Genomics of Cellulolytic Clostridia and Development of Rational Metabolic Engineering Strategies

by

Robert Carlo Carere

A Thesis Submitted to the Faculty of Graduate Studies of
The University of Manitoba
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Department of Biosystems Engineering
University of Manitoba
Winnipeg

Genomics of Cellulolytic Clostridia and Development of Rational

Metabolic Engineering Strategies

by

Robert Carlo Carere

Ph.D., University of Manitoba, 2013

Supervisory Committee

Dr. David B. Levin (Department of Biosystems Engineering, University of Manitoba)
Supervisor

Dr. Nazim Cicek (Department of Biosystems Engineering, University of Manitoba)
Departmental Member

Dr. Ivan Oresnik (Department of Microbiology, University of Manitoba)
Departmental Member

Dr. Frank Robb (Department of Microbiology and Immunology, University of Maryland)
External Examiner

Supervisory Committee

Dr. David B. Levin (Department of Biosystems Engineering, University of Manitoba)
Supervisor

Dr. Nazim Cicek (Department of Biosystems Engineering, University of Manitoba)
Co-supervisor or Departmental Member

Dr. Ivan Oresnik (Department of Microbiology, University of Manitoba)
Departmental Member

Dr. Frank Robb (Department of Microbiology and Immunology, University of Maryland)
External Examiner

Abstract

Consolidated bioprocessing, a process in which cellulase production, substrate hydrolysis, and fermentation occur simultaneously, offers the potential for lower biofuel production costs than traditional approaches and is an economically attractive near-term goal for fermentative production of ethanol and/or hydrogen (H_2) as biofuels. Current yields fall short of theoretical maxima, vary considerably between species, and are influenced by the highly branched metabolic pathways utilized by fermentative organisms. For fermentative ethanol/ H_2 production to become practical, yields must be increased either through intelligent species selection, a manipulation of culture conditions, or via the implementation of rational metabolic engineering strategies. A comparative genomics approach amoungst select members of the Firmicutes, Euryarchaeota, and Thermotogae was used to identify genes relevent to ethanol and H_2 production. Growth, end-product synthesis, enzyme activities and the associated transcription of select genes were studied in the cellulolytic anaerobe, *Clostridium thermocellum* ATCC 27405, during batch fermentation of cellobiose to determine the effect of elevated N_2 and H_2 sparging on end-product distribution. The absence of genes encoding acetaldehyde dehydrogenase and bifunctional acetaldehyde/alcohol dehydrogenase (AdhE) correlates with elevated H_2 yields and low ethanol production. The type(s) of encoded hydrogenases appear to have minimal impact on H_2 production in organisms that do not encode ethanologenic pathways, however, they do influence reduced end-product yields in those that do. We also find that while gas sparging can be used to effectively shift carbon and electron flow, the observed shifts at the pyruvate

branch-point are likely principally influenced by the availability of reduced electron carriers (NAD, NADP, ferredoxin) and thermodynamic considerations. Finally, both electrotransformation and conjugative plasmid protocols were developed and evaluated for thermophilic species *C. thermocellum* and *Thermoanaerobacter pseudethanolicus* 39E, and the mesophilic bacterium, *Clostridium termitidis* CT1112. The efficiency of transformation for *C. thermocellum strain* ATCC 27405 is consistently low whereas transformation frequencies were ~100-fold higher in *C. termitidis*. Observed frequencies of plasmid transfer, via conjugation, were similar in both *C. thermocellum* and *C. termitidis* suggesting the transfer of single stranded DNA may circumvent aggressive restriction methylation systems.

Table of Contents

List of Abbreviations

2-KLG	2-keto-L-gluconic acid
ABC	Applied Genomics Research in Bioproducts or Crops
ack	Acetate kinase
adh	Alcohol dehydrogenase
AdhE	Acetaldehyde/alcohol dehydrogenase
aldh	Aldehyde dehydrogenase
AMP	Adenosine monophosphate
ARDI	Agricultural and Rural Development Initiative
asRNA	Anti-sense RNA
ATCC	American type culture collection
atk	Acetate thiokinase
ATP	Adenosine triphosphate
BK	Butyrate kinase
BSA	Bovine serum albumin
cat	Chloramphenicol acetyltransferase
CBP	Consolidated bioprocessing
CCR	Carbon catabolite repression
CFCs	Chloroflurocarbons
CO	Carbon monoxide
CO_2	Carbon dioxide
COGs	Clusters of orthologous groups
D	Dilution

DTT	Dithiothreitol
E:A	Ethanol-to-acetate ratio
Ech	Enery conserving hydrogenase
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbant assay
EPB	Electroporation buffer
Fd	Ferredoxin
FDP	Fructose-1,6-bisphosphate
FHL	Formate hydrogen lyase
GADPH	Glyceraldehyde-3-phosphate dehydrogenase
GAPFOR	Glyceraldehyde-3-phosphate:ferredoxin oxidoreductase
GHG	Greenhouse gasses
GTP	Guanosine triphosphate
H_2	Hydrogen gas
H_2ase	Hydrogenase
HPLC	High pressure liquid chromatography
IMG	Integrated microbial genomics
IPCC	Intergovernmental Panel on Climate Change
KO	KEGG ontology
ldh	Lactate dehydrogenase
malE	Malic enzyme
Mbh	Membrane bound hydrogenase
mdh	Malate dehydrogenase

mls	Macrolide-lincosamide-streptogramin
MV	Methylviologen
N_2	Nitrogen gas
NAD(P)H	Nicotinamide adenine dinucleotide (phosphate)
ND	Not determined
nfo	NADH:ferredoxin oxidoreductase
NSERC	Natural Sciences and Engineering Research Council of Canada
OAADC	Oxaloacetate decarboxylase
OD	Optical density
$p(H_2)$	Hydrogen gas partial pressure
PBS	Phophate buffered saline
PCR	Polymerase chain reaction
pdh	Pyruvate dehydrogenase
PEP	Phosphoenol pyruvate
PEPCK	Phosphoenolpyruvate carboxykinase
PFK	Phosphofructokinase
pfl	Pruvate formate lyase
PFOR	Pyruvate:ferredoxin oxidoreductase
PPDK	Pyruvate phosphate dikinase
PP_i	Pyrophosphate
PPK	Pyruvate phosphate kinase
pta	Phosphotransacetylase
PTB	Phosphotransbutyrylase

RM	Restriction methylation
RNF	Rhodobacter nitrogen fixing
RT	Reverse transcription
SSF	Simultaneous Saccharification and Fermentation

List of Tables

List of Figures

List of Supplementary Tables and Figures

Acknowledgments

The support and help of many individuals is gratefully acknowledged. Firstly, I would like to thank Dr. David Levin for providing me an opportunity to pursue this research, for his mentorship, optimism and unwavering support throughout the duration of my studies. I would like to thank all members of our lab, past and present, for their friendship and support. Both Tom Rydzak, for his assistance with thermodynamic calculations, genomic meta-analysis and enzyme activity assays and Tobin Verbeke, for his assistance with genomic analysis, merit special recognition. Parveen Sharma assisted with the development of a conjugation protocol and vectors. Rumana Islam, Lauren Magnusson, Warren Blunt, Chris Dartialh, Nathan Wrana, Marcel Taillefer, Sadhana Lal, John Schellenberg, , Ryan Sestric and Valory Agbor were all supportive of my studies. I would also, at this time, like to extend my appreciation to my supervisory committee members, Dr. Nazim Cicek and Dr. Ivan Oresnik for their support and advice.

My sincerest thanks go to my colleagues and collaborators both at the University of Manitoba and abroad. In particular I would like to thank Dr. Richard Sparling for his ongoing mentoring advice and technical assistance. I would also like to thank Pin-Ching Maness and Katherine Chou at NREL (Golden, CO) for graciously hosting me during my time there and with their assistance in my transformation studies.

I am grateful of the administrative assistance provided by Evelyn Fehr and Debby Watson in the Biosystems Engineering office. This work was supported by funds provided by the Natural Sciences and Engineering Research Council of Canada (NSERC), through a Strategic Programs grant (STPGP 306944-04), by Genome Canada,

through the Applied Genomics Research in Bioproducts or Crops (ABC) program for the grant titled, “Microbial Genomics for Biofuels and CoProducts from Biorefining Processes”, and by the Province of Manitoba, Agricultural and Rural Development Initiative (ARDI), grant 09-986.

Dedication

I would like to dedicate this work to my amazing partner, Kathryn Dawson. For her unbelievable patience and support, her humour and capacity to illuminate the important things in my life. Thank you.

Chapter 1: Third Generation Biofuels via Direct Cellulose Fermentation

1.1 Abstract

Consolidated bioprocessing (CBP) is a system in which cellulase production, substrate hydrolysis, and fermentation are accomplished in a single process step by cellulolytic microorganisms. CBP offers the potential for lower biofuel production costs due to simpler feedstock processing, lower energy inputs, and higher conversion efficiencies than separate hydrolysis and fermentation processes, and is an economically attractive near-term goal for "third generation" biofuel production. In this chapter, production of third generation biofuels from cellulosic feedstocks will be addressed with respect to the metabolism of cellulolytic bacteria and the development of strategies to increase biofuel yields through metabolic engineering will be discussed.

1.2 Introduction

With world energy consumption predicted to increase 54 % between 2001 and 2025, considerable focus is being directed towards the development of sustainable and carbon neutral energy sources to meet future needs [1]. Biofuels are an attractive alternative to current petroleum-based fuels as they can be utilized as transportation fuels with little change to current technologies and have significant potential to improve sustainability and reduce greenhouse gas emissions. Liquid (ethanol, biodiesel) or gaseous (methane or hydrogen) biofuels are derived from organic materials such as starch, oilseeds and animal fats, or cellulose.

Cellulose is the most abundant biopolymer in the world [2]. Discarded cellulosic biomass derived from forestry, agriculture, and municipal sources are potential feedstocks for the synthesis of biofuels (and other value-added bioproducts) that could

displace fossil fuel consumption and reduce greenhouse gas emissions [3, 4]. The crystalline structure of cellulose, however, makes it difficult to hydrolyze. Conventional production of ethanol or H_2 from cellulose via fermentation involves a complex process of pre-treatment including: (i) cellulase production, (ii) hydrolysis of cellulose and hemicellulose (if present), followed by (iii) fermentation of hexose sugars generated by cellulose hydrolysis and pentose sugars generated by hemicellulose hydrolysis (if present). Current strategies to produce fuel ethanol from cellulose, referred to as "second-generation" biofuels, utilize simultaneous saacharification and fermentation (SSF) or simultaneous saacharification and co-fermentation (SSCF) [5, 6]. Both SSF and SSCF require extensive pre-treatment of the cellulosic feedstock by steam-explosion and/or acid treatment, followed by addition of exogenously produced cocktails of cellulolytic enzymes to hydrolyse cellulose chains and release the glucose monomers required for fermentation.

Consolidated bioprocessing (CBP) is an alternative processing strategy in which cellulase production, substrate hydrolysis, and fermentation are accomplished in a single process step by microorganisms that express cellulolytic (and hemicellulolytic) enzymes [5-7]. CBP offers the potential for lower biofuel production costs due to simpler feedstock processing, lower energy inputs (and therefore better energy balance), and higher conversion efficiencies than SSF based processes. CBP is an economically attractive near-term goal for process for "third generation" biofuel production [5-7].

Successful development of "third generation" biofuels depends heavily on a detailed understanding of the metabolism of cellulolytic bacteria. In this paper the application of metabolic engineering strategies to improve biofuel yields (ethanol and

hydrogen) will be discussed. Particular emphasis will be placed on the utilization of cellulose as an available feedstock and the metabolism of cellulolytic Clostridia.

1.3 Biofuels and Fossil Fuels

1.3.1 Environmental Concerns

The extraction, processing and combustion of fossil fuels contributes to the pollution of soil, air and water and thus is hazardous to the environment and to public health [8]. In 1996, the Intergovernmental Panel on Climate Change (IPCC) reported the earth's mean surface temperature had increased 0.2 °C per decade since 1975 and asserted anthropogenic greenhouse gasses (GHG) as the primary contributor to the observed change. Greenhouse gasses include carbon dioxide (CO_2), methane (CH_4), nitrous oxide and various engineered chemicals including chloroflurocarbons (CFCs). These gasses prevent heat reflected from the earth's surface from escaping into space and therefore contribute to a global warming [9].

CO_2 emissions contribute to the majority of the documented heat trapping capacity of the atmosphere and worldwide CO_2 concentrations continue to increase by 0.5 % annually. The combustion of fossil fuels is by far the largest contributor to the observed CO_2 increase and in general accounts for approximately 60 % of greenhouse warming resulting from anthropogenic sources [9].

In response to growing concerns regarding environmental degradation, biofuels have emerged as an attractive alternative to conventional petroleum-based fuels. The utilization of fuels derived from biomass can significantly reduce greenhouse gas emissions. In 1998 Brown *et al.* reported that the utilization of bioethanol as an automotive fuel could reduce CO_2 emissions by 60-90 % relative to conventional

petroleum fuels. CO_2 released during the combustion of bioethanol is recycled through the photosynthetic process resulting in no net increase to CO_2 levels. The combustion of neat ethanol (E100), however, contributes to the emission of aldehydes, particularly acetaldehyde, which is 2 to 4 times lower in gasoline emissions.

Alternatively, the utilization of hydrogen as a transportation fuel would significantly reduce GHG emissions. Hydrogen fuel cells directly convert chemical energy to electrical energy without the need of combustion. In theory this conversion can achieve efficiencies approaching 90 %. While H_2 fuels cell produce only water, it is noteworthy that the biological production of H_2 gas does involve production of large quantities of CO_2 [10-12].

1.3.2 Bioethanol

Bioethanol is the most widely used liquid biofuel and is most commonly used, in the United States as an E10 gasoline additive. In Brazil, ethanol-powered and flexible-fuel vehicles are manufactured for operation with hydrated ethanol, an azeotrope of ethanol (around 93 % v/v) and water (7 %) [13-15]. As a result of this, coupled to the development of domestic deep-water oil sources, Brazil has achieved complete self-sufficiency in oil [9]. Fuel ethanol can be used in a variety of ways. Ethanol is commonly used as an oxygenated fuel additive to reduce emissions of carbon monoxide, nitrous oxides and hydrocarbons [16]. Numerous common ethanolic motor-fuel formulations are being used with increasing frequency (Table 1.1). Ethanol has a higher octane rating than petroleum fuels enabling combustion engines to run at higher compression ratios and thus give superior net performance [16]. In addition, ethanol

exhibits higher vapour pressure and heat of vaporization than gasoline and therefore increased power outputs are observed while using ethanol [17].

While the Unitied States Environmetal Protection Agency (EPA) has approved the utilization of high content blends of ethanol, specifically E15, within vehicles produced after 2001, several factors, including concerns over the potential liabilities associated with vehicle misfueling, and infrastructure costs are likely to prevent widespread usage of these fuels in the near future [18]. In 2012, ethanol increased from 8 % of U.S. gasoline consumption by volume in 2009 to over 10 %. Total production of bioethanol within the U.S. in 2011 was nearly 52.7 billion liters. During the 2010/2011 agricultural marketing year, it was estimated that over 40 % of the national corn crop and 14 % of soybean crop was used for the production of bioethanol. Exports of bioethanol, from the U.S. increased dramatically in 2011-2012 as Brazil experienced a poor sugar harvest. Furthermore, a major drought in the summer of 2012 within the midwestern U.S. that resulted in increased bioethanol costs is illustrative of an emerging industry that is, unlike the petroleum industry, sensitive to environmental influence [18].

The production of cellulosic ethanol continues to fall far short of the targets outlined by the Energy Independence and Security Act of 2007. Even after the EPA issued waivers that reduced mandated commercialization values, no significant production was reported in 2011 and 2012 values likely will continue to fall far short of the reduced targets [18].

Table 1.1. Common ethanolic motor-fuel formulations [16]

Fuel[a]	Ethanol content (% v/v)
Hydrous ethanol (Alcohol)(Brazil)	95.5
E85 (North America)	85
Gasoline (Brazil)	24
E10 (gasohol)(North America)	10
Oxygenated fuel (USA)	7.6
Reformulated gasoline (USA)	5.7
Biodiesel' (Sweden)	15

[a]Hydrous ethanol contains 4.5 % (v/v) water. The other formulations contain hydrocarbons and a trace of water

Bioethanol is produced by converting sugars directly from crops like sugarcane or sugar beets, indirectly through starch from corn, wheat, potatoes, or cassava, or through cellulose from biomass, into ethanol via fermentation followed by distillation [19]. Improving ethanol yield is critical to the establishment of a viable industry and is fundamentally tied to increasing throughput and reducing costs associated with production.

During the fermentation of sugarcane, up to 10 % of available sugar can be diverted away from ethanol and converted to glycerol and succinic acid. Attempts to improve ethanol yields by reducing the synthesis of these unwanted metaboites has resulted in reduced yeast competitiveness within fermentors [16]. Glycerol production is coupled to acid, ethanol, and temperature induced stress conditions. The synthesis of glycerol occurs in response to osmotic stress and therefore likely has an essential role in cell viability. Although yeasts with improved properties such as ethanol and temperature tolerances have been genetically engineered, such strains are not yet used widely by the fuel ethanol industry [17].

1.3.3 Biohydrogen

Hydrogen gas (H_2) is a clean fuel that possesses a high energy content per unit weight (122 KJ g^{-1}) and does not contribute particulate or greenhouse gas emissions into the atmosphere upon combustion [20]. Hydrogen can be produced by nuclear or fossil fuel mediated electrolysis of water, coal gasification, and steam reformation of natural gas or through the action of biological systems. Production of H_2 using fermentative biological processes is potentially the most attractive of these strategies as it is not as

energy intensive as other means and could potentially utilize refuse or agricultural waste-streams as the raw material [3, 11]. Periodic crises in the supply and price of fossil fuels has drawn attention to the fact that renewable energy sources are the only long-term solution to future energy requirements. Consequently, improving H_2 yields, reducing production costs and developing technologies to exploit this fuel are being actively investigated. Attainable biohydrogen yields and the potential for increasing these yields in the future are not well understood. Several recent studies focusing on saccharolytic species of clostridia have begun to investigate potential H_2 yields.

The biological production of H_2 is fundamentally dependent on hydrogen evolving enzymes. Hydrogenases constitute a family of enzymes found throughout the bitota. These enzymes catalyze the reversible oxidation of hydrogen gas: $H_2 \leftrightarrow 2H^+ + 2e$. While the host organisms of a hydrogen-producing hydrogenase, such as the CpI hydrogenase in *Clostridium pasteurianum*, will typically use it to dispose of excess electrons, considerable research has recently focused on the potential for these enzymes to produce hydrogen as an energy source. If harnessed properly, hydrogenases and/or hydrogenase-containing organisms could be used to supply affordable and renewable H_2 to be used as an energy fuel.

Polymerase chain reaction (PCR) targeting 16S rDNA and known hydrogenase nucleotide sequences has been used to confirm the presence of clostridial species within anaerobic biohydrogen fermenting reactors [21]. Samples taken from the exponential-phase of growth and subsequent reverse-transcriptase (RT) PCR and DNA sequence analyses revealed that the H_2 producing strains either expressed a *Clostridium pasteurianum*-like or *Clostridium saccharobutylicum*-like hydrogenase [21].

The potential for H_2 production by *Clostridium thermolacticum* during continuous fermentation of lactose was explored using a waste-stream generated by the cheese industry. Approximately 150,000 tonnes of milk permeate are produced each year within Switzerland as a by-product of cheese production. Given its high biochemical oxygen demand for biodegradation, the effluent must be treated prior to release into the environment [22]. Hydrogen formed during lactose fermentation by *C. thermolacticum* was measured during continuous culture at different dilution rates and pH levels. Although acetate was the major metabolite formed as a result of lactose fermentation, H_2 was evolved as a by-product. Collet *et al.* [22] found that H_2 production was maximized at high dilution rates by maintaining a pH above 7.0. They concluded that lactose fermentation by *C. thermolacticum* represents a cheap method of biohydrogen production that makes use of an otherwise polluting waste-stream.

1.4 Cellulose Feedstocks and Cellulolytic Bacteria

Biofuels may be produced from the sugars of a variety of different feedstocks, including food crops and cellulosic substrates. Food crops such as sugarcane and sugar beets contribute to 60 % of the world's bioethanol production and contain large amounts of sucrose [17]. Other crops, including corn and cereal crops like wheat, contain starch that is then converted into glucose during fermentation. The expansion of biofuels production, particularly in the United States, together with increased worldwide demand for grains and increased energy costs, has led to drastically higher grain prices.

The total worldwide harvest of wheat and coarse grains for the 2012-2013 growing season is forecast to reach 1.76 billion tonnes, a figure 2.2 % short of predicted worldwide consumption estimates (1.80 billion tonnes). Although this represents a 4.7 %

decrease over the 2011-2012 growing season, harvests are estimated to recover in 2013-2014 by 8.1 % and continue to increase by a moderate 1.6 % (per year) through 2017-2018 [23]. In the US, competition for land and other resources used to produce corn for fuel ethanol already has led to restrictions in land use for oilseed production and consequently higher prices associated with lower production levels. The pursuit of lower biofuels production costs and competition with traditional food crops has led many to consider cellulosic substrates as a potential feedstock. It is expected that the consumption of grain for industrial purposes (including ethanol production) will increase slightly through 2012-2013, and then be followed decrease as lignocellulosic derived biofuels, in North America begin to take precendence

1.4.1 Cellulose

Cellulose is the most abundant bio-polymer on earth; an estimated 7.5×10^{10} tons are annually synthesized through photosynthetic processes [2]. Found primarily in plant cell walls, cellulose is embedded in a hetero-matrix composed of xylan, other hemi-celluloses and lignin. Specifically, cellulose is a linear, insoluble biopolymer composed of repeating β–D-glucopyranose residues linked by β-1,4 glycosidic bonds. In contrast to other glucan polymers, such as starch, the repeating unit of cellulose is not glucose, but cellobiose, a disaccharide. Cellulose exhibits a high degree of polymerization: the individual glucan chains, or cellodextrins, can reach lengths of greater than 25,000 glucose residues [24].

