimer, with a subunit mass of 120 kDa (Furdui and Ragsdale, 2002). The gene encoding this enzyme, Moth\_0064, is a fusion of three COGs, COG0674, COG1014, COG1013, which are annotated as alpha, gamma, and beta subunits. Five other sets of genes belonging to the same COGs as PFOR (Supplementary Table S4) could encode authentic pyruvate:ferredoxin oxidoreductases, but they could also encode other proteins in the oxoglutamate oxidoreductase family, like 2-ketoisovalerate oxidoreductase, indolepyruvate oxidoreductase, and 2-ketoglutarate oxidoreductase. Many of the residues that are known to be important in the well-characterized PFOR from *Desulfovibrio africanus* are conserved in the PFOR-like gene clusters of *M. thermoacetica*. These include residues that bind iron-sulfur clusters, thiamine pyrophosphate, and pyruvate. Conservation of these residues suggests that these *M. thermoacetica* gene clusters likely encode enzymes that perform similar functions to PFOR. Variation in the residue that forms a hydrophobic interaction with the methyl group of pyruvate in the *D. africanus* enzyme could be due to different substrate specificity of the protein products of these different gene clusters. For example, the replacement of this residue with a glycine residue could accommodate a substrate larger than pyruvate.

### Genes involved in utilization of electron acceptors other than CO<sub>2</sub>

*M. thermoacetica* can use electron acceptors other than CO<sub>2</sub> during heterotrophic growth. When nitrate and CO<sub>2</sub> are both supplied, nitrate is used preferentially (Seifritz et al., 1993). In medium containing a carbon source for heterotrophic growth, addition of nitrate stimulates growth, and reducing equivalents from H<sub>2</sub> and one-carbon substrates are used to make ammonium, while acetate production is decreased. Work by Fröstl et al. and Arendsen et al. (Fröstl et al., 1996; Arendsen et al., 1999) suggest that carbon incorporation through

the Wood-Ljungdahl pathway is repressed by nitrate. Nitrite (Seifritz et al., 2003), dimethylsulfoxide and thiosulfate (Beaty and Ljungdahl, 1991) are also electron sinks for *M. thermoacetica*. Some of these electron acceptors could be reduced by oxidoreductases similar to DMSO and TMAO reductases found in *Escherichia coli* and many other organisms. These proteins use molybdopterin cofactors, are often membrane associated, and in some cases are transported to the periplasm by twin-arginine transport (TAT) systems.

*M. thermoacetica* genes similar to the *E. coli* DMSO and TMAO reductases have been identified. The *E. coli* DMSO reductase has three subunits. DmsA is the molybdopterin-containing catalytic subunit (COG0243), DmsB is an iron-sulfur electron transfer subunit (COG0437), and DmsC is a membrane anchor subunit (COG3302). TMAO reductase has a cytochrome subunit instead of DmsC and B-like subunits. *M. thermoacetica* has eight genes belonging to COGs 0243 and 3383, which are made up of molybdopterin-containing proteins, shown in Table 7. Five of these genes are adjacent to COG0437 proteins. One homolog of the *E. coli* DmsC subunit is also present in *M. thermoacetica*. No oxidoreductase subunit homologous to the pentaheme cytochrome c family TorC subunit of the TMAO reductase in *E. coli* (gi 2506330, COG3005) (Gon et al., 2001) was found in *M. thermoacetica*. This subunit connects TMAO reductase in *E. coli* with respiration via the quinone pool. Thus, in *M. thermoacetica*, it is likely that one of the electron transfer proteins discussed above in the *Anaerobic Respiration* section couples TMAO and DMSO-derived electrons to the quinone pool.

Analysis of the N-terminal sequences of the molybdopterin oxidoreductases shows that three of the genes include likely TAT signal sequences. *M. thermoacetica* encodes TatA and TatC proteins (Moth\_1278 and Moth\_1279), but does not have a *tatB* gene. This is also true of *Bacillus subtilis*, which uses a system composed of A and C subunits (Jongbloed et al., 2004). Two genes similar to *torD*, flanking a gene cluster that encodes three molybdopterin-containing oxidoreductases, two of which have TAT sequences, might be chaperones for the twin-arginine transport system (Hatzixanthos et al., 2005; Genest et al., 2006).

In addition to the molybdopterin oxidoreductases, *M. thermoacetica* has additional genes that could be involved in growth on electron acceptors other than CO<sub>2</sub>, e.g., sulfite and nitrate (Supplementary Table 5). One gene cluster has three genes that have about 80 percent sequence identity with the dissimilatory sulfite reductase (*dsrAB*) and the gene *dsrD*, of unknown function, from *Desulfotomaculum thermocisternum* (Larsen et al., 1999). Directly upstream from these genes are a gene similar to the membrane component of *E. coli* nitrate reductase, *narI*, a gene similar to *nrfD* which is a transmembrane component of *E. coli* nitrite reductase, two predicted iron-sulfur containing proteins, and a cytochrome c. The dissimilatory sulfite reductase does not have any likely transmembrane regions, but the additional genes might serve to connect the reductase to a membrane-bound respiratory pathway. One other possible sulfite reductase gene cluster, Moth\_1629-32, has *dsrAB* genes, and two genes similar to *dsrC* from *Archaeoglobus fulgidus*, whose function is not fully resolved. The genes in this cluster have only about 30 percent sequence identity to the most closely related genes in GenBank. The four genes in this second cluster have no predicted transmembrane domains.

### Expanded gene groups in *M. thermoacetica* based on comparative COG analysis

The *M. thermoacetica* genome was compared with those of *T. ethanolicus* and *T. tengcongensis* to obtain insights into *M. thermoacetica*'s wide metabolic range. We analyzed the placement of *M. thermoacetica*, *T. ethanolicus*, and *T. tengcongensis* genes in COGs to identify functions with increased representation in *M. thermoacetica*. Supplementary Table S6 shows some of the COGs with increased representation along with the number of genes belonging to each COG. The largest expansions in *M. thermoacetica*, relative to the other

Thermoanaerobacteriaceae genome sequences, are in COGs that could encode proteins involved in electron transfer to or from growth substrates, including molybdopterin oxidoreductases in the dimethylsulfoxide reductase family detailed above, and in the xanthine dehydrogenase family (Kisker et al., 1999), with subunits in COGs 1529, 1319, and 2080. The xanthine dehydrogenase subunits might be part of a pathway of purine degradation including amidohydrolase genes (COG0402), which are adjacent to two of the xanthine dehydrogenase genes. The COGs containing heterodisulfide reductase and sulfite reductase discussed above are absent from *T. ethanolicus* and *T. tengcongensis*, except for one subunit A homolog in *T. ethanolicus*. *M. thermoacetica* has thirteen genes representing two COGs of proteins that are likely to contain iron-sulfur clusters, and that seem to have diverse functions in the *M. thermoacetica* genome. COG0247 is part of the possible glycolate oxidase proposed above as the homolog of *E. coli glcF*. COG0247 sequences are also adjacent to a heterodisulfide reductase A gene (MoTh\_1453) that is unlinked to any other heterodisulfide reductase subunits, and to one of the sulfite reductase gene clusters. A gene encoding a COG0437 protein is part of the same sulfite reductase gene cluster, and other COG0437 proteins are predicted to be encoded by genes adjacent to many of the molybdopterin oxidoreductase genes, and is found in two gene clusters with aldehyde-ferredoxin oxidoreductases (COG2414).

*M. thermoacetica* has four genes in COG1251, which contains NAD(P)H dependent nitrite reductase subunits similar to *E. coli*'s cytoplasmic nirB. However, *M. thermoacetica* lacks the gene encoding the other subunit of the enzyme found in *E. coli*, nirD (COG2146). One COG1251 gene is within an apparent glutamate synthase operon (MoTh\_1289-94), while another is adjacent to a gene cluster containing COG0437, an aldehyde-ferredoxin oxidoreductase, and a gene similar to *E. coli*'s MoaD, which is involved in molybdopterin biosynthesis (MoTh\_0721-24). The third COG1251 gene is in a gene cluster with COG0243 and the lone heterodisulfide reductase A gene (MoTh\_1452-54), while the fourth is next to a hypothetical protein sequence of unknown function.

Another large expansion is in COGs related to methyl transfer reactions. Included in these COGs are the methyltransferase of the Wood-Ljungdahl pathway (Figure 2) and other genes that could feed methyl groups into this pathway, including *mtvA* and *mtvC* homologs, and COG0407 which includes *mtvB* (Table 6). Four COG0407 genes are positioned close to *mtvA* and *mtvC* homologs, and four are adjacent to two putative sugar phosphate permeases that are members of family 1 of the major facilitator superfamily, in COG2271, which is also expanded in *M. thermoacetica*.

Another expanded COG group contains components of ABC-type nitrate/sulfonate/bicarbonate transport systems. These are represented by four gene clusters, each containing an ATPase, a permease component, and a periplasmic component. Each of the *Thermoanaerobacter* species encodes one complete transporter from these three COGs.

### Genes involved in cofactor synthesis

The only vitamin required for *M. thermoacetica*'s growth is nicotinic acid, a precursor of NAD and NADP (Lundie and Drake, 1984). *M. thermoacetica* appears to have genes encoding the phosphoribosyl transferase (MoTh\_1836), adenylyltransferase (MoTh\_0564), and glutamine dependent NAD synthetase (MoTh\_0553) needed for conversion of nicotinic acid to NAD (MoTh\_1836, MoTh\_0564, and MoTh\_0553).

Since only one vitamin is required for growth, genes encoding pathways for synthesis of all other cofactors should be present in the genome. Two metabolic routes for the synthesis of tetrahydrofolate are the *de novo* synthesis pathway (Supplementary Figure S5), and a salvage pathway. In *E. coli*, the *de novo* synthesis pathway requires the activities of the

proteins encoded by the genes *folE*, *folB*, *folK*, *folP*, *folC* and *folA* to produce the final product, 6-(S)-tetrahydrofolate. *M. thermoacetica* has no *folA* gene, but the isoenzyme encoded by *ydgB* (*folM*) is present in three copies. The activities of two other enzymes from this pathway have been measured in *E. coli*, but the genes that encode the proteins are not yet known, so no comparison is possible with the *M. thermoacetica* genome. The functions of the tetrahydrofolate salvage pathway in *M. thermoacetica*, which regenerates tetrahydrofolate from 5- or 10-formyl-tetrahydrofolate could be accomplished by methylene-tetrahydrofolate dehydrogenase/methenyl-tetrahydrofolate cyclohydrolase of the Wood-Ljungdahl pathway (Moth\_1516) and Moth\_2045, a phosphoribosylglycinamide formyl-transferase.

