



Analysis of the Core Genome and Pan-Genome of Autotrophic Acetogenic Bacteria

Jongoh Shin¹, Yoseb Song¹, Yujin Jeong¹ and Byung-Kwan Cho^{1,2*}

¹ Systems and Synthetic Biology Laboratory, Department of Biological Sciences and KI for the BioCentury, Korea Advanced Institute of Science and Technology, Daejeon, South Korea, ² Intelligent Synthetic Biology Center, Daejeon, South Korea

OPEN ACCESS

Edited by:

Sabine Kleinstuber,
Helmholtz Centre for Environmental
Research, Germany

Reviewed by:

Michael Köpke,
LanzaTech, USA
Guillaume Bruant,
National Research Council Canada,
Canada

*Correspondence:

Byung-Kwan Cho
bcho@kaist.ac.kr

Specialty section:

This article was submitted to
Microbiotechnology, Ecotoxicology
and Bioremediation,
a section of the journal
Frontiers in Microbiology

Received: 01 February 2016

Accepted: 12 September 2016

Published: 28 September 2016

Citation:

Shin J, Song Y, Jeong Y and
Cho B-K (2016) Analysis of the Core
Genome and Pan-Genome
of Autotrophic Acetogenic Bacteria.
Front. Microbiol. 7:1531.
doi: 10.3389/fmicb.2016.01531

Acetogens are obligate anaerobic bacteria capable of reducing carbon dioxide (CO₂) to multicarbon compounds coupled to the oxidation of inorganic substrates, such as hydrogen (H₂) or carbon monoxide (CO), via the Wood-Ljungdahl pathway. Owing to the metabolic capability of CO₂ fixation, much attention has been focused on understanding the unique pathways associated with acetogens, particularly their metabolic coupling of CO₂ fixation to energy conservation. Most known acetogens are phylogenetically and metabolically diverse bacteria present in 23 different bacterial genera. With the increased volume of available genome information, acetogenic bacterial genomes can be analyzed by comparative genome analysis. Even with the genetic diversity that exists among acetogens, the Wood-Ljungdahl pathway, a central metabolic pathway, and cofactor biosynthetic pathways are highly conserved for autotrophic growth. Additionally, comparative genome analysis revealed that most genes in the acetogen-specific core genome were associated with the Wood-Ljungdahl pathway. The conserved enzymes and those predicted as missing can provide insight into biological differences between acetogens and allow for the discovery of promising candidates for industrial applications.

Keywords: acetogens, comparative genomics, conserved pathway, CO₂ fixation, Wood-Ljungdahl pathway

INTRODUCTION

In recent decades, demands for fossil fuel-derived chemicals and energy have rapidly increased, along with concerns about climate change. Currently, ~80% of world energy is generated via fossil fuel processing, which is responsible for 40% of CO₂ emissions and global warming (Spigarelli and Kawatra, 2013; Saeidi et al., 2014). Although several methods for replacing fossil fuels have been proposed (Naik et al., 2010), lack of environmental and economic sustainability have demonstrated the technological inability to derive a solution to the climate and energy crisis. As an alternative approach, the gas fermentation process has received attention; it utilizes a unique metabolism in acetogenic bacteria (acetogens), which convert CO₂ to biofuels (Henstra et al., 2007; Bengelsdorf et al., 2013; Latif et al., 2014).

Acetogens are a physiologically defined group of bacteria that synthesize acetyl-CoA as a central metabolic intermediate from chemolithoautotrophic substrates, such as CO/CO₂ or H₂/CO₂, through acetogenesis (Drake, 1994). Acetogenesis constitutes an appropriate type of microbial metabolism for the substitution of fossil fuels owing to its ability to convert single carbon (C₁) compounds, such as CO and CO₂, via the reductive acetyl-CoA pathway to acetyl-CoA, which is

referred to as the Wood-Ljungdahl pathway. Owing to this physiological trait, acetogens play key roles in the global carbon cycle (McInerney and Bryant, 1981) by performing the production of large volumes of acetic acid ($>10^{12}$ kg annually; Wood and Ljungdahl, 1991). Moreover, acetogens have been engineered as a novel platform for conversion of waste gasses, such as industrial synthesis gas or syngas, from gasification of biomass into useful multicarbon chemicals (Schiel-Bengelsdorf and Dürre, 2012). This strategy has many advantages over traditional thermochemical processes, such as Fischer-Tropsch synthesis, including operation at lower temperature, lower pressure, higher tolerance of impurities, and flexible syngas-composition utilization (Spigarelli and Kawatra, 2013).

Though acetogens are present in at least 23 different genera (Drake et al., 2006), comprehensive analysis of genes and proteins involved in acetogenesis indicated that acetogens contain conserved physiological properties. The most important shared feature is the conversion of CO_2 to formate via fixation and to acetyl-CoA, which can be used as a metabolic intermediate for biomass and byproduct synthesis. To elucidate these properties, the biochemistry of the Wood-Ljungdahl pathway and energy conservation systems has been extensively studied (Drake et al., 2008; Ragsdale and Pierce, 2008). In recent years, the enzymatic reactions associated with acetogenesis have been well characterized, especially in *Clostridium autoethanogenum* (Wang et al., 2013; Mock et al., 2015), *Moorella thermoacetica* (Huang et al., 2012; Mock et al., 2014), and *Acetobacterium woodii* (Schuchmann and Müller, 2012; Schuchmann and Muller, 2013; Bertsch et al., 2015).

In addition to the understanding of acetogenesis, elucidation of the molecular mechanisms associated with acetogens has undergone tremendous progress as a result of genome sequencing. The genome sequences of acetogens represent useful information to aid the search for novel enzymes/pathways, generating hypotheses related to energy conservation systems, and accessing evolutionary relationships between species that have not previously been characterized biochemically. For example, studies focusing on construction of *in silico* genome-scale mathematical models, as well as transcriptomics and proteomics investigation of the Wood-Ljungdahl pathway and related energy conservation systems, were undertaken primarily owing to the availability of genome-sequence information (Nagarajan et al., 2013; Islam et al., 2015; Marcellin et al., 2016).

Given the increased volume of genomic information, comparative genomic analysis of acetogens is possible. Among currently available comparative genomic approaches, pan-genome analysis is widely used to construct a framework for estimating genomic diversity of entire repertoires and identifying core genomes (shared by all strains), dispensable genomes (existing in two or more strains), and specific (unique to single strain) gene pools for a species (Tettelin et al., 2005). Conserved and alternative pathways across species provide insight into the biological differences between species (Kelley et al., 2003), allow the discovery of promising target proteins for industrial applications, and create hypotheses regarding missing genes or possible alternatives to current metabolic pathways. Moreover,

these findings increase the understanding of genetic differences and related reactions.

In this review, we specifically addressed recent studies on the complete genomes and conserved genes associated with CO/CO_2 utilization in diverse acetogens. We focused on pathways essential for autotrophic growth, discussed the main features and conservation of metabolic pathways, and addressed the structural differences and relationships between acetogens.

THE CORE GENOME OF ACETOGENS: WHICH GENETIC CHARACTERISTICS ARE SHARED AMONG ACETOGENS?

Currently, >100 acetogens have been isolated from diverse habitats (Drake et al., 2006). With advances in sequencing technology along with increased biotechnological interest in acetogens, the number of sequenced acetogen genomes has increased every year since the first genome was sequenced. Recently, eight complete genomes (34.7%) were published in 2015, containing five *de novo* sequencing and three resequencing genomes (Table 1). In response to the diversely isolated environments and culture conditions, the features of the genomes vary. The length of acetogen genomes range from ~ 2.4 to ~ 5.7 Mb, with an average length of 3.8 Mb and having GC content between 29.1% and 55.8% (average: 38.5%; Table 1). Analysis of sequence annotations revealed that on average, 85.6% of the genomes consist of coding sequences, with approximately 1.1 coding sequence per kb.

Based on these complete acetogen genomes, comprehensive genome analysis is possible to understand the functionality and specificity conserved among autotrophic acetogenic bacteria (Hayashi et al., 2001; Ohnishi et al., 2001). For this purpose, we selected 14 strains that have been experimentally confirmed as capable of converting acetyl-CoA from CO/CO_2 and, thus, from inorganic carbon through the Wood-Ljungdahl pathway (Table 1). Although *Carboxydotherrmus hydrogenoformans* and *Thermacetogenium phaeum* are carboxydophilic hydrogenogenic and syntrophic acetate-oxidizing bacteria, respectively, unlike model acetogens, their acetogenic growth has been reported (Hattori et al., 2000, 2005; Henstra and Stams, 2011; Haddad et al., 2013). On the other hand, the capability of *Clostridium sticklandii* DSM 519 for autotrophic growth on C_1 substrates via the Wood-Ljungdahl pathway was not confirmed (Fonknechten et al., 2010); therefore, this strain was excluded in this analysis.

For downstream analysis, 14 complete acetogen genome sequences were obtained from the National Center for Biotechnology Information database¹ (Table 1). Pan-Genomes Analysis Pipeline (PGAP-1.12; Zhao et al., 2012) identified functional genes presented in all strains (core genome), two or more strains (dispensable genomes), and unique strains (specific genomes; Tettelin et al., 2005). For comparative analysis, the MultiParanoid method was used to analyze cluster orthologs and inparalogs shared by multiple genomes based

¹<http://ftp.ncbi.nih.gov/genomes/genbank/bacteria>

TABLE 1 | Characteristics of the complete genomes of acetogens.

Strain	Isolation	Temperature	Genome size (bp)	G+C (%)	Number of genes	Number of CDSs	Accession no.	Reference
<i>Acetobacterium woodii</i> DSM 1030*	Mud	30°C	4,044,777	39.3	3,649	3,521	CP002987	Poehlein et al., 2012
<i>Acetohalobium arabaticum</i> DSM 5501*	Lagoons	37°C	2,469,596	36.6	2,396	2,286	CP002105	Sikorski et al., 2010
<i>Carboxydotherrmus hydrogeniformans</i> Z-2901*	Hotspring	78°C	2,401,520	42	2,495	2,406	CP000141	Wu et al., 2005
<i>Clostridium acetium</i> DSM 1496*	Mud	30°C	4,201,318	35.3	3,847	3,705	CP009687	Poehlein et al., 2015b
<i>Clostridium autoethanogenum</i> DSM 10061*	Rabbit faeces	37°C	4,352,205	31.1	3,983	3,741	CP006763	Brown et al., 2014
<i>Clostridium autoethanogenum</i> DSM 10061	Rabbit faeces	37°C	4,352,446	31.1	4,069	3,964	CP012395	Humphreys et al., 2015
<i>Clostridium carboxidivorans</i> P7*	Lagoons	37°C	5,732,880	29.9	5,167	5,004	CP011803	Li et al., 2015
<i>Clostridium ljungdahlii</i> DSM 13528*	Chicken yard waste	37°C	4,630,065	31.1	4,234	4,081	CP001666	Köpke et al., 2010
<i>Clostridium scatologenes</i> ATCC 25775*	Soil	37°C	5,749,410	29.6	5,183	4,974	CP009933	Zhu et al., 2015
<i>Clostridium sticklandii</i> DSM 519	Mud	37°C	2,715,461	33.3	2,625	2,476	FP565809	Fonknechten et al., 2010
<i>Eubacterium limosum</i> KIST 612*	Anaerobic digester fluid	37°C	4,316,707	47.5	4,089	3,966	CP002273	Roh et al., 2011
<i>Eubacterium limosum</i> SA11	Sheep lumen	37°C	4,150,332	47.4	3,902	3,805	CP011914	Kelly et al., 2016
<i>Moorella thermoacetica</i> ATCC 39073*	Horse faeces	55°C	2,628,784	55.8	2,613	2,463	CP000232	Pierce et al., 2008
<i>Moorella thermoacetica</i> DSM 521	Horse faeces	55°C	2,527,564	55.9	2,518	2,405	CP012369	Poehlein et al., 2015a
<i>Moorella thermoacetica</i> DSM 2955	Horse faeces	55°C	2,623,349	55.8	2,609	2,508	CP012370	Bengelsdorf et al., 2015
<i>Clostridium difficile</i> 630*	Human intestine	37°C	4,290,252	29.1	3,971	3,756	AM180355	Sebalhia et al., 2006
<i>Clostridium difficile</i> CD196	Human intestine	37°C	4,110,554	28.7	3,526	3,487	FN538970	Stabler et al., 2009
<i>Clostridium difficile</i> M120	Human intestine	37°C	4,047,729	28.7	3,707	3,502	FN665653	He et al., 2010
<i>Clostridium difficile</i> 630	Human intestine	37°C	4,274,806	29	3,972	3,794	CP010905	Riedel et al., 2015
<i>Clostridium difficile</i> 630 Deltaerm	Human intestine	37°C	4,293,049	29.1	3,990	3,816	LN614756	van Eijk et al., 2015
<i>Thermoacetogenium phaeum</i> DSM 12270*	sludge	60°C	2,939,057	53.9	2,894	2,766	CP003732	Oehler et al., 2012
<i>Thermoanaerobacter kivui</i> LKT-1*	Lake sediment	65°C	2,397,824	35	2,425	2,198	CP009170	Hess et al., 2014
<i>Treponema primitia</i> ZAS-2*	Termitite	30°C	4,059,867	50.8	3,536	3,427	CP001843	Rosenthal et al., 2011