Cellulose produced by plants is composed of both highly amorphous regions containing large voids and other irregularities as well as tightly packed crystalline regions. Cellulose, because it is resistant to most forms of degradation, accumulates

within the environment. It has been estimated that approximately half of the carbon fixed annually within terrestrial ecosystems is stored as cellulose. Cellulose synthesis is primarily associated with plants, however some animals, bacteria and fungal species can also produce the polymer [25].

Organisms that are capable of degrading the polymer and utilizing it as a source of carbon are ecologically very important. Cellulose is generally degraded into H_2O and CO_2 in aerobic systems while in anaerobic systems CH_4 and H_2 are also produced. Although, most cellulose is degraded in aerobic environments, 5 to 10 % is degraded under anaerobic conditions by a range of physiologically diverse bacteria. Cellulolytic species are found within the phyla Thermotogae, Proteobacteria, Actinobacteria, Spirochaetes, Firmicutes, Fibrobacteres and Bacteroids. Of these, approximately 80 % of the isolated cellulolytic bacteria are found within phyla Firmicutes and Actinobacteria [26]. The majority of the gram-positive cellulolytic bacteria are found within Firmicutes and belong to the class Clostridia and the genus *Clostridium.* These bacteria are ubiquitous in anaerobic soil environments, form endospores and often digest cellulose via an exocellular enzymatic complex called a cellulosome, converting cellulose into various oligosaccharides.

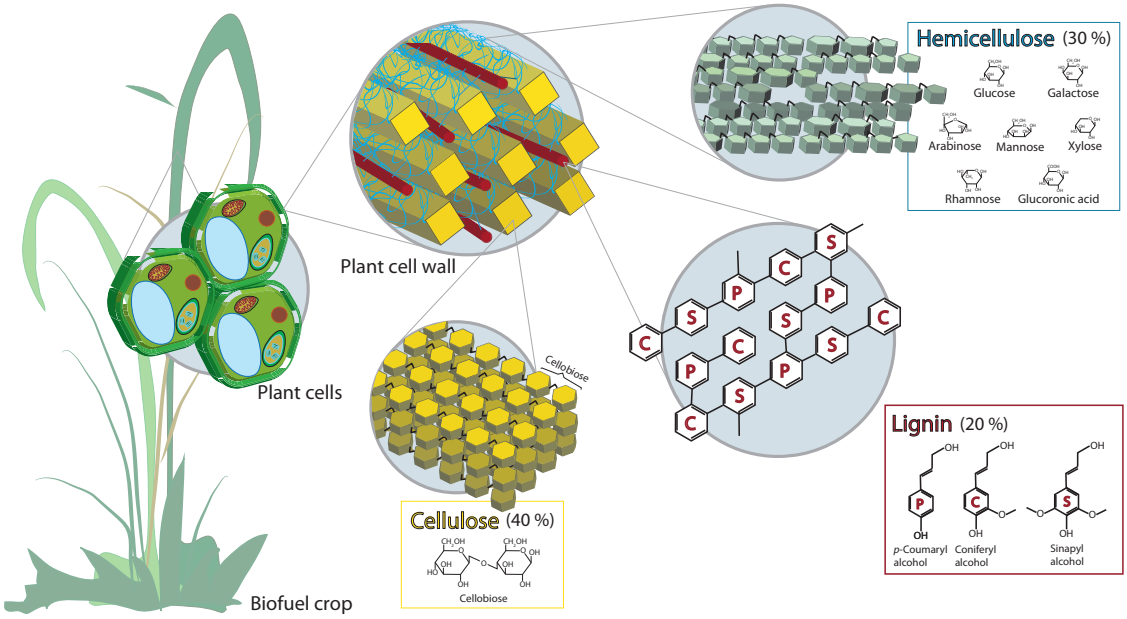


Figure 1.1. Diagrammatic representation of lignocellulose.

1.4.2 Detection and Enumeration of Cellulolytic Bacteria

As the primary and rate-limiting step involved in waste degradation, cellulose breakdown is an important step in the development of strategies to treat municipal solid waste [27]. O'Sullivan *et al.* [28] investigated enriched cellulolytic microbial communities within anaerobic batch reactors in order to determine the dynamics of cellulolytic bacterial populations during the fermentation process. In fermentation experiments conducted under mesophilic temperatures, 80 % of the cellulose was solubilised within 20 days [28].

Fluorescence *in situ* hybridization (FISH) analysis revealed the bacteria grew as surface-attached biofilms. Fluctuations in relative abundance of the three clostridial groups did not have a significant impact on the rate of cellulose degradation. Over the initial incubation period, the sum of the three target groups of bacteria accounted for

more than 99 % of the total bacteria present. After cellulose depletion, however, the proportion of the target bacteria was reduced to 13 %. In general, *C. thermocellum*-like bacteria were always the most abundant group throughout the study, while *C. stercorarium* and *B. cellulosolvens*-like bacteria also accounted for a significant proportion of the community. The lack of evidence linking clostridial population dynamics to the rate of cellulose degradation suggests that cellulose solubilisation rates are dependent on the amount of colonized surface area and biofilm architecture.

Within cellulose rich environments, non-cellulolytic bacteria can affect the metabolic flux of cellulose hydrolyzing microorganisms. In 1977, the fermentation of cellulose and cellobiose by *Clostridium thermocellum* was investigated in the presence and absence of *Methanobacterium thermoautotrophicum* [29]. It was hypothesized that methanogenic bacteria are able to act as electron sinks by making it energetically favourable for cellulose fermenting bacteria to dispose of H_2 instead of other reduced products such as ethanol [29]. Known as interspecies hydrogen transfer, H_2 produced by cellulose fermenting bacteria is oxidized to methane by methanogenic microorganisms. Co-cultures of *C. thermocellum* and *M. thermoautotrophicum* produced more H_2 and acetic acid and less ethanol than within *C. thermocellum* monocultures. Furthermore, when grown on cellulose, the co-cultures exhibited a shorter lag before growth, and cellulase activity appeared earlier within the co-culture than in the monoculture. The conversion of H_2 to methane was complete within the co-culture and the majority of the methane produced was derived from the reduction of CO_2 rather than acetate conversion [29].

When grown on cellobiose, the methanogen caused only very small changes to the fermentation balance of *C. thermocellum*. The absence of free H_2 within the co-culture grown on cellulose indicated that methanogenesis was limited by the rate of H_2 evolution by *C. thermocellum*. The increased growth rate of *C. thermocellum* in the cellobiose cultures was sufficient to result in an accumulation of H_2. In this instance, methanogenesis within cellobiose co-cultures was limited by the growth rate of *M. thermoautotrophicum* [29].

Weimer and Zeikus [29] concluded that the altered fermentation patterns observed within *C. thermocellum* grown in co-culture with *M. thermoautotrophicum* was in general agreement with the concept of interspecies H_2 transfer. The absence of significant metabolic interactions between the clostridia and the methanogen when grown on cellobiose can be explained in terms of ecological significance.

In nature, the decomposition of organic matter is limited by the rate of the degradation of insoluble biopolymers such as cellulose. In this regard, soluble intermediates of anaerobic digestion, including glucose, cellobiose, and acetate are normally found in low concentrations and generally have a smaller environmental impact than readily available insoluble substrates such as cellulose. From a kinetic point of view, interspecies H_2 transfer that influences the rate at which insoluble biopolymers are degraded may be of greater environmental consequence than transfers involving mixed-cultures grown on soluble substrates such as cellobiose [29].

1.5 The Cellulosome

1.5.1 Cellulolytic Enzymes

Through the secretion of cellulases as single enzymes, as single polypeptides with multiple cellulosic domains or as extracellular multienzyme complexes, microorganisms have developed several strategies to digest cellulose. The cellulosome, which was first described within *C. thermocellum* by Lamed *et al*. [30], is a multi-component cellulolytic exocellular complex of proteins that mediates cellulose binding and degradation. Similar to other cellulose degrading strategies, the cellulosome hydrolyzes the biopolymer to its building block, the disaccharide cellobiose. The cellulosome is a specialized structure that plays a significant physiological role. On the cell surface they appear as polycellulosomal aggregates promoting the adherence of the bacterium to the cellulose fibre [31]. Cellulosomes are able to hydrolyze both amorphous, and highly ordered crystalline cellulose. The degradation of cellulose is accomplished through the action of enzymes that include endo-1,4-β-glucanases, exo-1,4-β-glucanases (cellobiohydrolases), and cellobiase (β-glucosidase).

Endoglucanases are able to hydrolyze amorphous cellulose, carboxymethylcelluose, and phosphoric acid-swollen cellulose, producing soluble oligosaccharides that are subsequently degraded into cellobiose and glucose through the action of β-glucosidase. Cellobiohydrolases degrade cellulose by cleaving cellobiose units from the non-reducing end of a cellulose fibre [31]. A major component of the cellulosome is a large non-enzymatic polypeptide that binds and supports the enzymatic subunits of the cellulosome to the cell. This protein, referred to as scaffoldin or CipC, has no catalytic activity and is modularly organized with numerous distinct cohesion domains

and cellulose binding motifs. Some cohesion domains of the scaffoldin protein permit specific binding to the dockerin domains of the various enzymatic components of the cellulosome. Other cohesion domains mediate cell attachment via cell-surface attached dockerin domains. Cellulose attachment can be mediated by cellulose binding motifs present on the scaffoldin protein or from motifs present on the various enzymatic components (Figure 1.1).

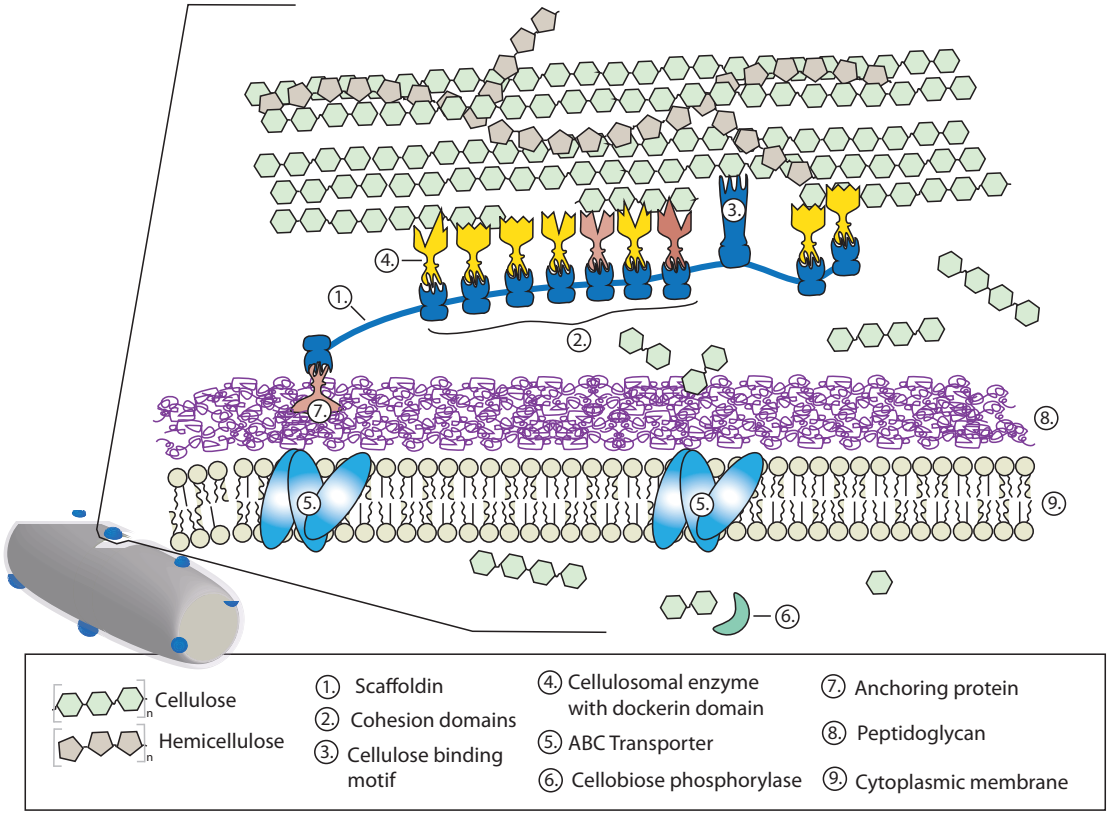


Figure 1.2. Diagrammatic representation of the cellulosome of *Clostridium thermocellum* ATCC 27405. The scaffoldin protein (1, CipC) is linked to the cell wall via an anchoring protein. Cohesion domains (2) located on the scaffoldin protein mediate attachment to dockerin domains of various enzymatic components (4). Binding to cellulose is accomplished via cellulose binding motifs (3) associated with both the scaffoldin protein and some enzymatic components (adapted from [32]).

1.5.2 Mechanisms of Cellulose Degradation

A number of studies have been dedicated to enhancing our understanding of the cellulosome. The specific and direct attachment to the substrate permits efficient competition with other non-cellulosome utilizing cellulose degrading bacteria. Close proximity of the cellulose to the cell aids in minimizing the diffusion of soluble cello-oligosaccharides into the extracellular environment and insures an efficient uptake into the cell. From an enzymatic point of view, the cellulosome promotes concerted activity and synergism of the enzymatic components. The action of the endo-1,4-β-glucanases produce novel non-reducing ends for enzymatic attack by cellobiohydrolases while β-glucosidase converts soluble cello-oligosaccharides into cellobiose to be imported into the cell. Although first described within *Clostridium thermocellum*, numerous other clostridial species have been shown to utilize a cellulosome as their cellulose degrading strategy. These bacteria are ubiquitous in anaerobic soil environments and include *C. cellulolyticum*, *C. acetobutylicum*, *C. cellobioparum* and *C. papyrosolvens* [24].

Ultra-structural observations using electron microscopy, Fourier-transform IR spectroscopy and X-ray diffraction analysis have reaffirmed that the synergistic action of the various enzymes of the cellulosome provides an advantage over cellulose degrading strategies that employ free cellulase systems [25]. In aerobic fungi, various cellulolytic enzymes are excreted as separate entities, each displaying a specific enzymatic function. In contrast, the cellulosome consists of discrete multi-functional enzymatic subunits physically attached to a central scaffoldin protein. Studies have shown that the combined cellulases from *Clostridium thermocellum* are at least 50 times more active in cotton

degradation than the extracellular cellulase system of the fungus *Trichoderma reesei* [33, 34].

It has been hypothesized that the rate-determining step for cellulose degradation is the initial attack. Once a cellulose fibre has undergone initial enzymatic attack, degradation occurs in a processive manner until the fibre has been completely degraded. This explains why some intact fibres are still observed after near complete solubilisation of available cellulose, it is possible that some fibres represent a sub-population that are highly resistant to the action of cellulases.

1.5.3 Regulation of Cellulase Synthesis

Regulation of cellulase synthesis is an important feature for the physiology of cellulolytic clostridia. The synthesis of a single cellulosome requires a substantial initial investment of ATP. The cellulosome of *C. thermocellum* may be composed of more than 20 different catalytic subunits in addition to the scaffoldin protein. The regulation of expression for these proteins plays a critical role in determining cellulose hydrolysis and cell growth rates. It has been shown that the degradation of crystalline cellulose is significantly decreased within cultures that are grown concomitantly on cellobiose [35]. This suggests the expression of cellulase proteins is inhibited through a carbon catabolite repression (CCR) mechanism in which the presence of cellobiose negates the requirement for continued assembly of cellulosomes.

Recently, an enzyme-linked immunosorbant assay (ELISA) based protocol was developed using an antibody raised against a common cohesion domain of the scaffoldin protein. In this study, cellulase production was inferred based on the quantity of scaffoldin observed [36]. Cellulase expression was measured in cultures grown on

Avicel and cellobiose within both batch and continuous culture conditions. ELISA results revealed that a nine-fold greater quantity of cellulase was detected within cultures grown on Avicel compared to cellobiose grown cells. The inverse correlation between cellobiose concentrations and cellulase adds support to the theory that the assembly of cellulosomes is mediated by a carbon catabolite repression mechanism [37].

1.6 Metabolism of Cellulolytic Clostridia

1.6.1 Sugar Uptake

The metabolism of cellulolytic clostridia has best been studied within *Clostridium cellulolyticum*. *C. cellulolyticum* is a mesophilic, cellulolytic bacterium that was originally isolated from decaying grass. As with most truly cellulolytic clostridia, *C. cellulolyticum* degrades cellulose via a cellulosome and releases soluble oligosaccharides, principally cellobiose. The first studies of metabolic activity focused mainly on *C. cellulolyticum* behaviour such as colonization. These studies showed that following colonization of a cellulose fibre, the catalytic components of the cellulosome began the depolymerisation process and consequently soluble sugars, such as glucose and cellodextrins (i.e. from cellobiose to celloheptose) were released [38]. This early work indicated that the released cello-oligosaccharides remain in close proximity to the cell between the cellulosome and the cell wall before uptake. The uptake of these sugars is a highly efficient process that utilizes an ATP-binding cassette transport system [39]. Uptake into the cell of the soluble sugars often occurs within seconds after their initial release from the cellulose fibre [39]. Following uptake, sugars are processed within the cytosol by cellobiose phosphorylase and cellodextrin phosphorylase proteins, producing glucose-6-phosphate residues ready for entry into the glycolytic pathway.

1.6.2 Cellulose Fermentation and Metabolic Fluxes in *C. cellulolyticum*

As within any cell, carbon and electron flow within *C. cellulolyticum* are closely linked. Clostridia generate their ATP through substrate-level phosphorylation and the final electron acceptors are organic molecules. During glycolysis NADH is generated during the conversion of glyceraldehyde-3-phosphate to 1,3-diphosphoglycerate and afterwards a second oxidation occurs when pyruvate is converted to acetyl-CoA by pyruvate:ferredoxin oxidoreductase (PFOR). This reaction leads to the formation of reduced ferredoxin after which electrons may be transferred to a Fd-dependent hydrogenase resulting in H_2 production [24]. NADH generated during the Embden-Meyerhof-Parnas pathway can be oxidized by NADH:Fd reductase to generate fresh reducing equivalents for the catabolic process. In conditions of high carbon flux, pyruvate overflow necessitates the oxidation of reducing equivalents through the generation of less-reduced metabolites such as ethanol and lactate (Figure 1.2; [40]).

Guedon *et al.* [41] demonstrated that metabolite yields in *C. cellulolyticum* depend strongly on the initial cellulose concentration and that early growth arrest was linked to pyruvate overflow. In this study, carbon flow in *C. cellulolyticum* was investigated within batch and continuous cultures using synthetic media with cellobiose as the sole carbon source. By measuring the specific rates of NAD(P)H production and consumption, the specific rates of H_2 production, the oxidation/reduction ratio and the main products of cellobiose fermentation, Guedon *et al.* [41] were able to determine that cellobiose catabolism exceeded the rate of pyruvate catabolism. This suggested that pyruvate oxidation via PFOR was a limiting step in the glycolytic pathway. Furthermore, pyruvate overflow was tightly coupled to an observed inhibition of growth.

Previous work had suggested that growth of *C. cellulolyticum* was limited by an imbalance in the specific rates of NADH production and consumption. Intracellular NADH/NAD^+ ratios as high as 57:1 had been observed within chemostat grown cultures and it was concluded that the reduction of NAD^+ during ethanol, lactate, and H_2 synthesis was limiting [42]. Although the lack of available reducing equivalents affected the pronounced growth inhibition observed, the critical bottleneck that occurred during glucose catabolism in *C. cellulolyticum* was the conversion of pyruvate to acetyl-CoA via PFOR. Guedon *et al.* [41] contend that *C. cellulolyticum* is not adapted to utilize carbon sources or other nutrients in excess. In natural ecosystems, habitats rarely have all nutrients in saturating quantities. Therefore, populations of cellulolytic clostridia are unlikely to ever face such a bottleneck in nature.

In contrast, a chemically defined medium contains all required nutrients for growth and therefore growth inhibitions as a result of bottlenecks in the glycolytic pathway are more likely to occur. In spite of the fact that many bacteria have developed the ability, spontaneously or through adaptation, to adjust to growth on carbon-rich media, this behaviour is not generally observed in cellulolytic bacteria [43]. Rather than adapt to habitats of high carbon availability, it is likely that *C. cellulolyticum* has evolved to optimize it's catabolic pathways to exploit pools of low available carbon sources [41].

Continuous culturing of *C. cellulolyticum* on synthetic medium containing cellobiose as the sole carbon source demonstrated that the flow of electrons from the glycolytic pathway was balanced by the production of lactate, ethanol, and H_2 gas. At low levels of carbon loading, pyruvate was oxidized preferentially to acetate and lactate, maximizing ATP formation. Conversely, high levels of carbon flux led to pyruvate

overflow and to increased levels of ethanol and lactate production. Pyruvate overflow represents an inability of PFOR to support the carbon flow arriving at pyruvate from the catabolism of cellobiose. Under these conditions, electron flow from glycolysis was balanced primarily by ethanol and lactate production [41]. For the reason that lactate production was only observed within conditions of pyruvate overflow, it can be inferred that there is no competition between the lactate producing and the acetyl-CoA producing pathways.

This preference for pathways generating ATP over pathways designed to oxidize reducing equivalents further supports the theory that *C. cellulolyticum* rarely finds nutrient substances in high concentrations in nature. Although the optimization of cellobiose catabolism under nutrient poor conditions has been proposed for *C. cellulolyticum*, other cellulolytic species exhibit considerably different fermentation profiles. In relation to biofuels production, the metabolism of cellulolytic bacteria is highly dynamic and unique to each species. End-product yields are dependant on numerous factors (pH, carbon accessibility, redox balance, gas pressure, product concentration) and strategies aiming to metabolically engineer novel strains of bacteria for increased biofuels production should be based on a detailed understanding of the genetic, enzymatic and thermodynamic mechanisms that direct carbon flow.