*M. thermoacetica* has a complete pathway of cobalamin biosynthesis. Cobalamin is a central cofactor in the autotrophic metabolism of *M. thermoacetica*, linking the methyl and carbonyl branches of the Wood-Ljungdahl pathway, with the methylated corrinoid iron-sulfur protein serving as a substrate of the acetyl-CoA synthase reaction. *M. thermoacetica* contains over twenty different types of cobalamin or precursors that total 300 – 700 nmol per gram of cells (Ljungdahl et al., 1966). The levels of corrinoids and their structural diversity in anaerobic bacteria have been reviewed (Stupperich et al., 1990). Both the anaerobic branch of corrin ring synthesis and the common branch of the adenosyl-cobalamin pathway are intact. Although several genes in the aerobic pathway of corrin ring synthesis are also present in the *M. thermoacetica* genome, these genes are not specific to the aerobic route, and genes that are unique to the aerobic pathway are absent in *M. thermoacetica*. Almost all of the cobalamin biosynthesis genes are located within a single gene cluster (Moth\_1087-1104) in the *M. thermoacetica* genome. ABC transporters, which are encoded by *cbiM*, *N*, *Q*, and *O*, mediate high affinity transport of cobalt ions for B<sub>12</sub> synthesis. Three gene clusters (Moth\_1219 to Moth\_1222, Moth\_1861 to Moth\_1863 and Moth\_2427 to Moth\_2429) have combinations of *cbiM*, *N*, *Q* and *O* genes.

As expected from the numbers of DMSO reductase and xanthine oxidase family proteins encoded in the *M. thermoacetica* genome, this organism has genes that could be used to synthesize molybdenum cofactors (Supplementary Figure 6). Most of these were found based on their homology to *E. coli* and *Arabidopsis thaliana* genes (Schwartz, 2005). *M. thermoacetica* has homologs of all *E. coli* molybdenum cofactor synthesis genes except for *moaE*, which encodes the large subunit of molybdopterin synthase. Some *M. thermoacetica* homologs of *mogA*, *moaC*, *moaA*, and *moeA* are found in one cluster of genes (Moth\_2132 to Moth\_2137), along with a gene belonging to pfam03473, which is a family of predicted sulfur-carrying proteins which are sometimes associated with molybdenum cofactor sulfurylase domains, as in molybdopterin synthase. Another gene cluster (Moth\_1819 to Moth\_1824) has a gene homologous to *mobB* from *E. coli*, the function of which is not known, another *moeA* homolog, a gene that matches the c-terminal half of *moaA*, and three subunits of a possible ABC-type molybdate transporter, homologous to *modABC* from *E. coli*.

### Transporter Genes

Two hundred ten transport genes, constituting approximately one hundred thirty transport systems, were identified in the *M. thermoacetica* genome (Supplementary Table 7). These include six alpha-helical membrane channels for the transport of anions, cations and small solutes. The genome encodes sixty-one secondary permeases, which transport a variety of substrates that include amino acids, sugars, anions, cations, nucleotides, and toxic drugs. Sixteen belong to the Major Facilitator Superfamily (MFS), two are RND-type transporters, six are DMT-type drug and metabolite transporters, and three belong to the MOP superfamily. One hundred three proteins encoded in the genome constitute approximately thirty-five ABC-type transport systems, which are predicted to transport anions, cations,

peptides, amino acids, vitamins, heavy metal ions, and toxic drugs. The genome also encodes four polypeptides that form a single fructose-type sugar-transporting phosphotransferase system. Three cation-transporting P-type ATPases were also recognized in the genome. Approximately nineteen putative transporters with unknown substrates were also identified in the genome.

### Motility Genes

*M. thermoacetica* is not motile, although stains show peritrichous flagella (Fontaine et al., 1942). Thirty-five *M. thermoacetica* genes encode proteins forming entire components and substructure of flagella, such as motor, switch complex, basal body, and filament (Supplementary Figure S7). These proteins fall into at least six groups: hook-associated proteins (FlgK and FlgL), basal and hook proteins (FlgE), capping proteins (FliD), biosynthesis secretory proteins (FlhA, FlhB, FliF, FliH and FliI), flagella formation proteins, motor proteins (FliG and FliM) and basal proteins (FlgC and FlgB). However, genes encoding the accessory flagellar assembly proteins, FlgA, FlhC, FlhD, and FliT, could not be identified in the *M. thermoacetica* genome. The organization of the genes for the general component of the flagellar proteins export apparatus appears to be similar to that found in *Bacillus subtilis* (Macnab, 2004). A similar gene cluster exists in *Thermoanaerobacter tengcongensis*, which also does not have functional flagella in cultures, but seems to have many of the genes for flagellar assembly and chemotaxis (Xue et al., 2001; Bao et al., 2002). From the genome annotation, *M. thermoacetica* also has an elaborate cascade of chemotaxis genes, including 6 chemoreceptors, together with a complete set of flagellar genes, most located within three large clusters of about 70 genes (Moth\_0740-7, Moth\_0764-91, Moth\_0797-Moth\_0805). The only exception is CheZ, which is not found in the *Moorella* or *Thermoanaerobacter tengcongensis* genomes. CheZ, a protein known to accelerate dephosphorylation of the response regulator CheY, has only been found in a few non-thermophilic gram negative bacteria, such as *E. coli* (Blattner et al., 1997), *P. aeruginosa* (Stover et al., 2000), and *V. cholerae* (Heidelberg et al., 2000), and it is not essential for flagellar motility and its chemotaxis (Scharf et al., 1998). The presence of genes involved in flagellar structure and movement in both non-motile *M. thermoacetica* and *T. tengcongensis* suggests either that they might be inactive or activated only under certain environmental conditions or that they have lost function only recently before divergence of these two species. In the proteomics study, two key gene products of the flagella formation, motor proteins (FliG and FliM) and basal proteins (FlgC and FlgB) were not detected, supporting the suspicion that incomplete gene expression may cause the absence of flagella in the cultured cells (Wang et al., 2004). (Wang et al., 2004)

### Discussion

This paper describes the first genomic sequence of an acetogen. The genomic results clearly indicate that the trait of acetogenesis is a phylogenetic, not a metabolic trait. To identify a set of genes that are responsible for the acetogenic lifestyle, the *M. thermoacetica* genome was compared to those of all Firmicute, Chlorflexi, and Deltaproteobacterial sequences available at IMG. Surprisingly, the only *M. thermoacetica* genes with homologs in the other bacterial genomes that also have acetyl-CoA synthase, but lack homologs in any of the genome sequences without acetyl-CoA synthase, are acetyl-CoA synthase itself and the two subunits of the corrinoid iron-sulfur protein (CFeSP). While the genes encoding the H<sub>4</sub>folate-dependent enzymes in the Eastern part of the pathway are scattered around the genome, the “core set of Wood-Ljungdahl genes” (*acsA-acsE*) in the Western part of the pathway are co-localized. Being required for general C-1 metabolism in all organisms, the Eastern branch genes are ubiquitous; thus, we suggest that acquisition of the *acs* gene cluster by horizontal transfer might be sufficient for gaining the ability to utilize the Wood-

Ljungdahl pathway. For some organisms, this would confer the ability to grow as a homoacetogen; for others (sulfate reducers and acetoclastic methanogens), the ability to oxidize acetate. Acquisition of the core Wood-Ljungdahl genes would allow a wide variety of organisms, including archaea and a broad range of bacterial phyla (Firmicute, Chlorflexi, and Deltaproteobacteria), to fix CO and CO<sub>2</sub> by the reductive acetyl-CoA pathway. The simple strategy of the Wood-Ljungdahl pathway of successively joining two one-carbon compounds to make a two-carbon compound has been envisioned to be the earliest form of metabolism (Martin and Russell, 2007).

The genomic analysis also revealed a second CODH (Moth\_1972), which has not been characterized biochemically. Other genomes contain multiple CODHs, like *C. hydrogenoformans*, which contains five CODHs (Wu et al., 2005). The only genes that could be identified to encode a phosphotransacetylase in *M. thermoacetica* (Moth\_1181 and Moth\_0864), are homologous to the recently characterized novel enzyme from *Salmonella enterica* called *pduL* (Liu et al., 2007); thus, it is important to test this function by performing biochemical assays on these putative phosphotransacetylases.

Perhaps not surprisingly, given the metabolic diversity of acetogenic microbes, the largest COG expansions in *M. thermoacetica*, relative to the other Thermoanaerobacteriaceae genome sequences, involve electron transfer proteins. How electron donors and acceptors interact with respiratory pathways in *M. thermoacetica* also remains an open question. Although the enzymes in the Wood-Ljungdahl pathway have been characterized, it has been unclear how acetogens gain energy, since there is no net ATP synthesis by substrate-level phosphorylation. Homologs of NADH dehydrogenase I (Table 3) and archaeal heterodisulfide reductase (Table 5) are present in the genome. Other genes identified that may play important roles in electron transfer linked proton translocation are Moth\_2184-90, with significant homology to the genes encoding the subunits of *E. coli*'s hydrogenase 4 (Table 4), which has been proposed to have proton translocating ability (Andrews et al., 1997). These various sets of genes could encode partial or complete membrane-bound complexes that would function in anaerobic respiration by generating a transmembrane proton gradient for use by the F<sub>1</sub>F<sub>0</sub>ATP synthase (Moth\_2377-84). Localization of genes encoding homologs of heterodisulfide reductase (Moth\_1194-Moth\_1196), a hydrogenase (Moth\_1193), and 5,10-methylene-H<sub>4</sub>folate reductase (Moth\_1191) near the *acs* gene cluster are consistent with a suggestion that the reductase and a hydrogenase might link to CODH to generate a proton motive force, driving anaerobic respiration in acetogens (Thauer et al., 1977).

*M. thermoacetica* contains eight genes encoding molybdopterin-containing proteins, that are likely to encode the reductases that allow growth on nitrate (Fröstl et al., 1996), dimethylsulfoxide, and thiosulfate (Beatty and Ljungdahl, 1991), as well as reductases for other electron acceptors not yet identified (Table 7). These molybdopterin-dependent proteins would then link to the membrane-bound respiratory systems described in the previous paragraph. The genomic information described here should help identify the specific membrane-associated proteins that are recruited to allow growth on these various electron acceptors. It is likely that these electron transfer components are differentially expressed growth condition, and the large number of them might be part of *M. thermoacetica*'s strategy for fine-tuning its growth to specific environmental conditions. The presence of so many different sets of genes that are likely to encode proteins involved in energy generation by electron and proton transport (for example, the presence of three heterodisulfide reductase gene clusters) could be part of this strategy. That cytochrome *b*, which is found in membranes of *M. thermoacetica* cells grown using the Wood-Ljungdahl pathway, is down-regulated in cells grown on nitrate (Arendsen et al., 1999) indicates that

the respiratory pathway used in connection with the Wood-Ljungdahl pathway is different than the one used during growth on nitrate, at least in some components.

Other expanded gene groups that were identified in *M. thermoacetica* include methyltransferases, which are required for growth on various methyl group donors (methanol, methoxylated aromatics), and ABC-type nitrate/sulfonate/bicarbonate transport systems, which also perhaps reflect the ability of acetogens to grow on a wide variety of substrates. Methyltransferase genes that link to the Wood-Ljungdahl pathway (besides *acsCDE*) include *mtaABC* (Moth\_1208-Moth\_1209) encoding the methanol:methyl-H<sub>4</sub>folate methyltransferase system, *mtvABC* (Moth\_1316-1318 and Moth\_385- Moth\_387), which transfers the methyl group of methoxylated aromatic compounds to H<sub>4</sub>folate, and several other methyltransferase genes whose specific biochemical function have not been assigned.