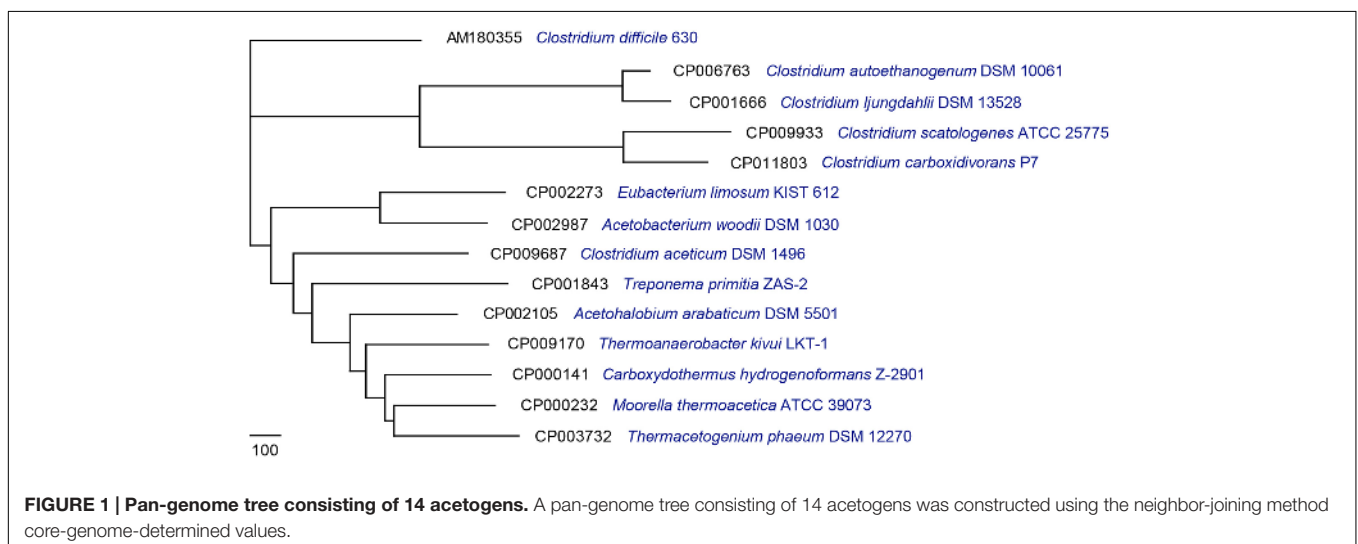
Genome sequences analyzed in this paper are indicated in asterisk (*).

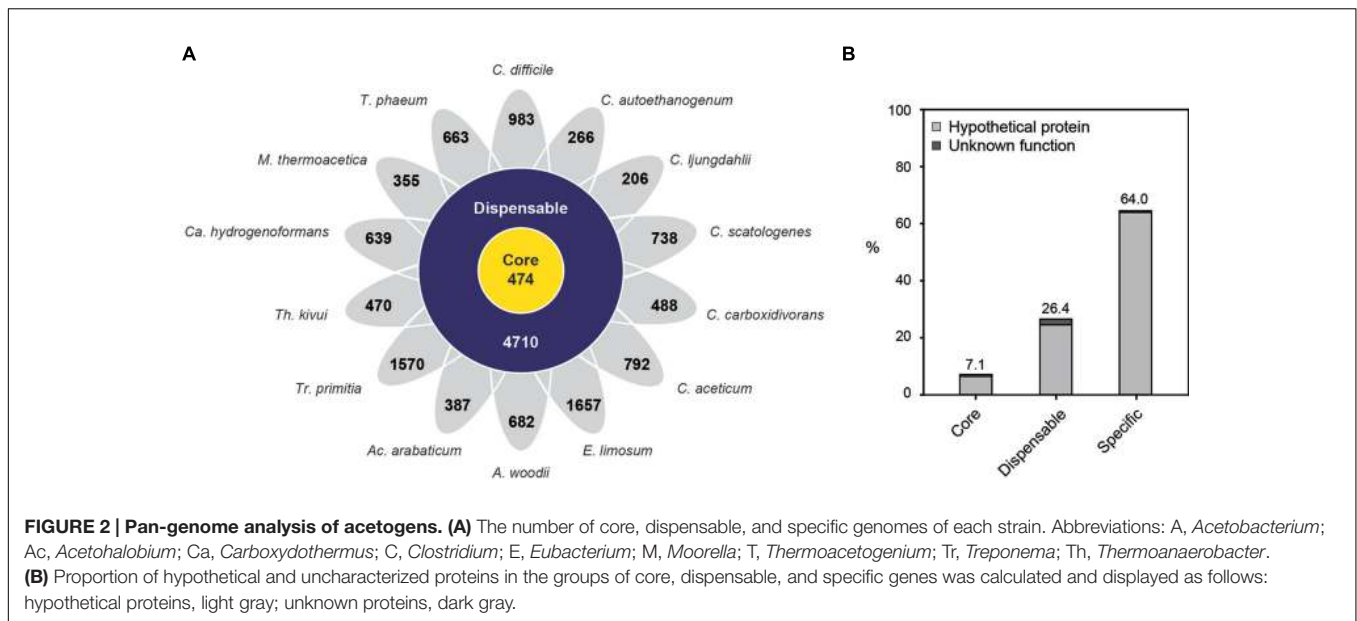
on sequence similarity (Alexeyenko et al., 2006; Zhao et al., 2012). Additionally, BLASTP was used to determine similarities between protein sequences and filter results by setting minimum scores at 50 and E -values to 10^{-10} . The obtained result was clustered using the Markov cluster algorithm (Enright et al., 2002). To understand the evolutionary relationships among these acetogens, a pan-genome tree was constructed (Figure 1) based on the pan-genome dataset and neighbor-joining method (Zhao et al., 2012). All sister groups were clustered by the same genera or optimal temperature conditions. In contrast to the 16S-based phylogenetic tree (Bengelsdorf et al., 2013), the strain exhibiting the least amount of evolutionary change from a common ancestor was *Clostridium difficile*. *M. thermoacetica* (strain AMP) was previously reported to show atypical hydrogenogenic metabolism (Jiang et al., 2009), and the pan-genome tree also showed evolutionary closeness among *Ca. hydrogenoformans*, *T. phaeum*, and *M. thermoacetica* (Figure 1). These results suggested that functional gene composition of *M. thermoacetica* is similar to *Ca. hydrogenoformans*.

According to comparative genome analysis, a total of 15,079 orthologous groups with 50,178 genes were identified, consisting of 474 core gene groups with 12,457 genes, 4710 dispensable gene groups with 27,825 genes, and 9896 specific genes identified (Figure 2A; Supplementary Table S6). Core genes were well annotated, with 92.9% of genes. However, the number of specific genes in each organism varied from 206 to 1657, with 64.0% of the specific genes identified as having hypothetical functions (Figure 2B). Additionally, the number of specific genes did not correlate with the size of the genome, which is in contrast to the correlation between the number of genes and the size of the genome. For example, the genome of *Clostridium ljungdahlii* is the third largest (4.6 Mb), but its number of specific genes is 206, which is the least number of genes in the set. Additionally, 266 specific genes, which was the second least number of genes in the set, were found in *C. autoethanogenum*, having the fourth largest (4.3 Mb) genome.

To decipher the 474 core genes of the 14 acetogenic bacteria, functionally grouped networks of enriched categories were generated for the biological interpretation of core genes using ClueGo version 2.2.4 (Saito et al., 2012), which is a widely used Cytoscape version 3.3.0 (Shannon et al., 2003) plugin. For this analysis, *C. autoethanogenum* data was used as the standard, because *C. autoethanogenum* was recently confirmed systematically by transcriptome and proteome analysis of the Wood-Ljungdahl pathway (Marcellin et al., 2016). Gene Ontology (GO) terms (GO:0030634; Biological Process, carbon fixation by acetyl-CoA pathway) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (M00377; Pathway module, Wood-Ljungdahl pathway) were manually added along with the published experimental evidence (Marcellin et al., 2016) (Supplementary Table S1).

As a result, 95 GO terms were significantly enriched and categorized into 10 groups according to their kappa scores (Figure 3A). Overall, highly connected groups were assigned to adenosine triphosphate (ATP) binding, macromolecule modification and sulfate transport, cellular macromolecule metabolic process, and regulation of cellular process as group-leading terms (Figure 3A). Additionally, five sub-groups were involved in membrane component, monocarboxylic acid binding, transcription-factor binding, and transport and plasma membrane (Figure 3A; Supplementary Table S2). Therefore, GO analysis showed that the core genome was significantly correlated with a number of essential cellular functions, similar to most bacteria (Gil et al., 2004). To examine the acetogenic characteristics, core genome was trimmed by non-acetogenic core genome, which contains five non-acetogens phylogenetically close to 14 selected acetogenic bacteria (Supplementary Figure S1). Based on enrichment p -values, 27 GO terms and 8 KEGG pathways were enriched (Supplementary Table S3) and functionally categorized into 12 groups (Supplementary Figure S2). The most linked functional groups were assigned to cysteine and methionine metabolism, monobactam biosynthesis, small molecule biosynthetic process, Mo-molybdopterin cofactor





biosynthetic process, iron chelate transport, and the Wood-Ljungdahl pathway. This result is in agreement with related acetogenesis and cofactor biosynthetic pathways involved in the Wood-Ljungdahl pathway.