1.6.3 Cellulose Fermentation and Metabolic Fluxes in *C. thermocellum*

C. thermocellum is a thermophilic (optimum growth at 60 ^{o}C) bacterium that utilizes cellulose as a sole carbon source and carries out mixed product fermentation, synthesizing various amounts of lactate, formate, acetate, ethanol, H_2, and CO_2, under different growth conditions [7]. As previously described, *C. thermocellum* expresses a

cellulosome on its surface, and displays the highest known rate of cellulose degradation [6, 44, 45]. Fermentation of dilute-acid pre-treated hardwood or avicel (crystalline cellulose) in batch and continuous cultures of *C. thermocellum* were reported by Lynd *et al*. [44]. In batch cultures, ethanol and acetate were the main products, with more ethanol produced than acetate (g/g ethanol:acetate ratios were 1.8:1). Lactate synthesis was also observed at low levels as the cells entered stationary phase. In continuous cultures of *C. thermocellum*, ethanol:acetate (g/g) ratios were ~ 1:1, and substrate conversion ranged from 0.86 at a dilution rate (*D*) of 0.05 h^{-1} to 0.48 at a *D* of 0.167 h^{-1} for avicel and 0.75 at a *D* of 0.05 h^{-1} to 0.43 a *D* of 0.11 h^{-1} for pretreated wood [44].

Hydrogen and soluble end-product synthesis patterns by *C. thermocellum* in batch cultures were also investigated by Levin *et al*. [4], using either cellobiose or cellulosic substrates (α-cellulose, shredded filter paper, and delignified wood fibers) at low (0.1 g L^{-1}), medium (1.1 g L^{-1}), and high (4.5 g L^{-1}) initial substrate concentrations. Cellulosic substrates produced higher total amounts of H_2 in high substrate concentration cultures, but better H_2 yields were observed at both low and medium substrate concentrations. Delignified wood fiber was the most effective substrate, providing an average yield of 1.6 mol H_2 mol^{-1}-glucose equvalent. Ethanol, acetate and formate were produced during exponential growth, while lactate synthesis coincided with a decrease in pH to ~ 6.3 [4, 46, 47]. Batch cultures in which the initial pH was set at 6.3 (with cellobiose as the sole carbon source) produced only lactate and displayed very slow growth [48]. On average, the ratio of ethanol:acetate was lower (1:1.3) than that reported by Lynd *et al* [44], and stayed roughly constant under all growth conditions tested [4]. H_2 production and yields were similar or greater in cultures containing cellulosic substrates compared with

cellobiose, and H_2 production increased with the concentration of cellulosic substrate. Thus, it appears that *C. thermocellum* metabolism is not limited by carbon flow (unlike *C. cellulolyticum*, where metabolic flux is greatly influenced by carbon flow into the cell).

Shifts in metabolic end-product synthesis patterns can be induced by the addition of acetone, sodium azide or by increasing the hydrostatic pressure of the bioreactor by adding exogenous H_2 [49] or N_2 gas [50]. End-product mediated metabolic shifts were also observed in *C. thermocellum* [48]. Addition of ethanol to the growth medium at the initiation of the fermentation process resulted in significant increases in H_2 and acetate (54 % and 25 % respectively). Formate addition increased H_2 and ethanol by ~10 % each and decreased acetate production. Addition of H_2 decreased CO_2 and increased formate production, whereas acetate increased ethanol and decreased formate production. The addition of 10 % (v/v, 1 atm) carbon monoxide (CO) in the bioreactor headspace led to decreased synthesis of H_2, CO_2, and acetate, but significantly increased ethanol synthesis.

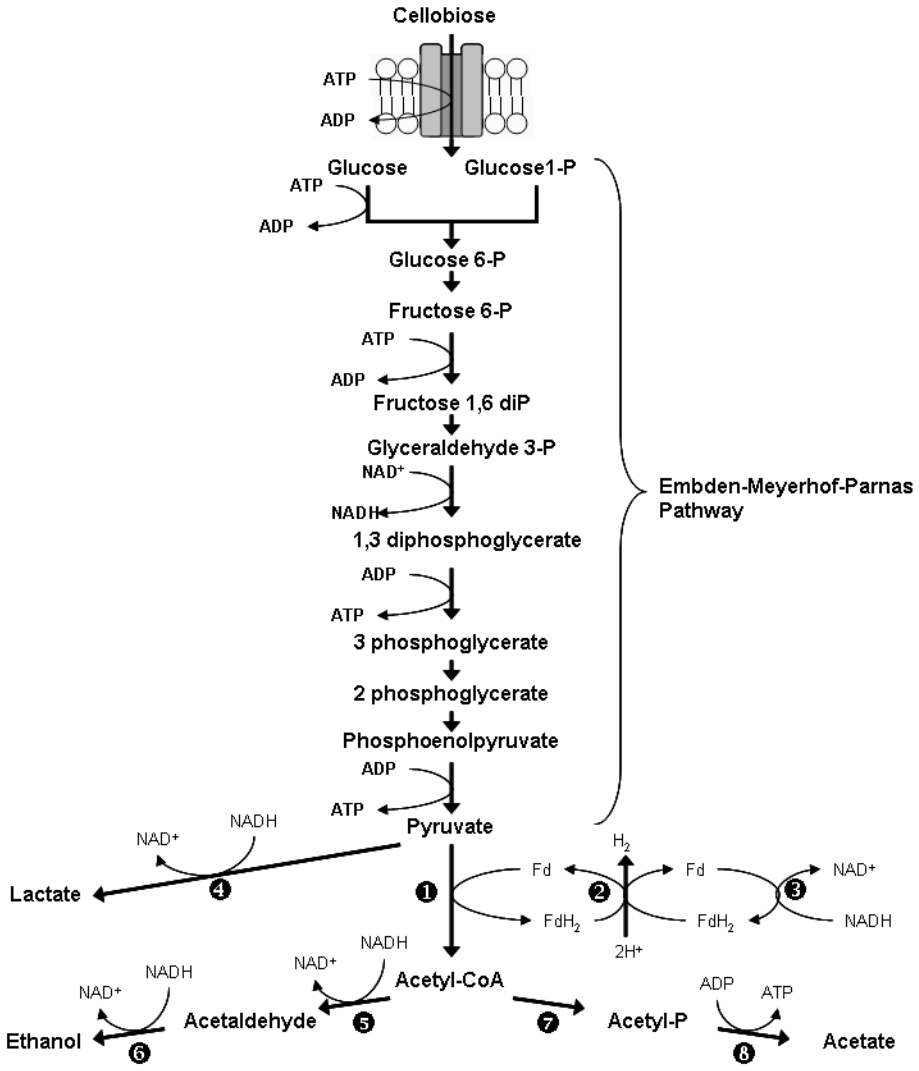


Figure 1.3. The catabolic pathway of cellobiose in *C. cellulolyticum*. 1) pyruvate-ferredoxin oxidoreductase; 2: hydrogenase; 3) NADH-ferredoxin oxidoreductase; 4) lactate dehydrogenase; 5) acetaldehyde dehydrogenase; 6: alcohol dehydrogenase; 7) phosphotransacetylase; 8) acetate kinase. [24].

1.7 Metabolic Engineering

The application of recombinant DNA techniques to direct metabolism towards the production of industrially valuable substrates is an emerging field of study. Metabolic engineering seeks to "improve" cellular function through the modulation of enzymatic, transport, or other regulatory functions of the cell. In contrast to traditional strain improvement approaches that involve mutagenesis followed by the screening of colonies for a desired phenotype, metabolic engineering often involves the introduction of heterologous genes or regulatory elements that are employed to confer novel metabolic configurations [51]. The introduction of genes encoding heterologous proteins, however, does not guarantee attainment of the desired phenotype. The expressed, heterologous protein must avoid proteolysis, fold and assemble correctly, and obtain appropriate prosthetic groups and/or post-translational modifications [51]. In addition to these potential barriers, successful metabolic engineering requires a detailed understanding of the factors influencing anabolic and catabolic flux (pH, redox potential, gas partial pressure, end-product inhibition) and a strategy to best exploit cellular function.

The cloning and expression of heterologous genes can serve several purposes including: 1) extending existing pathway to produce novel products, 2) engineering arrays of enzymatic activities that synthesize novel structures, 3) shifting metabolic flux towards synthesis of desired end-products and 4) accelerating a rate determining step [51].

1.7.1 Completion of Partial Pathways for Novel Product Synthesis

The spectrum of substrate utilization and product synthesis is a reflection of the genetic and metabolic diversity that exists in nature. Many bacterial species, however,

are inefficient from an applied perspective as oftentimes native pathways leading to the synthesis of valuable substrate are incomplete as a result of transposition, horizontal gene transfer or mutation events. One established strategy pertaining to metabolic engineering for industrially valuable substrate production involves the completion of pre-existing pathways. As an example, the final precursor to ascorbic acid synthesis is 2-keto-L-gluconic acid (2-KLG). Conventional industrial synthesis of 2-KLG involves 2 successive fermentations: i) Glucose → 2,5-diketo-D-gluconic acid (in *Erwinia herbicola*), followed by, ii) 2,5-diketo-D-gluconic acid → 2-KLG (in *Corynebacterium sp.*). A *Corynebacterium sp*. 2,5-diketo-D-gluconic acid reductase gene was cloned into *E. herbicola* resulting in a novel strain capable of fermentation of glucose into 2-KLG in a single stage [51].

1.7.2 Engineering Metabolic Shifts to Increase Synthesis of Desired End-Products

The introduction of genes encoding secondary metabolites within a host capable of producing its own different secondary metabolites can result in the construction of an array of enzymatic activities capable of producing novel products [51]. Antibiotics mederrhodins A and B, dihydrogranatirhodin, 2-nonrerythromycins A, B, C and D, and isovaleryl spiramycin are all produced within recombinant strains of *Streptomyces* engineered utilizing this strategy [51].

The manipulation of metabolic fluxes towards synthesis of a desired end-product is likely the strategy most amenable to biofuels production. The synthesis of desired products typically involves metabolic flow past forks where intermediates may enter alternative pathways [7, 40, 52, 53]. Maximizing the synthesis of valuable products

therefore requires the desired pathways to be made a priority and alternative pathways be minimized to the extent possible without inhibiting cell viability.

Over-expression of an enzyme, or of enzymes, catalyzing a reaction towards the desired fork is a common strategy in metabolic engineering [11]. Heterologous gene expression of pyruvate decarboxylase and alcohol dehydrogenase from *Zymomonas mobilis* within *C. cellulolyticum* was shown to increase cell density by 180 % and cellulose consumption by 150 % compared to wild type strains [54]. Acetate and ethanol were shown to increase by 93 % and 53 % respectively and H_2 yields increased by more than 75 %. High carbon flux had previously been shown to cause pyruvate overflow as a result of the inability of PFOR to metabolize pyruvate to acetyl-CoA. This overflow was shown to lead to cell growth inhibition and a flux towards increased ethanol and lactate production [41]. The introduction of these genes was hypothesized to increase NAD^+ cycling via the alcohol dehydrogenase, which in turn promoted flux towards ATP producing pathways and, as a consequence, increased cell growth [Guedon *et al.* [54]. These results demonstrate that fermentation of cellulose can be improved by using genetically engineered strains of clostridia designed to favour production of industrially valuable metabolites.

More frequently antisense RNA (asRNA) strategies are being utilized in bacteria to direct metabolic flow [55]. In contrast to the conventional introduction of a heterologous gene resulting in the expression of a novel protein, the aim of antisense RNA is to specifically down-regulate expression of a native protein and thus redirect metabolism. Down-regulation is achieved through the 1) inhibition of translation due to the duplex RNA structure impeding access to a ribosome binding site, 2) rapid

degradation of the mRNA due to RNA-duplex specific RNases, and 3) inhibition of mRNA transcription as a result of premature termination [55]. Antisense RNA strategies offer a number of advantages over traditional gene inactivation techniques. Most notably, asRNA avoids the hazard of introducing lethal mutations, as complete inhibition of protein expression is unlikely. Furthermore, asRNA strategies may be used to inducibly repress protein expression through the utilization of inducible promoters for asRNA transcription [56].

The application of asRNA technology as a strategy to influence primary metabolism was first demonstrated by van den Berg *et al.* [57] during which they examined the role of a periplasmic hydrogenase within *Desulfovibrio vulgaris* during lactate fermentation. More recently asRNA has been used to direct metabolic flow within *Clostridium acetobutylicum* by reducing levels of enzymes responsible for butyrate formation [55]. Butyrate is produced from the intermediate butyryl coenzyme-A in two steps: i) Butyryl coenzyme-A → Butyryl phosphate (via phosphotransbutyrylase (PTB), followed by ii) Butyryl phosphate + ADP → Butyrate + ATP (via butyrate kinase (BK)).

The genes *ptb* and *buk*, encoding PTB and BK respectively, are organized in a single *ptb-buk* operon. Previous studies have shown butyrate to influence induction of solventogenesis and it was therefore hypothesized that the down-regulation of PTB and BK would impact primary metabolism [55]. It was found that recombinant strains exhibited 70 % and 80 % reduced PTB and BK activities respectively compared to the wild type strain. Growth yields of the recombinant strains were 28 % less than controls. While levels of acid production were not affected in the *ptb* targeted strain (pRD1),

acetone and butanol yields were 96 % and 75 % lower respectively compared with the control strain. Finally, both lactate yields and lactate dehydrogenase activity were higher (~100 % and ~ 300 % respectively) [55].

Metabolic engineering of *C. acetobutylicum* strains ATCC 824 (pRD4) and ATCC 824 (pRD1) demonstrated that asRNA can be used down-regulate specific protein production and alter the metabolism of *C. acetobutylicum.* These studies, however, also suggest that the elaborate nature of metabolic regulatory systems makes the down-regulation of a specific pathway unlikely to reduce synthesis of a specific product [55]. These findings further stress the importance that design decisions should be based on a detailed understanding of the genetic, enzymatic and thermodynamic mechanisms that direct carbon flow. While relatively simple strategies may be successful in the development of novel bacterial strains, the dynamic nature of metabolism makes accurate forecasts of change problematic.

1.8 Conclusions

Biofuels are an attractive alternative to current petroleum-based fuels as they can be utilized as transportation fuels with little change to current technologies and have significant potential to improve sustainability and reduce greenhouse gas emissions. The transition from an economy based on fossil fuels to an economy based on biofuels depends on the improvement of biofuel yields via the successful development of suitable microorgansims capable of efficiently fermenting a variety of sugars while simultaneously displaying tolerance to high end-product concentrations. Cellulose represents an attractive feedstock for biofuels production because of its abundance, low cost and ability to be efficiently degraded by cellulolytic bacteria, fungi and some archea.

Metabolic engineering of bacterial and yeast species have already demonstrated that the construction of novel strains using recombinant DNA technologies can confer advantageous traits in regards to biofuels production. Success, however, will be dependent upon design decisions based on a detailed understanding of the extremely complex genetic, enzymatic, and thermodynamic mechanisms that direct carbon flow. In combination with other strategies including (meta)genomics, biodiversity studies and systems biology, metabolic engineering is a promising approach to the improvement of biofuel yields and the establishment of renewable, non-polluting energy sources.

1.9 Research Objectives

The **long-term objective** of this research were to characterize cellulose-degrading Clostridia with respect to their potential to produce biofuel chemicals, specifically H_2 and ethanol, during fermentative growth. *Clostridium thermocellum,* strain ATCC 27405, served as a model organism for these analysis as it has been well characterized physiologically, has a completely sequenced genome, grows optimally at moderately thermophilic temperatures and is able to efficiently hydrolyze lignocellulosic substrates. Of all known cellulose degrading microorganisms, *C. thermocellum* displays the highest rate of cellulose hydrolysis [44, 45].

The **short-term research objectives** presented in this thesis were to: 1) Identify genes relevant to pyruvate catabolism, ethanol production and H_2 metabolism. Using a comparative genomics approach, *C. thermocellum* was compared with select species from Phyla Firmicutes, Euryarchaeota, and Thermotogae. It was hypothesized that reported differences in product yields would be reflected in the compliment of genes encoding enzymes that participate in pyruvate metabolism and reductive end-product synthesis; 2)

The second research objective was to investigate how carbon and electron flux through the pyruvate branch point can be influenced by changes in a reactor environment during fermentation of cellobiose by *C. themocellum*. It was hypothesized that gas sparging, alternatively with N_2 or H_2, would significantly effect fermentation product yields that would be reflected in the expression of key genes involved in biofuels production. The final (3) research objective was to address the current lack of transformation methodologies by developing and evaluating both electroporation and conjugation protocols suitable for both mesophilic and thermophilic Clostridial species.

1.10 Thesis Outline

The preceeding introductory review was based on the manuscript, authored by Carere *et al*., titled "Third Generation Biofuels via Direct Cellulose Fermentation" [58]. The second chapter of this thesis reports on the genomics of cellulolytic Clostridia and seeks to examine the link between reported biofuel chemical yields with genome content and reaction thermodynamics. A comparative meta-analysis of select Firmicutes, Euryarchaeota, and Thermotogae was performed and includes reported end-product yields, and genes encoding pyruvate metabolism and reductive end-product synthesis pathways. The contents of this chapter were published by Carere *et al*. in the manuscript "Linking genome content to biofuel production yields: A meta-analysis of major catabolic pathways among select H_2 and ethanol-producing bacteria" (BMC microbiology).

Chapter three examines how metabolic flux in *Clostridium thermocellum* ATCC 27405 may be influenced during batch fermentation of cellobiose by elevated N_2 and H_2 gas sparging. End-product profiles, the transcription of key genes involved in pyruvate

catabolism, as well as hydrogen and ethanol synthesis and associated enzyme activities, are reported.

Chapter four addresses the lack of efficient transformation strategies amoung Clostridia and reports on the development of both electrotransformation and conjugative plasmid transfer protocols into recipient thermophilic species *C. thermocellum* (strains ATCC 27405 and DSM 1313), *Thermoanaerobacter pseudethanolicus* 39E, and the mesophilic, cellulolytic bacterium, *Clostridium termitidis* CT1112. The role of host restriction methylation systems is discussed as a likely obstacle to achieve high transformation efficiency/frequency.

Finally, chapter five summarizes the contributions of this thesis to the field of cellulolytic Clostridia and considers how the techniques employed and the findings reported can be applied to future research including the development of rational metabolic engineering techniques. The research objectives and hypotheses stated in the preceeding section are briefly discussed.

Chapter 2: Linking Genome Content to Biofuel Production Yields: A Meta-Analysis of Major Catabolic Pathways Among Select H_2 and Ethanol-Producing Microorganisms

2.1 Abstract

Fermentative bacteria offer the potential to convert lignocellulosic waste-streams into biofuels such as hydrogen (H_2) and ethanol. Current fermentative H_2 and ethanol yields, however, are below theoretical maxima, vary greatly among organisms, and depend on the extent of metabolic pathways utilized. For fermentative H_2 and/or ethanol production to become practical, biofuel yields must be increased. A comparative meta-analysis of (i) reported end-product yields, and (ii) genes encoding pyruvate metabolism and end-product synthesis pathways was undertaken in order to both identify suitable biomarkers for screening a microorganism's potential of H_2 and/or ethanol production, and to identify targets for metabolic engineering to improve biofuel yields. Our interest in H_2 and/or ethanol optimization restricted our meta-analysis to organisms with sequenced genomes and limited branched end-product pathways. These included members of the Firmicutes, Euryarchaeota, and Thermotogae. Bioinformatic analysis revealed that the absence of genes encoding acetaldehyde dehydrogenase and bifunctional acetaldehyde/alcohol dehydrogenase (AdhE) in *Caldicellulosiruptor*, *Thermococcus*, *Pyrococcus*, and *Thermotoga* species coincide with high H_2 yields and low ethanol production. Organisms containing genes (or activities) for both ethanol and H_2 synthesis pathways (i.e. *Caldanaerobacter subterraneus* subsp. *tengcongensis*, *Ethanoligenens harbinense*, and *Clostridium* species) had relatively uniform mixed product patterns. The absence of hydrogenases in *Geobacillus* and *Bacillus* species did

not correlate with high ethanol production, but rather high lactate production. Only *Thermoanaerobacter pseudethanolicus* produced relatively high ethanol and low H_2 yields. This may be attributed to the presence of genes encoding proteins that promote NADH production. Lactate dehydrogenase and pyruvate:formate lyase activities are not conducive to ethanol and/or H_2 production. While the type(s) of encoded hydrogenases appear to have little impact on H_2 production in organisms that do not encode ethanol producing pathways, they do influence reduced end-product yields in those that do**.** Here we show that the complement of genes encoding pathways involved in pyruvate catabolism and end-product synthesis can be used to approximate potential end-product distribution patterns. A number of genetic biomarkers for streamlining ethanol and H_2 producing capabilities were identified. By linking genome content, reaction thermodynamics, and end-product yields, potential targets for the optimization of either ethanol or H_2 yields through metabolic engineering are proposed.

2.2 Introduction

Fuel derived from waste-stream lignocellulosic biomass via consolidated bioprocessing is a renewable and carbon-neutral alternative to current petroleum-based fuels [59-61]. Consequently, considerable effort is being made to characterize species capable of efficiently converting lignocellulosic substrates into biofuels. An ideal biofuel-producing microorganism should posses several key features, including: (i) high yields of the desired product, (ii) simultaneous utilization of sugars (cellulose, hemicellulose, pectin), and (iii) growth at elevated temperatures, and (iv) low product inhibition.