We were unable to predict the pathways by which *M. thermoacetica* metabolizes the two-carbon organic acids oxalate, glyoxylate, and glycolate (Seifritz et al., 1999). The most reduced compound among these is glycolate, which is oxidized to glyoxylate in *E. coli* (Pellicer et al., 1996) by an FMN- and H<sub>2</sub>O<sub>2</sub>-dependent glycolate oxidase, which is encoded by the *glcDEF* genes that are homologous to genes located in two gene clusters in *M. thermoacetica*. However, it seems unlikely that the anaerobe *M. thermoacetica* uses an oxidase mechanism. Homologs of malate synthase (GlcB), which catalyzes formation of malate from glyoxylate and acetyl-CoA, or glyoxylate carboligase (converts two mol of glyoxylate to form tartronate semialdehyde), were not found in *M. thermoacetica*, which is consistent with the lack of enzymatic activities for these proteins in extracts of cells grown on glyoxylate (Seifritz et al., 2002). Perhaps malate is formed from glyoxylate and acetyl-CoA by a glyoxylate reductase homolog (Moth\_1954). Homologs of proteins in the pathway of oxalate metabolism in *Oxalobacter formigenes*, which begins with oxalyl-CoA formation, were also missing from the *M. thermoacetica* genome. A hint for the oxalate metabolic pathway is that a 33 kDa protein is induced when *M. thermoacetica* is grown on oxalate (Daniel and Drake, 1993); however, no sequence information for this protein has yet been described. Postgenomic studies will likely uncover how acetogens use the *glc* gene cluster and/or other genes to metabolize oxalate, glyoxylate, and glycolate under anaerobic conditions. Identifying the mechanism of growth on oxalate should be particularly interesting, because during oxalate-dependent growth, *M. thermoacetica* uses CO<sub>2</sub> preferentially over nitrate as a terminal electron acceptor. Oxalate is the only substrate for *M. thermoacetica* for which this preference has been found.

Knowing the complete genome sequence of *M. thermoacetica* will be a valuable tool to help understand the growth of this organism and the acetogenic lifestyle. The genome sequence enables protein expression studies and microarrays, which should address important questions arising from annotation of this genome sequence. For example, growth of *M. thermoacetica* on oxalate and glyoxylate, followed by 2D electrophoresis analysis of proteins could uncover shifts in protein expression that would uncover what pathways are used. Protein and gene expression studies, on cells grown on different electron acceptors, could help identify the pathways of anaerobic respiration in *M. thermoacetica*. Such experiments will uncover how electrons are passed from electron donor to electron acceptor during growth, and how *M. thermoacetica* switches among different donors and acceptors, depending on what substrates are available in its environment. Similar studies may reveal how *M. thermoacetica* grows in different environments and in different ecological settings; for example, how its metabolism and gene expression patterns change when an acetogen is part of a microbial community or in coculture with a methanogen or sulfate reducer that can act as a H<sub>2</sub> sink. In turn, proteomic and microarray studies should help us improve the



annotation of the genome, by allowing us to identify specific biochemical functions and metabolic roles for genes of unknown or partially known function.

## Experimental Procedures

### DNA extraction and purification

*M. thermoacetica* cells were grown at 55 °C in the presence of 100 mM glucose (18 g/liter) under 100% CO<sub>2</sub> in the medium described by Ljungdahl and Andreesen (Ljungdahl and Andreesen, 1978), containing yeast extract and tryptone. Another sample of DNA was prepared from cells grown at 58 °C on 200 mM methanol under 100 % CO<sub>2</sub> in Drake's medium (Seifritz et al., 1993). The DNA was then isolated according to the procedure of Marmur (Marmur, 1961) and stored frozen at -20 °C in 10 mM Tris/HCl, pH 8.0 and 1 mM EDTA at a concentration of 438 µg/ml.

### Genome sequencing

The genome of *M. thermoacetica* was sequenced at the Joint Genome Institute (JGI) using a combination of 3 kb, 6 kb, 7kb, and 37kb (fosmid) DNA libraries. All general aspects of library construction and sequencing performed at the JGI can be found at <http://www.jgi.doe.gov/>. Draft assemblies were based on 41897 total reads. All four libraries provided 18x coverage of the genome. The Phred/Phrap/Consed software package ([www.phrap.com](http://www.phrap.com)) was used for sequence assembly and quality assessment (Han and Chain, 2006). After the shotgun stage, reads were assembled with parallel phrap (High Performance Software, LLC). Possible mis-assemblies were corrected with Dupfinisher (Abraham et al., 2006) or transposon bombing of bridging clones (Epicentre Biotechnologies, Madison, WI). Gaps between contigs were closed by editing in Consed, custom primer walk, and PCR amplification (Roche Applied Science, Indianapolis, IN). A total of 3345 additional reactions were necessary to close gaps and to raise the quality of the finished sequence. The completed genome sequence of *M. thermoacetica* contains 45242 reads, achieving an average of 8-fold sequence coverage per base with an error rate less than 1 in 100,000

### Annotation

Automated gene modeling was completed by combining results from Critica, Generation, and Glimmer modeling packages, and comparing the translations to GenBank's nonredundant database using the basic local alignment search tool for proteins (BLASTP). Predicted coding sequences were subjected to manual analysis using JGI's gene models quality assessment pipeline. The predicted functions were further analyzed by using the Integrated Microbial Genomes (IMG) annotation pipeline (<http://img.jgi.doe.gov>) (Markowitz et al., 2006).

### Comparative analysis

Comparative analysis of *M. thermoacetica* ATCC 39073, *Thermoanaerobacter tengcongensis* MB4, and *Thermoanaerobacter ethanolicus* 39E was performed by using a set of tools available at IMG. Unique and orthologous *M. thermoacetica* genes were identified by using BLASTp (cutoff scores of  $E < 10^{-2}$  and 20% identity and reciprocal hits with cutoffs of  $E < 10^{-5}$  and 30%, respectively). The transporter database (TCDB; <http://www.tcdb.org>) (Saier et al., 2006) was used to identify all the transporters in the genome of *M. thermoacetica*. Twin arginine transport signal sequences were predicted by the TatP server at <http://www.cbs.dtu.dk/services/TatP/> (Bendtsen et al., 2005). Metabolic pathways were constructed with reference to MetaCyc (Caspi et al., 2006) and PubMed. Identification numbers given in the text for individual genes are IMG locus tag numbers, which can be used to search for genes at the Integrated Microbial Genomes website.

## Phylogenetic trees

The phylogenetic tree was made using pre-aligned 16S rRNA sequences derived from the Ribosomal Database site (<http://rdp.cme.msu.edu/index.jsp>) (Cole et al., 2007). Manual alignment adjustments were made as needed with the assistance of the BioEdit multiple alignment tool of Hall (<http://www.mbio.ncsu.edu/BioEdit/bioedit.html>). The refined multiple alignment was used as input for generation of a phylogenetic tree using the program package version 3.1 (Kumar et al., 2004). The Neighbor Joining (Saitou and Nei, 1987), Minimum Evolution (Nei et al., 1998; Takahashi and Nei, 2000), UPGMA (Nei and Kumar, 2000), and Maximum Parsimony (Fitch, 1971) methods with the default settings were employed to obtain phylogenetic trees. Both the distance- and character-based trees yielded similar clusters and arrangements of taxa.

## Nucleotide sequence accession number

The sequence data described here have been deposited in GenBank under accession number CP000232.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We are grateful to the DOE Joint Genome Institute for selecting *M. thermoacetica* for sequencing and to many people at JGI who were involved in the sequencing effort. The work was supported by a grant from NIH (GM39451, SWR). We thank Thomas Morton for extracting one of the DNA samples used in these studies.

## References

- Abraham LO, Li Y, Zamble DB. The metal- and DNA-binding activities of *Helicobacter pylori* NikR. *J Inorg Biochem.* 2006; 100:1005–1014. [PubMed: 16343630]
- Adrian NR, Arnett CM. Anaerobic Biodegradation of Hexahydro-1,3,5-trinitro-1,3,5-triazine (RDX) by *Acetobacterium malicum* Strain HAAP-1 Isolated from a Methanogenic Mixed Culture. *Current Microbiology.* 2004; 48:332–340. [PubMed: 15060728]
- Andreesen JR, Ljungdahl LG. Formate dehydrogenase of *Clostridium thermoaceticum*: incorporation of selenium-75, and the effects of selenite, molybdate, and tungstate on the enzyme. *Journal of Bacteriology.* 1973; 116:867–873. [PubMed: 4147651]
- Andreesen JR, Schaupp A, Neurauter C, Brown A, Ljungdahl LG. Fermentation of glucose, fructose, and xylose by *Clostridium thermoaceticum*: effect of metals on growth yield, enzymes, and the synthesis of acetate from CO<sub>2</sub>. *Journal of Bacteriology.* 1973; 114:743–751. [PubMed: 4706193]
- Andrews SC, Berks BC, McClay J, Ambler A, Quail MA, Golby P, Guest JR. A 12-cistron *Escherichia coli* operon (hyf) encoding a putative proton- translocating formate hydrogenlyase system. *Microbiology.* 1997; 143:3633–3647. [PubMed: 9387241]
- Arendsen AF, Soliman MQ, Ragsdale SW. Nitrate-dependent regulation of acetate biosynthesis and nitrate respiration by *Clostridium thermoaceticum*. *Journal of Bacteriology.* 1999; 181:1489–1495. [PubMed: 10049380]
- Balch WE, Schoberth S, Tanner RS, Wolfe RS. *Acetobacterium*, a new genus of hydrogen-oxidizing, carbon-dioxide-reducing, anaerobic bacteria. *International Journal of Systematic Bacteriology.* 1977; 27:355–361.
- Bao Q, Tian Y, Li W, Xu Z, Xuan Z, Hu S, et al. A Complete Sequence of the *Thermoanaerobacter tengcongensis* Genome. *Genome Res.* 2002; 12:689–700. [PubMed: 11997336]
- Barker HA, Kamen MD. Carbon dioxide utilization in the synthesis of acetic acid by *Clostridium thermoaceticum*. *Proceedings of the National Academy of Sciences.* 1945; 31:219–225.