To further investigate unique core genes found in acetogens, the core genome was filtered using genomes of non-acetogenic anaerobic bacteria. In this analysis, the complete genome of *Clostridium butyricum* KNU-L09 was used, which is a strictly anaerobic, non-acetogenic bacteria that is phylogenetically similar to *C. difficile* 630 (Supplementary Figure S1). According to the functional annotation network of the acetogen-specific core genome, five KEGG pathways and five GO terms were specifically enriched (Figure 3B; Supplementary Table S4). Acetogen-specific functional networks consisted of 13 genes annotated as methionine synthase, CO dehydrogenase/acetyl-CoA synthase (CODH/ACS), ferredoxins, and a subunit of formylmethanofuran dehydrogenase. Thus, acetogen-specific functional networks were involved in specific molecular functions, such as iron-sulfur cluster-binding transferase activity and dihydropteroate-synthase activity, and biological processes, such as carbon fixation by the acetyl-CoA pathway and the pteridine-containing compound metabolic process. Interestingly, 12 of the 13 genes (92.3%) were highly associated with the Wood-Ljungdahl pathway. Of the 12 genes, six were located in a single gene cluster encoding the Wood-Ljungdahl pathway (CAETHG_1606-CAETHG_1621), while the other six genes were additional copies of those genes. Another gene specifically conserved in acetogens was the tungsten-containing formylmethanofuran dehydrogenase subunit E (*fwdE*), which catalyzes the first reduction of CO₂ in methanogens (Hochheimer et al., 1998). However, the other genes encoding tungsten formylmethanofuran dehydrogenase (*fwdABCD*), which often form an operon with *fwdE*, were absent in all 14 acetogen genomes. This protein encoded by *fwdE* contains a zinc-β-ribbon domain, suggesting that it

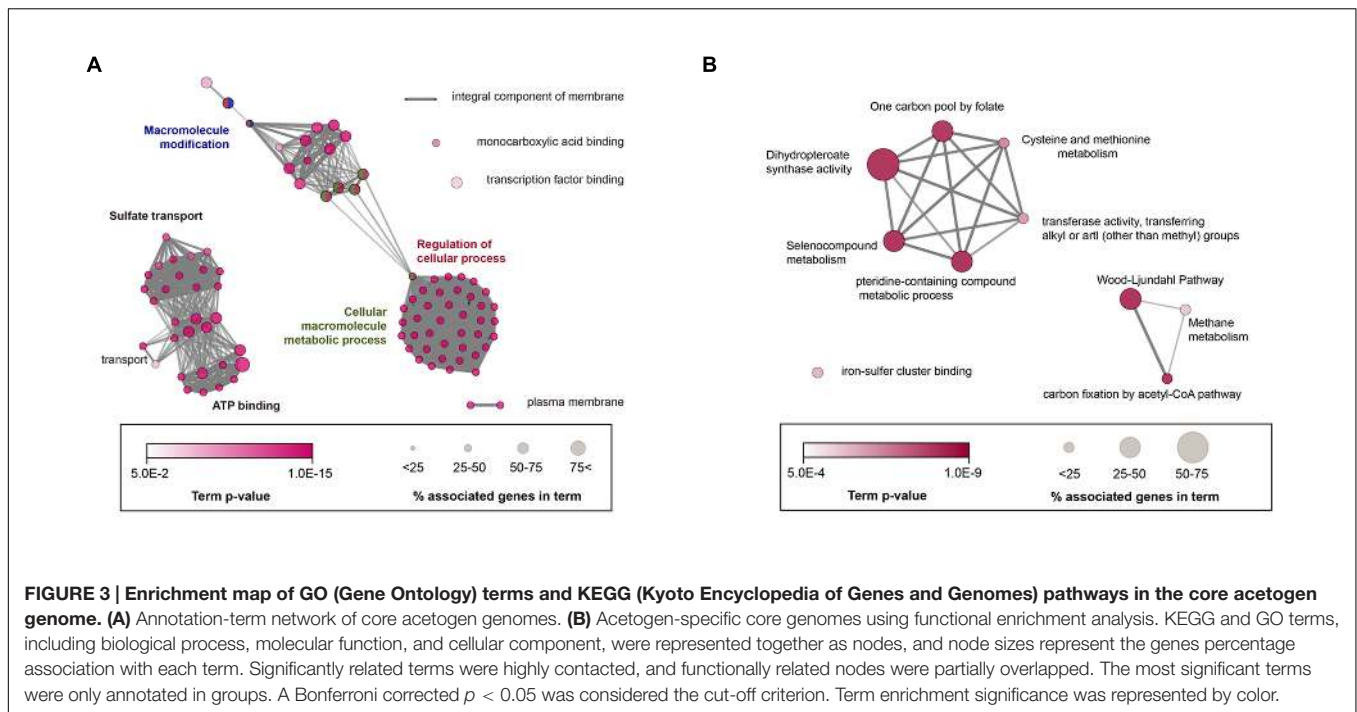
plays a role in transcriptional regulation as a DNA-binding protein; however, its exact role in acetogenesis remains unclear.

BIOSYNTHESIS OF ACETATE FROM CO/CO₂: THE WOOD-LJUNGDAHL PATHWAY

Based upon the analysis of the acetogen-specific core genome, the genes related to the Wood-Ljungdahl pathway were highly conserved as hallmarks of acetogens. This pathway involves the reduction of two CO₂ molecules into one acetyl-CoA with several coenzymes and electron carriers (Drake and Daniel, 2004; Ragsdale, 2008), and it is highly interconnected with energy conservation systems to overcome the same thermodynamically unfavorable reaction. Nevertheless, the pathway is the most efficient of the all CO₂-fixation pathways, including the Calvin cycle, the reductive tricarboxylic acid cycle, and the hydroxypropionate cycle (Fast and Papoutsakis, 2012). Moreover, the arrangement of genes related to the Wood-Ljungdahl pathway was well conserved with phylogenetic correlation in their genomes (Poehlein et al., 2015c). In this review, the Wood-Ljungdahl pathway was functionally separated into three core groups. The first core group encodes enzymes responsible for reducing CO₂ to formate. The second core group consists of the methyl- and the carbonyl-branch enzymes. The last core group is composed of acetate-producing genes.

THE WOOD-LJUNGDAHL PATHWAY CORE GROUP I: CO₂ TO FORMATE

The first reaction of acetogenesis is the reduction of CO₂ to formate by two-electron reduction, which is catalyzed



by selenocysteine- or non-selenocysteine-containing formate dehydrogenase (FDH) in a ferredoxin- or NADH-dependent reaction (Ljungdahl and Andreesen, 1978; Gollin et al., 1998; Schuchmann and Muller, 2013; Wang et al., 2013). Genes associated with the reaction are well conserved in all acetogens. According to genome-comparison analysis, two genes encoding selenocysteine-containing FDH (*fdhF*) and FDH-accessory protein (*fdhD*) are well conserved in core group I (Figure 4A). Despite conservation of *fdhF* and *fdhD*, a number of *fdh* gene copies are different in all of the genomes. For instance, *fdhF* and *fdhD* were located as a single gene cluster in the *C. difficile* genome. However, three copies of *fdhF* were found in *C. ljungdahlii* and *C. autoethanogenum*. Similar to the genes encoding seleno-containing FDH, the genes encoding non-selenocysteine residues containing FDH are also well conserved in the acetogen genomes. Although the selenoproteins are mutant forms of FDH that differ only in the presence of selenium instead of sulfur at the active site, seleno-containing FDHs exhibit higher catalytic rates relative to non-selenocysteine FDHs (Stadtman, 1991; Matson et al., 2010). However, non-selenocysteine FDH may be useful for acetogenesis in selenium-free environments.

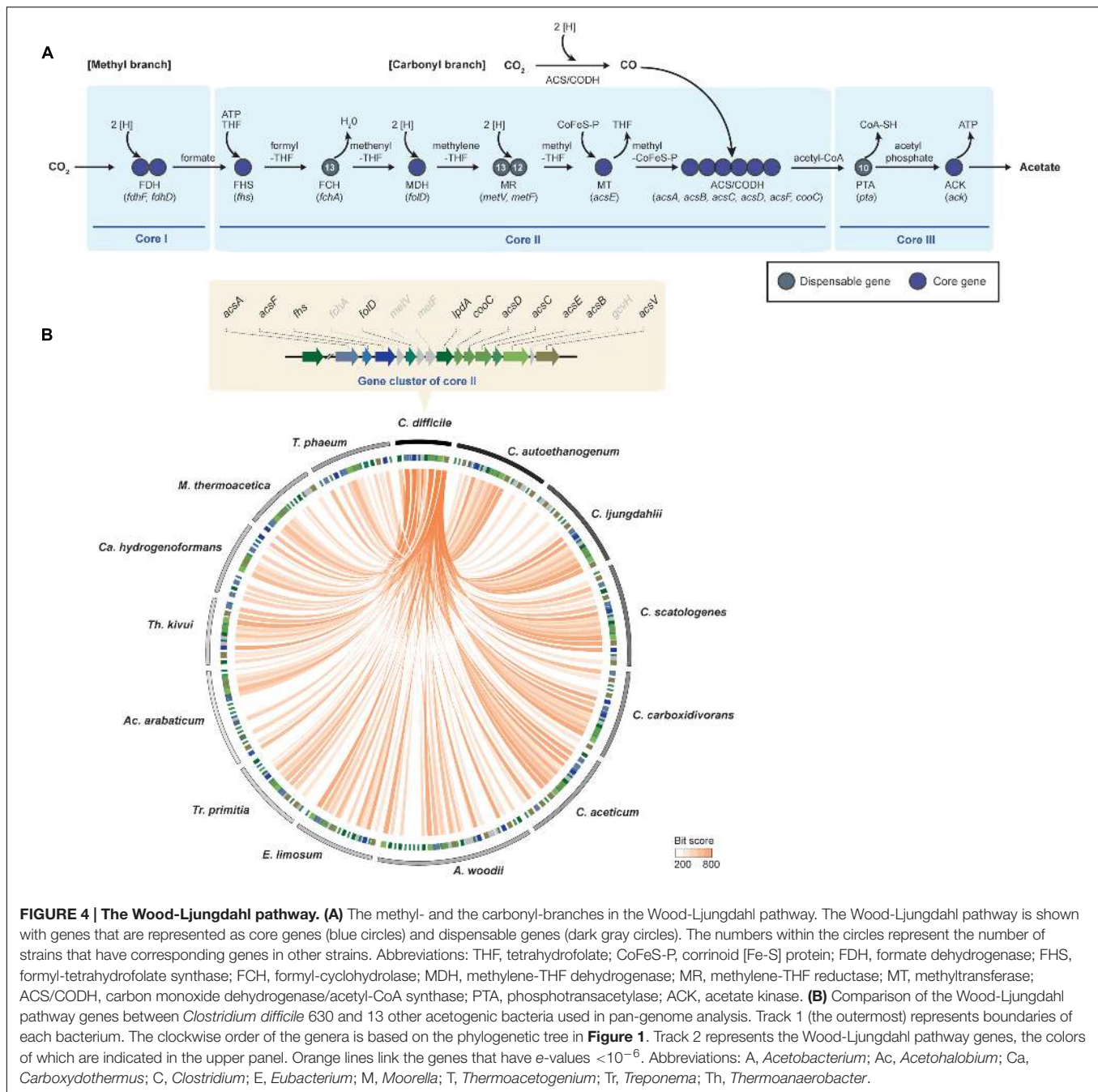
Although the *fdh* genes are highly conserved, electron-delivery systems involved in this reaction differ, owing to the diversity of electron acceptors associated with FDH (Schuchmann and Müller, 2014). For example, *A. woodii* and *Clostridium aceticum* have four or three hydrogenase modules, respectively, which are located in a gene cluster with the selenocysteine-containing *fdh* genes (Poehlein et al., 2012, 2015c; Schuchmann and Muller, 2013). In this process, *A. woodii* uses H_2 as an electron donor for CO_2 reduction, referred to as hydrogen-dependent CO_2 reductase, which can be energetically more advantageous as compared with utilizing energy intermediates by not expending

a substrate for the chemiosmotic gradient (Schuchmann and Muller, 2013). *C. autoethanogenum* and *C. ljungdahlii* also have complexes of ferredoxin and NAD-dependent [FeFe]-hydrogenases for CO_2 reduction, which are located near an *fdh* gene cluster encoding selenocysteine-containing FDH (Nagarajan et al., 2013; Wang et al., 2013).