Recent studies have focused on the characterization of numerous cellulose and hemicellulose degrading species of bacteria [24, 62, 63]. To fully exploit the biofuel producing potential of these organisms, several genomes have been sequenced and are now available for analysis (http://genome.jgi-psf.org/). While some hemicellulolytic or cellulolytic microorganisms are capable of hydrogen (H_2) or ethanol production via fermentation, end-product yields typically are far lower than their maximum theoretical values (4 mol H_2 or 2 mol ethanol mol^{-1}- glucose) when cells are grown in pure culture. This is due to the presence of branched catabolic pathways that divert carbon and/or electrons away from a particular desired end-product [11]. Strategies that optimize yields for a single biofuel (H_2 or ethanol) can only be developed through a detailed knowledge of the relationships between genome content, gene and gene product expression, pathway utilization, and end-product synthesis patterns.

Given that our primary focus is to optimize ethanol and/or H_2 yields, we restricted our meta-analysis to sequenced organisms with limited branched end-product pathways

(i.e. organisms that do not produce butyrate, butanol, propionate, propanol, and acetoin) for which end-product data was available. These included members of the Firmicutes (*Clostridium*, *Caldicellulosiruptor*, *Thermoanaerobacter*, *Caldanaerobacter*, *Ethanoligenens*, *Geobacillus*, and *Bacillus* species), Euryarchaeota (*Thermococcus* and *Pyrococcus* species), and Thermotogae (*Thermotoga* species). A list of species analyzed and corresponding GenBank accession numbers are summarized in Table 2.1. With the exception of *Caldanaerobacter subterraneus* subsp. *tengcongensis*, *Thermoanaerobacter pseudethanolicus*, *Pyrococcus furiosus*, *Geobacillus thermoglucosidasius*, and *Bacillus cereus*, all organisms were capable of cellulose and/or xylan saccharification.

Table 2.1. Ethanol and H_2 producing organisms included in meta-analysis of end-product yields and genome content.

Organism	Synonyms	Taxon ID	GenBank #	Sequencing Center	Phyla	C sources
Caldicellulosiruptor saccharolyticus DSM 8903		351627	NC_009437	DOE Joint Genome Institute	F	S,C,X
Caldicellulosiruptor besci Z-1320	*Anaerocellum thermophilum*; Z-1320	521460	NC_012036	DOE Joint Genome Institute	F	S,C,X
Pyrococcus furiosus DSM 3638		186497	AE009950	Univ of Maryland, Univ of Utah	E	S,C,X
Thermococcus kodakaraensis KOD1		69014	NC_006624	Kwansei Gakuin Univ, Kyoto University	E	S
Thermotoga neapolitana DSM 4359	ATCC 49049; JCM 10099; NS-E	309803	NC_011978	Genotech corp.	T	S,C
Thermotoga petrophila RKU-1		390874	NC_009486	DOE Joint Genome Institute	T	S,C,X
Thermotoga maritima MSB8	DSM 3109	243274	NC_000853	J. Craig Venter Institute	T	S,C,X
Caldanaerobacter subterraneus subsp. *tengcongensis* MB4	*Thermoanaerobacter tencongensis*	273068	NC_003869	Beijing Genomics Institute, The Institute of Microbiology, China	F	S
Ethanoligenens harbinense YUAN-2T	DSM 18485	663278	NC_014828	DOE Joint Genome Institute	F	S,C
Clostridium cellulolyticum H10		394503	NC_011898	DOE Joint Genome Institute	F	S,C,X
Clostridium phytofermentans ISDg	ATCC 700394	357809	NC_010001	DOE Joint Genome Institute	F	S,C,X
Clostridium thermocellum ATCC 27405	DSM 1237	203119	NC_009012	DOE Joint Genome Institute, University of	F	S,C,X

				Rochester		
Clostridium thermocellum DSM 4150	JW20	492476	ABVG00000000	DOE Joint Genome Institute	F	S,C,X
Thermoanaerobacter pseudethanolicus 39E	ATCC 33223	340099	NC_010321	DOE Joint Genome Institute	F	S,X
Geobacillus thermoglucosidasius C56-YS93		634956	NC_015660	DOE Joint Genome Institute	F	S
Bacillus cereus ATCC 14579	DSM 31	226900	NC_004721	Integrated Genomics Inc.	F	S

National Center for Biotechnology Information taxon IDs, GenBank accession numbers, corresponding sequencing centers responsible for the generation of the genome sequences data analyzed in this study are provided. Phyla (F; Firmicutes: E;Euryarchaeota: T; Thermotogae), and polymeric carbon sources degraded (S; starch: C; cellulose: X; xylose) by each organism are indicated

The various metabolic branches involved in pyruvate formation from phosphoenol pyruvate (PEP) and the subsequent catabolism of pyruvate into end-products was the focus of this analysis. Although studies comparing the ethanol- and H_2-producing potential of several cellulose degrading bacteria have been previously published [64-66], a comprehensive comparison of the major biofuel producing pathways at the genome level has not yet been reported. Here, a comparison of the genes encoding proteins involved in (i) pyruvate metabolism, (ii) ethanol synthesis, and (iii) H_2 metabolism is presented in order to rationalize reported end-product yields. Results indicate that the presence or absence of specific genes dictating carbon and electron flow towards end-products may be used to infer end-product synthesis patterns and help develop informed metabolic engineering strategies for optimization of ethanol and H_2 yields. Furthermore, certain genes may be suitable biomarkers for screening novel microorganisms' potential for biofuels production.

2.3 Methods

2.3.1 Comparative Analysis of Genome Annotations

All sequence data and gene annotations were accessed using the Joint Genome Institute's Integrated Microbial Genomes (IMG) database [67]. Gene annotations presented in this paper reflect the numbering of the final assembly or most recent drafts available (July, 2012). Comparative analyses were performed using the IMG database. In brief, analyses of all genomes (Table 2.1) were conducted using three annotation databases independently: i) Clusters of Orthologs Groups (COGs)[68], ii) KEGG Orthology assignments (KO) [69], and (iii) TIGRFAMs [70]. Genes identified using a single database were cross-referenced against the others to identify genes of interest.

Functional annotations of the identified genes were evaluated on a case-by-case basis and decisions regarding the annotation accuracy were made using a combination of manual analysis of genomic context, literature searches, and functional prediction through RPS-BLAST using the Conserved Domain Database website [71].

Hydrogenases were classified based on phylogenetic relationships of hydrogenase large subunits according to Calusinska *et al*. [72]. The evolutionary history was inferred using the Neighbor-Joining method [73]. The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa analyzed [74]. The evolutionary distances were computed using the Poisson correction method [75] and are in the units of the number of amino acid substitutions per site. The analysis involved 50 amino acid sequences. All ambiguous positions were removed for each sequence pair. There were a total of 863 positions in the final dataset. Evolutionary analyses were conducted in MEGA5 [76]. Thermodynamic calculations were performed using values provided by Thauer *et al.* [77] and the CRC Handbook of Chemistry and Physics [77, 78]. BioEdit v.7.0.9.0 [79] was used to perform ClustalW sequence alignments.

2.4 Results and Discussion

2.4.1 Survey of End-Product Yields

A literature survey of end-product yields (normalized to mol end-product mol^{-1}-hexose equivalent) of the species surveyed in this study is summarized in Table 2.2. While it is difficult to perform a direct comparison of end-product yields from available literature due to different growth conditions employed (ex. growth substrate, carbon loading, reactor conditions, etc.), and further difficult to validate these data due to incomplete end-product quantifications and lack of corresponding carbon balances and

O/R ratios, it still provides a good approximation of molar end-product yields based on substrate utilization. Calculated end-product yields reveal that the *Caldicellulosiruptor*, *Pyrococcus*, *Thermococcus*, and *Thermotoga* species surveyed, produced, in most cases, near-maximal H_2 yields with concomitant CO_2 and acetate production, and little or no ethanol, formate, and lactate [80-96]. It is important to note that while some studies [85-87, 90, 91, 95] report lower overall end-product yields, likely due to a large amount of carbon flux being directed towards biomass production under a given growth condition, H_2:ethanol ratios remain high. *Cal. subterraneus* subsp. *tengcongensis*, *E. harbinense*, and *Clostridium* species displayed mixed end-product fermentation patterns, with comparatively lower H_2, CO_2, and acetate yields, higher ethanol yields, and generally low formate and lactate yields [66, 97-103]. *Ta. pseudethanolicus* produced the highest ethanol yields of the organisms surveyed with little concomitant H_2, acetate, and lactate production, and no formate synthesis [104-106]. *G. thermoglucosidasius* and *B. cereus* produced the highest lactate and formate yields, moderate ethanol and acetate yields, and low H_2 and CO_2 yields [107, 108].

Table 2.2. Summary of end-product yields, optimal growth temperatures, total molar reduction values of H_2 + ethanol (RV_{EP}), and growth conditions employed

Organism	Growth temp (°C)	End Products (mol/mol hexose equivalent)							Growth condition	Ref
		H_2	CO_2	Acetate	Ethanol	Formate	Lactate	RV_{EP}		
Ca. saccharolyticus DSM 8903	70	4.0	1.8	NR	ND	ND	ND	4.0	Cont., 1.1 g l^{-1} glucose (D=0.09 h^{-1})	[80]
		3.6	1.5	1.6	ND	ND	ND	3.6	Cont., 4.1 g l^{-1} glucose (D=0.1 h^{-1})	[80]
		3.5	NR	2.1	NR	NR	NR	3.5	Batch, 10 g l^{-1} sucrose	[81]
		2.5	1.4	1.4	ND	ND	0.1	2.5	Batch, 10 g l^{-1} glucose	[82]
Ca. bescii DSM 6725	75	✓	✓	✓	NR	NR	✓	NA		[83, 84]
P. furiosus DSM 3638	100	3.8	1.9	1.5	0.1	NR	NR	4.0	Cont, cellobiose (D=0.45 h^{-1})	[85][A]
		3.5	1.0	1.4	ND	NR	ND	3.5	Batch, 1.9 g l^{-1}, maltose	[86][A]
		2.9	1.9	0.8	0.1	NR	ND	3.1	Batch, 2 g l^{-1} maltose	[87][B]
		2.8	0.9	1.2	ND	NR	ND	2.8	Batch, 3.5 g l^{-1}, cellobiose	[86][A]
		2.6	1.4	1.0	ND	NR	NR	2.6	Cont, maltose (D=0.45 h^{-1})	[85][A]
Th. kodakaraensis KOD1	85	3.3	1.8	1.1	NR	NR	NR	3.3	Cont, starch (D=0.2 h^{-1})	[88][C]
T. neapolitana DSM 4359	80-85	3.8	2.0	1.8	ND	NR	0.1	3.8	Batch, 2.5 g l^{-1} glucose	[89]
		3.2	NR	1.9	NR	NR	NR	3.2	Batch (N_2 sparged), 7.0 g l^{-1} glucose	[90]

		2.4	NR	1.1	NR	NR	0.7	2.4	Batch, 1.1 g l^{-1} glucose	[91]
		1.8	NR	1.0	NR	NR	NR	1.8	Batch, 7.5 g l^{-1} glucose	[96]
		1.8	NR	1.5	NR	NR	NR	1.8	Batch, 7.0 g l^{-1} glucose	[90]
T. petrophila RKU-1	80	3.7	0.4	1.8	NR	NR	0.3	3.7	Batch, 1 g l^{-1} glucose	[92]
T. maritima MSB8	80	4.0	2.0	2.0	NR	ND	NR	4.0	Batch, 2 g l^{-1} glucose	[94]
		2.2	1.1	1.0	ND	NR	0.3	2.2	Batch, 3 g l^{-1} glucose	[95]
		1.7	NR	1.0	NR	NR	NR	1.7	Batch, 7.5 g l^{-1} glucose	[96]
Cal. subterraneus subsp. *tengcongensis* MB4	75	2.8	NR	1.4	0.6	NR	ND	4.0	Cont, starch (D=0.27 h^{-1})	[98]
		NR	NR	2.0	ND	NR	ND	NA	Cont (N_2 sparged), glucose (D=0.24 h^{-1})	[98]
		0.3	1.5	1.0	0.7	NR	ND	1.7	Batch, 4 g l^{-1} glucose	[97]
E. harbinense YUAN-3T	35	2.8	✓	0.7	1.1	ND	ND	5.0	Batch, 20 g l^{-1} glucose	[99]
C. cellulolyticum H10	37	1.6	1.0	0.8	0.3	ND	NR	2.2	Batch, 5 g l^{-1} cellulose	[100]
		1.8	1.1	0.8	0.4	ND	NR	2,6	Batch, 5 g l^{-1} cellobiose	[100]
C. phytofermentans ISDg	35-37	Major	Major	0.6	1.4	0.1	0.3	NA	Batch, 34 g l^{-1} cellobiose	[101]
		1.0	0.9	0.6	0.5	0.1	NR	2.0	Batch, 5 g l^{-1} cellulose	[100]
		1.6	1.2	0.6	0.6	ND	NR	2.8	Batch, 5 g l^{-1} cellobiose	[100]
C. thermocellum ATCC 27405	60	0.8	1.1	0.7	0.8	0.3	ND	2.4	Batch, 1.1 g l^{-1} cellobiose	[66]
		1.0	0.8	0.8	0.6	0.4	0.4	2.2	Batch, 4.5 g l^{-1} cellobiose	[102]

C. thermocellum DSM 4150	60	1.8	1.7	0.9	0.8	ND	0.1	3.4	Batch, 2 g l^{-1} glucose	[103]
		0.6	1.8	0.3	1.4	ND	0.2	3.4	Batch, 27 g l^{-1} cellobiose	[103]
Ta. Pseudethanolicus ATCC 33223	65	0.1	2.0	0.1	1.8	NR	0.1	3.7	Batch, 8 g l^{-1} glucose	[106]
		NR	NR	NR	1.6	NR	<0.1	3.2	1 g l^{-1} xylose	[104]
		NR	NR	0.4	1.0	NR	<0.1	2.0	Batch, 20 g l^{-1} xylose	[105]
		NR	NR	0.2	0.4	NR	1.1	0.8	Batch, 20 g l^{-1} glucose	[105]
G. thermoglucosidasius M10EXG[D]	60	NR	NR	0.6	0.4	1.0	0.9	0.8	Batch, 10 g l^{-1} glucose	[108]
B cereus ATCC 14579	35	NR	0.1	0.2	0.2	0.3	1.1	0.4	Batch, 3.6 g l^{-1} glucose	[107]

[A]~0.5 mol alanine per mol-hexose produced on cellobiose and maltose;

[B]Produces H_2, CO_2, volatile fatty acids, and NH_3 on peptides in the absence of carbon source;

[C]~0.5 mol alanine per mol-hexose produced on starch;

[D]Only *G. thermoglucosidasuis* strain C56-TS93 has been sequenced but no end-product data is available. Strain M10EXG was used for end-product yield comparisons instead;

.

Abbreviations: NR, not reported; ND, not detected; NA, not applicable; Major, reported as major product without absolute values; ✓, reported as present with no values indicated; Cont, continuous culture; D, dilution rate.

While reported yields vary considerably for each organism, it is important to note that different growth conditions may influence end-product yields through regulation of gene and gene product expression [98, 109], and modulation of metabolic flux and intracellular metabolite levels [110, 111] that may act as allosteric regulators [112, 113]. Variations in fermentation conditions, including substrate availability/dilution rates [102, 109-111, 114-117], substrate composition [110, 118-123], media composition [111], pH [124], gas partial pressures [90, 98, 125, 126], growth phase [113], and accumulation of end-products [103, 118, 125, 127, 128] have been shown to influence end-product yields. Hence, while genome content alone cannot be used to predict end-product yields with accuracy, it can reflect end-product distribution profiles.

2.4.2 Genome Comparison of Pyruvate Metabolism and End-Product Synthesis Pathways

The assemblage of genes encoding proteins involved in pyruvate metabolism and end-product synthesis dictate, in part, how carbon and electron flux is distributed between the catabolic, anabolic, and energy producing pathways of the cell. The flow of carbon and electrons from PEP towards end-products may be separated into branch-points or nodes which include (i) the PEP/oxaloacetate/pyruvate node, (ii) the pyruvate/lactate/acetyl-CoA node, (iii) the acetyl-CoA/acetate/ethanol node, and the (iv) ferredoxin/NAD(P)H/H_2 node [129]. Several different enzymes may be involved in the conversion of intermediate metabolites within these nodes. These enzymes, and the presence of corresponding genes encoding these proteins in each of the organisms surveyed, are summarized in Figure 2.1. The oxidation of electron carriers (NADH and/or reduced ferredoxin) is required for maintaining glycolytic flux and leads to the

ultimate production of reduced products (ethanol, lactate, and H_2). Thus, distribution of carbon and electron flux among different pathways can influence levels of reduced electron carrier pools, which in turn can dictate end-product distribution patterns. Genome content can be used to resolve the relationship between carbon and electron flux with end-product distribution.

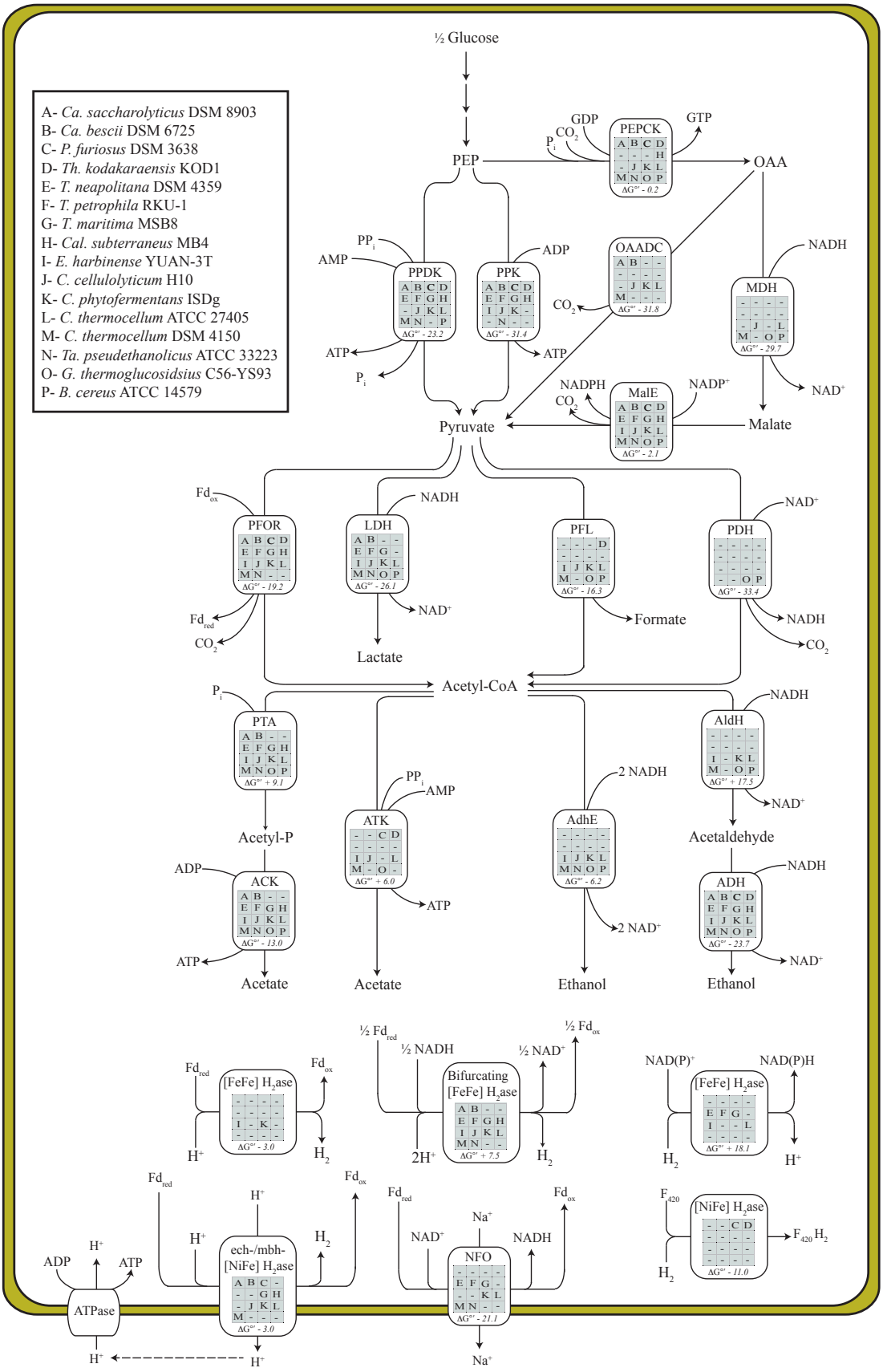

A- Ca. saccharolyticus DSM 8903
B- Ca. bescii DSM 6725
C- P. furiosus DSM 3638
D- Th. kodakaraensis KOD1
E- T. neapolitana DSM 4359
F- T. petrophila RKU-1
G- T. maritima MSB8
H- Cal. subterraneus MB4
I- E. harbinense YUAN-3T
J- C. cellulolyticum H10
K- C. phytofermentans ISDg
L- C. thermocellum ATCC 27405
M- C. thermocellum DSM 4150
N- Ta. pseudethanolicus ATCC 33223
O- G. thermoglucosidsius C56-YS93
P- B. cereus ATCC 14579
½ Glucose
PEP
PEPCK
GDP
GTP
OAA
OAADC
MDH
NADH
NAD+
PPDK
PPK
ADP
AMP
ATP
MalE
NADPH
NADP+
Malate
Pyruvate
PFOR
LDH
PFL
PDH
Formate
Lactate
Acetyl-CoA
PTA
Acetyl-P
ACK
Acetate
ATK
AldH
Acetaldehyde
AdhE
ADH
Ethanol
2 NADH
2 NAD+
[FeFe] H2ase
Bifurcating [FeFe] H2ase
[NiFe] H2ase
ech-/mbh-[NiFe] H2ase
NFO
ATPase

Figure 2.1. Comparison of putative gene products involved in pyruvate metabolism and end-product synthesis among select hydrogen and ethanol-producing species. Presence of putative gene products are indicated in matrix with respective letters corresponding to selected organism (see legend). Numbers indicate standard free energies of reaction ($\triangle G^{\circ\prime}$) corresponding to a particular enzyme. Abbreviations: PEPCK, phosphoenolpyruvate carboxykinase; OAADC, oxaloacetate decarboxylase; MDH, malate dehydrogenase; ME, malic enzyme; PPK, pyruvate kinase; PPDK, pyruvate phosphate dikinase; LDH, lactate dehydrogenase; PFL, pyruvate formate lyase; PFOR, pyruvate:ferredoxin oxidoreductase; PDH, pyruvate dehydrogenase; ADH, alcohol dehydrogenase; ALDH, acetaldehyde dehydrogenase; ADHE, bifinctional acetaldehyde/alcohol dehydrogenase; ACK, acetate kinase; PTA, phosphotransacetylase; LDH, lactate dehydrogenase; NFO, NADH:Fd oxidoreductase.