- Beaty, SP.; Ljungdahl, LG. Growth of *Clostridium thermoaceticum* on methanol, ethanol, propanol, and butanol in medium containing either thiosulfate or dimethylsulfoxide. Abstracts of the Annual Meeting of the American Society for Microbiology; 1991. p. 236
- Bendtsen J, Nielsen H, Widdick D, Palmer T, Brunak S. Prediction of twin-arginine signal peptides. BMC Bioinformatics. 2005; 6:167. [PubMed: 15992409]
- Blattner FR, Plunkett G III, Bloch CA, Perna NT, Burland V, Riley M, et al. The Complete Genome Sequence of *Escherichia coli* K-12. Science. 1997; 277:1453–1462. [PubMed: 9278503]
- Blokesch M, Paschos A, Bauer A, Reissmann S, Drapal N, Böck A. Analysis of the transcarbamoylation-dehydration reaction catalyzed by the hydrogenase maturation proteins HypF and HypE. European Journal of Biochemistry. 2004; 271:3428–3436. [PubMed: 15291820]
- Boynton ZL, Bennett GN, Rudolph FB. Cloning, sequencing, and expression of genes encoding phosphotransacetylase and acetate kinase from *Clostridium acetobutylicum* ATCC 824. Appl Environ Microbiol. 1996; 62:2758–2766. [PubMed: 8702268]
- Braun M, Mayer F, Gottschalk G. *Clostridium aceticum* (Wieringa), a microorganism producing acetic acid from molecular hydrogen and carbon dioxide. Archives of Microbiology. 1981; 128:288–293. [PubMed: 6783001]
- Breznak, JA. Acetogenesis from carbon dioxide in termite guts. In: Drake, HL., editor. Acetogenesis. New York: Chapman and Hall; 1994. p. 303-330.
- Caspi R, Foerster H, Fulcher CA, Hopkinson R, Ingraham J, Kaipa P, et al. MetaCyc: a multiorganism database of metabolic pathways and enzymes. Nucl Acids Res. 2006; 34:D511–516. [PubMed: 16381923]
- Chassard C, Bernalier-Donadille A. H<sub>2</sub> and acetate transfers during xylan fermentation between a butyrate-producing xylanolytic species and hydrogenotrophic microorganisms from the human gut. FEMS Microbiology Letters. 2006; 254:116–122. [PubMed: 16451188]
- Clark JE, Ljungdahl LG. Purification and properties of 5,10-methenyltetrahydrofolate cyclohydrolase. J Biol Chem. 1982; 257:3833–3836. [PubMed: 7061514]
- Clark JE, Ljungdahl LG. Purification and properties of 5,10-methylenetetrahydrofolate reductase, an iron-sulfur flavoprotein from *Clostridium formicoaceticum*. J Biol Chem. 1984; 259:10845–10889. [PubMed: 6381490]
- Cole JR, Chai B, Farris R, Wang Q, Kulam-Syed-mohideen AS, McGarrell D, et al. The ribosomal database project (RDP-II): introducing *myRDP* space and quality controlled public data. Nucleic Acids Res. 2007; 35:D169–D172. [PubMed: 17090583]
- Collins M, Lawson P, Willems A, Cordoba J, Fernandez-Garayzabal J, Garcia P, et al. The phylogeny of the genus *Clostridium*: proposal of five new genera and eleven new species combinations. Int J Syst Bacteriol. 1994; 44:812–826. [PubMed: 7981107]
- Daniel SL, Drake HL. Oxalate- and glyoxylate-dependent growth and acetogenesis by *Clostridium thermoaceticum*. Applied and Environmental Microbiology. 1993; 59:3062–3069. [PubMed: 16349048]
- Das A, Ljungdahl L. Composition and primary structure of the F1F0 ATP synthase from the obligately anaerobic bacterium *Clostridium thermoaceticum*. J Bacteriol. 1997; 179:3746–3755. [PubMed: 9171425]
- Das A, Hugenholtz J, Van Halbeek H, Ljungdahl LG. Structure and function of a menaquinone involved in electron transport in membranes of *Clostridium thermoautotrophicum* and *Clostridium thermoaceticum*. J Bacteriol. 1989; 171:5823–5829. [PubMed: 2808299]
- Das A, Fu ZQ, Tempel W, Liu ZJ, Chang J, Chen L, et al. Characterization of a corrinoid protein involved in the C1 metabolism of strict anaerobic bacterium *Moorella thermoacetica*. Proteins: Structure, Function, and Bioinformatics. 2007; 67:167–176.
- Deppenmeier U. The membrane-bound electron transport system of *Methanosarcina* species. Journal of Bioenergetics and Biomembranes. 2004; 36:55–64. [PubMed: 15168610]
- Drake HL. Demonstration of hydrogenase in extracts of the homoacetate-fermenting bacterium *Clostridium thermoaceticum*. Journal of Bacteriology. 1982; 150:702–709. [PubMed: 7040339]
- Drake HL, Daniel SL. Acetogenic bacteria: what are the in situ consequences of their diverse metabolic versatilityes? BioFactors. 1997; 6:13–24. [PubMed: 9233536]

- Drake HL, Hu SI, Wood HG. Purification of five components from *Clostridium thermoaceticum* which catalyze synthesis of acetate from pyruvate and methyltetrahydrofolate. Properties of phosphotransacetylase. *J Biol Chem.* 1981; 256:11137–11144. [PubMed: 7287757]
- Drake HL, Kusel K, Matthies C. Ecological consequences of the phylogenetic and physiological diversities of acetogens. *Antonie Van Leeuwenhoek.* 2002; 81:203–213. [PubMed: 12448719]
- Drake HL, Gossner AS, Daniel SL. Old Acetogens, New Light. *Ann N Y Acad Sci.* 2008; 1125:100–128. [PubMed: 18378590]
- Drake, HL.; Daniel, SL.; Matthies, C.; Kusel, K. Acetogenesis, acetogenic bacteria, and the acetyl-CoA pathway: Past and current perspectives. In: Drake, HL., editor. *Acetogenesis.* New York: Chapman and Hall; 1994. p. 3-60.
- Ferry JG. Enzymology of one-carbon metabolism in methanogenic pathways. *FEMS Microbiol Rev.* 1999; 23:13–38. [PubMed: 10077852]
- Fischer F, Lieske R, Winzer K. Biologische gasreaktionen. II. Gber die bildung von essigs ure bei der biologischen umsetzung von kohlenoxyd und kohlens ure mit wasserstoff zu methan. *Biochem Z.* 1932; 245:2–12.
- Fitch WM. Towards defining the course of evolution: Minimum change for a specific tree topology. *Systematic Zoology.* 1971; 20:406–416.
- Fontaine FE, Peterson WH, McCoy E, Johnson MJ, Ritter GJ. A new type of glucose fermentation by *Clostridium thermoaceticum* n sp. *Journal of Bacteriology.* 1942; 43:701–715. [PubMed: 16560531]
- Fox JD, He Y, Shelver D, Roberts GP, Ludden PW. Characterization of the region encoding the CO-induced hydrogenase of *Rhodospirillum rubrum.* *J Bacteriol.* 1996; 178:6200–6208. [PubMed: 8892819]
- Fröstl JM, Seifritz C, Drake HL. Effect of nitrate on the autotrophic metabolism of the acetogens *Clostridium thermoautotrophicum* and *Clostridium thermoaceticum.* *Journal of Bacteriology.* 1996; 187:4597–4603.
- Furdui C, Ragsdale SW. The Roles of Coenzyme A in the Pyruvate:Ferredoxin Oxidoreductase Reaction Mechanism: Rate Enhancement of Electron Transfer from a Radical Intermediate to an Iron-Sulfur Cluster. *Biochemistry.* 2002; 41:9921–9937. [PubMed: 12146957]
- Gencic S, Grahame DA. Nickel in subunit beta of the acetyl-CoA decarbonylase/synthase multienzyme complex in methanogens: Catalytic properties and evidence for a binuclear Ni-Ni site. *J Biol Chem.* 2003; 278:6101–6110. [PubMed: 12464601]
- Genest O, Seduk F, Ilbert M, Méjean V, Iobbi-Nivol C. Signal peptide protection by specific chaperone. *Biochemical and Biophysical Research Communications.* 2006; 339:991–995. [PubMed: 16337610]
- Gill SR, Pop M, Deboy RT, Eckburg PB, Turnbaugh PJ, Samuel BS, et al. Metagenomic analysis of the human distal gut microbiome. *Science.* 2006; 312:1355–1359. [PubMed: 16741115]
- Gon S, Giudici-Orticoni M-T, Mejean V, Iobbi-Nivol C. Electron Transfer and Binding of the c-Type Cytochrome TorC to the Trimethylamine N-Oxide Reductase in *Escherichia coli.* *J Biol Chem.* 2001; 276:11545–11551. [PubMed: 11056172]
- Göbner A, Devereux R, Ohnemüller N, Acker G, Stackebrandt E, Drake HL. *Thermicanus aegyptius* gen. nov., sp. nov., isolated from oxic soil, a fermentative microaerophile that grows commensally with the thermophilic acetogen *Moorella thermoacetica.* *Applied and Environmental Microbiology.* 1999; 65:5124–5133. [PubMed: 10543831]
- Gottwald M, Andreesen JR, LeGall J, Ljungdahl LG. Presence of Cytochrome and Menaquinone in *Clostridium formicoaceticum* and *Clostridium thermoaceticum.* *J Bacteriol.* 1975; 122:325–328. [PubMed: 1123319]
- Greening RC, Leedle JAZ. Enrichment and isolation of *Acetitomaculum ruminis*, gen. nov., sp. nov.: acetogenic bacteria from the bovine rumen. *Archives of Microbiology.* 1989; 151:399–406. [PubMed: 2500921]
- Han, CS.; Chain, P. Finishing repeat regions automatically with Dupfinisher. In: Valafar, H., editor. *Proceeding of the 2006 International Conference on Bioinformatics and Computational Biology.* CSREA Press; 2006. p. 141-146.