THE WOOD-LJUNGDAHL PATHWAY CORE GROUP II: FORMATION OF ACETYL-CoA

Formate is subsequently converted to acetyl-CoA by a series of reactions catalyzed by the enzymes of the methyl branch of the Wood-Ljungdahl pathway. Core group II was composed of all key enzymes in the methyl and carbonyl branches (Figure 4A). In the methyl branch, formyl-tetrahydrofolate (THF) synthase (FHS) converts formate to formyl-THF by investing one molecule of ATP. For the next two steps, formyl-THF cyclohydrolase (FCH) and methylene-THF dehydrogenase (MDH) consecutively catalyze the converted THF into methenyl-THF, then to methylene-THF, which is then converted to methyl-THF and methyl-CoFeSP by using methylene-THF reductase (MR, two subunits of methylene-THF reductase; *metV* and *metF*) and methyltransferase (MT, two subunits of corrinoid/Fe-S protein; *acsC* and *acsD*, methyltransferase: *acsE*), respectively. For the carbonyl branch, CO_2 becomes CO via catalysis by the CODH/ACS complex (CODH: *acsA*, *acsF*, and *acoC*; ACS: *acsB*). Using the same enzyme, the two molecules, methyl-CoFeSP and CO, combine into acetyl-CoA.

Nine genes encoding FHS, MDH, MT, CODH, and ACS were well conserved in all 14 acetogens. However, two genes



that encode FCH and two MR subunits were determined to be dispensable genes. One of the four dispensable genes, *fchA*, is responsible for converting formyl-THF into methyl-THF. In order to perform a similarity search of *fchA* throughout the other genomes, the *fchA* sequence from *C. difficile* was used, and it was determined that *fchA* from 13 acetogen genomes was highly conserved, although the enzyme was only absent in the *M. thermoacetica* genome (Pierce et al., 2008). According to a previous study, in *M. thermoacetica*, the cyclization of formyl-THF and the reduction of methenyl-THF were observed being catalyzed by MDH by substituting FCH (O'Brien et al., 1973;

Pierce et al., 2008), which is not a core gene in the Wood-Ljungdahl pathway. Although the *fchA* gene is not a core gene set, the biochemical reaction associated with conversion of formyl-THF to methylene-THF is a conserved step in all acetogens for acetogenesis.

Other dispensable genes included *metF* and *metV* that encode MR. These redox enzymes contain iron-sulfur clusters and utilize reduced forms of electron carriers (ferredoxin or NADH) as electron donors. They reduce methylene-THF to methyl-THF using different enzyme complexes (Clark and Ljungdahl, 1984; Park et al., 1991). In this step, enzymatic diversity denoted

by related-subunit compositions was reported among acetogens (Mock et al., 2014; Bertsch et al., 2015; Jeong et al., 2015). In *A. woodii*, a trimeric enzyme-complex system was detected for methyl-THF conversion, consisting of *metF*, *metV*, and *rnfC2* (Bertsch et al., 2015). In the gene cluster, RnfC2 accepts an electron from the reduced form of NADH and then transfers the electron to reduce methylene-THF. However, the MR gene cluster consists of a heterohexameric complex with electron-bifurcating heterodisulfide reductase (*hdrA*, *hdrB*, and *hdrC*), *metV*, and *mvhD* in *M. thermoacetica* (Mock et al., 2014). Additionally, the heterohexameric complex does not catalyze NADH-dependent methylene-THF reduction, but utilizes some form of second-electron acceptor. Although genes of redox enzymes were highly conserved, a configuration of actual enzymatic reactions will be quite different. According to the results of the comparative analysis, only *metV* is absent in *Acetohalobium arabaticum*, and both genes encoding MR are missing in *Treponema primitia*. In other bacteria, *Thermus thermophilus* HB8 and *Escherichia coli* K12 utilize only *metF* to catalyze the methylene tetrahydrofolate reductase reaction (Guenther et al., 1999; Igari et al., 2011). Perhaps the conversion of 5,10-methylenetetrahydrofolate to 5-methyltetrahydrofolate in *Ac. arabaticum* may function as an MR reaction in *Escherichia coli* and *T. thermophilus* containing only *metF*. The *Ac. arabaticum metF* gene consists of methylenetetrahydrofolate reductase and methylene-tetrahydrofolate reductase C-terminal domains and is 663 base pairs longer than the *A. woodii metF* gene. Given the presence of the *metV* domains, the *metF* gene in *Ac. arabaticum* is capable of solely catalyzing MR reactions to reduce methylene-THF. However, alternative pathways for the missing subunits involved in the MR reaction in *Tr. primitia* remain unknown.

The last dispensable gene in core group II is *gcvH*, encoding glycine-cleavage system H protein in the glycine cleavage/synthesis pathway, whose functional role in the Wood-Ljungdahl pathway remains unclear. The glycine cleavage/synthesis pathway consists of four proteins; however, only *gcvH* and *lpdA*, which encodes dihydrolipoamide dehydrogenase, are acetogens. All of the genes encoding this pathway are found in *C. sticklandii* (Fonknechten et al., 2010). Although the genes encoding the complete Wood-Ljungdahl pathway are present in the genome, *C. sticklandii* is unable to utilize CO₂ as a substrate. One proposed hypothesis is that due to the presence of all glycine cleavage/synthesis complexes, an efficient electron acceptor substitutes for the role of CO₂, which leads to shutdown of the methyl-branch of the Wood-Ljungdahl pathway (Fonknechten et al., 2010). Although *lpdA* is conserved in 14 acetogens, *gcvH* is absent in core group II due to the risk of shutting down the Wood-Ljungdahl pathway.

Aside from enzymatic diversity, conserved genes from core group II showed a tendency to co-localize in the genomes (Figure 4B). Although acetogens are phylogenetically diverse, conserved genes encoding FHS or CODH/ACS complexes are co-localized in acetogen genomes (Bruant et al., 2010; Poehlein et al., 2015c). In the least evolutionarily changed *C. difficile* genome (Figure 1), the Wood-Ljungdahl pathway enzymes are located in one gene cluster (Figure 4B), which has been reported (Bruant et al., 2010; Köpke et al., 2013). Although two copies of *lpdA*

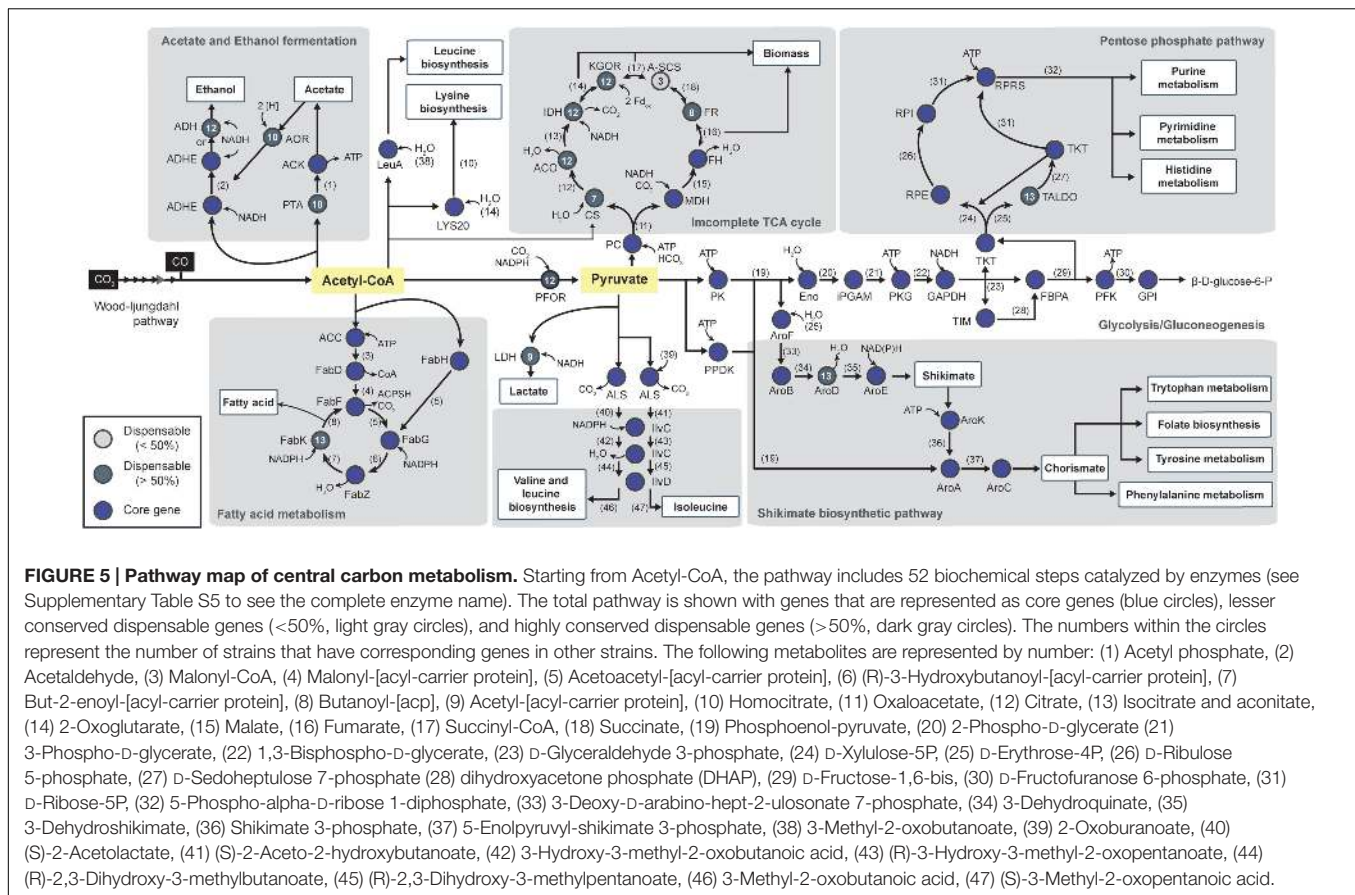
were found, only one copy of each core gene was detected. In all *Clostridium* genera of acetogenic bacteria, the Wood-Ljungdahl pathway gene cluster with the same order of genes was conserved (Figure 4B). Beside the *Clostridium* genera, the methyl- and carbonyl-branch-encoding genes presented as multiple copies. *A. woodii* and *Eubacterium limosum* are phylogenetically related and contain two gene clusters encoding the Wood-Ljungdahl pathway, which is composed of both the methyl and the carbonyl branches. Additionally, duplication of *acsE* explains the rapid growth rate under autotrophic conditions in both strains (Blach et al., 1977; Tschuch and Pfennig, 1984; Sharak Genthner and Bryant, 1987). Interestingly, throughout all 14 acetogens, *acsB*, *acsC*, *acsD*, *acsE*, and *acsF* genes were always located as a gene cluster (Figure 4B). Thus, the highly conserved CODH/ACS complex indicated that the complex functions most efficiently when the genes form a gene cluster. Under such circumstances, gene clusters reflect evolutionary changes in pathways and associated taxonomy, while the phylogenetic tree describes the evolution of acetogenic bacteria.