2.4.3 Genes Involved in Pyruvate Synthesis

All organisms considered in this study utilize the Embden-Meyerhof-Parnas pathway for conversion of glucose to PEP with the following notable variations. Alignments key residues of phosphofructokinase (PFK) according to Bapteste *et al.* [130, 131], suggest that *Cal. subterraneus* subsp. *tengcongensis*, *E. harbinense*, *G. thermoglucosidasius*, and *B. cereus* encode an ATP-dependent PFK, while *Thermotoga*, *Caldicellulosiruptor*, *Clostridium*, and *Thermoanaerobacter* species encode both an ATP-dependent PFK, as well as a pyrophosphate (PP_i)-dependent PFK [130, 131] (Supplementary Table 1). *P. furiosus* and *Th. kodakaraensis,* alternatively, encode both ADP-dependent hexokinase ADP-dependent and PFK. Furthermore, while bacteria catalyze the oxidation of glyceraldehyde-3-P to 3-phosphoglycerate (yielding NADH and ATP) with glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and phosphoglycerate kinase (PGK), archea (*P. furiosus* and *Th. kodakaraensis*) catalyze the same reaction via glyceraldehyde-3-phosphate:ferredoxin oxidoreductase (GAPFOR). This enzyme reduces ferredoxin (Fd) rather than NAD^+ and does not produce ATP [132].

In contrast to the generally conserved gene content required for the production of PEP, a number of enzymes may catalyze the conversion of PEP to pyruvate [129] (Figure 2.1; Table 2.3). PEP can be directly converted into pyruvate via an ATP-dependent pyruvate kinase (PPK), or via an AMP-dependent pyruvate phosphate dikinase (PPDK). All strains considered in this review encode both *ppk* and *ppdk*, with the exception of *C. thermocellum* strains, which do not encode a *ppk*, and *E. harbinense*, *G. thermoglucosidasius*, and *B. cereus*, which do not encode *ppdk*. Given that the formation of ATP from ADP and P_i is more thermodynamically favorable than from AMP and PP_i

($\triangle G°'$ = 31.7 vs. 41.7 kJ mol^{-1}), production of pyruvate via PPK is more favorable than via PPDK [77].

Table 2.3. Genes encoding proteins involved in interconversion of phosphenolpyruvate and pyruvate.

Organism	Gene						
	eno	***ppk***	***ppdk***	***pepck***	***oaadc***	***mdh***	***malE***
Standard free energy (△G°´)	ND	-31.4	-23.2	-0.2	-31.8	-29.7	-2.1
Ca. saccharolyticus DSM 8903	Athe_1403	Athe_1266	Athe_1409	Athe_0393	Athe_1316-1319		Athe_1062
Ca. bescii DSM 6725	Csac_1950	Csac_1831	Csac_1955	Csac_0274	Csac_2482-2485		Csac_2059
P. furiosus DSM 3638	PF0215 PF1641	PF1188	PF0043	PF0289			PF1026
Th. kodakaraensis KOD1	TK1497 TK2106	TK0511	TK0200 TK1292	TK1405			TK1963
T. neapolitana DSM 4359	CTN_1698	CTN_0477	CTN_0413				CTN_0126
T. petrophila RKU-1	Tpet_0050	Tpet_0716	Tpet_0652				Tpet_0379
T. maritima MSB8	TM0877	TM0208	TM0272				TM0542
Cal. subterraneus subsp. *tengcongensis* MB4[A]	TTE1759	TTE1815	TTE0164 TTE0981	TTE1783			TTE2332
E. harbinense YUAN-3T	Ethha_2662	Ethha_0305					Ethha_0739
C. cellulolyticum H10	Ccel_2254	Ccel_2569	Ccel_2388	Ccel_0212	Ccel_1736-1738	Ccel_0137	Ccel_0138
C. phytofermentans ISDg	Cphy_3001	Cphy_0741 Cphy_2900	Cphy_0651	Cphy_3853	Cphy_2433-2434		Cphy_0409

C. thermocellum ATCC 27405	Cthe_0143		Cthe_1253 Cthe_1308	Cthe_2874	Cthe_0699-0701	Cthe_0345	Cthe_0344
C. thermocellum DSM 4150			CtherDRAFT_1661 CtherDRAFT_1896	CtherDRAFT_1742	CtherDRAFT_0819-0822	Yes[A]	Yes[A]
Ta. pseudethanolicus 39E	Teth39_0735	Teth39_0684	Teth39_1358 Teth39_2098	Teth39_0711			Teth39_0337
G. thermoglucosidasius C56-YS93	Geoth_0446	Geoth_0898		Geoth_0811		Geoth_0904 Geoth_3508	Geoth_1713 Geoth_2444
B.cereus ATCC 14579	BC5135	BC3323 BC4599	BC3087	BC4762		BC4592 BC2959	BC0580 NAD) BC1741 (NAD) BC4604 (NADP)

[A]Genes have been verified by PCR amplification (*unpublished*)

Abbreviations: *eno*, enolase; *ppk*, pyruvate kinase; *ppdk*, pyruvate phosphate dikinase; *pepck*, phosphoenolpyruvate carboxykinase; *oaadc*, oxaloacetate decarboxylase; *mdh*, malate dehydrogenase; *malE*, malic enzyme

Flux balance analysis integrated with RNAseq data suggests higher carbon and electron flux in *C. thermocellum* ATCC 27405 is directed through enzymes capable of direct, rather than indirect, conversion of PEP to pyruvate [133]. However, *C. cellulolyticum* mutation studies suggests that a portion of PEP can also be converted to pyruvate via a "*malate shunt*" [134]. This PPK/PPDK bypass system utilizes either (i) phosphoenolpyruvate carboxykinase (PEPCK), malate dehydrogenase (MDH), and malic enzyme (MalE), or (ii) PEPCK and oxaloacetate decarboxylase (OAADC), for the interconversion of PEP and pyruvate (Figure 2.1). While PEPCK provides a pathway for energy conservation via ATP (or GTP) production, MDH and MalE permit transhydrogenation from NADH to $NADP^+$ [127], generating additional reducing equivalents required for biosynthesis. *G. thermoglucosidasius*, *B. cereus*, *C. thermocellum* (ATCC 27405), and *C. cellulolyticum* contain *pepck*, *mdh* and *malE* suggesting that they are capable of transhydrogenation using these proteins. Although the draft genome of *C. thermocellum* DSM 4150 does not include genes encoding MDH and MalE, we have verified their presence via PCR amplification (unpublished results). Deletion of *mdh* in *C. cellulolyticum* resulted in significant increases in lactate, and to a lesser extent ethanol yields, and reduced acetate production when grown on cellulose demonstrating carbon and electron flux through MDH in wild type strains [134]. It seems evident that in the absence of MDH, transhydrogenation was reduced, and thus the resulting increase in NADH:NADPH ratios promote lactate and ethanol production, while decreasing NADPH levels for biosynthesis.

A number of organisms analyzed encode *pepck* and *oaadc* (*Ca. bescii*, *Ca. saccharolyticus*, *C. cellulolyticum*, *C. phytofermentans*, and *C. thermocellum*), also

allowing for indirect conversion of PEP to pyruvate via an oxaloacetate intermediate. While the redirection of carbon and electron flux through this pathway likely has little effect on product yields, synthesis of GTP, versus ATP, may promote transcription and protein synthesis. Finally, *Cal. subterraneus*, *E. harbinense*, *P. furiosus*, *Th. kodakaraensis*, *Ta. pseudethanolicus*, and *Thermotoga* species do not encode all of the proteins required for a "malate shunt" and consequentially the catalysis of PEP to pyruvate must be achieved via PPK and/or PPDK.

2.4.4 Genes Involved in Pyruvate Catabolism

The pyruvate/lactate/acetyl-CoA node plays an important role in regulating carbon flux and electron distribution and dramatically affects end-product distribution. The NADH-dependent reduction of pyruvate to lactate via fructose-1,6-bisphosphate activated lactate dehydrogenase (LDH) [112] diverts reducing equivalents away from biofuels such as H_2 and ethanol. Alternatively, the oxidative decarboxylation of pyruvate to acetyl-CoA via pyruvate dehydrogenase (*pdh*) or pyruvate:ferreodoxin oxidoreductase (*pfor*) generate NADH and reduced Fd, respectively. These reducing equivalents may then be oxidized during the production of H_2 or ethanol (Figure 2.1). Pyruvate may also be catabolised to acetyl-CoA via pyruvate:formate lyase (*pfl*) yielding formate in the process. In some enterobacteria, formate is further oxidized to CO_2, releasing H_2, through the action of a multisubunit formate hydrogen lyase (FHL) complex [135]. However, FHL was not encoded in any of the organisms analysed.

With the exception of *Cal. subterraneus* subsp. *tengcongensis*, *P. furiosus*, and *Th. kodakaraensis*, *ldh* genes were identified in all organisms studied (Table 2.4). Surprisingly, while the production of lactate from pyruvate is highly favorable

thermodynamically ($\triangle G°' = -26.1$ kJ mol^{-1}), only *B. cereus*, *G. thermoglucosidasius*, and, under some conditions, *Ta. pseudethanolicus* and *T. neapolitana* produce high yields of lactate (> 0.5 mol mol-$glucose^{-1}$). In all other organisms surveyed, lactate production was either a minor end-product, not detected, or not reported under the reported growth conditions (Table 2.2). This suggests that the presence of *ldh* cannot be used to predict lactate production.

Table 2.4. Genes encoding proteins directly involved in pyruvate catabolism.

Organism	Gene			
	ldh	***pdh***	***pfor***	***pfl***
Standard free energy (△G°′)	-26.1	-33.4	-19.2	-16.3
Ca. saccharolyticus DSM 8903	Csac_1027		Csac_1458-1461 Csac_2248-2249	
Ca. bescii DSM 6725	Athe_1918		Athe_0874-0877 Athe_1708-1709	
P. furiosus DSM 3638			PF0965-PF0967, PF0971	
Th. kodakaraensis KOD1			TK1978, TK1982-1984	TK0289
T. neapolitana DSM 4359	CTN_0802		CTN_0680-CTN_0683	
T. petrophila RKU-1	Tpet_0930		Tpet_0905-Tpet_0908	
T. maritima MSB8	TM1867		TM0015-TM0018	
Cal. subterraneus subsp. *tengcongensis MB4*			TTE0445 TTE0960	
E. harbinense YUAN-3T, DSM 18485	Ethha_1350 Ethha_2705		Ethha_0231-0234	Ethha_1657
C. cellulolyticum H10	Ccel_2485		Ccel_0016 Ccel_1164	Ccel_2224 Ccel_2582
C. phytofermentans ISDg	Cphy_1117		Cphy_0603	Cphy_1174 Cphy_1417

	Cphy_1232		Cphy_3558	Cphy_2823
C. thermocellum ATCC 27405	Cthe_1053		Cthe_2390-2393 Cthe_2794-2797 Cthe_3120	Cthe_0505
C. thermocellum DSM 4150	CtherDRAFT_2943		CtherDRAFT_0414-0417 CtherDRAFT_1182-1185 CtherDRAFT_1311	CtherDRAFT_2234
Ta. pseudethanolicus 39E	Teth39_1997		Teth39_0289 Teth39_1842	
G. thermoglucosidasius C56-YS93	Geoth_3351	Geoth_0237-0239 Geoth_1595-1597 Geoth_2366-2368 Geoth_2479-2480 Geoth_2860-2863		Geoth_3895
B.cereus ATCC 14579	BC1924 BC4870 BC4996	BC3970-3973		BC0491

Abbreviations: *ldh*, lactate dehydrogenase; *pdh*, pyruvate dehydrogenase; *pfor*, pyruvate:ferredoxin oxidoreductase; *pfl*, pyruvate formate lyase

LDH is, in fact, allosterically activated by fructose-1,6-bisphosphate in *C. thermocellum* ATCC 27405, *Ca. saccharolyticus*, and *Thermoanaerobacter brockii* [112, 113, 118, 136]. While enzyme assays reveal high LDH activity in *C. thermocellum* [66, 128], most studies report only trace amounts of lactate. Islam *et al*. [102], however, demonstrated that lactate production was triggered in stationary-phase batch cultures only under excess cellobiose conditions. In *Thermoanaerobacter brockii,* Ben-Bassat *et al*. reported elevated lactate production as a consequence of accumulated intracellular fructose-1,6-bisphosphate when cultures were grown on glucose compared to starch [118]. Finally, Willquist and van Niel [113] reported that LDH in *Ca. saccharolyticus* was activated by FDP and ATP, and inhibited by NAD^+ and PP_i. An increase in fructose-1,6-bisphosphate, NADH:NAD^+ ratios, and ATP:PP_i ratios was observed during the transition from exponential to stationary phase in *Ca. saccharolyticus* cultures, and was accordingly accompanied by lactate production [113].

All organisms analyzed encode either *pdh* or *pfor*, but not both (Table 2.4). While *G. thermoglucosidasius* and *B. cereus* encode *pdh*, all other organisms analyzed encode *pfor*. Although *Caldicellulosiruptor*, *Clostridia*, and *Thermoanaerobacter* species studied appear to encode a putative *pdh*, there has been no enzymatic evidence to support the presence of PDH in these species. Thus far, only PFOR activity has been verified in *C. cellulolyticum* [114, 116] and *C. thermocellum* [66, 128]. The putative E1, E2, and E3 subunits of the *pdh* complex (Csac_0874-0872) in *Ca. saccharolyticus* were designated simply as a keto-acid dehydrogenase by van de Werken *et al*. [137]. Similarly, while genes encoding a putative *pdh* (Teth_0790-0793) are present in *Ta. pseudethanolicus*, genomic context strongly supports that this putative *pdh* is part of an

acetoin dehydrogenase complex, despite the absence reported acetoin production. In *Clostridia* species, putative *pdh*'s (Cthe_3449-3450, Cthe_1543) may actually encode 2-oxo acid dehydrogenase complexes, which share a common structure and homology to pyruvate dehydrogenase. These include 2-oxoglutarate dehydrogenase, branched-chain alpha-keto acid dehydrogenase, acetoin dehydrogenase complex, and the glycine cleavage complex. All organisms that encode a *pfor* also encode a Fd-dependent H_2ase, bifurcating H_2ase, and/or an NFO, and are thus capable of reoxidizing reduced Fd produced by PFOR. Conversely, *G. thermoglucosidasius* and *B. cereus*, which encode *pdh* but not *pfor*, do not encode enzymes capable of reoxidizing reduced Fd, and thus do not produce H_2. While the presence of PDH allows for additional NADH production that could be used for ethanol production, *G. thermoglucosidasius* and *B. cereus* end-product profiles suggest that this NADH is preferentially rexodized through lactate production rather than ethanol production. Pyruvate decarboxylase, a homotetrameric enzyme that catalyzes the decarboxylation of pyruvate to acetaldehyde was not encoded by any of the species considered in this study.

Given the requirement of reduced electron carriers for the production of ethanol/H_2, the oxidative decarboxylation of pyruvate via PDH/PFOR is favorable over PFL for the production of these biofuels. Genome analyses revealed that a number of organisms, including *P. furiosus*, *Ta. pseudethanolicus, Cal. subterraneus* subsp. *tencongensis*, and all *Caldicellulosiruptor* and *Thermotoga* species considered, did not encode PFL. In each of these species, the production of formate has neither been detected nor reported. Unfortunately, many studies do not report formate production, despite the presence of PFL. This may be a consequence of the quantification methods

used for volatile fatty acid detection. When formate is not produced, the total oxidation value of 2 CO_2 per mole glucose (+4), must be balanced with the production of H_2 and/or ethanol. Thus, the "total molar **r**eduction **v**alues of reduced e**n**d-**p**roducts (H_2 + ethanol)", termed RV_{EP}, should be -4, providing that all carbon and electron flux is directed towards end-product formation and not biosynthesis. Indeed, RV_{EP}'s were usually greater than 3.5 in organisms that do not encode *pfl* (*T. maritima*, *Ca. saccharolyticus*), and below 3.5 in those that do encode *pfl* (*C. phytofermentans*, *C. thermocellum*, *G. thermoglucosidasius*, and *B. cereus*; Table 2.2). In some studies, RV_{EP}'s were low due to a large amount of carbon and electron flux directed towards biosynthesis. In *G. thermoglucosidasius* and *B. cereus* RV_{EP}'s of H_2 plus ethanol ranged from 0.4 to 0.8 due to higher reported formate yields. The large differences in formate yields between organisms that encode *pfl* may be due to regulation of *pfl*. In *Escherichia.coli* [138, 139] and *Streptococcus bovis* [140, 141], *pfl* expression has been shown to be negatively regulated by AdhE. Thus presence of *pfl* alone is not a good indicator of formate yields.

2.4.5 Genes Involved in Acetyl-CoA Catabolism, Acetate Production, and Ethanol Production

The acetyl-CoA/acetate/ethanol node represents the third major branch-point that dictates how carbon and electrons flow towards end-products (Figure 2.1). Acetyl-CoA may be converted to acetate, with the concomitant production of ATP, either indirectly through an acetyl phosphate intermediate using phosphotransacetylase (*pta*) and acetate kinase (*ack*), or directly via acetate thiokinase *(atk*). Although both reactions produce ATP, the former uses ADP and P_i whereas the latter uses AMP and inorganic PP_i as

substrates for ATP synthesis. As a result, acetate production via *pta* and *ack* is more thermodynamically favorable than via *atk* ($\triangle G^{\circ\prime}$ = -3.9 vs. +6.0 kJ/mol, respectively) which is typically used for acetate assimilation. Of the organisms surveyed, *E. harbinense*, *G. thermodenitrificans*, *C. cellulolyticum*, both *C. thermocellum* strains, and *G. thermoglucosidasius* contain all three genes capable of converting pyruvate to acetate (Table 2.5). Conversely, *Cal. subterraneus* subsp. *tengcongensis*, *Thermotoga* and *Caldicellulosiruptor* species, *C. phytofermentans*, *Ta. pseudethanolicus*, and *B. cereus* encode only *pta* and *ack*, whereas *P. furiosus* and *Th. kodakaraensis* encode only *atk*.

Alternatively, acetyl-CoA may be converted into ethanol, during which 2 NADH (or NADPH) are oxidized, either directly via a fused acetaldehyde/alcohol dehydrogenase encoded by *adhE*, which has been proposed to be the key enzyme responsible for ethanol production [142, 143], or indirectly through an acetaldehyde intermediate via acetaldehyde dehydrogenase (*aldH*) and alcohol dehydrogenase (*adh*). While all organisms surveyed encoded multiple class IV Fe-containing ADHs (Table 2.5), the functions of these ADHs may vary with respect to substrate specificity (aldehyde length and substitution), coenzyme specificity (NADH vs. NADPH), and the catalytic directionality favored (ethanol formation vs. consumption) [66, 113-115, 128, 144-147]. Although there are reports of *in silico* determinations of substrate and cofactor specificity amongst ADHs, in our experience such resolutions are problematic [148, 149]. Often times, the gene neighborhoods of identified ADHs were suggestive that the physiological role of many enzymes was not ethanol production. This is evident in *Ca. saccharolyticus*, which does not produce ethanol despite reported NADPH-dependent ADH activity [113].

Table 2.5. Genes encoding proteins involved in end-product synthesis from acetyl-CoA.