- Harmsen H, Van Kuijk B, Plugge C, Akkermans A, De Vos W, Stams A. Syntrophobacter fumaroxidans sp. nov., a syntrophic propionate- degrading sulfate-reducing bacterium. *Int J Syst Bacteriol.* 1998; 48:1383–1387. [PubMed: 9828440]
- Hatzixanthis K, Clarke TA, Oubrie A, Richardson DJ, Turner RJ, Sargent F. Signal peptide-chaperone interactions on the twin-arginine protein transport pathway. *Proceedings of the National Academy of Sciences.* 2005; 102:8460–8465.
- Hedderich R. Energy-Converting [NiFe] Hydrogenases From Archaea and Extremophiles: Ancestors of Complex I. *Journal of Bioenergetics and Biomembranes.* 2004; 36:65–75. [PubMed: 15168611]
- Heidelberg JF, Elsen JA, Nelson WC, Clayton RA, Gwinn ML, Dodson RJ, et al. DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature.* 2000; 406:477–484. [PubMed: 10952301]
- Heise R, Muller V, Gottschalk G. Sodium dependence of acetate formation by the acetogenic bacterium *Acetobacterium woodii*. *J Bacteriol.* 1989; 171:5473–5478. [PubMed: 2507527]
- Himes RH, Wilder T. Formyltetrahydrofolate synthetase. Effect of pH and temperature on the reaction. *Arch Biochem Biophys.* 1968; 124:230. [PubMed: 5661603]
- Hugenholtz J, Ivey DM, Ljungdahl LG. Carbon monoxide-driven electron transport in *Clostridium thermoautotrophicum* membranes. *J Bacteriol.* 1987; 169:5845–5847. [PubMed: 3680181]
- Jeon WB, Cheng J, Ludden PW. Purification and characterization of membrane-associated CooC protein and its functional role in the insertion of nickel into carbon monoxide dehydrogenase from *Rhodospirillum rubrum*. *Journal of Biological Chemistry.* 2001; 42:38602–38609. [PubMed: 11507093]
- Jongbloed JDH, Grieger U, Antelmann H, Hecker M, Nijland R, Bron S, van Dijk JM. Two minimal Tat translocases in *Bacillus*. *Molecular Microbiology.* 2004; 54:1319–1325. [PubMed: 15554971]
- Kerby R, Zeikus JG. Growth of *Clostridium thermoaceticum* on H<sub>2</sub>/CO<sub>2</sub> or CO as energy source. *Current Microbiology.* 1983; 8:27–30.
- Kisker C, Schindelin H, Baas D, Rétey J, Merkenstock RU, Kroneck PMH. A structural comparison of molybdenum cofactor-containing enzymes. *FEMS Microbiology Reviews.* 1999; 22:503–521. [PubMed: 9990727]
- Kumar S, Tamura K, Nei M. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Briefings in Bioinformatics.* 2004; 5:150–163. [PubMed: 15260895]
- Kusel K, Wagner C, Drake HL. Enumeration and metabolic product profiles of the anaerobic microflora in the mineral soil and litter of a beech forest. *FEMS Microbiology Ecology.* 1999; 29:91–103.
- Ladapo J, Whitman W. Method for Isolation of Auxotrophs in the Methanogenic Archaeobacteria: Role of the Acetyl-CoA Pathway of Autotrophic CO<sub>2</sub> Fixation in *Methanococcus maripaludis*. *PNAS.* 1990; 87:5598–5602. [PubMed: 11607093]
- Larsen Ø, Lien T, Birkeland NK. Dissimilatory sulfite reductase from *Archaeoglobus profundus* and *Desulfotomaculum thermocisternum*: phylogenetic and structural implications from gene sequences. *Extremophiles.* 1999; 3:63–70. [PubMed: 10086846]
- Le Van TD, Robinson JA, Ralph J, Greening RC, Smolenski WJ, Leedle JAZ, Schaefer DM. Assessment of Reductive Acetogenesis with Indigenous Ruminant Bacterial Populations and *Acetitomaculum ruminis*. *Appl Environ Microbiol.* 1998; 64:3429–3436. [PubMed: 9726893]
- Leadbetter JR, Schmidt TM, Graber JR, Breznak JA. Acetogenesis from H<sub>2</sub> plus CO<sub>2</sub> by spirochetes from termite guts. *Science.* 1999; 283:686–689. [PubMed: 9924028]
- Li LF, Ljungdahl L, Wood HG. Properties of nicotinamide adenine dinucleotide phosphate-dependent formate dehydrogenase from *Clostridium thermoaceticum*. *Journal of Bacteriology.* 1966; 92:405–412. [PubMed: 16562128]
- Liesack W, Bak F, Kreft JU, Stackebrandt E. *Holophaga foetida* gen. nov., sp. nov., a new, homoacetogenic bacterium degrading methoxylated aromatic compounds. *Arch Microbiol.* 1994; 162:85–90. [PubMed: 8085918]
- Liu S, Sufflita JM. H<sub>2</sub>-CO<sub>2</sub>-Dependent Anaerobic O-Demethylation Activity in Subsurface Sediments and by an Isolated Bacterium. *Appl Environ Microbiol.* 1993; 59:1325–1331. [PubMed: 16348928]

- Liu Y, Leal NA, Sampson EM, Johnson CLV, Havemann GD, Bobik TA. PduL Is an Evolutionarily Distinct Phosphotransacylase Involved in B12-Dependent 1,2-Propanediol Degradation by *Salmonella enterica* Serovar Typhimurium LT2. *J Bacteriol.* 2007; 189:1589–1596. [PubMed: 17158662]
- Ljungdahl LG. The Autotrophic Pathway of Acetate Synthesis in Acetogenic Bacteria. *Annual Review of Microbiology.* 1986; 40:415–450.
- Ljungdahl, LG. The acetyl-CoA pathway and the chemiosmotic generation of ATP during acetogenesis. In: Drake, HL., editor. *Acetogenesis*. New York: Chapman and Hall; 1994. p. 63-87.
- Ljungdahl LG, Andreesen JR. Tungsten, a component of active formate dehydrogenase from *Clostridium thermoacetum*. *FEBS Lett.* 1975; 54:279–282. [PubMed: 1132514]
- Ljungdahl LG, Andreesen JR. Formate dehydrogenase, a selenium-tungsten enzyme from *Clostridium thermoacetum*. *Methods in Enzymology.* 1978; 53:360–372. [PubMed: 713844]
- Ljungdahl LG, Irion E, Wood HG. Role of corrinoids in the total synthesis of acetate from CO<sub>2</sub> by *Clostridium thermoacetum*. *Federal Proceedings.* 1966; 25:1642.
- Lovell CR, Przybyla A, Ljungdahl LG. Cloning and expression in *Escherichia coli* of the *Clostridium thermoacetum* gene encoding thermostable formyltetrahydrofolate synthetase. *Arch Microbiol.* 1988; 149:280–285. [PubMed: 2833195]
- Lovell CR, Przybyla A, Ljungdahl LG. Primary structure of the thermostable formyltetrahydrofolate synthetase from *Clostridium thermoacetum*. *Biochemistry.* 1990; 29:5687–5694. [PubMed: 2200509]
- Lundie LL Jr, Drake HL. Development of a minimally defined medium for the acetogen *Clostridium thermoacetum*. *Journal of Bacteriology.* 1984; 159:700–703. [PubMed: 6746575]
- Lutz S, Jacobi A, Schlenz V, Böhm R, Sawers G, Böck A. Molecular characterization of an operon (hyp) necessary for the activity of the three hydrogenase isoenzymes in *Escherichia coli*. *Molecular Microbiology.* 1991; 5:123–135. [PubMed: 1849603]
- Macbeth TW, Cummings DE, Spring S, Petzke LM, Sorenson KS Jr. Molecular Characterization of a Dechlorinating Community Resulting from In Situ Biostimulation in a Trichloroethene-Contaminated Deep, Fractured Basalt Aquifer and Comparison to a Derivative Laboratory Culture. *Appl Environ Microbiol.* 2004; 70:7329–7341. [PubMed: 15574933]
- Mackie, RI.; Bryant, MP. Acetogenesis and the rumen: syntrophic relationships. In: Drake, HL., editor. *Acetogenesis*. New York: Chapman and Hall; 1994. p. 331-364.
- Macnab RM. Type III flagellar protein export and flagellar assembly. *Biochim Biophys Acta.* 2004; 1694:207–217. [PubMed: 15546667]
- Markowitz VM, Korzeniewski F, Palaniappan K, Szeto E, Werner G, Padki A, et al. The integrated microbial genomes (IMG) system. *Nucl Acids Res.* 2006; 34:D344–348. [PubMed: 16381883]
- Marmur J. A procedure for the isolation of deoxyribonucleic acid from microorganisms. *Journal of Molecular Biology.* 1961; 3:208–218.
- Martin W, Russell MJ. On the origin of biochemistry at an alkaline hydrothermal vent. *Philos Trans R Soc Lond B Biol Sci.* 2007; 362:1887–1925. [PubMed: 17255002]
- Maupin-Furlow JA, Ferry JG. Analysis of the CO dehydrogenase/acetyl-coenzyme A synthase operon of *Methanosarcina thermophila*. *J Bacteriol.* 1996; 178:6849–6856. [PubMed: 8955306]
- McGuire JJ, Rabinowitz JC. Studies on the mechanism of formyltetrahydrofolate synthetase. The *Peptococcus aerogenes* enzyme. *J Biol Chem.* 1978; 253:1079. [PubMed: 624720]
- Meuer J, Bartoschek S, Koch J, Künkel A, Hedderich R. Purification and Catalytic Properties of Ech Hydrogenase from *Methanosarcina barkeri*. *European Journal of Biochemistry.* 1999; 265:235–335.
- Moore MR, O'Brien WE, Ljungdahl LG. Purification and characterization of nicotinamide adenine dinucleotide- dependent methylenetetrahydrofolate dehydrogenase from *Clostridium formicoaceticum*. *Journal of Biological Chemistry.* 1974; 249:5250–5253. [PubMed: 4153026]
- Muller V. Energy conservation in acetogenic bacteria. *Appl Environ Microbiol.* 2003; 69:6345–6353. [PubMed: 14602585]
- Naidu D, Ragsdale SW. Characterization of a Three-Component Vanillate O-Demethylase from *Moorella thermoacetica*. *J Bacteriol.* 2001; 183:3276–3281. [PubMed: 11344134]