THE WOOD-LJUNGDAHL PATHWAY CORE GROUP III: ACETYL-CoA TO ACETATE

All acetogens have an ability to produce acetate via acetogenesis as a core feature (Drake et al., 2008). In many acetogenic bacteria, phosphotransacetylase (*pta*) and acetate kinase (*ack*) genes were found as a single operon, similar to that observed in *C. ljungdahlii*, and *C. autoethanogenum* (Köpke et al., 2010; Brown et al., 2014). In the 14 acetogen genomes, the *ack* gene was categorized as a core gene, but the *pta* gene was classified as a dispensable gene. The acetate-production operon, which consisted of the *pta* and *ack* genes, was found in *C. autoethanogenum*, *C. ljungdahlii*, *Clostridium scatologenes*, *Clostridium carboxidivorans*, *Thermoanaerobacter kivui*, *Ca. hydrogeniformans*, and *T. phaeum*. However, in *A. woodii* and *Tr. primitia*, the *ack* and *pta* genes were scattered in the genomes and not located as a gene cluster. Additionally, the *pta* gene was unidentified in four acetogen genomes: *C. difficile*, *C. acetivum*, *E. limosum*, and *M. thermoacetica*. It was suggested that an alternative protein for *pta* is phosphotransbutyrylase (*ptb*; Köpke et al., 2013; Poehlein et al., 2015b) and butyrate kinase (*buk*), which are located on a single operon and can bind to both acetyl-CoA and butyryl-CoA, or propanediol utilization protein (*pduL*), which exhibits transacetylase function (Pierce et al., 2008; Köpke et al., 2010; Poehlein et al., 2015b). In contrast to *pta*, the *ack* gene was found as a single copy and exhibited high similarity in all strains, except *Ac. arabaticum*, which has two *ack* genes.

CENTRAL INTERMEDIATES OF AUTOTROPHIC GROWTH: ACETYL-CoA AND PYRUVATE

As an essential cellular function in all bacteria, biomass and byproducts must be derived from acetyl-CoA. For bacterial



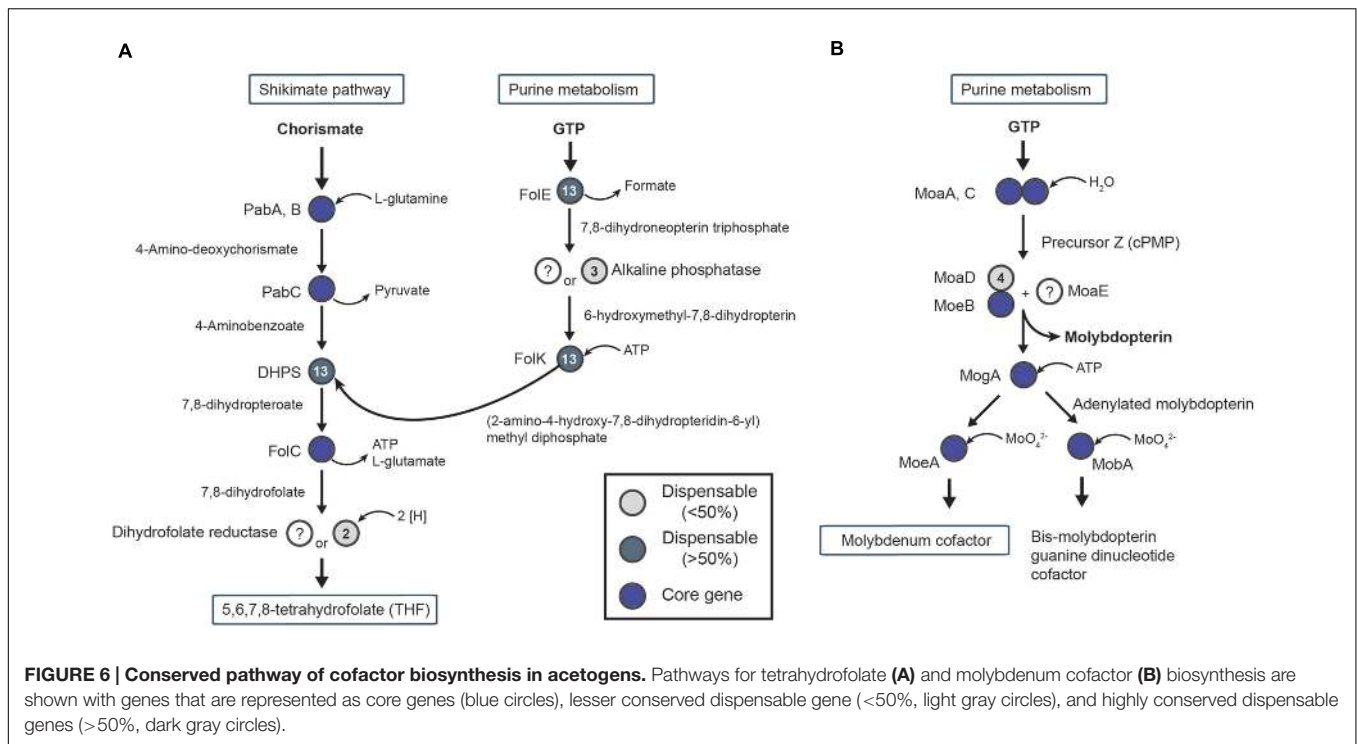
growth under autotrophic conditions, the central precursor can only be synthesized from C_1 compounds via the Wood-Ljungdahl pathway, which plays an important role in cell proliferation. According to a previous study, the proportion of carbon flux toward biomass was predicted as 5% of total carbon flux during autotrophic fermentation (Fast and Papoutsakis, 2012).

Acetate and ethanol are common products generated by acetogenic fermentation, and the production of acetate coupled to ATP synthesis is associated with the Wood-Ljungdahl pathway. Following acetate production, acetate is reduced to acetaldehyde via an aldehyde:ferredoxin oxidoreductase reaction with reduced ferredoxin, and the corresponding gene is categorized as a dispensable gene. Acetyl-CoA can also be converted to acetaldehyde by bifunctional aldehyde/alcohol dehydrogenase (Leang et al., 2013), which was conserved in all 14 acetogens. Additional reduction of acetaldehyde can generate ethanol by the same aldehyde/alcohol dehydrogenase or alcohol dehydrogenase (Figure 5; Supplementary Table S5). Although the alcohol dehydrogenase or aldehyde/alcohol dehydrogenase enzymes responsible for ethanol production are encoded in their genomes, ethanol production was reported in only four strains under autotrophic conditions. Three strains, *C. autoethanogenum* (Köpke et al., 2011), *C. ljungdahlii* (Köpke et al., 2010), and *C. carboxidivorans* (Liou et al., 2005; Bruant et al., 2010), are capable of producing ethanol as the main product, and *C. scatologenes* (Liou et al., 2005) is able to

produce ethanol at low levels. Although genetic mechanisms for ethanol production are present, ethanol production by other strains was not reported under autotrophic conditions. Possible explanations are that these strains lack functional efficiency of the aldehyde:ferredoxin oxidoreductase reaction (putative formaldehyde:Fd oxidoreductase) or presence of bioenergetic constraints (Bertsch and Müller, 2015; Mock et al., 2015).

In addition to alcohol production, acetyl-CoA can be used for fatty acid, leucine, and lysine biosynthesis in one of the most conserved pathways in bacteria. Acetyl-CoA can be utilized directly for fatty acid biosynthesis by seven conserved genes. Although six of the genes were classified as core genes, enoyl-acyl carrier-protein reductase (*fabK*, EC 1.3.1.9) was identified as being dispensable due to its being absent in *Tr. primitia* (Figure 5).

To biosynthesize nucleic acids, amino acids, and essential cofactors, three-carbon pyruvate was used as a central metabolite in several pathways for autotrophic growth (Bar-Even et al., 2012). For this, pyruvate was interconverted from acetyl-CoA by pyruvate:ferredoxin oxidoreductase (Charon et al., 1999). Although highly important, pyruvate:ferredoxin oxidoreductase gene was not classified as a core gene. In the cases of *Ca. hydrogeniformans* Z-2901 and *T. phaeum* DSM 12270, the pyruvate:ferredoxin oxidoreductase gene was not identified in the genomes. For the alternate reaction, formate C-acetyltransferase gene (pyruvate formate lyase, *tph_c09600* and *CHY_0877*)



present in the genome can be utilized for converting one acetyl-CoA with one formate to one pyruvate (Oehler et al., 2012).

To supply carbon skeletons, pyruvate reacts through reductive or oxidative branches of the incomplete tricarboxylic acid cycle, similar to most anaerobic bacteria. Specifically, the reductive branch was highly conserved throughout the acetogens (Figure 5). Initially, oxaloacetate, which is derived from pyruvate, was converted to fumarate via the reductive branch. Following this reaction, fumarate reductase, which was conserved in eight strains, synthesizes succinate from fumarate. However, all genes encoding the oxidative branch were classified as dispensable genes. The citrate synthase gene was located in only seven strains (Figure 5; Supplementary Table S5), while other enzymes, such as isocitrate dehydrogenase and 2-oxoglutarate synthase, were conserved, except in *Tr. primitia*, *Th. kivui*, *C. ljungdahlii*, and *C. autoethanogenum*. Among the acetogens, the least conserved enzyme associated with the tricarboxylic acid cycle was succinyl-CoA synthetase. In all acetogens, succinyl-CoA synthetases were located with the incomplete tricarboxylic acid cycles, which were composed of formations, with one direction leading to the formation of 2-oxoglutarate or succinyl-CoA from citrate and the other direction leading to the formation of fumarate or succinate from acetyl-CoA.

Central metabolic pathways, such as the glycolysis pathway, the pentose phosphate pathway, and the shikimate biosynthetic pathway, were highly conserved in all acetogens for nucleotide and amino acid biosynthesis (Figure 5). To produce the pentose phosphate for RNA and DNA precursors, the pentose phosphate pathway and gluconeogenesis must be utilized with related core genes. The shikimate pathway was also used in early steps for biosynthetic production of cofactors

(folate), electron-transfer components (quinones), and aromatic amino acids (phenylalanine, tryptan, and tryptophan). All parts of these pathways were conserved, except for *aroD* genes, which were absent in the *Tr. primitia* genome (Figure 5; Supplementary Table S5). For the production of valine, leucine, and isoleucine from acetyl-CoA, acetolactate synthase, ketol-acid reductoisomerase (IlvC), and dihydroxy-acid dehydratase (IlvD) are required, which were conserved in all 14 acetogens (Figure 5). Following acetyl-CoA conversion, these conserved enzymes convert pyruvate into branched-chain amino acids.

COFACTOR BIOSYNTHETIC PATHWAYS

Several enzyme-cofactor interactions are heavily involved in the Wood-Ljungdahl pathway, including THF, corrinoid iron-sulfur protein, and molybdopterin cofactor, which play key roles in one-carbon transfer for synthesizing acetyl-CoA from CO₂/H₂ (Drake, 1994; Ragsdale, 2008; Ragsdale and Pierce, 2008). Under the circumstances, genes encoding enzymes involved in the biosynthesis of cofactors should be present in the genome for pure cultures of CO/CO₂-dependent chemolithotrophs without supplementation of the required cofactors.