Organism	gene					
	pta	***ack***	***atk***	***aldH***	***adh***	***adhE***
Standard free energy (△G°′)	9.1	-13.0	6.0	17.5	-23.7	-6.2
Ca. saccharolyticus DSM 8903	Csac_2041	Csac_2040			Csac_0407 Csac_0554 Csac_0622 Csac_0711 Csac_1500	
Ca. bescii DSM 6725	Athe_1494	Athe_1493			Athe_0928 Athe_0224	
P. furiosus DSM 3638			PF1540 PF1787		PF0075 PF0608	
Th. kodakaraensis KOD1			TK0465 TK0665		TK1008 TK1569	
T. neapolitana DSM 4359	CTN_0945 CTN_1440	CTN_0411			CTN_0257 CTN_0369 CTN_0385 CTN_0580 CTN_1655 CTN_1756	
T. petrophila RKU-1	Tpet_1042 Tpet_1615	Tpet_0650			Tpet_0007 Tpet_0107 Tpet_0484 Tpet_0508	

					Tpet_0563 Tpet_0614 Tpet_0813	
T. maritima MSB8	TM1130 TM1755	TM0274			TM0111 TM0298 TM0412 TM0436 TM0820 TM0920	
Cal. subterraneus subsp. *tengcongensis MB4*	TTE1482	TTE1481			TTE0313 TTE0695 TTE0696 TTE1591	
E. harbinense YUAN-3T, DSM 18485	Ethha_2711	Ethha_2004	Ethha_1333	Ethha_0578 Ethha_0625	Ethha_0051 Ethha_0580 Ethha_1164 Ethha_2217 Ethha_2239	Ethha_1385
C. cellulolyticum H10	Ccel_2137	Ccel_2136	Ccel_0494 Ccel_1469		Ccel_0894 Ccel_1083 Ccel_3337	Ccel_3198
C. phytofermentans ISDg	Cphy_1326	Cphy_1327		Cphy_0958 Cphy_1178 Cphy_1416 Cphy_1428 Cphy_2418 Cphy_2642	Cphy_1029 Cphy_1421 Cphy_2463 Cphy_2463	Cphy_3925

				Cphy_3041		
C. thermocellum ATCC 27405	Cthe_1029	Cthe_1028	Cthe_0551	Cthe_2238	Cthe_0101 Cthe_0394 Cthe_2579	Cthe_0423
C. thermocellum DSM 4150	CtherDRAFT_2741	CtherDRAFT_2742	CtherDRAFT_2349	CtherDRAFT_1042	CtherDRAFT_0189 CtherDRAFT_0616 CtherDRAFT_2833	CtherDRAFT_1096
Ta. pseudethanolicus 39E	Teth39_1296	Teth39_1295			Teth39_0220 Teth39_1597 Teth39_1979	Teth39_0206
G. thermoglucosidasius C56-YS93	Cthe_3862	Geoth_0875	Geoth_0855 Geoth_0879 Geoth_2349	Geoth_0268 Geoth_0652 Geoth_3494	Geoth_1572 Geoth_1941 Geoth_0631	Geoth_3879
B. cereus ATCC 14579	BC5387	BC4637		BC2832 BC3555 BC1285	BC0802 BC2529 BC2220	BC4365

Abbreviations: *pta*, phosphotransacetylase; *ack*, acetate kinase; *atk*, acetate thiokinase; *aldH*, acetaldehyde dehydrogenase; *adh*, alcohol dehydrogenase; *adhE*; bifunctional acetylaldehyde/alcohol dehydrogenase

P. furiosus, *Th. kodakaraensis*, and all *Thermotoga* and *Caldicellulosiruptor* species do not encode *adhE* or *aldH*, and therefore produce negligible or no ethanol. Given the absence of ethanol producing pathways in these species, reducing equivalents are disposed of through H_2 production via H_2ases and/or lactate production via LDH. Surprisingly, while *Cal. subterraneus* subsp. *tengcongensis* also does not appear to encode *aldH* or *adhE*, NADPH-dependent AldH and both NADH and NADPH-dependent ADH activities, as well as ethanol production, have been reported by Soboh *et al*. [98]. Similarly, *Caldicellulosiruptor obsidiansis*, which does not encode *aldH* or *adhE*, does produce trace levels of ethanol, suggesting that the various encoded ADHs may have broad substrate specificities [150]. Although *C. cellulolyticum* and *Ta. pseudethanolicus* do not encode *aldH*, they do encode *adhE*, and thus are capable of ethanol production. Of the organisms surveyed, only *G. thermoglucosidasius* and *C. cellulolyticum* encoded *aldH* and *adh* but no *adhE*, and produced moderate amounts of ethanol (~0.4 mol per mol hexose). Conversely, a number of organisms (*E. harbinense*, *C. phytofermentans*, both *C. thermocellum* strains, *G. thermoglucosidasius*, and *B. cereus*) encoded *aldH*, *adh*, and *adhE*, all of which produce ethanol.

2.4.6 Hydrogenases

In addition to disposal of reducing equivalents via alcohol and organic acid production, electrons generated during the conversion of glucose to acetyl-CoA can be used to produce molecular hydrogen via a suite of [FeFe] and/or [NiFe] H_2ases. The incredible diversity of H_2ases has been extensively reviewed by Vignais *et al*. and Calusinska *et al*. [72, 151, 152]. H_2ases may be (i) monomeric or multimeric, (ii) can catalyze the reversible production of H_2 using various electron donors, including reduced

Fd and NAD(P)H, or (iii) can act as sensory H_2ases capable of regulating gene expression [153]. While most H_2ases can reversibly shuttle electrons between electron carriers and H_2, they are typically committed to either H_2-uptake or evolution, depending on reaction thermodynamics and the requirements of the cell *in vivo* [151]. While Fd-dependent H_2 production remains thermodynamically favorable at physiological concentrations ($\triangle G°'$ ~ -3.0 kJ mol^{-1}), potential production of H_2 from NAD(P)H ($\triangle G°'$ = +18.1 kJ mol^{-1}) becomes increasingly unfavorable with increasing hydrogen partial pressure [154]. Hence, Fd-dependent H_2ases are associated with H_2 evolution, whereas NAD(P)H-dependent H_2ases are more likely to catalyze H_2 uptake. Recent characterization of a heterotrimeric "bifurcating" H_2ase from *Thermotoga maritma* demonstrated that it can simultaneously oxidize reduced Fd and NADH to H_2 ($\triangle G°'$ ~ +7.5 kJ mol^{-1}), which drives the endergonic production of H_2 from NADH by coupling it to the exergonic oxidation of reduced Fd [155].

With the exception of *G. thermoglucosidasius* and *B. cereus,* which did not contain putative H_2ase genes, the genomes of all of the organisms surveyed encode multiple H_2ases. These H_2ases were classified based on i) the phylogenetic relationship of H_2ase large subunits (Supplementary Figures 1 and 2), according to Calusinska *et al*. [72], ii) H_2ase modular structure, and iii) subunit composition, based on gene neighbourhoods. Encoded [NiFe] H_2ases fell into 3 major subgroups including: (i) Fd-dependent, H_2-evolving, membrane-bound H_2ases (Mbh) and/or energy conserving [NiFe] H_2ases (Ech) capable of generating sodium/proton motive force (Group 4) [98], (ii) Soluble cofactor-dependent (F_{420} or NAD(P)H), bidirectional, cytoplasmic, heteromultimeric H_2ases (Group 3), and (iii) H_2-uptake, membrane bound H_2ases (Group

1) [152] (Supplementary Figure 1). Similarly, encoded [FeFe] H_2ases fell into 5 major subgroups including: (i) heterotrimeric bifurcating H_2ases, (ii) dimeric, NAD(P)H-dependent uptake H_2ases, (iii) monomeric, putatively Fd-dependent H_2ases, (iv) dimeric sensory H_2ases containing PAS/PAC sensory domains which may be involved in redox sensing, and (v) monomeric sensory H_2ases (Supplementary Figure 2). These sensory H_2ases are usually encoded upstream of trimeric bifurcating H_2ases (Table 2.6) and are often separated by a histidine/serine kinase suggesting a regulatory relationship between these two enzymes [72].

Table 2.6. Genes encoding putative hydrogenases, sensory hydrogenases, and NADH:Fd oxidoreductases using ferredoxin, coenzyme F_{420}, and NAD(P)H as electron carriers.

Organism	Hydrogenase and NADH:Fd Oxidoreductase Classification and Corresponding Genes						
	[NiFe] H_2ase		[FeFe] H_2ase				NFO
	Fd-dependent *ech* and *mbh* [G4]	**F_{420}-dependent** [G3] **and other** [G1]	**Bifurcating**	**Sensory**[A]	**NAD(P)H-dependent**	**Fd-dependent**	***rnf*-type**
Standard free energy (△G°′)*	-3.0	11	+7.5**	NA	18.1	18.1	-21.1***
Ca. bescii DSM 6725	Athe_1082-Athe_1087		Athe_1297-Athe_1299 [A1 TR(M3)]	Athe_1292 [D M2e]			
Ca. saccharolyticus DSM 8903	Csac_1534-Csac_1539		Csac_1862-Csac_1864 [A1 TR(M3)]	Csac_1857 [D M2e]			
P. furiosus DSM 3638	PF1423-PF1436	PF0891-PF0894 [G3] PF1329-PF1332 [G3]					
Th. kodakaraensis KOD1	TK2080-TK2093	TK2069-TK2072 [G3]					
T. neapolitana DSM 4359			CTN_1067-CTN1069 [TTH]	CTN_1071-CTN_1072 [CD(M2f)]	CTN_0485 [TTH]		CTN_0437-CTN_0442
T. petrophila RKU-1			Tpet_1367-Tpet_1369 [TTH]	Tpet_1371-Tpet_1372 [CD(M2f)]	Tpet_0723 [TTH]		Tpet_0675-Tpet_0680
T. maritima MSB8			TM1424-TM1426 [TTH]	TM1420-TM1422 [CD(M2f)]	TM0201 [TTH]		TM0244-TM0249

Cal.subterraneus subsp. *tengcongensis MB4*	TTE0123-TTE0134		TTE0892-TTE0894 A1 TR(M3)	TTE0887 **D M2e** TTE0697 **CD(M2f)**			
E. harbinense YUAN-3T			Ethha_2614-Ethha_2616 A8 TR(M3)	Ethha_0052 **CD(M2f)**	Ethha_2293 A7 D(M3)	Ethha_0031 B2 M2a	
C. cellulolyticum H10	Ccel_1686-Ccel_1691 Ccel_3363-Ccel_3371	Ccel_1070-Ccel_1071 G1	Ccel_2303-Ccel_2305 A8 TR(M3) Ccel_2232-Ccel_2234 A1 TR(M3) Ccel_2467-Ccel_2468 A1 TR(M3)	Ccel_2300-Ccel_2301 **CD(M2f)**		Ethha_2695 B3 M3a	
C. phytofermentans ISDg	Cphy_1730-Cphy_1735		Cphy_0087-Cphy_0089 A8 TR(M3) Cphy_3803-Cphy_3805 A1 TR(M3)	Cphy_0092-Cphy_0093 **CD(M2f)** Cphy_3798 **D M2e**	Cthe_3003-Cthe_3004	Cphy_2056 A5 M2c Cphy_0090 B1 M3a	Cphy_0211-Cphy_0216
C.thermocellum ATCC 27405	Cthe_3013-Cthe_3024		Cthe_0428-Cthe_0430 A8 TR(M3) Cthe_0340-Cthe_0342 A1 TR(M3)	Cthe_0425-Cthe_0426 **CD(M2f)** Cthe_0335 **D M2e**			Cthe_2430-Cthe_2435
C. thermocellum DSM 4150	CtherDRAFT_2162-CtherDRAFT_2173		CtherDRAFT_1101-CtherDRAFT_1103 A8 TR(M3) CtherDRAFT_2978 A1 TR(M3)	CtherDRAFT_1098-CtherDRAFT_1099 **CD(M2f)**	Yes[B]		CtherDRAFT_0369-CtherDRAFT_0375
Ta. pseudethanolicus 39E				Teth39_0221 **CD(M2f)**			Teth39_2119-Teth39_2124

	Teth39_1456-Teth39_1458 [A1] TR(M3)	Teth39_1463 [D M2e]
G. thermoglucosidasius C56-YS93		
B. cereus ATCC 14579		

[A]Group D M2e hydrogenases are poorly characterized and do not contain a PAS/PAC-sensory domain. However, given their proximity to protein kinases and bifurcating hydrogenases, and their phylogenetic proximity to group C D(M2f) sensory hydrogenases (Supplementary figure 2) we have classified them as sensory hydrogenases.

[B]Verified by microarray and proteomic analysis (*unpublished*). Characterization of hydrogenase specificity was based metallocenter composition ([NiFe] or [FeFe]), modular structure, subunit composition, and large (catalytic) subunit phylogeny according to Vignais *et al*. and Calusinska *et al*. [72, 151, 152]. Phylogenetic cluster groupings are indicated in superscript, and corresponding phylogenetic trees are provided in supplementary figures 1 and 2. Abbreviations: H_2ase, hydrogenase; NFO, NADH:ferredoxin oxidoreductase; *ech*, energy conserving hydrogenase; *mbh*, membrane bound hydrogenase; *rnf*, *Rhodobacter* nitrogen fixation.

With the exception of *P. furiosus* and *Th. kodakaranesis,* which encode only Fd-dependent and putative F_{420}-dependent [NiFe] H_2ases, all other H_2ase encoding organisms surveyed are capable of H_2ase-mediated oxidation/reduction of both Fd and NAD(P)H. This seems fitting given that *P. furiosus* and *Th. kodakaraensis* preferentially catalyze the oxidation of glyceraldedhyde-3-P via GAPFOR rather than GAPDH and PGK, and thus must reoxidize reduced Fd, rather than NADH, during fermentative product synthesis. All other H_2ase encoding organisms produce NADH during glycolysis and reduced Fd via PFOR. In these organisms, the oxidation of these electron carriers may be carried out using various different types of H_2ases. All of these species encoded at least a single putative bifurcating H_2ase (Table 2.6). The majority of these bifurcating H_2ases were found downstream of dimeric or monomeric sensory [FeFe] H_2ases that may be involved in their regulation (Table 2.6). Soboh *et al.* have demonstrated that NADH-dependent H_2ase activities in *Cal. subterraneus* subsp. *tengcongensis* are affected by H_2 partial pressures [98] suggesting possible regulation of these H_2ases via a two-component signal transduction mechanism in response changes in redox levels [72, 153]. It is important to note that these NADH-dependent H_2ase activities may reflect bifurcating H_2ase activities given that *Cal. subterraneus* subsp. *tengcongensis* encodes only a Fd-dependent and a putative bifurcating H_2ase, and no NAD(P)H-dependent H_2ases.

While *Ta. pseudethanolicus* only encodes a bifurcating H_2ase, all other organisms that encode a bifurcating H_2ase also encode Fd-dependent H_2ases. Putative Fd-dependent, [NiFe] Ech/Mbh-type H_2ases were identified in the genomes of *Cal. subterraneus* subsp. *tengcongensis*, *P. furiosus*, *Th. kodakaraensis*, and all *Caldicellulosiruptor* and *Clostridium* species (Table 2.6). A pair of putative Fd-

dependent [FeFe] H_2ases were identified in both *E. harbinense* and *C. phytofermentans*. With the exception *Ta. pseudethanolicus*, *Cal. subterraneus* subsp. *tengcongensis*, and *Caldicellulosiruptor* species, all organisms surveyed containing a bifurcating H_2ase also appear to be capable of NADH and/or NADPH oxidation using NADH/NADPH-dependent H_2ases. As with ADHs, however, we could not determine H_2ase cofactor specificity exclusively using *in silico* sequence analysis, stressing the importance for activity characterization of enzyme substrate specificity. While *C. cellulolyticum* achieves NAD(P)H oxidation using a putative H_2-uptake [NiFe] H_2ases, *E. harbinense*, *Thermotoga* species, and *C. thermocellum* ATCC 27405 achieve this using [FeFe] H_2ases. Although the draft genome of *C. thermocellum* DSM 4150 does not encode an NAD(P)H-dependent H_2ase, our proteomic and microarray data reveal the presence of Cthe_3003/Cthe_3004 homologues (Rydzak, *unpublished results*).

In addition to H_2ases-mediated electron transfer between Fd and/or NADH and H_2, electrons may be transferred directly between Fd and NAD(P)H via an Rnf-like (Rhodobacter nitrogen fixation) NADH:ferredoxin oxidoreductase (NFO), a membrane-bound enzyme complex capable of generating a sodium motive force derived from the energy difference between reduced Fd and NADH. Only *Thermotoga* species, *C. phytofermentans*, *C. thermocellum*, and *Ta. pseudethanolicus* encode putatively identified NFO. Proteomic analysis of *C. thermocellum*, however, revealed low, or no, expression of NFO subunits, suggesting it does not play a major factor in electron exchange between Fd and NADH [156].

While the presence/absence of genes encoding pathways that lead to reduced fermentation products (i.e. formate, lactate, and particularly ethanol) is a major

determinant of H_2 yields, we can make some inferences with respect to H_2 yields based on the types of H_2ases encoded. Given the thermodynamic efficiencies of H_2 production using different cofactors, we can say that Fd-dependent H_2ases are conducive for H_2 production while NAD(P)H-dependent H_2ases are not. However, organisms that do not encode ethanol-producing pathways (i.e. *Caldicellulosiruptor* and *Thermotoga* species) may generate high intracellular NADH:NAD^+ ratios, making NADH-dependent H_2 production thermodynamically feasible under physiological conditions. Conversely, in organisms capable of producing both H_2 and ethanol (*Ethanoligenens*, *Clostridium*, and *Thermoanaerobacter* species), the presence of Fd-dependent H_2ases appears to be beneficial for H_2 production. For example, *E. harbinense* and *Clostridium* species, which encode Fd-dependent, as well as bifurcating and NAD(P)H-dependent H_2ases, produce much higher H_2 yields when compared to those of *Ta. pseudethanolicus*, which encodes only one bifurcating H_2ase and no Fd or NAD(P)H-dependent H_2ases. Interestingly, organisms that do not encode H_2ases (*G. thermoglucosidasius* and *B. cereus*) produce low ethanol and high lactate (and/or formate yields), suggesting that H_2 production can help lower NADH:NAD^+ ratios, and thus reduce flux through LDH.

2.5 Influence of Overall Genome Content on End-Product Profiles

The presence and absence of genes encoding proteins involved in pyruvate metabolism and end-product synthesis may be used as an indicator of end-product distribution. By comparing genome content to end-product yields, we identified key markers that influence ethanol and H_2 yields. These include (i) MDH (ii) LDH, (iii) PFL vs. PFOR and/or PDH (iv) Aldh and AdhE, and (v) bifurcating, Fd-dependent, and NAD(P)H dependent H_2ase.

While it is difficult to elucidate how differences in "malate shunt" genes affect end-product synthesis patterns by comparing reported yields, eliminating MDH has been shown to increase lactate and ethanol production, and decrease acetate production in *C. cellulolyticum* [134]. The elimination of this transhydrogenation pathway may increase $NADH:NAD^+$ ratios for reduced end-product synthesis and reduce $NADPH:NADP^+$ ratios for biosynthesis. While the presence of LDH is not a good predictor of lactate yields, LDH, when active, diverts reducing equivalents away from H_2 and ethanol. In contrast to PFL, PFOR and PDH produce additional reducing equivalents (reduced Fd and NADH, respectively), and thus promote reduced end-product synthesis. Organisms that do not encode *pfl* generally produce more ethanol and H_2 (based on sum redox value) compared to those that do encode *pfl*. Of the organisms surveyed, those that did not encode (or express) both *adhE* and *aldH* produced near-maximal H_2 yields and little to no ethanol.

While the type(s) of encoded H_2ases appear to have little impact in organisms that do not encode ethanol producing pathways, they may influence reduced end-product yields in those that do. For example, *Ta. pseudethanolicus*, which encodes an *adhE*, NFO, and a single bifurcating H_2ase, but no discernable Fd or NAD(P)H-dependent H_2ases, generates low H_2 and near-optimal ethanol yields. The inability to oxidize reduced Fd via Fd-dependent H_2ases may elevate reduced Fd levels, which in turn can be used by NFO to produce additional NADH for ethanol synthesis. Interestingly, in the absence of H_2ases, lactate production was favoured over ethanol production, suggesting that H_2 production can help lower $NADH:NAD^+$ ratios, and thus reduce flux through LDH.

Given the impact that MDH, PFL, Aldh, AdhE, and the different H_2ases have on end-product yields, screening for these biomarkers can streamline ethanol and H_2 producing potential of sequenced and novel organisms through *in silico* gene mining and the use of universal primers, respectively. Furthermore, understanding how end-product yields are affected by (i) the framework of genes encoding pathways catalyzing pyruvate into end-products, and (ii) thermodynamic efficiencies of these reactions, we can begin to develop informed metabolic engineering strategies for optimization of either ethanol or H_2 (Figure 2.2). For example, in order to optimize either ethanol or H_2, we would recommend elimination of *ldh* and *pfl* in order to allow accumulation of additional reducing equivalents. Given that ethanol and H_2 compete for reducing equivalents, elimination of one product should direct carbon/and or electron flux towards the other.