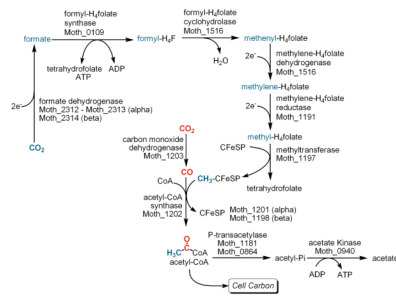
- Nei, M.; Kumar, S. *Molecular Evolution and Phylogenetics*. New York: Oxford University Press; 2000.
- Nei M, Kumar S, Takahashi K. The optimization principle in phylogenetic analysis tends to give incorrect topologies when the number of nucleotides or amino acids used is small. *Proceedings of the National Academy of Sciences of the United States of America*. 1998; 95:12390–12397. [PubMed: 9770497]
- Neumann A, Engelmann T, Schmitz R, Greiser Y, Orthaus A, Diekert G. Phenyl methyl ethers: novel electron donors for respiratory growth of *Desulfitobacterium hafniense* and *Desulfitobacterium* sp. strain PCE-S. *Archives of Microbiology*. 2004; 181:245–249. [PubMed: 14758469]
- Nonaka H, Keresztes G, Shinoda Y, Ikenaga Y, Abe M, Naito K, et al. Complete Genome Sequence of the Dehalorespiring Bacterium *Desulfitobacterium hafniense* Y51 and Comparison with *Dehalococcoides ethenogenes* 195. *J Bacteriol*. 2006; 188:2262–2274. [PubMed: 16513756]
- Núñez FM, Pellicer MT, Badía J, Aguilar J, Baldomà L. The gene *yghK* linked to the *glc* operon of *Escherichia coli* encodes a permease for glycolate that is structurally and functionally similar to L-lactate permease. *Microbiology*. 2001; 147:1069–1077. [PubMed: 11283302]
- O'Brien, WE.; Brewer, JM.; Ljungdahl, LG. Chemical, physical and enzymatic comparisons of formyltetrahydrofolate synthetases from thermo- and mesophilic clostridia. In: Zuber, H., editor. *Proceedings of the International Symposium on Enzymes and Proteins from Thermophilic Microorganisms, Structure and Functions*. Basel, Switzerland: Birkhauser-Verlag; 1975. p. 249
- Odelson DA, Breznak JA. Volatile fatty acid production by the hindgut microbiota of xylophagous termites. *Appl Environ Microbiol*. 1983; 45:1602–1613. [PubMed: 16346296]
- Ohshima T, Nunoura-Kominato N, Kudome T, Sakuraba H. A novel hyperthermophilic archaeal glyoxylate reductase from *Thermococcus litoralis*. Characterization, gene cloning, nucleotide sequence and expression in *Escherichia coli*. *Eur J Biochem*. 2001; 268:4740–4747. [PubMed: 11532010]
- Ollivier B, Caumette P, Garcia JL, Mah RA. Anaerobic Bacteria from Hypersaline Environments. *Microbiological Reviews*. 1994; 58:27–38. [PubMed: 8177169]
- Park, EY.; Clark, JE.; DerVartanian, DV.; Ljungdahl, LG. 5,10-methylenetetrahydrofolate reductases: iron-sulfur-zinc flavoproteins of two acetogenic clostridia. In: Miller, F., editor. *Chemistry and Biochemistry of Flavoenzymes*. Boca Raton, FL: CRC Press; 1991. p. 389-400.
- Pellicer MT, Badía J, Aguilar J, Baldomà L. *glc* locus of *Escherichia coli*: characterization of genes encoding the subunits of glycolate oxidase and the *glc* regulator protein. *Journal of Bacteriology*. 1996; 178:2051–2059. [PubMed: 8606183]
- Pester M, Brune A. Hydrogen is the central free intermediate during lignocellulose degradation by termite gut symbionts. *ISME J*. 2007; 1:551–565. [PubMed: 18043656]
- Peters V, Conrad R. Methanogenic and Other Strictly Anaerobic Bacteria in Desert Soil and Other Oxic Soils. *Appl Environ Microbiol*. 1995; 61:1673–1676. [PubMed: 16535011]
- Ragsdale SW. The Eastern and Western branches of the Wood/Ljungdahl pathway: how the East and West were won. *BioFactors*. 1997; 6:3–11. [PubMed: 9233535]
- Ragsdale SW, Ljungdahl LG. Purification and properties of NAD-dependent 5,10-methylenetetrahydrofolate dehydrogenase from *Acetobacterium woodii*. *J Biol Chem*. 1984; 259:3499–3503. [PubMed: 6608524]
- Roberts DL, James-Hagstrom JE, Garvin DK, Gorst CM, Runquist JA, Baur JR, et al. Cloning and expression of the gene cluster encoding key proteins involved in acetyl-CoA synthesis in *Clostridium thermoaceticum*: CO dehydrogenase, the corrinoid/FeS protein, and methyltransferase. *Proceedings of the National Academy of Sciences*. 1989; 86:32–36.
- Rosencrantz D, Rainey FA, Janssen PH. Culturable Populations of *Sporomusa* spp. and *Desulfovibrio* spp. in the Anoxic Bulk Soil of Flooded Rice Microcosms. *Applied and Environmental Microbiology*. 1999; 65:3526–3533. [PubMed: 10427044]
- Saier MH Jr, Tran CV, Barabote RD. TCDB: the Transporter Classification Database for membrane transport protein analyses and information. *Nucl Acids Res*. 2006; 34:D181–186. [PubMed: 16381841]
- Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*. 1987; 4:406–425. [PubMed: 3447015]

- Salmassi TM, Leadbetter JR. Analysis of genes of tetrahydrofolate-dependent metabolism from cultivated spirochaetes and the gut community of the termite *Zootermopsis angusticollis*. *Microbiology*. 2003; 149:2529–2537. cover illustration. [PubMed: 12949177]
- Sazanov LA, Hinchliffe P. Structure of the Hydrophilic Domain of Respiratory Complex I from *Thermus thermophilus*. *Science*. 2006; 311:1430–1436. [PubMed: 16469879]
- Scharf BE, Fahrner KA, Berg HC. CheZ Has No Effect on Flagellar Motors Activated by CheY13DK106YW. *J Bacteriol*. 1998; 180:5123–5128. [PubMed: 9748445]
- Schauder R, Preuß A, Jetten MS, Fuchs G. Oxidative and reductive acetyl CoA/carbon monoxide dehydrogenase pathway in *Desulfobacterium autotrophicum*. *Arch Microbiol*. 1988; 151:84–89.
- Schaupp A, Ljungdahl LG. Purification and properties of acetate kinase from *Clostridium thermoaceticum*. *Archives of Microbiology*. 1974; 100:121–129. [PubMed: 4447427]
- Schink B. Energetics of syntrophic cooperation in methanogenic degradation. *Microbiol Mol Biol Rev*. 1997; 61:262–280. [PubMed: 9184013]
- Schink B, Thiemann V, Laue H, Friedrich MW. Desulfotignum phosphitoxidans sp. nov., a new marine sulfate reducer that oxidizes phosphite to phosphate. *Arch Microbiol*. 2002; 177:381–391. [PubMed: 11976747]
- Schwartz G. Molybdenum cofactor biosynthesis and deficiency. *Cellular and Molecular Life Sciences*. 2005; 62:2792–2910. [PubMed: 16261263]
- Seifritz C, Drake HL, Daniel SL. Nitrite as an Energy-Conserving Electron Sink for the Acetogenic Bacterium *Moorella thermoacetica*. *Current Microbiology*. 2003; 46:329–333. [PubMed: 12732959]
- Seifritz C, Daniel SL, Gößner A, Drake HL. Nitrate as a preferred electron sink for the acetogen *Clostridium thermoaceticum*. *Journal of Bacteriology*. 1993; 175:8008–8013. [PubMed: 8253688]
- Seifritz C, Fröstl JM, Drake HL, Daniel SL. Glycolate as a metabolic substrate for the acetogen *Moorella thermoacetica*. *FEMS Microbiology Letters*. 1999; 170:399–405.
- Seifritz C, Fröstl JM, Drake HL, Daniel SL. Influence of nitrate on oxalate- and glyoxylate-dependent growth and acetogenesis by *Moorella thermoacetica*. *Archives of Microbiology*. 2002; 178:457–464. [PubMed: 12420166]
- Seshadri R, Adrian L, Fouts DE, Eisen JA, Phillippy AM, Methe BA, et al. Genome Sequence of the PCE-Dechlorinating Bacterium *Dehalococcoides ethenogenes*. *Science*. 2005; 307:105–108. [PubMed: 15637277]
- Setzke E, Hedderich R, Heiden S, Thauer R. H<sub>2</sub>: heterodisulfide oxidoreductase complex from *Methanobacterium thermoautotrophicum*. Composition and properties. *Eur J Biochem*. 1994; 220:139–148. [PubMed: 8119281]
- Shannon KW, Rabinowitz JC. Isolation and characterization of the *Saccharomyces cerevisiae* *MIS1* gene encoding mitochondrial C<sub>1</sub>-tetrahydrofolate synthase. *J Biol Chem*. 1988; 263:7717. [PubMed: 2836393]
- So CM, Phelps CD, Young LY. Anaerobic Transformation of Alkanes to Fatty Acids by a Sulfate-Reducing Bacterium, Strain Hxd3. *Appl Environ Microbiol*. 2003; 69:3892–3900. [PubMed: 12839758]
- Soboh B, Linder D, Hedderich R. Purification and catalytic properties of a CO-oxidizing:H<sub>2</sub>-evolving enzyme complex from *Carboxydotherrmus hydrogenoformans*. *Eur J Biochem*. 2002; 269:5712–5721. [PubMed: 12423371]
- Spormann AM, Thauer RK. Anaerobic acetate oxidation to CO<sub>2</sub> by *Desulfotomaculum acetoxidans* - Demonstration of enzymes required for the operation of an oxidative acetyl-CoA/carbon monoxide dehydrogenase pathway. *Arch Microbiol*. 1988; 150:374–380.
- Staben C, Rabinowitz JC. Nucleotide sequence of the *Saccharomyces cerevisiae* *ADE3* gene encoding C<sub>1</sub>-tetrahydrofolate synthase. *J Biol Chem*. 1986; 261:2986.
- Stover CK, Pham XQ, Erwin AL, Mizoguchi SD, Warrener P, Hickey MJ, et al. Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. 2000; 406:959–964.
- Strous M, Pelletier E, Mangenot S, Rattei T, Lehner A, Taylor MW, et al. Deciphering the evolution and metabolism of an anammox bacterium from a community genome. *Nature*. 2006; 440:790–794. [PubMed: 16598256]