First, THF is important for the transformation of methyl-tetrahydrofolate following reduction of CO₂. For THF synthesis, the *de novo* synthesis pathway begins with chorismate and guanosine triphosphate from the shikimate pathway and purine metabolism, respectively. All required genes were present in the core-gene set, except for two genes (Figure 6A): dihydrofolate reductase (DHR) and alkaline phosphate. Specifically, DHR was missing in most of the acetogens. A possible alternative enzyme

is an oxygen-insensitive nitroreductase (Tph_c13060) for DHR (Oehler et al., 2012). The nitroreductase genes are core genes in acetogens, and studies of oxygen-insensitive nitroreductase reported evidence of DHR activity (Vasudevan et al., 1992).

In the steps of formate synthesis, selenocysteine FDH requires the molybdopterin cofactor to catalyze the reduction of CO₂ to formate (Ragsdale and Pierce, 2008). The biosynthetic pathway associated with the molybdopterin cofactor is shown in **Figure 6B**. The first steps, catalyzed by MoaA and MoaC, use guanosine triphosphate to synthesize the precursor Z, followed by molybdopterin synthesis by MoaD, MoeB, and MoaE (**Figure 6B**). Interestingly, the gene encoding MoaE was not reported in any acetogens, including *M. thermoacetica* (Pierce et al., 2008). A predicted alternative enzyme is cysteine desulfurase (EC 2.8.1.7), which was located in all 14 acetogen genomes and uses a sulfur donor, such as MoaD, for molybdopterin synthesis (Mihara et al., 2002).

Cobalamin is a central cofactor in the Wood-Ljungdahl pathway, given that acetyl-CoA synthase reactions are cobalamin dependent. Although pathways for cobalamin biosynthesis were reported in *M. thermoacetica* (Pierce et al., 2008), the pathway has not been fully elucidated. The genes encoding cobalamin biosynthesis are located as a large gene cluster in the genome (Köpke et al., 2010; Oehler et al., 2012; Poehlein et al., 2012). Two distinct cobalamin-biosynthesis pathways were reported as an anaerobic and an aerobic pathway (Rodionov et al., 2003). Comparative genome analysis indicated that the aerobic pathway was absent in all acetogen genomes; however, the *cobJ*, *cobM*, *cobH*, and *cobB* genes were highly conserved. Nevertheless, the anaerobic cobalt-insertion pathway was conserved in six strains (*A. woodii*, *E. limosum*, *C. autoethanogenum*, *C. ljungdahlii*, *C. scatologenes*, and *Th. kivui*). Previously, the ability to produce vitamin B₁₂ under autotrophic or methylotrophic conditions was evaluated in two strains (Stupperich et al., 1988; Lebloas et al., 1994). However, sirohydrochlorin cobaltochelate (*cbiK*) and precorrin-3 synthase (*cbiL*) genes were missing in two strains (*C. acetivum* and *C. difficile*). In the case of the others, two more genes were missing from the anaerobic cobalt-insertion pathway (Oehler et al., 2012). Such genes only found in individual strains may exist due to the dependency on vitamin B₁₂ during autotrophic growth.

PERSPECTIVES AND CONCLUSION

Acetogens inhabit diverse environments, temperatures, and pH conditions (Drake et al., 2006). Correspondingly, the genomes of acetogens comprise highly diverse metabolic and energy conservation systems (Schuchmann and Müller, 2014; Poehlein et al., 2015b). For example, an F₀F₁-type ATP synthase, a conserved energy generating component, was conserved with seven subunits in 13 strains, except for *E. limosum* (Supplementary Table S5). However, ion specificity for gradient-driven phosphorylation is quite different between the strains due to the sequence motif present in the gamma subunit (Krah et al., 2010). Normally, the gamma subunit binds H⁺ at a site between the carboxyl oxygen of a carboxylate and a backbone carbonyl

of another amino acid (Pogoryelov et al., 2009). For Na⁺, four amino acid residues are conserved: Gln32, Val63, Ser66, and Thr 67 (Murata et al., 2005). Although subunit α and β were well conserved with high similarity, the ion-binding subunit gamma was diverse, with relatively low similarity throughout the acetogens, possibly due to the variations in environmental conditions.

Despite this genetic diversity, the Wood-Ljungdahl pathway, a central metabolic pathway, and cofactor-biosynthetic pathways are highly conserved to promote autotrophic growth. Together, these data and previously reported results (Becerra et al., 2014) suggested that the ability to perform acetogenesis was obtained by genetic transfer of core genes associated with the Wood-Ljungdahl pathway and remains interconnected with its own inherent metabolic and energy conservation systems. Similarly, gene-set enrichment analysis revealed that acetogens do not share special gene sets, with the exception of the Wood-Ljungdahl pathway and *fwdE*.

Additionally, we predicted missing enzymes and suggested possible alternative enzymes based on the information from each genome. This information can aid in understanding the basic model of acetogens. Although we predicted the conserved pathways associated with individual strains, several key pathways remain unclear and require biochemical confirmation. Furthermore, the mechanisms involved in chemolithoautotrophic growth, systematic energy conservation, and precisely regulating carbon and energy flux also remain unknown. Also, the reconstruction of genome-scale models will be also required for the prediction of phenotypes and biosynthesis of value-added products of interest from syngas. In order for this to happen, the small differences found in conserved and alternative biochemical pathways can be used to optimize the genetic network to efficiently utilize the optimal enzymes or to convert optimal non-acetogenic microorganisms into novel acetogens.

AUTHOR CONTRIBUTIONS

JS and B-KC conceived and designed the analyses. JS, YS, and YJ performed the analyses. JS and B-KC wrote the paper. All authors approved the final manuscript.

FUNDING

This work was supported by the Intelligent Synthetic Biology Center of Global Frontier Project 2011-0031957 of the National Research Foundation of Korea (NRF), funded by the Ministry of Science, ICT, and Future Planning.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fmicb.2016.01531>

REFERENCES

- Alexeyenko, A., Tamas, I., Liu, G., and Sonnhammer, E. L. L. (2006). Automatic clustering of orthologs and inparalogs shared by multiple proteomes. *Bioinformatics* 22, e9–e15. doi: 10.1093/bioinformatics/btl213
- Bar-Even, A., Noor, E., and Milo, R. (2012). A survey of carbon fixation pathways through a quantitative lens. *J. Exp. Bot.* 63, 2325–2342. doi: 10.1093/jxb/err417
- Becerra, A., Rivas, M., García-Ferris, C., Lazcano, A., and Peretó, J. (2014). A phylogenetic approach to the early evolution of autotrophy: the case of the reverse TCA and the reductive acetyl-CoA pathways. *Int. Microbiol.* 17, 91–97. doi: 10.2436/20.1501.01.211
- Bengelsdorf, F. R., Poehlein, A., Esser, C., Schiel-Bengelsdorf, B., Daniel, R., and Dürre, P. (2015). Complete genome sequence of the acetogenic bacterium *Moorella thermoacetica* DSM 2955T. *Genome Announc.* 3, e1157-15. doi: 10.1128/genomeA.01157-15
- Bengelsdorf, F. R., Straub, M., and Dürre, P. (2013). Bacterial synthesis gas (syngas) fermentation. *Environ. Technol.* 34, 1639–1651. doi: 10.1080/09593330.2013.827747
- Bertsch, J., and Müller, V. (2015). Bioenergetic constraints for conversion of syngas to biofuels in acetogenic bacteria. *Biotechnol. Biofuels* 8:210. doi: 10.1186/s13068-015-0393-x
- Bertsch, J., Öppinger, C., Hess, V., Langer, J. D., and Müller, V. (2015). Heterotrimeric NADH-oxidizing methylenetetrahydrofolate reductase from the acetogenic bacterium *Acetobacterium woodii*. *J. Bacteriol.* 197, 1681–1689. doi: 10.1128/JB.00048-15
- Blach, W. E., Schoberth, S., Tanner, R. S., and Wolfe, R. S. (1977). *Acetobacterium*, a new genus of hydrogen-oxidizing, carbon dioxide-reducing, anaerobic bacteria. *Int. J. Sys. Bacteriol.* 27, 355–361. doi: 10.1099/00207713-27-4-355
- Brown, S. D., Nagaraju, S., Utturkar, S., De Tissera, S., Segovia, S., Mitchell, W., et al. (2014). Comparison of single-molecule sequencing and hybrid approaches for finishing the genome of *Clostridium autoethanogenum* and analysis of CRISPR systems in industrial relevant Clostridia. *Biotechnol. Biofuels* 7:40. doi: 10.1186/1754-6834-7-40
- Bruant, G., Lévesque, M.-J., Peter, C., Guiot, S. R., and Masson, L. (2010). Genomic analysis of carbon monoxide utilization and butanol production by *Clostridium carboxidivorans* strain P7. *PLoS ONE* 5:e13033. doi: 10.1371/journal.pone.0013033
- Charon, M.-H., Volbeda, A., Chabriere, E., Pieulle, L., and Fontecilla-Camps, J. C. (1999). Structure and electron transfer mechanism of pyruvate:ferredoxin oxidoreductase. *Curr. Opin. Struct. Biol.* 9, 663–669. doi: 10.1016/S0959-440X(99)00027-5
- Clark, J. E., and Ljungdahl, L. G. (1984). Purification and properties of 5,10-methylenetetrahydrofolate reductase, an iron-sulfur flavoprotein from *Clostridium formicoaceticum*. *J. Biol. Chem.* 259, 10845–10849.
- Drake, H. L. (1994). “Acetogenesis, acetogenic bacteria, and the Acetyl-CoA “Wood/Ljungdahl” pathway: past and current perspectives,” in *Acetogenesis, Chapman and Hall*, eds H. I. Drake (New York, NY: Springer), 3–60.
- Drake, H. L., and Daniel, S. L. (2004). Physiology of the thermophilic acetogen *Moorella thermoacetica*. *Res. Microbiol.* 155, 869–883. doi: 10.1016/j.resmic.2004.10.001
- Drake, H. L., Gößner, A. S., and Daniel, S. L. (2008). Old acetogens, new light. *Ann. N. Y. Acad. Sci.* 1125, 100–128. doi: 10.1196/annals.1419.016
- Drake, H. L., Küsel, K., and Matthies, C. (2006). “Acetogenic prokaryotes,” in *The Prokaryotes - Prokaryotic Physiology and Biochemistry*, eds E. Rosenberg, E. F. DeLong, S. Lory, E. Stackebrandt, and F. Thompson (New York, NY: Springer), 354–420.
- Enright, A. J., Van Dongen, S., and Ouzounis, C. A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res.* 30, 1575–1584. doi: 10.1093/nar/30.7.1575
- Fast, A. G., and Papoutsakis, E. T. (2012). Stoichiometric and energetic analyses of non-photosynthetic CO₂-fixation pathways to support synthetic biology strategies for production of fuels and chemicals. *Curr. Opin. Chem. Eng.* 1, 380–395. doi: 10.1016/j.coche.2012.07.005
- Fonknechten, N., Chaussonnerie, S., Tricot, S., Lajus, A., Andreesen, J. R., Perchat, N., et al. (2010). *Clostridium sticklandii*, a specialist in amino acid degradation: revisiting its metabolism through its genome sequence. *BMC Genomics* 11:555. doi: 10.1186/1471-2164-11-555
- Gil, R., Silva, F. J., Peretó, J., and Moya, A. (2004). Determination of the core of a minimal bacterial gene set. *Microbiol. Mol. Biol. Rev.* 68, 518–537. doi: 10.1128/MMBR.68.3.518-537.2004
- Gollin, D., Li, X. L., Liu, S. M., Davies, E. T., and Ljungdahl, L. G. (1998). “Acetogenesis and the primary structure of the NADP-dependent formate dehydrogenase of *Clostridium thermoaceticum*, a tungsten-selenium-iron protein,” in *Proceedings of the Fourth International Conference on Carbon Dioxide Utilization, Advances in Chemical Conversions for Mitigating Carbon Dioxide* (Amsterdam: Elsevier), 303–308.
- Guenther, B. D., Sheppard, C. A., Tran, P., Rozen, R., Matthews, R. G., and Ludwig, M. L. (1999). The structure and properties of methylenetetrahydrofolate reductase from *Escherichia coli* suggest how folate ameliorates human hyperhomocysteinemia. *Nat. Struct. Biol.* 6, 359–365. doi: 10.1038/7594
- Haddad, M., Cimpoaia, R., Zhao, Y., and Guiot, S. R. (2013). Growth profile of *Carboxydotherrmus hydrogenoformans* on pyruvate. *AMB Express* 3, 60. doi: 10.1186/2191-0855-3-60
- Hattori, S., Galushko, A. S., Kamagata, Y., and Schink, B. (2005). Operation of the CO dehydrogenase/acetyl coenzyme A pathway in both acetate oxidation and acetate formation by the syntrophically acetate-oxidizing bacterium *Thermacetogenium phaeum*. *J. Bacteriol.* 187, 3471–3476. doi: 10.1128/JB.187.10.3471-3476.2005
- Hattori, S., Kamagata, Y., Hanada, S., and Shoun, H. (2000). *Thermacetogenium phaeum* gen. nov., sp. nov., a strictly anaerobic, thermophilic, syntrophic acetate-oxidizing bacterium. *Int. J. Syst. Evol. Microbiol.* 50, 1601–1609. doi: 10.1099/00207713-50-4-1601
- Hayashi, T., Makino, K., Ohnishi, M., Kurokawa, K., Ishii, K., Yokoyama, K., et al. (2001). Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genomic comparison with a laboratory strain K-12. *DNA Res.* 8, 11–22. doi: 10.1093/dnares/8.1.147
- He, M., Sebahia, M., Lawley, T. D., Stabler, R. A., Dawson, L. F., Martin, M. J., et al. (2010). Evolutionary dynamics of *Clostridium difficile* over short and long time scales. *Proc. Natl. Acad. Sci. U.S.A.* 107, 7527–7532. doi: 10.1073/pnas.0914322107
- Henstra, A. M., Sipma, J., Rinzema, A., and Stams, A. J. (2007). Microbiology of synthesis gas fermentation for biofuel production. *Curr. Opin. Biotechnol.* 18, 200–206. doi: 10.1016/j.copbio.2007.03.008
- Henstra, A. M., and Stams, A. J. M. (2011). Deep conversion of carbon monoxide to hydrogen and formation of acetate by the anaerobic thermophile *Carboxydotherrmus hydrogenoformans*. *Int. J. Microbiol.* 2011, 641582–641584. doi: 10.1155/2011/641582
- Hess, V., Poehlein, A., Weghoff, M. C., Daniel, R., and Müller, V. (2014). A genome-guided analysis of energy conservation in the thermophilic, cytochrome-free acetogenic bacterium *Thermoanaerobacter kivui*. *BMC Genomics* 15:1139. doi: 10.1186/1471-2164-15-1139
- Hochheimer, A., Hedderich, R., and Thauer, R. K. (1998). The formylmethanofuran dehydrogenase isoenzymes in *Methanobacterium wolfei* and *Methanobacterium thermoautotrophicum*: induction of the molybdenum isoenzyme by molybdate and constitutive synthesis of the tungsten isoenzyme. *Arch. Microbiol.* 170, 389–393. doi: 10.1007/s002030050658
- Huang, H., Wang, S., Moll, J., and Thauer, R. K. (2012). Electron bifurcation involved in the energy metabolism of the acetogenic bacterium *Moorella thermoacetica* growing on glucose or H₂ plus CO₂. *J. Bacteriol.* 194, 3689–3699. doi: 10.1128/JB.00385-12
- Humphreys, C. M., McLean, S., Schatschneider, S., Millat, T., Henstra, A. M., Annan, F. J., et al. (2015). Whole genome sequence and manual annotation of *Clostridium autoethanogenum*, an industrially relevant bacterium. *BMC Genomics* 16:1085. doi: 10.1186/s12864-015-2287-5
- Igari, S., Ohtaki, A., Yamanaka, Y., Sato, Y., Yohda, M., Odaka, M., et al. (2011). Properties and crystal structure of methylenetetrahydrofolate reductase from *Thermus thermophilus* HB8. *PLoS ONE* 6:e23716. doi: 10.1371/journal.pone.0023716
- Islam, M. A., Zengler, K., Edwards, E. A., Mahadevan, R., and Stephanopoulos, G. (2015). Investigating *Moorella thermoacetica* metabolism with a genome-scale constraint-based metabolic model. *Integr. Biol.* 7, 869–882. doi: 10.1039/c5ib00095e
- Jeong, J., Bertsch, J., Hess, V., Choi, S., Choi, I.-G., Chang, I. S., et al. (2015). Energy conservation model based on genomic and experimental analyses of a