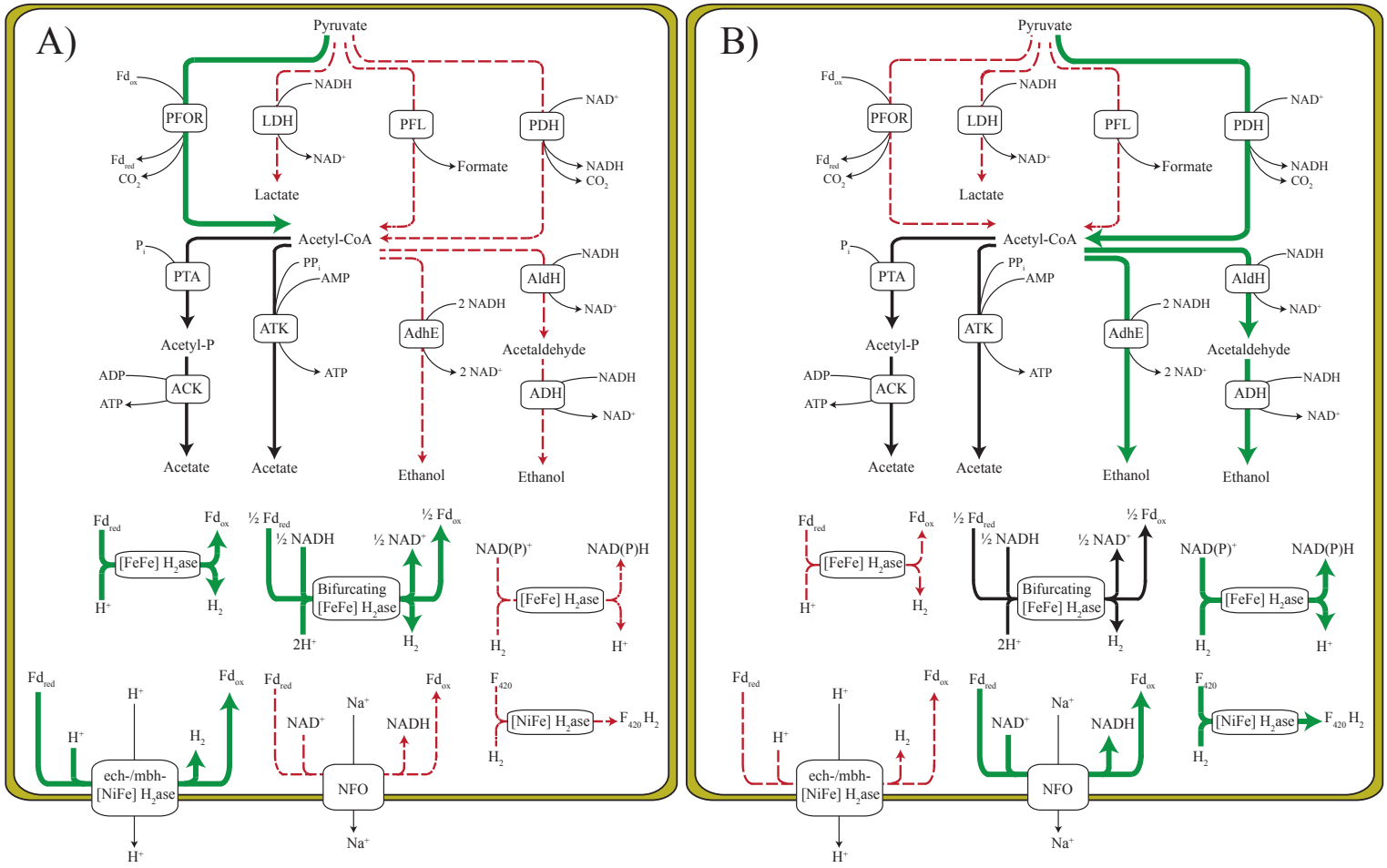

A)
B)
Pyruvate
PFOR
LDH
PFL
PDH
Lactate
Formate
Acetyl-CoA
PTA
ATK
AdhE
AldH
ADH
Acetyl-P
ACK
Acetate
Ethanol
Acetaldehyde
[FeFe] H2ase
Bifurcating [FeFe] H2ase
[NiFe] H2ase
ech-/mbh- [NiFe] H2ase
NFO

Figure 2.2. Differentiation between fermentation pathways that favor (A) hydrogen and (B) ethanol production based on comparative genomics and end-product profiles. Pathways that favor (green lines), disfavor (broken red lines), and appear to have little impact (black lines) on production of H_2 or ethanol are indicated. Correlation of reaction thermodynamics and genome content with reported end-product yields suggest that reduction, and subsequent reoxidation, of ferredoxin via PFOR and Fd-dependent (and/or bifurcating) H_2ases, respectively, support H_2 production. Alternatively, reduction, of NAD^+ via PDH (and/or NADH generating uptake H_2ases) generate NADH conducive for ethanol production. Abbreviations (see figure 2.1)

.

For optimization of H_2 yields (Figure 2.2a), deletion of *aldH* and *adhE* is likely most effective. Although conversion of pyruvate to acetyl-CoA is more thermodynamically favorable using PDH versus PFOR ($\triangle G°'$ = -33.4 vs. -19.2 kJ mol^{-1}), production of H_2 from NADH is highly unfavorable compared to the use of reduced Fd ($\triangle G°'$ = +18.1 vs. -3.0 kJ mol^{-1}). This in turn demonstrates that reduction of Fd via PFOR and subsequent H_2 production via a Fd-dependent H_2ase ($\triangle G°'$ = -21.2 kJ mol^{-1}) is more favorable than NADH production via PDH and subsequent H_2 production via NAD(P)H-dependent H_2ases ($\triangle G°'$ = -15.3 kJ mol^{-1}). Therefore, we propose that conversion of pyruvate to acetyl-CoA via PFOR is favorable for H_2 production, and *pdh* (and *pfl*) should be deleted. Given that 2 NAD^+ (per glucose) are reduced during glycolysis, the presence of bifurcating H_2ases, which could simultaneously oxidize the NADH generated during glycolysis and the reduced Fd produced via PFOR, would be required to theoretically achieve maximal H_2 yields. A Fd-dependent H_2ase would also be conducive for H_2 production during times when reducing equivalents generated during glycolysis are redirected towards biosynthetic pathways, resulting in a disproportionate ratio of reduced ferredoxin to NAD(P)H.

Alternatively, in organisms such as *P. furiosus* and *Th. kodakaraensis*, which generate high levels of reduced Fd and low levels of NADH, the presence of strictly Fd-dependent H_2ases, rather than bifurcating H_2ases, would be favorable for H_2 production. In all cases, NFO and NAD(P)H-dependent H_2ases should be deleted to prevent oxidation of reduced Fd and uptake of H_2, respectively, which would generate NAD(P)H.

The metabolic engineering strategies employed for optimization of ethanol (Figure 2.2b) are much different than those used for the production of H_2. First, *adhE*

and/or *aldH* and *adh* genes that encode enzymes with high catalytic efficiencies in the direction of ethanol formation should be heterologously expressed. Given that ethanol production is NAD(P)H dependent, NAD^+ reduction should be maximized while Fd reduction should be eliminated. Through deletion of *pfl* and *pfor*, and expression of *pdh*, up to 4 NADH can be generated per glucose, allowing for the theoretical maximum of 2 mol ethanol per mol glucose to be produced. To prevent NADH reoxidation, lactate and H_2 production should be eliminated by deleting *ldh* and NAD(P)H-dependent H_2ases. While this strategy is theoretically sound, low AldH/Adh catalytic efficiencies could cause NADH:NAD^+ ratios to rise high enough to impede glycolysis. In these situations, the presence of NFO or NAD(P)H-dependent H_2ases may alleviate these high ratios, at the cost of reduced ethanol yields, through generation of reduced Fd or via H_2 production, respectively.

While some attempts to increase H_2 and/or ethanol yields through genetic engineering have been successful in a number of lignocellulolytic organisms (reviewed elsewhere; [157]) engineering of strains discussed here has only been marginally successful. Heterologous expression of *Zymomonas mobilis* pyruvate decarboxylase and Adh in *C. cellulolyticum* increased cellulose consumption and biomass production, and decreased lactate production and pyruvate overflow due to a more efficient regulation of carbon and electron flow at the pyruvate branchpoint [158]. However, despite higher levels of total ethanol produced, ethanol yields (per mol hexose consumed) actually decreased when compared to the wild-type strain. Similarly, deletion of PTA in *C. thermocellum* drastically reduced acetate production, but had minimal impact on lactate or ethanol production [159]. This suggests that genome content alone cannot exclusively

dictate the extent of end-product yields observed in literature, and thus growth conditions must be optimized in order to moderate regulatory mechanisms that direct carbon and electron flux. This could only be attained through a thorough understanding of regulatory mechanisms that mediate gene and gene-product expression and activity levels under various growth conditions through a combination of genomics, transcriptomics, proteomics, metabolomics, and enzyme characterization.

2.6 Conclusions

Fermentative bacteria offer the potential to convert biomass into renewable biofuels such as H2 and ethanol through consolidated bioprocessing. However, these bacteria display highly variable, branched catabolic pathways that divert carbon and electrons towards unwanted end products (i.e. lactate, formate). In order to make fermentative H2 and/or ethanol production more economically feasible, biofuel production yields must be increased in lignocellulolytic bacteria capable of consolidated bioprocessing. While the cellulolytic and, to a lesser extent, H2 and ethanol producing capabilities of cellulolytic bacteria have been reviewed [64, 65, 100], a comprehensive comparison between genome content and corresponding end-product distribution patterns has not been reported. Although reported end-product yields vary considerably in response to growth conditions, which may influence gene and gene product expression and metabolic flux, this study demonstrates that the complement of genes encoding pyruvate catabolism and end-product synthesis pathways can be used to approximate potential end-product distribution patterns.

A number of genetic biomarkers, including (i) MDH (ii) LDH, (iii) PFL vs. PFOR and/or PDH (iv) Aldh and AdhE, and (V) bifurcating, Fd-dependent, and NAD(P)H

dependent H2ases, that can be used for streamlining H2 and/or ethanol producing capabilities in sequenced and novel isolates were identified. By linking genome content, reaction thermodynamics and end-product yields, potential targets for the optimization of ethanol or H_2 yields via metabolic engineering were considered. Deletion of LDH and PFL could potentially increase both H2 and ethanol yields by maximizing available pools of reducing equivalents. The deletion of ethanologenic pathways (aldH, adh, adhE), increasing flux through PFOR, overexpression of Fd-dependent H2ases, and the elimination of potential H2-uptake (NAD(P)H-dependent) H2ases would likely result in increased H2 production. Alternatively, directing flux through PDH would be beneficial for ethanol production. Although gene and gene-product expression, functional characterization, and metabolomic flux analysis remains critical in determining pathway utilization, insights regarding how genome content affects end-product yields can be used to direct metabolic engineering strategies and streamline the characterization of novel species with potential industrial applications.

Chapter 3: Effect of Gas Sparging on Pyruvate Catabolism during Batch Fermentation of *Clostridium thermocellum* ATCC 27405

3.1 Abstract

Growth, end-product synthesis, enzyme activities, and transcription of genes associated with pyruvate catabolism, H_2 synthesis, and ethanol production were studied in the cellulolytic anaerobe, *Clostridium thermocellum* ATCC 27405, during batch fermentation of cellobiose to determine the effect of elevated N_2 and H_2 gas sparging on metabolism using a 14 liter (L) fermenter with a 7 L working volume. End-product synthesis profiles revealed increases in acetate, CO_2, and H_2 production under high N_2 sparging with respect to standard N_2 gas flow rates. Alternatively, elevated H_2 sparging shifted carbon and electron flow towards the production of formate and ethanol. Transcription levels and enzyme activities of key proteins involved in pyruvate catabolism and end-product synthesis (pyruvate:ferredoxin oxidoreductase, pyruvate:formate lyase, lactate dehydrogenase, NAD- and NADP-dependent alcohol dehydrogenase, methyl viologen- ferredoxin, NAD- and NADP-dependent hydrogenase) revealed no significant differences under the conditions tested, with the exception of the *adhE* and *adhY*. The results presented here suggest that while gas sparging can be effectively used to shift carbon and electron flow, the observed shifts at the pyruvate branch-point are principally influenced by the availability of reduced electron carriers (NAD, NADP, ferredoxin) and thermodynamic considerations.

3.2 Introduction

Conventional biofuel production from cellulose involves a complex process of feedstock pre-treatment including: (i) cellulase production, (ii) hydrolysis of cellulose/hemicellulose, followed by the (iii) fermentation of liberated sugars. Consolidated bioprocessing (CBP) is a processing strategy in which each of these processes are accomplished in a single-step by microorganisms that express cellulolytic (and hemicellulolytic) enzymes [5-7]. CBP offers the potential for lower biofuel production costs due to simpler feedstock processing, lower energy inputs, and higher conversion efficiencies than traditional simultaneous saccharification and fermentation (SSF) based processes. The exploitation of cellulolytic, thermophilic anaerobes has been proposed for CBP biofuel production, as they are capable of the direct utilization of raw lignocellulosic substrates, typically display high metabolic rates, and their elevated growth temperatures are amenable to end-product recovery [127].

Successful development of "third generation" biofuels, however, depends heavily on a detailed understanding of the metabolism of cellulolytic bacteria. Growth conditions can significantly influence end-product yields through regulation of gene and gene product expression [98, 109], and modulation of metabolic flux and intracellular metabolite levels [110, 111]. Variations in substrate availability/dilution rate [102, 109-111, 114-117], substrate composition [110, 118-123], media composition [111], pH [124], gas partial pressure [90, 98, 125, 126], growth phase [113], and end-product accumulation [103, 118, 125, 127, 128] are all known to dramatically influence end-product yields.

Clostridium thermocellum ATCC 27405 is an anaerobic, moderately thermophilic bacterium that is able to efficiently degrade cellulose and ferment the soluble sugars released into fuel chemicals ethanol and H_2 in addition to acetate, lactate, formate and CO_2 [59, 62, 66, 102]. As a consequence of highly branched catabolic pathways, yields of fuel chemicals fall far below theoretically maximal values (2 mol-ethanol mol-glucose^{-1} or 4 mol-H_2 mol-glucose^{-1}). While recent advances in Clostridial genetics have enabled the implementation of rational metabolic engineering strategies, published reports of knockout *C. thermocellum* strains continue to display sub-optimal yields [159].

In 2012 Li *et al*. examined the effect of N_2 and H_2 sparging on the production of ethanol and volatile fatty acids, however the authors did not report formate production or examine the transcriptional or translational relationship of genes and/or proteins associated with pyruvate catabolism and product formation [134]. A similar investigation into the effect of gas sparging (with N_2 or H_2) during fermention by *Caldanaerobacter subterraneus* subsp. *tengcongensis* (formerly *Thermoanaerobacter tengcongensis*) reported significant changes in end-product distribution but did not include an examination of potential transcriptional regulation [98]. Therefore, in this study the extent to which changes in end-product distribution are reflected by changes in enzyme activities and gene transcription is investigated. Specifically, how *C. thermocellum* ATCC 27405 growth, end-product distribution, enzyme activity, and the transcription of select genes involved in pyruvate catabolism, as well as H_2 and ethanol metabolism are affected during batch fermentation of cellobiose under elevated N_2 and H_2 sparging rates is reported.

3.3 Methods

3.3.1 Microorganism and Media

Clostridium thermocellum strain 27405 was obtained from the American Type Culture Collection (ATCC) and used for all growth experiments. Fresh cultures were maintained by routinely transferring 10 % (v/v) mid-exponential phase cultures into freshly prepared anaerobic 1191 media. Sealed Corning (1.2 L) or serum bottles (162 mL; Bellco Glass Co., New Jersey, USA) were prepared containing 500 mL or 50 mL 1191 media respectively and gassed:degassed with N_2 for 3 cycles prior to the addition of 2 mM anaerobic Na_2S and autoclaving. After cooling, sterile cellobiose was added to a final concentration of 2 g L^{-1}.

Complex 1191 media (pH 7.2) was prepared as previously described [102] and contained (per 1000 mL distilled deionized water) KH_2PO_4, 1.5 g; $Na_2HPO_4 \bullet 12H_2O$, 4.2 g; NH_4Cl, 0.5 g; $MgCl_2 \bullet 6H_2O$, 0.18 g; yeast extract (BD 212750), 2.0 g; 2 mM Na_2S, 1.0 g; 0.25 mg/ml resazurin, 1.0 mL; vitamin solution, 0.5 mL; mineral solution, 1 mL. Vitamin solution contained the following (per 1000 mL): biotin, 20 mg; *p*-aminobenzoic acid, 50 mg; folic acid, 20 mg; nicotinic acid, 50 mg; thiamine, 50 mg; riboflavin, 50 mg; lipoic acid, 50 mg; cyanocobalamin, 10 mg. Mineral solution contained (per 1000 mL): trisodium nitrilacetate, 20.2 g; $FeCl_3$-$6H_2O$, 2.1 g; $CoCl_2$-$6H_2O$, 2.0 g; $MnCl_2$-$4H_20$, 1.0; $ZnCl_2$, 1.0 g; $NiCl_2$-$6H_2O$, 1.0 g; $CaCl_2$-$2H_2O$, 0.5 g; $CuSO_4$-$2H_2O$, 0.5 g; Na_2MoO_4-$2H_2O$, 0.5 g. All chemicals were reagent grade and purchased from Sigma Chemical Co. (St. Louis, MO). Gasses were purchased from Welder's Supply (Winnipeg, Mb).

3.3.2 Experimental Design

Batch culture growth was carried out using a 14 L New Brunswick Scientific Bioflow 110 fermentor with a working volume of 7.0 L. Reactor temperature was maintained at 60 °C and pH was monitored in real time. Fermentation experiments were conducted under low gas sparging rates with N_2 (2 mL s^{-1}) and under conditions of high gas sparging rates (20 mL s^{-1}) of either N_2 or H_2. An additional experiment, under low N_2 sparging conditions, was performed during which pH was maintained at 7.18 by addition of 1 M KOH. Fermentation broth was continuously mixed (250 rpm) by an axial flow impellor under direction of the fermentor control module to maintain a homogenous suspension.

3.3.3 Analytical Procedures

Cell-growth was monitored spectrophotometerically by measuring the optical density (Biochrom, Novaspec II) at 600 nm (OD_{600}) throughout growth in the bioreactor. Samples (1.5 mL) were collected during fermentation and centrifuged in a Sorvall Legend Micro 21 microcentrifuge (14,000 x g, 10 min at 4 °C, Thermo Scientific) to isolate supernatant for end-product analysis and cell pellet for subsequent protein determinations. Separate 50 mL samples were isolated for RNA extraction and RT-qPCR analysis (see below). These samples were processed immediately following harvest by centrifugation in a Sorvall RC6 centrifuge with a SH-3000 rotor at 4500 x g for 10 minutes at 4 °C before discarding the supernatant and re-suspension of the cell pellet in RNA*later* solution (Ambion, Carlsbad CA) as per the manufactures recommended protocol. An additional 2.0 L of culture was harvested from the bioreactor for enzyme activity analysis at mid-exponential phase of growth (OD_{600nm} ~ 0.5).

3.3.4 Protein Determination

Protein determinations of *C. thermocellum* cultures were performed as described by Rydzak *et al*. 2009 [66]. Briefly, cell pellets were washed with 0.9 % NaCl before resuspension in 1 mL 0.2 M NaOH and incubation in a boiling water bath for 10 min. After cooling, samples were centrifuged (as above) and supernatants were collected for protein assays. Total protein content was determined by using a modified Bradford [160] method and standards prepared from bovine serum albumin (BSA). A PowerWave-XS single channel microtitre plate reader running KCjunior software (BIO-TEK Instruments Inc., Winooski, VT) was used for all absorbance readings (595 nm). Biomass composition ($C_4H_7O_2N$) was calculated based on a stoichiometric conversion of cellobiose of into cell material and protein content as described by Guedon *et al*. [161] and an average composition of cell material of 101 g mol^{-1}.

3.3.5 Gas Analysis

Product gas compositions (H_2 and CO_2) were determined using a Multiple Gas Analyzer #1 Gas Chromatograph System Model 8610-0070 (SRI Instruments, Torrance, CA), using argon as the carrier gas and equipped with a thermal conductivity detector capable of detecting concentrations between 200 and 500 ppm. Hydrogen measurements were conducted with a stainless steel (3.2 mm x 1.8 m) column packed with a 13 x molecular sieve. CO_2 was detected with a silica gel packed stainless steel column (3.2 mm x 1.8 m). Product gas concentrations were quantified by comparing observed peak areas to those of known standards (H_2 and CO_2) that had been prepared the same day. Ambient temperature, atmospheric pressure, gas solubility in water [162] and bicarbonate equilibrium (for CO_2) [163] was each considered when calculating gas production.

3.3.6 Cellobiose and Organic End-Product Analysis

Culture supernatants (1.5 mL) were collected and stored at -20 °C until required for analysis. Lactate, acetate, formate and pyruvate were measured by high-pressure liquid chromatography (HPLC; Dionex, ICS 3000) using an anion-exchange IonPac AS11-HC analytical column (Dionex Corporation, Sunnyvale, CA, USA). Cellobiose was measured by HPLC using an anion-exchange CarboPac-PA1 analytical column (4 X 250 mm). Ethanol production was quantified spectrophotometrically by measuring NADH oxidation by alcohol dehydrogenase at 340 nm using an Enzymatic BioAnalysis UV-Test kit (R-Biopharm AG, Darmstadt, Germany) modified for use in microtitre plate readers. Standard curves for acids, sugars and ethanol were all prepared with 1191 media.

3.3.7 Preparation of Cell Extracts and Enzyme Assays

The preparation of *C. thermocellum* cell extracts and enzyme assays was performed as previously described by Rydzak *et al* [66]. All cell manipulations were performed under anaerobic conditions in a Thermo Forma Anaerobic Chamber (Model 1025, Thermo Fisher Scientific, Waltham, MA, USA) with an atmosphere of 10 % hydrogen, 90 % nitrogen. Cells harvested from the bioreactor were collected by centrifugation (4500 x g, 20 min) at 4 °C using a Sorvall RC6 centrifuge equipped with an SH-3000 rotor in sealed centrifuge bottles (Nalgene, Rochester, NY) and the supernatant was discarded. Cell pellets were next suspended in Balch tubes containing either 10 mL cell suspension buffer (HEPES, 50 mM, pH 7.0; DTT, 10 mM; glycerol, 10 %) or reactivation buffer (potassium phosphate buffer, 50 mM, pH 7.0; S-adenosyl-L-methionine, 0.6 mM; pyruvate, 10 mM; DTT, 2.85 mM; $Fe(NH_4)_2(SO_4)_2$-$6H_2O$, 2 mM;

methyl viologen, 1 mM) and sealed in the anaerobic chamber prior to being regassed with N_2 to ensure anaerobiosis. Cells suspended in reactivation buffer were used immediately while cells suspended in activation buffer may have been stored at -70 °C prior to use. Cells were lysed by passing the suspension through a French pressure cell three times at 20,000 kg m^{-2} that had previously been rinsed three times with Na_2S to remove oxygen and then centrifuged (4500 x g, 20 min) to remove unlysed cells. Supernatants were collected in fresh balch tubes and sealed and re-gassed as described above. Protein content was determined by the Bradford method (as described above).