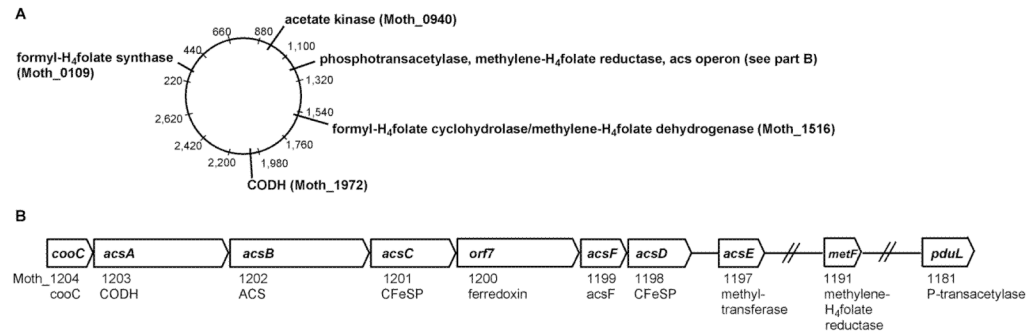
- Stupperich E, Eisinger HJ, Schurr S. Corrinoids in anaerobic bacteria. *FEMS Microbiology Letters*. 1990; 87:355–360.
- Stupperich E, Hammel KE, Fuchs G, Thauer RK. Carbon monoxide fixation into the carboxyl group of acetyl coenzyme A during autotrophic growth of *Methanobacterium*. *FEBS Lett*. 1983; 152:21–23. [PubMed: 6840273]
- Sun AY, Ljungdahl L, Wood HG. Total Synthesis of Acetate from CO<sub>2</sub> II. Purification and Properties of Formyltetrahydrofolate Synthetase from *Clostridium thermoaceticum*. *J Bacteriol*. 1969; 98:842–844. [PubMed: 5784233]
- Takahashi K, Nei M. Efficiencies of fast algorithms of phylogenetic inference under the criteria of maximum parsimony, minimum evolution, and maximum likelihood when a large number of sequences are used. *Molecular Biology and Evolution*. 2000; 17:1251–1258. [PubMed: 10908645]
- Thauer RK, Jungermann K, Decker K. Energy conservation in chemotrophic anaerobic bacteria. *Bacteriological Reviews*. 1977; 41:100–180. [PubMed: 860983]
- Tholen A, Brune A. Localization and In Situ Activities of Homoacetogenic Bacteria in the Highly Compartmentalized Hindgut of Soil-Feeding Higher Termites (*Cubitermes* spp.). *Applied and Environmental Microbiology*. 1999; 65:4497–4505. [PubMed: 10508081]
- Wang J, Xue Y, Feng X, Li X, Wang H, Li W, et al. An analysis of the proteomic profile for *Thermoanaerobacter tengcongensis* under optimal culture conditions. *Proteomics*. 2004; 4:136–150. [PubMed: 14730678]
- Warnecke F, Luginbuhl P, Ivanova N, Ghassemian M, Richardson TH, Stege JT, et al. Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature*. 2007; 450:560–565. [PubMed: 18033299]
- Wiegel J, Ljungdahl LG. *Thermoanaerobacter ethanolicus* gen. nov., spec. nov., a new, extreme thermophilic, anaerobic bacterium. *Archives of Microbiology*. 1981; 128:343–348.
- Wieringa KT. Over het verdwijnen van waterstof en koolzuur onder anaerobe voorwaarden. *Antonie van Leeuwenhoek*. 1936; 3:263–273.
- Wieringa KT. The formation of acetic acid from carbon dioxide and hydrogen by anaerobic spore-forming bacteria. *Antonie van Leeuwenhoek*. 1940; 6:251–262.
- Wood HG. Fermentation of 3,4-C<sup>14</sup>- and 1-C<sup>14</sup>- labeled glucose by *Clostridium thermoaceticum*. *Journal of Biological Chemistry*. 1952; 199:579–583. [PubMed: 13022664]
- Wood, HG.; Ljungdahl, LG. Autotrophic character of the acetogenic bacteria. In: Shively, JM.; Barton, LL., editors. *Variation in Autotrophic Life*. New York: Academic Press; 1991. p. 201-250.
- Wu M, Ren Q, Durkin AS, Daugherty SC, Brinkac LM, Dodson RJ, et al. Life in Hot Carbon Monoxide: The Complete Genome Sequence of *Carboxydotherrmus hydrogeniformans* Z-2901. *PLoS Genetics*. 2005; 1:e65. [PubMed: 16311624]
- Xue Y, Xu Y, Liu Y, Ma Y, Zhou P. *Thermoanaerobacter tengcongensis* sp. nov., a novel anaerobic, saccharolytic, thermophilic bacterium isolated from a hot spring in Tengcong, China. *Int J Syst Evol Microbiol*. 2001; 51:1335–1341. [PubMed: 11491330]
- Yamamoto I, Saiki T, Liu S, Ljungdahl L. Purification and properties of NADP-dependent formate dehydrogenase from *Clostridium thermoaceticum*, a tungsten-selenium-iron protein. *J Biol Chem*. 1983; 258:1826–1832. [PubMed: 6822536]
- Ye Q, Roh Y, Carroll SL, Blair B, Zhou J, Zhang CL, Fields MW. Alkaline Anaerobic Respiration: Isolation and Characterization of a Novel Alkaliphilic and Metal-Reducing Bacterium. *Appl Environ Microbiol*. 2004; 70:5595–5602. [PubMed: 15345448]





**Figure 2. The Wood-Ljungdahl pathway of CO<sub>2</sub> fixation**

Specific *M. thermoacetica* genes predicted or known to encode each step in the pathway are identified by locus tags, which can be used to find gene information at [img.jgi.doe.gov](http://img.jgi.doe.gov). The C-1 groups are color coded with the C-1 unit metabolized through the Eastern or methyl branch of the pathway shown in blue and the C-1 group in the Western or carbonyl branch shown in red.



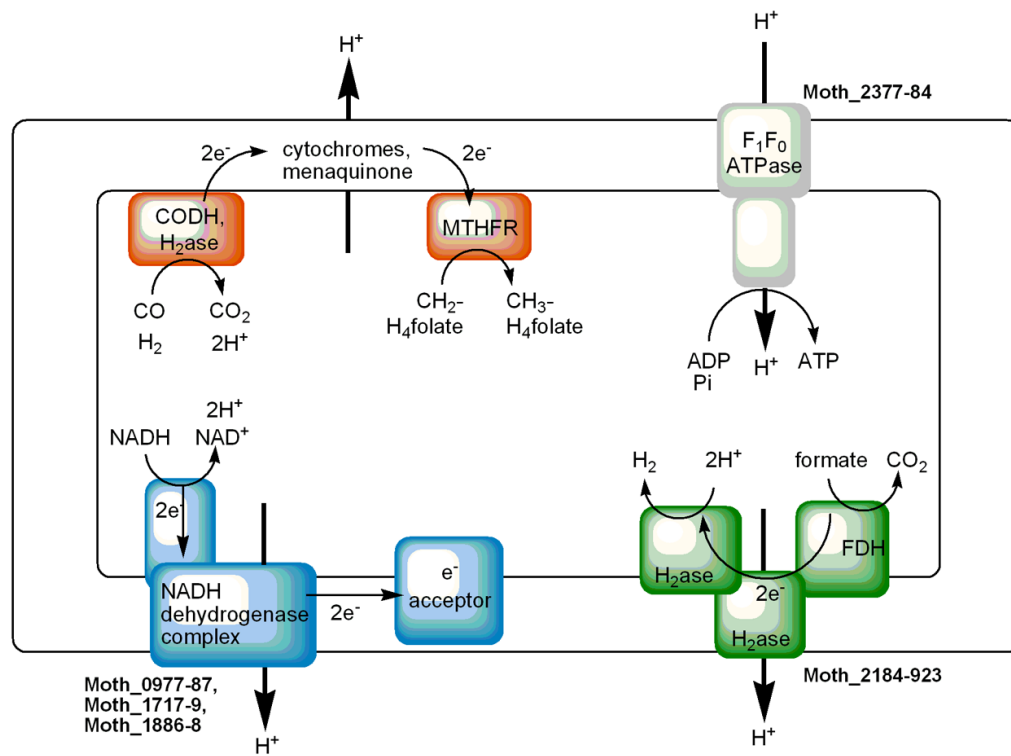
**Figure 3. Arrangement of Wood-Ljungdahl pathway genes in the *acs* gene cluster and in the chromosome**

(A) Arrangement of Wood-Ljungdahl pathway genes on the circular chromosome of *M. thermoacetica*. The numbering shows kilobase pairs from the origin of replication. (B) shows the *acs* gene cluster that contains core Wood-Ljungdahl pathway genes discussed in the text.



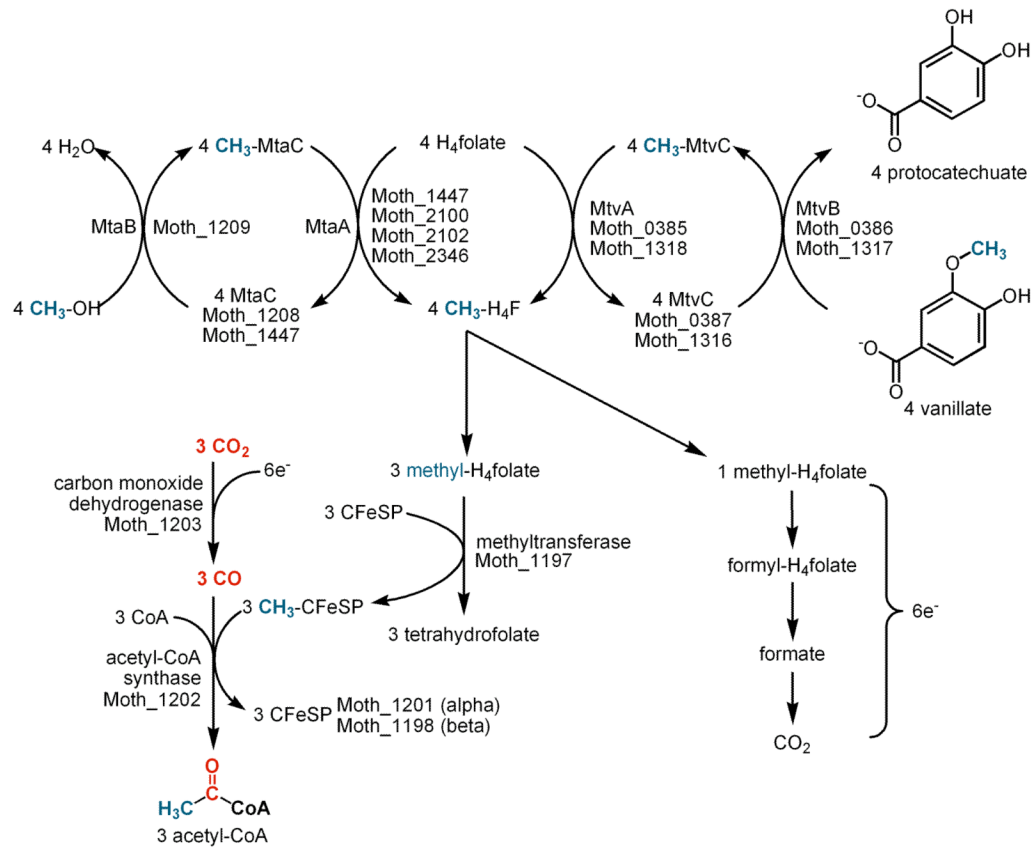
**Figure 4. Wood-Ljungdahl pathway gene clusters described in the text**

Abbreviations are as follows: *cooC*: CODH chaperone; *acsA*: CODH; *acsB*: ACS; *acsC*: corrinoid iron-sulfur protein large subunit; *orf7*: unknown function; *acsF*: ACS chaperone similar to *cooC*; *acsD*: corrinoid iron-sulfur protein small subunit; *acsE*: methyltransferase; *hdrC*: heterodisulfide reductase subunit C; *hdrB*: heterodisulfide reductase subunit B; *hdrA*: heterodisulfide reductase subunit A; *d*: similar to archaeal F420 reducing hydrogenase, delta subunit (COG1908); *clp*: similar to clp protease; *u*: unknown function; *fhs*: formyl-H<sub>4</sub>folate synthase; *ftcd*: formimido-H<sub>4</sub>folate cyclodeaminase; *X*: zinc-finger protein of unknown function; *metF*: methylene-H<sub>4</sub>folate reductase; *acoL*: dihydrolipoamide dehydrogenase; *gcvH*: glycine cleavage system H protein; *fd*: ferredoxin; *pulE*: Type II secretory pathway, ATPase PulE; *folD*: formyl-H<sub>4</sub>folate cyclohydrolase/methylene-H<sub>4</sub>folate dehydrogenase; *methfr/acsD*: methylene-H<sub>4</sub>folate reductase/corrinoid iron-sulfur protein small subunit fusion; *Y*: NADPH-dependent oxidoreductase (COG0493); *Z*: molybdopterin oxidoreductase (COG3383).