- carbon monoxide-utilizing, butyrate-forming acetogen, *Eubacterium limosum* KIST612. *Appl. Environ. Microbiol.* 81, 4782–4790. doi: 10.1128/AEM.00675-15
- Jiang, B., Henstra, A.-M., Paulo, P. L., Balk, M., van Doesburg, W., and Stams, A. J. M. (2009). Atypical one-carbon metabolism of an acetogenic and hydrogenogenic *Moorella thermoacetica* strain. *Arch. Microbiol.* 191, 123–131. doi: 10.1007/s00203-008-0435-x
- Kelley, B. P., Sharan, R., Karp, R. M., Sittler, T., Root, D. E., Stockwell, B. R., et al. (2003). Conserved pathways within bacteria and yeast as revealed by global protein network alignment. *Proc. Natl. Acad. Sci. U.S.A.* 100, 11394–11399. doi: 10.1073/pnas.1534710100
- Kelly, W. J., Henderson, G., Pacheco, D. M., Li, D., Reilly, K., Naylor, G. E., et al. (2016). The complete genome sequence of *Eubacterium limosum* SA11, a metabolically versatile rumen acetogen. *Stand. Genomic Sci.* 11:26. doi: 10.1186/s40793-016-0147-9
- Köpke, M., Held, C., Hujer, S., Liesegang, H., Wiezer, A., Wollherr, A., et al. (2010). *Clostridium ljungdahlii* represents a microbial production platform based on syngas. *Proc. Natl. Acad. Sci. U.S.A.* 107, 13087–13092. doi: 10.1073/pnas.1004716107
- Köpke, M., Mihalcea, C., Liew, F., Tizard, J. H., Ali, M. S., Conolly, J. J., et al. (2011). 2,3-butanediol production by acetogenic bacteria, an alternative route to chemical synthesis, using industrial waste gas. *Appl. Environ. Microbiol.* 77, 5467–5475. doi: 10.1128/AEM.00355-11
- Köpke, M., Straub, M., and Dürre, P. (2013). *Clostridium difficile* is an autotrophic bacterial pathogen. *PLoS ONE* 8:e62157. doi: 10.1371/journal.pone.0062157
- Krah, A., Pogoryelov, D., Langer, J. D., Bond, P. J., Meier, T., and Faraldo-Gómez, J. D. (2010). Structural and energetic basis for H⁺ versus Na⁺ binding selectivity in ATP synthase Fo rotors. *Biochim. Biophys. Acta Bioenergetics* 1797, 763–772. doi: 10.1016/j.bbabi.2010.04.014
- Latif, H., Zeidan, A. A., Nielsen, A. T., and Zengler, K. (2014). Trash to treasure: production of biofuels and commodity chemicals via syngas fermenting microorganisms. *Curr. Opin. Biotechnol.* 27, 79–87. doi: 10.1016/j.copbio.2013.12.001
- Leang, C., Ueki, T., Nevin, K. P., and Lovley, D. R. (2013). A genetic system for *Clostridium ljungdahlii*: a chassis for autotrophic production of biocommodities and a model homoacetogen. *Appl. Environ. Microbiol.* 79, 1102–1109. doi: 10.1128/AEM.02891-12
- Lebloas, P., Loubiere, P., and Lindley, N. D. (1994). Use of unicarbon substrate mixtures to modify carbon flux improves vitamin B12 production with the acetogenic methylotroph *Eubacterium limosum*. *Biotechnol. Lett.* 16, 129–132. doi: 10.1007/BF01021658
- Li, N., Yang, J., Chai, C., Yang, S., Jiang, W., and Gu, Y. (2015). Complete genome sequence of *Clostridium carboxidivorans* P7(T), a syngas-fermenting bacterium capable of producing long-chain alcohols. *J. Biotechnol.* 211, 44–45. doi: 10.1016/j.jbiotec.2015.06.430
- Liou, J. S.-C., Balkwill, D. L., Drake, G. R., and Tanner, R. S. (2005). *Clostridium carboxidivorans* sp. nov., a solvent-producing clostridium isolated from an agricultural settling lagoon, and reclassification of the acetogen *Clostridium scatologenes* strain SL1 as *Clostridium drakei* sp. nov. *Int. J. Syst. Evol. Microbiol.* 55, 2085–2091. doi: 10.1099/ijms.0.63482-0
- Ljungdahl, L. G., and Andreesen, J. R. (1978). Formate dehydrogenase, a selenium-tungsten enzyme from *Clostridium thermoacetum*. *Methods Enzymol.* 53, 360–372.
- Marcellin, E., Behrendorff, J. B., Nagaraju, S., DeTissera, S., Segovia, S., Palfreyman, R. W., et al. (2016). Low carbon fuels and commodity chemicals from waste gases – systematic approach to understand energy metabolism in a model acetogen. *Green Chem.* 18, 3020–3028. doi: 10.1039/C5GC02708J
- Matson, E. G., Zhang, X., and Leadbetter, J. R. (2010). Selenium controls transcription of paralogous formate dehydrogenase genes in the termite gut acetogen, *Treponema primitia*. *Environ. Microbiol.* 12, 2245–2258. doi: 10.1111/j.1462-2920.2010.02188.x
- McInerney, M. J., and Bryant, M. P. (1981). “Basic principles of bioconversions in anaerobic digestion and methanogenesis,” in *Biomass Conversion Processes for Energy and Fuels*, eds S. S. Sofer and O. R. Zaborsky (New York, NY: Springer).
- Mihara, H., Kato, S. I., Lacourciere, G. M., Stadtman, T. C., Kennedy, R. A. J. D., Kurihara, T., et al. (2002). The *iscS* gene is essential for the biosynthesis of 2-selenouridine in tRNA and the selenocysteine-containing formate dehydrogenase H. *Proc. Natl. Acad. Sci. U.S.A.* 99, 6679–6683. doi: 10.1073/pnas.102176099
- Mock, J., Wang, S., Huang, H., Kahnt, J., and Thauer, R. K. (2014). Evidence for a hexaheteromeric methylenetetrahydrofolate reductase in *Moorella thermoacetica*. *J. Bacteriol.* 196, 3303–3314. doi: 10.1128/JB.01839-14
- Mock, J., Zheng, Y., Mueller, A. P., Ly, S., Tran, L., Segovia, S., et al. (2015). Energy conservation associated with ethanol formation from H₂ and CO₂ in *Clostridium autoethanogenum* involving electron bifurcation. *J. Bacteriol.* 197, 2965–2980. doi: 10.1128/JB.00399-15
- Murata, T., Yamato, I., Kakinuma, Y., Leslie, A. G. W., and Walker, J. E. (2005). Structure of the rotor of the V-Type Na⁺-ATPase from *Enterococcus hirae*. *Science* 308, 654–659. doi: 10.1126/science.1110064
- Nagarajan, H., Sahin, M., Nogales, J., Latif, H., Lovley, D. R., Ebrahim, A., et al. (2013). Characterizing acetogenic metabolism using a genome-scale metabolic reconstruction of *Clostridium ljungdahlii*. *Microb. Cell Fact.* 12:118. doi: 10.1186/1475-2859-12-118
- Naik, S. N., Goud, V. V., Rout, P. K., and Dalai, A. K. (2010). Production of first and second generation biofuels: a comprehensive review. *Renew. Sust. Energ. Rev.* 14, 578–597. doi: 10.1016/j.rser.2009.10.003
- O'Brien, W. E., Brewer, J. M., and Ljungdahl, L. G. (1973). Purification and characterization of thermostable 5,10-methylenetetrahydrofolate dehydrogenase from *Clostridium thermoacetum*. *J. Biol. Chem.* 248, 403–408.
- Oehler, D., Poehlein, A., Leimbach, A., Müller, N., Daniel, R., Gottschalk, G., et al. (2012). Genome-guided analysis of physiological and morphological traits of the fermentative acetate oxidizer *Thermacetogenium phaeum*. *BMC Genomics* 13:723. doi: 10.1186/1471-2164-13-723
- Ohnishi, M., Kurokawa, K., and Hayashi, T. (2001). Diversification of *Escherichia coli* genomes: are bacteriophages the major contributors? *Trends Microbiol.* 9, 481–485. doi: 10.1016/S0966-842X(01)02173-4
- Park, E. Y., Clark, J. E., DerVartanian, D. V., and Ljungdahl, L. G. (1991). “5,10-methylenetetrahydrofolate reductases: iron-sulfur-zinc flavoproteins of two acetogenic clostridia,” in *Chemistry and Biochemistry of Flavoenzymes*, Vol. 1, ed. F. Miller (Boca Raton, FL: CRC Press), 389–400.
- Pierce, E., Xie, G., Barabote, R. D., Saunders, E., Han, C. S., Detter, J. C., et al. (2008). The complete genome sequence of *Moorella thermoacetica* (f. *Clostridium thermoacetum*). *Environ. Microbiol.* 10, 2550–2573. doi: 10.1111/j.1462-2920.2008.01679.x
- Poehlein, A., Bengelsdorf, F. R., Esser, C., Schiel-Bengelsdorf, B., Daniel, R., and Dürre, P. (2015a). Complete genome sequence of the type strain of the acetogenic bacterium *Moorella thermoacetica* DSM 521T. *Genome Announc.* 3:e1159-15. doi: 10.1128/genomeA.01159-15
- Poehlein, A., Bengelsdorf, F. R., Schiel-Bengelsdorf, B., Gottschalk, G., Daniel, R., and Dürre, P. (2015b). Complete genome sequence of *rnf*- and cytochrome-containing autotrophic acetogen *Clostridium acetatum* DSM 1496. *Genome Announc.* 3:e786-15. doi: 10.1128/genomeA.00786-15
- Poehlein, A., Cebulla, M., Ilg, M. M., Bengelsdorf, F. R., Schiel-Bengelsdorf, B., Whited, G., et al. (2015c). The complete genome sequence of *Clostridium acetatum*: a missing link between *rnf*- and cytochrome-containing autotrophic acetogens. *mBio* 6:e1168-15. doi: 10.1128/mBio.01168-15
- Poehlein, A., Schmidt, S., Kaster, A.-K., Goenrich, M., Vollmers, J., Thürmer, A., et al. (2012). An ancient pathway combining carbon dioxide fixation with the generation and utilization of a sodium ion gradient for ATP synthesis. *PLoS ONE* 7:e33439. doi: 10.1371/journal.pone.0033439
- Pogoryelov, D., Yildiz, O., Faraldo-Gómez, J. D., and Meier, T. (2009). High-resolution structure of the rotor ring of a proton-dependent ATP synthase. *Nat. Struct. Mol. Biol.* 16, 1068–1073. doi: 10.1038/nsmb.1678
- Ragsdale, S. W. (2008). Enzymology of the wood-Ljungdahl pathway of acetogenesis. *Ann. N. Y. Acad. Sci.* 1125, 129–136. doi: 10.1196/annals.1419.015
- Ragsdale, S. W., and Pierce, E. (2008). Acetogenesis and the Wood-Ljungdahl pathway of CO₂ fixation. *Biochim. Biophys. Acta* 1784, 1873–1898. doi: 10.1016/j.bbapap.2008.08.012
- Riedel, T., Bunk, B., Thürmer, A., Spröer, C., Brzuszkiewicz, E., Abt, B., et al. (2015). Genome resequencing of the virulent and multidrug-resistant reference strain *Clostridium difficile* 630. *Genome Announc.* 3, e276-15. doi: 10.1128/genomeA.00276-15
- Rodionov, D. A., Vitreschak, A. G., Mironov, A. A., and Gelfand, M. S. (2003). Comparative genomics of the vitamin B12 metabolism and regulation in prokaryotes. *J. Biol. Chem.* 278, 41148–41159. doi: 10.1074/jbc.M305837200