**Figure 5. Possible energy conserving pathways**

Shown are pathways for growth by the Wood-Ljungdahl pathway (orange), using formate hydrogenlyase (green), or with alternate electron donors and acceptors, using NADH dehydrogenase I (blue). Under the right growth condition for each system, the system would generate a transmembrane proton gradient that could be used to drive ATP synthesis by the F<sub>1</sub>F<sub>0</sub>ATPase (grey).



**Figure 6. Methyl groups from methanol and vanillate linked to the Wood-Ljungdahl pathway**  
Demethylation of vanillate or methanol provides pre-formed methyl groups to the Wood-Ljungdahl pathway. Three moles of acetyl-CoA are formed from four moles of methanol or vanillate, because oxidation of one methyl group provides six electrons that reduce three moles of CO<sub>2</sub> to three moles of CO, which can be used in the synthesis of three acetyl-CoA molecules. The C-1 units are labeled as in Figure 2.

**Table 1**General features of the *M. thermoacetica* genome.

Genome size (base pairs)	2628784
G+C Content	0.56
Open reading frames	2615
Predicted protein encoding sequences	2523
Predicted sequences encoding RNA genes	92
Coding density	97.90%
rRNA	3
tRNA	51
Genes with function prediction (percentage)	1834 (71.17%)
Number of genes in COG (percentage)	2008 (76.79%)
Number of genes in Pfam (percentage)	2001 (76.52%)
Genes encoding signal peptides (percentage)	159 (6.08%)
Genes encoding transmembrane proteins (percentage)	570 (21.8%)
Genes with orthologs (percentage)	2326 (90.26%)
Genes with paralogs (percentage)	1361 (52.81%)



**Table 2**Cytochromes present in the *M. thermoacetica* genome.

Locus tag	Annotation	COG
Moth_0452	Formate dehydrogenase, gamma subunit (cytochrome b) <sup>1</sup>	COG2864
Moth_1605	Cytochrome c family protein	no cog
Moth_2095	Cytochrome bd ubiquinol oxidase, subunit I	COG1271
Moth_2096	Cytochrome d ubiquinol oxidase, subunit II	COG1294

<sup>1</sup>This formate dehydrogenase is different than the selenocystine-containing formate dehydrogenase identified experimentally by Li, Ljungdahl, and Gollin in genbank (gi 1658401)

Table 3

Comparison of Moth\_0977-87 with genes encoding *C. hydrogenoformans*, *R. rubrum*, and *E. coli* NADH dehydrogenase I and *M. barkeri* F-420 dependent hydrogenase.

<i>M. thermoacetica</i>	<i>C. hydrogenoformans</i>	<i>R. rubrum</i>	<i>E. coli</i>	<i>M. barkeri</i>	COG
Moth_0977	CHY_1415	Rru_A1555	b2288/	Mbar_A3412	0838
Moth_0978	CHY_1416	Rru_A1556	b2287	Mbar_A3411	0388
Moth_0979	CHY_1417	Rru_A1557	NH	Mbar_A3410	0852
Moth_0980	CHY_1418	Rru_A1558	b2286	Mbar_A3409	0649
NG	NG	Rru_A1559-61/	b2285-3/	NG	1905, 1894, 1034
Moth_0981	CHY_1419	Rru_A1562	b2282	Mbar_A3408	1005
Moth_0982	CHY_1420	Rru_A1563	b2281	Mbar_A3407	1143
Moth_0983	CHY_1421	Rru_A1564	b2280/	Mbar_A3406	0839
NG	NG	NG	NG	Mbar_A3405	0713
Moth_0984	CHY_1422	Rru_A1565	b2279	Mbar_A3404	1009
Moth_0985	CHY_1423	Rru_A1566	b2278	Mbar_A3403	1009
Moth_0986	CHY_1424	Rru_A1567	b2277	Mbar_A3402	1008
Moth_0987	CHY_1425	Rru_A1568	b2276	Mbar_A3401	1007
Subunit designation					
Moth_0977	A				
Moth_0978	B				
Moth_0979	C				
Moth_0980	D				
NG	EFG				
Moth_0981	H				
Moth_0982	I				
Moth_0983	J				
NG					
Moth_0984	K				
Moth_0985	L				
Moth_0986	M				
Moth_0987	N				

<sup>1</sup>These genes do not have reciprocal best hits with genes in the *M. thermoacetica* gene cluster (Moth\_0977 to Moth\_0987)

NG No homologous gene is present in this gene cluster.

NH No gene in this organism is a BLAST hit with the corresponding *M. thermoacetica* gene. The BLAST cut-off is  $e\text{-value} < 1e^{-2}$ .

Table 4

Comparison of Moth\_2184-92 with hydrogenases from *C. hydrogeniformans* and *E. coli*.

<i>M. thermoacetica</i>	<i>C. hydrogeniformans</i>	<i>E. coli</i> hydrogenase 3	<i>E. coli</i> hydrogenase 4
Moth_2184	CHY_1830 <i>cooL</i>	b2719 <i>hycG</i>	b2489 <sup>I</sup> <i>hyI</i>
Moth_2185	CHY_1829 <i>cooX</i>	b2720 <sup>I</sup> <i>hycF</i>	b2488 <i>hyH</i>
Moth_2186	CHY_1827 <i>cooH</i>	b2721 <i>hycE</i>	b2487 <sup>I</sup> <i>hyG</i>
Moth_2187	CHY_1832 <sup>I</sup> <i>cooM</i>	NH	b2486 <i>hyD</i>
Moth_2188	CHY_1832 <sup>I</sup> <i>cooM</i>	b2723 <sup>I</sup> <i>hycC</i>	b2485 <i>hyF</i>
Moth_2189	NH	NH	b2484 <i>hyE</i>
Moth_2190	CHY_1831 <i>cooK</i>	b2722 <sup>I</sup> <i>hycD</i>	b2483 <i>hyC</i>
Moth_2191	CHY_1832 <sup>I</sup> <i>cooM</i>	b2723 <sup>I</sup> <i>hycC</i>	b2482 <i>hyB</i>
Moth_2192	CHY_1825 <sup>I</sup> <i>cooF</i>	b2724 <sup>I</sup> <i>hycB</i>	b2481 <i>hyA</i>
	COG		
Moth_2184	3260		
Moth_2185	1143		
Moth_2186	3261		
Moth_2187	1009		
Moth_2188	0651		
Moth_2189	4237		
Moth_2190	0650		
Moth_2191	0651		
Moth_2192	1142		

<sup>I</sup> These genes do not have reciprocal best hits with genes in the *M. thermoacetica* gene cluster (Moth\_2184 to Moth\_2192)

NH No BLAST hit with e-value < 1e<sup>-2</sup> in this gene cluster

**Table 5**Heterodisulfide reductase gene clusters in the *M. thermoacetica* genome.

<b>Locus tag</b>	<b>Annotation</b>	<b>COG</b>
Moth_0439	Heterodisulfide reductase subunit C-like (sequence is shorter)	COG1150
Moth_0440	Heterodisulfide reductase subunit B	COG2048
Moth_0441	Heterodisulfide reductase subunit A	COG1148
Moth_0442 <sup>1</sup>	Similar to archaeal F420 reducing hydrogenase, delta subunit	COG1908
Moth_0809	Heterodisulfide reductase subunit C	COG1150
Moth_0810	Heterodisulfide reductase subunit B	COG2048
Moth_0811	Heterodisulfide reductase selenocysteine containing subunit A	COG1148
Moth_0812	Heterodisulfide reductase selenocysteine containing subunit A	COG1148
Moth_0813 <sup>1</sup>	Selenocysteine containing methyl viologen-reducing hydrogenase, delta subunit	COG1908
Moth_1193 <sup>1</sup>	Similar to archaeal F420 reducing hydrogenase, delta subunit	COG1908
Moth_1194	NADPH-dependent glutamate synthase beta chain and related oxidoreductases (COG0493), Heterodisulfide reductase subunit A (COG1148)	COG1148
		COG0493
Moth_1195	Heterodisulfide reductase subunit B	COG2048
Moth_1196	Heterodisulfide reductase subunit C	COG1150
Moth_1452	Unknown function	COG0247
Moth_1453	Heterodisulfide reductase A	COG1148
Moth_1454 <sup>1</sup>	FAD-dependent oxidoreductase/NAD(P)H dependent reductase	COG1251

<sup>1</sup>These *M. thermoacetica* genes are annotated by similarity to COGs 1908 and 1251, not by alignment with specific protein sequences.

**Table 6**MtvA, MtvB, and MtvC homologs in the *M. thermoacetica* genome.

<b>Locus tag</b>	<b>Annotation</b>	<b>COG</b>
Moth_0385	MtvA	COG1410
Moth_0386	MtvB	COG0407
Moth_0387	MtvC	COG5012
Moth_1197	AcsE (methyl-THF corrinoid FeS protein methyltransferase)	COG1410
Moth_1206	MtvC homolog	COG5012
Moth_1207	MtvA homolog	COG1410
Moth_1208	MtaC	COG5012
Moth_1209	MtaB	no COG
Moth_1213	MtvC homolog	COG1410
Moth_1311	MtvC homolog	COG5012
Moth_1314	MtvB plus an N-terminal domain	COG0407
Moth_1316	MtvC	COG5012
Moth_1317	MtvB	COG0407
Moth_1318	MtvA	COG1410
Moth_1447	MtaA/MtaC	COG0407, COG 5012
Moth_2100	MtaA	COG0407
Moth_2102	MtaA	COG0407
Moth_2114	MtvC homolog/histidine kinase fusion protein	COG5012, COG2972
Moth_2115	MtvA homolog	COG1410
Moth_2116	MtvC homolog	COG5012
Moth_2346	MtaA	COG0407

**Table 7**

Gene clusters containing predicted molybdopterin subunits in COGs 0243 and 3383.

<b>Locus tag</b>	<b>Annotation</b>	<b>COG</b>
Moth_0450	Molybdopterin subunit, TAT sequence	COG0243
Moth_0451	Iron sulfur cluster containing subunit	COG0437
Moth_0452	Cytochrome b556 subunit of formate dehydrogenase	COG2864
Moth_1379	TorD	COG3381
Moth_1381	Iron sulfur cluster containing subunit	COG0437
Moth_1382	Molybdopterin containing subunit	COG0243
Moth_1384	DmsC type membrane anchor subunit	COG3302
Moth_1385	Iron sulfur cluster containing subunit	COG0437
Moth_1386	Molybdopterin subunit, TAT sequence	COG0243
Moth_1388	Iron sulfur cluster containing subunit	COG0437
Moth_1389	Molybdopterin subunit, TAT sequence	COG0243
Moth_1390	TorD	COG3381
Moth_1906	Polysulfide reductase, integral membrane protein	COG5557
Moth_1907	Iron sulfur cluster containing subunit	COG0437
Moth_1908	Molybdopterin subunit, TAT sequence	COG0243
Moth_1909	Molybdopterin containing subunit	COG3383
Moth_2193	Se-containing formate hydrogen lyase subunit	COG3383
Moth_2312-13	Se-containing formate dehydrogenase subunit	COG3383
Moth_2314	Predicted NADPH dependent oxidoreductase	COG0493