- Roh, H., Ko, H.-J., Kim, D., Choi, D. G., Park, S., Kim, S., et al. (2011). Complete genome sequence of a carbon monoxide-utilizing acetogen, *Eubacterium limosum* KIST612. *J. Bacteriol.* 193, 307–308. doi: 10.1128/JB.01217-10
- Rosenthal, A. Z., Matson, E. G., Eldar, A., and Leadbetter, J. R. (2011). RNA-seq reveals cooperative metabolic interactions between two termite-gut spirochete species in co-culture. *ISME J.* 5, 1133–1142. doi: 10.1038/ismej.2011.3
- Saeidi, S., Amin, N. A. S., and Rahimpour, M. R. (2014). Hydrogenation of CO₂ to value-added products—a review and potential future developments. *J. CO₂ Util.* 5, 66–81. doi: 10.1016/j.jcou.2013.12.005
- Saito, R., Smoot, M. E., Ono, K., Ruscheinski, J., Wang, P.-L., Lotia, S., et al. (2012). A travel guide to Cytoscape plugins. *Nat. Methods* 9, 1069–1076. doi: 10.1038/nmeth.2212
- Schiel-Bengelsdorf, B., and Dürre, P. (2012). Pathway engineering and synthetic biology using acetogens. *FEBS Lett.* 586, 1–8. doi: 10.1016/j.febslet.2012.04.043
- Schuchmann, K., and Müller, V. (2012). A bacterial electron-bifurcating hydrogenase. *J. Biol. Chem.* 287, 31165–31171. doi: 10.1074/jbc.M112.395038
- Schuchmann, K., and Müller, V. (2013). Direct and reversible hydrogenation of CO₂ to formate by a bacterial carbon dioxide reductase. *Science* 342, 1382–1385. doi: 10.1126/science.1244758
- Schuchmann, K., and Müller, V. (2014). Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. *Nat. Rev. Microbiol.* 12, 809–821. doi: 10.1038/nrmicro3365
- Sebahia, M., Wren, B. W., Mullany, P., Fairweather, N. F., Minton, N., Stabler, R., et al. (2006). The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat. Genet.* 38, 779–786. doi: 10.1038/ng1830
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Sharak Genthner, B. R., and Bryant, M. P. (1987). Additional characteristics of one-carbon-compound utilization by *Eubacterium limosum* and *Acetobacterium woodii*. *Appl. Environ. Microb.* 53, 471–476.
- Sikorski, J., Lapidus, A., Chertkov, O., Lucas, S., Copeland, A., Glavina Del, et al. (2010). Complete genome sequence of *Acetohalobium arabaticum* type strain (Z-7288). *Stand Genomic Sci.* 3, 57–65. doi: 10.4056/sigs.1062906
- Spigarelli, B. P., and Kawatra, S. K. (2013). Opportunities and challenges in carbon dioxide capture. *J. CO₂ Util.* 1, 69–87. doi: 10.1016/j.jcou.2013.03.002
- Stabler, R. A., He, M., Dawson, L., Martin, M., Valiente, E., Corton, C., et al. (2009). Comparative genome and phenotypic analysis of *Clostridium difficile* 027 strains provides insight into the evolution of a hypervirulent bacterium. *Genome Biol.* 10:R102. doi: 10.1186/gb-2009-10-9-r102
- Stadtman, T. C. (1991). Biosynthesis and function of selenocysteine-containing enzymes. *J. Biol. Chem.* 266, 16257–16260.
- Stupperich, E., Eisinger, H. J., and Krautler, B. (1988). Diversity of corrinoids in acetogenic bacteria. P-Cresolylcobamide from *Sporomusa ovata*, 5-methoxy-6-methylbenzimidazolylcobamide from *Clostridium formicoaceticum* and vitamin B12 from *Acetobacterium woodii*. *Eur. J. Biochem.* 172, 459–464. doi: 10.1111/j.1432-1033.1988.tb13910.x
- Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., et al. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial pan-genome. *Proc. Natl. Acad. Sci. U.S.A.* 102, 13950–13955. doi: 10.1073/pnas.0506758102
- Tschech, A., and Pfennig, N. (1984). Growth yield increase linked to caffeate reduction in *Acetobacterium woodii*. *Arch. Microbiol.* 137, 163–167. doi: 10.1007/BF00414460
- van Eijk, E., Anvar, S. Y., Browne, H. P., Leung, W. Y., Frank, J., Schmitz, A. M., et al. (2015). Complete genome sequence of the *Clostridium difficile* laboratory strain 630Δerm reveals differences from strain 630, including translocation of the mobile element CTn5. *BMC Genomics* 16:31. doi: 10.1186/s12864-015-1252-7
- Vasudevan, S. G., Paal, B., and Armarego, W. L. F. (1992). Dihydropteridine reductase from *Escherichia coli* exhibits dihydrofolate reductase activity. *Biol. Chem. H-S* 373, 1067–1074.
- Wang, S., Huang, H., Kahnt, J., Mueller, A. P., Kopke, M., and Thauer, R. K. (2013). NADP-specific electron-bifurcating [FeFe]-Hydrogenase in a functional complex with formate dehydrogenase in *Clostridium autoethanogenum* grown on CO. *J. Bacteriol.* 195, 4373–4386. doi: 10.1128/JB.00678-13
- Wood, H. G., and Ljungdahl, L. G. (1991). “Autotrophic character of acetogenic bacteria,” in *Variations in Autotrophic Life*, eds L. L. B. Jessup and M. Shively (San Diego, CA: Academic Press).
- Wu, M., Ren, Q., Durkin, A. S., Daugherty, S. C., Brinkac, L. M., Dodson, R. J., et al. (2005). Life in hot carbon monoxide: the complete genome sequence of *Carboxydotherrmus hydrogenoformans* Z-2901. *PLoS Genet.* 1:e65. doi: 10.1371/journal.pgen.0010065
- Zhao, Y., Wu, J., Yang, J., Sun, S., Xiao, J., and Yu, J. (2012). PGAP: pan-genomes analysis pipeline. *Bioinformatics* 28, 416–418. doi: 10.1093/bioinformatics/btr655
- Zhu, Z., Guo, T., Zheng, H., Song, T., Ouyang, P., and Xie, J. (2015). Complete genome sequence of a malodorant-producing acetogen, *Clostridium scatologenes* ATCC 25775(T). *J. Biotechnol.* 212, 19–20. doi: 10.1016/j.jbiotec.2015.07.013

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2016 Shin, Song, Jeong and Cho. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.