All determinations of enzyme activity were performed at 25 °C under anaerobic conditions according to the specific reaction conditions described by Rydzak *et al* [66] with the exception of MDH and malic enzyme (not previously determined). MDH was measured in the direction of malate formation. The standard assay mixture contained 0.1 M Tris-HCl (pH 6.5), 10 mM DTT, 2 mM oxaloacetate, 0.3 mM NAD(P)H [164]. Malic enzyme was measured in the direction of pyruvate formation. The standard assay mixture contained 0.1 M Tris-HCl (pH 6.5), 10 mM DTT, 2 mM malate, 0.5 mM $NAD(P)^+$, 5 mM $MnCl_2$, 20 mM NH_4Cl [165]. LDH, MDH, ADH, malic enzyme, PFOR, and ALDH activities were all determined from extracts suspended in cell suspension buffer. Determinations of both PFL and hydrogenase (MV-, NADH-, NADPH-dependent) activity were measured from samples suspended in reactivation buffer that had been incubated at 37 °C for 90 min [166]. Specific activities were extrapolated from the linear portion of the reaction curve and are reported in nmol min^{-1} mg-$protein^{-1}$, which represents the conversion of substrate into specific products. One

unit of enzyme activity is defined as the amount of enzyme catalyzing the conversion of 1 nmol substrate min^{-1}.

3.3.8 RNA Extraction and RT-qPCR Analysis

Total RNA was isolated from freshly collected 50 mL samples of exponential phase *C. thermocellum* ATCC 27405 grown on 2 g L^{-1} cellobiose using the Invitrogen TRIzol Reagent kit (Invitrogen, Carlsbad CA). Cells were pelleted by centrifugation in a Sorvall RC6 centrifuge with a SH-3000 rotor at 4500 x g for 12 minutes at 4 °C prior to RNA extraction following the manufacturer's instructions for RNA extractions for cells grown in suspension. RNA pellets were dissolved in DEPC-treated RNase free H_2O containing RNase inhibitor (Invitrogen) and dithiothreitol (DTT) at concentrations of 0.5 U μL^{-1} and 1 mM, respectively. RNA samples were treated with DNase I (0.1 U μL^{-1}) for 15 min at 20°C in reactions containing 2 mM $MgCl_2$, 20 mM Tris (pH 8.4) and 50 mM KCl, prior to cDNA synthesis. Reactions were halted by the addition of ethylenediaminetetraacetic acid (EDTA) to a concentration of 2.5 mM and incubation at 65 °C for 10 min.

cDNA synthesis was accomplished with SuperScript II Reverse Transcriptase (Invitrogen), following the manufacture's recommended protocol with random hexamer primers. Each reaction was performed under the following conditions: between 1-5 μg of total RNA, 2.5 μM random hexamer primers, 0.5 mM dNTPs, 5 mM DTT, 1.25 mM $MgCl_2$, 0.5x RT buffer, 2.5 U μL SuperScript II Reverse Transcriptase and RNase H. An Eppendorf Mastercycler EP thermocycler was used for cDNA synthesis. Samples containing RNA, dNTPs and random hexamer primers were incubated at 65 °C for 5 min prior to the addition of RT buffer, DTT. Contents were next incubated for 2 min at 42 °C

before the addition of SuperScript II reverse transcriptase. Following incubation at 42 °C for 50 min, reactions were terminated by heating samples at 70 °C for 15 min. Finally, RNA complementary to cDNA was removed by addition of RNase H and incubation at 37°C for 20 min.

All primers used in qPCR reactions were designed using Primer3 software (http://primer3.sourceforge.net/) with amplicon sizes were between 185 and 250 bp in size and T_m values ranging from 58 to 60 °C. Primer nucleotide sequences are outlined in Table 3.1. All amplicons were cloned into plasmid pGEM-t (Promega, Madison, WI) following the manufacturers recommended protocol and insert identities were confirmed subsequently by DNA sequencing. Cloned vector DNA was extracted from *E. coli* DH5α cells using a QIAprep Spin miniprep kit (QIAGEN, Valencia, CA) and used to construct standard curves for qPCR analysis.

Table 3.1. Primers used for RT-qPCR analysis.

Gene Locus	Annotation	Accession no.	Forward Primer	Reverse Primer
Cthe_R0041	16s rDNA	CP000568	TGACGGGCGGTGTGTACAAGG	GGTGGGGACGACGTCAAATCA
Cthe_0344	*malE*	YP_001036775	CGGTGAGTGGAAGGGTAAAA	CGGTGAGTGGAAGGGTAAAA
Cthe_0345	*mdh*	YP_001036776	GCCGGAGCCAACAGAAAACCT	CGTCAACGCCCAATTTTTCGC
Cthe_1053	*ldh*	YP_001037478	CAAAGACTGTGCCGGATCCGA	GGCTGTCCAAAACCGTTCC
Cthe_2796	*pfor*	YP_001039188	GGGAAATGAAGCAGTGGCGGA	AGCCCCTGGGATGAAGTTGCC
Cthe_2392	*pfor*	YP_001038787	ACCGATGATGCCGATGTTGCC	AACAAAGGTCCTCCGGCTGCA
Cthe_3120	*pfor*	YP_001039508	GCAGGGGCATTGACGACCACT	TAATGGCCGAAAGATGCGAA
Cthe_0505	*pfl*	YP_001036935	CCGAAGCTTATGGCCACAGTG	GCAATCAGCCTGTCAACGCCA
Cthe_2238	*aldh*	YP_001038633	GACATGCATAGCTCCGGATT	ACAAACTTTTGCCCCCTCTT
Cthe_0423	*adhE*	YP_001036854	TTTCTTTGAGGCGGCAATGGC	CCGCAATAACCCCCAAAGGCT
Cthe_2579	*adhZ*	YP_001038972	TGCTGGTTACGGGAAAGAAC	CTCCTCCGATTCCGATGATA
Cthe_0394	*adhY*	YP_001036825	GACATTGAGGCACGGTCAAA	CTCCTGCGTTGGAATTGGTAA
Cthe_0101	*adh*	YP_001036535	TGTGCAGTTTGCCGTAAGAG	GCATTTTGACGCCATTTCTT
Cthe_0430	[FeFe] H_2ase	YP_001036861	AGGAGCCGGTGTGATATTTG	CGCTGACCACTGCAACTTTA
Cthe_3003	[FeFe] H_2ase	YP_001039392	CCGGAATACTGTTTGGTGCT	TCTTAAGGCCGCTGACAACT
Cthe_0342	[FeFe] H_2ase	YP_001036773	AGGAGCCGTCTGAACGACTA	TGTGTCAGCATGTGCAAGAA
Cthe_3020	[NiFe] H_2ase	YP_001039409	GCCAGTGCATATCCCTCATT	GGTTGGAGTACGAACCCTGA
Cthe_2430	*nfo*	YP_001038825	GTATTGCCAAAGGAGCAAGC	TAAGCACAGCTGCCACATTC

Ten (10) ng of cDNA (~2 µL) was used as a template for subsequent qPCR experiments. Reactions were run on a Stratagene Mx3005p thermocycler using 12 µL iTaq SYBR Green Supermix with Rox (2X reaction buffer, 0.4mM dNTPs, 50 U mL^{-1}, 6 mM $MgCl_2$, SYBR Green I dye, 1 µM ROX; BioRad, Mississauga, ON), 2.5 µL each of forward and reverse primer (final concentration 1 mM), and ddH_2O to a total volume of 25 µL. All qPCR reactions were run in 8-strip tubes with flat caps (Axygen, Union city, CA). qPCR reactions consisted of an initial 2 min denaturation step at 95 °C followed by 40 cycles of: 95 °C for 15 s, 55 °C for 45 s and 72 °C for 45 s. After cDNA amplification a thermal disassociation curve was generated by incubating samples at 95°C for 1 min followed by 30 s at 55 °C.

The transcription of genes examined in this study was divided into four classes. The first class included genes involved in a '*malate shunt*' and lactate production: malate dehydrogenase (*mdh*), malic enzyme (*malE*), and lactate dehydrogenase (*ldh*). The second class included four genes involved in the conversion of pyruvate to acetyl-CoA: pyruvate:ferredoxin oxidoreductase (*pfor*) and pyruvate formate lyase (*pfl*). For *pfor* analysis, primers were designed to amplify regions of the α–subunit. The third class included five genes potentially involved in alcohol metabolism and included three type IV alcohol dehydrogenases (*adh*), an acetaldehyde dehydrogenase (*aldh*), and a bifunctional acetaldehyde/alcohol dehydrogenase (*adhE*). The designations *adhZ* and *adhY*, put forth by Stevenson *et al*., 2005, to correspond to Cthe_2579 and Cthe_0394, respectively, and are maintained in this study to facilitate comparison [167]. The fourth class of genes analyzed included three [FeFe] hydrogenases, a [NiFe] hydrogenase, and

subunit C of NADH:ferredoxin oxidoreductase (*nfo*). 16S rRNA expression was used as a reference gene for all analysis. Relative expression of genes analyzed in this study was calculated using the Pfaffl method [168] with the 20 mL s^{-1} N_2 gas sparging condition used as the calibrator. The statistical significance of expression ratios was determined using the BootstRatio web application [169].

3.4 Results

3.4.1 Cell growth, pH and cellobiose consumption

Growth of *C. thermocellum* ATCC 27405 was monitored in batch cultures containing 2 g L^{-1} (5.9 mM) cellobiose in 1191 media continuously sparged with N_2 or H_2 (Figures 3.1 and 3.2). Although fresh mid-exponential phase cultures were used as source innocula (10 % v/v) for all experiments, in some instances a modest 2-3 h lag was observed. In batch cultures sparged with N_2 at 2 mL s^{-1} with no pH control, cells entered stationary phase within 8 h incubation at 60 °C. In contrast, cultures sparged with either 20 mL s^{-1} N_2 or H_2 or 2 mL s^{-1} N_2 with pH control entered stationary phase only following 12 h incubation. Cell densities measured at 600 nm (not shown) correlated well with biomass production as measured by protein content. Final cell biomass (Table 3.2) concentrations reached 3.3, 2,3, 2.75, and 2.5 mM for low N_2, low N_2 pH controlled, high N_2, and high H_2 sparged cultures, respectively. In all studies, cellobiose was completely utilized upon entry in stationary phase indicating cultures were carbon limited. Generation time of *C. thermocellum* during mid-exponential growth was determined to be 2.3 h per generation for cultures sparged at 2 mL s^{-1} and 3.0 h per generation for cultures sparged at 20 mL s^{-1}. Upon entry into stationary-phase, culture

pH was determined to be 6.57 for cultures sparged at 2 mL s^{-1} N_2, 6.77 for cultures sparged at 20 mL s^{-1} N_2 and 6.92 for cultures sparged at 20 mL s^{-1} H_2.

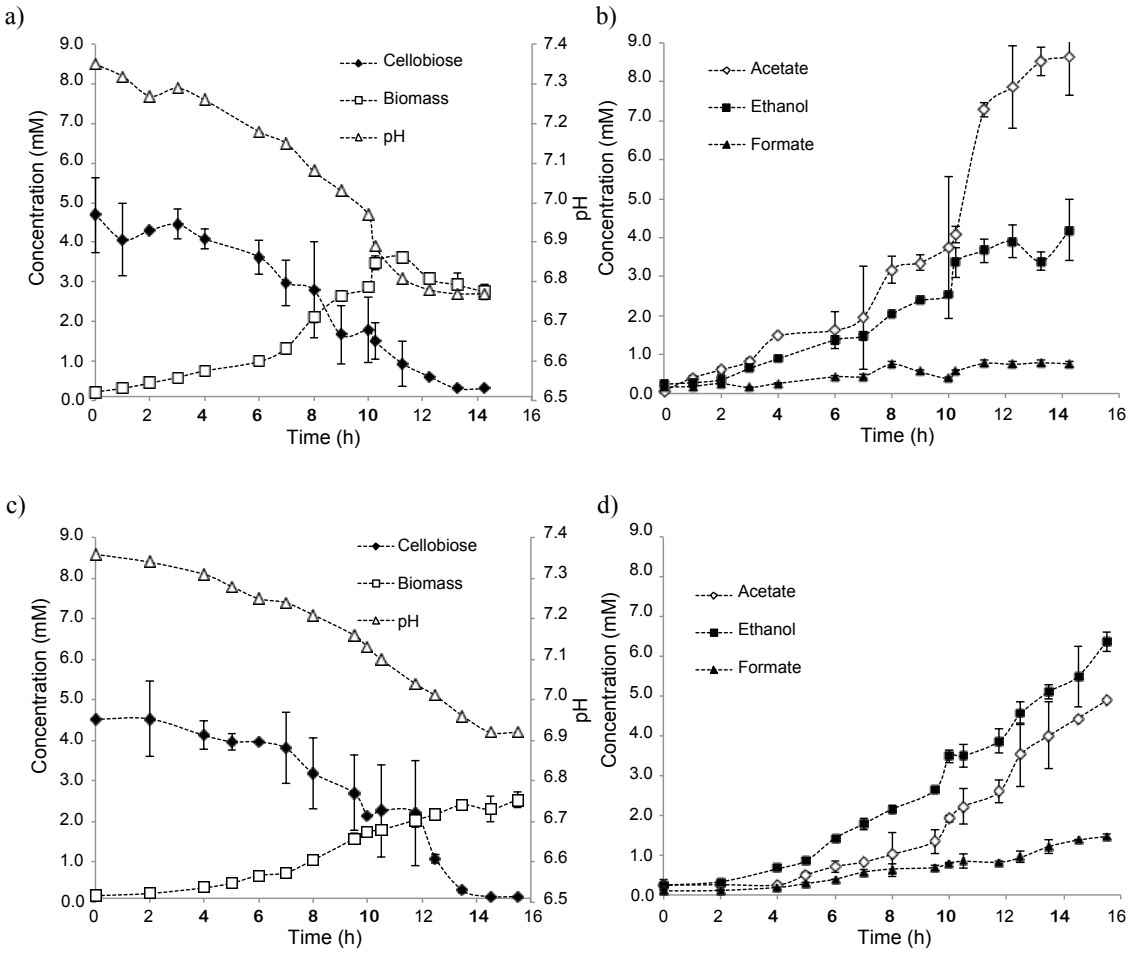


Figure 3.1. Growth and fermentation product synthesis of *Clostridium thermocellum* ATCC 27405 grown on 2 g L^{-1} cellobiose while sparged with N_2 (a and b), and H_2 gas (c and d) at 20 mL s^{-1}.

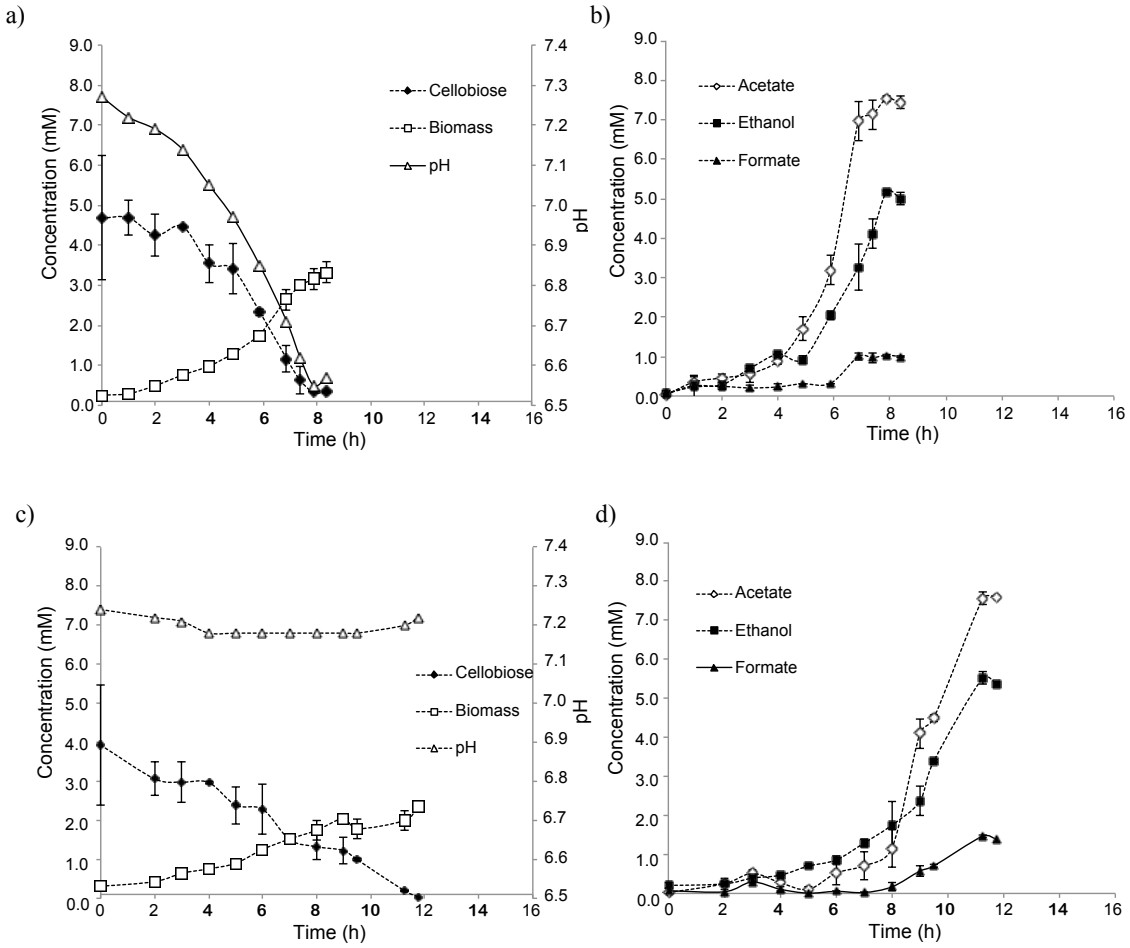


Figure 3.2. Growth and fermentation product synthesis of *Clostridium thermocellum* ATCC 27405 grown on 2 g L^{-1} cellobiose while sparged with N_2 gas (2 mL s^{-1}) without pH maintenance (a and b) and with pH maintenance (c and d; at pH 7.18) with 1M KOH.

3.4.2 End-Product Synthesis

Acetate, ethanol, formate, CO_2 and H_2 were the major end-products of cellobiose fermentation. Although both lactate dehydrogenase (LDH) enzyme activity and *ldh* transcription were detected at significant levels at each point of the growth conditions assayed, negligible lactate was produced (< 0.2 mM) during fermentation (discussed below).

When batch cultures of *C. thermocellum* were sparged with N_2 at 2 mL s^{-1}, 7.43 (± 0.16) mM acetate, 4.76 (± 0.16) mM ethanol, and 1.00 (± 0.03) mM formate were produced (Table 3.2). When the pH was maintained at 7.18 in N_2 sparged (2 mL s^{-1}) cultures, end-product profiles remained largely unchanged with 7.59 (± 0.08) mM acetate, 5.36 (± 0.12) mM ethanol, and 1.38 (± 0.13) mM formate produced. Cellobiose consumption and end-product synthesis throughout growth of cultures sparged with N_2 (2 mL s^{-1}) with and without pH maintenance are shown in Figure 3.2. An increase in acetate production of approximately 15 % was observed when gas sparging with N_2 was increased to 20 mL s^{-1}. Additionally, ethanol and formate yields were reduced by 14 % and 31 %, respectively, when compared to non-pH maintained cultures.

Sparging cultures with H_2 at 20 mL s^{-1} had the most dramatic affect on end-product distribution profiles. Compared to cultures of *C. thermocellum* sparged at 20 mL s^{-1} N_2, acetate production was reduced by 52 %, while ethanol and formate production increased by 34 % and 47 % respectively (Table 3.2). Cellobiose consumption and end-product synthesis throughout growth of cultures sparged with N_2 and H_2 at 20 mL s^{-1} are shown in Figure 3.1.

Ethanol:acetate (E:A) ratios are illustrative of how carbon and electron flux was affected by gas sparging with either N_2 or H_2. When sparged with N_2, ethanol-to-acetate ratios (mol/mol) of 0.6, 0.5 and 0.7 were calculated for 2 mL s^{-1}, 20 mL s^{-1} and pH maintained cultures respectively. In contrast, when *C. thermocellum* was sparged with H_2 at 20 mL s^{-1} an E:A ratio of 1.3 was observed.

As the bioreactor was subjected to continuous gas sparging and was openly vented, condensates were collected for end-product determinations. Condensate volume was greatest under high gas sparging rates with 148 mL and 250 mL collected from 20 mL s^{-1} N_2 and H_2 sparged cultures, respectively. Alternatively, 15 mL and 3.5 mL of condensate was collected from 2 mL s^{-1} N_2 gas sparged cultures with and without pH maintenance respectively. Acetate, lactate and formate concentrations were below detectable limits in each of the samples. Condensate ethanol was highest in the H_2 gas sparged samples (in mmoles) and is provided in Table 3.2.

Table 3.2. Effect of gas sparging on final end-product yields (mM) of *C. thermocellum* grown on 2 gL^{-1} cellobiose. Standard deviations of three replicates are indicated in brackets. Values indicated in bold represent calculated values assuming a carbon recovery ~ 90 % and C1/C2 ratio and an O/R index of 0.9-1.1.

Product (mM)	Reactor Condition:			
	20 mL s^{-1} N_2	2 mL s^{-1} N_2	pH	20 mL s^{-1} H_2
Biomass	2.75 (0.19)	3.33 (0.26)	2.34 (0.09)	2.53 (0.10)
Acetate	8.64 (0.98)	7.43 (0.16)	7.59 (0.08)	4.89 (0.04)
Formate	0.76 (0.08)	1.00 (0.03)	1.38 (0.13)	1.47 (0.06)
Lactate	< 0.005	< 0.005	< 0.005	< 0.005
Ethanol	4.19 (0.78)	4.76 (0.16)	5.36 (0.12)	6.37 (0.24)
Condensate Ethanol: mM	5.05 (0.88)	5.03 (0.20)	15.72 (1.49)	13.4 (1.04)
mmoles	0.747	0.018	0.235	3.34
H_2	**16.51**	**13.86**	**13.8**	**8.32**
CO_2	**12.07**	**11.43**	**11.56**	**9.79**
Generation time (h)	3.0	2.3	2.3	3.0
pH	6.77	6.57	7.18	6.92
E:A	0.5	0.6	0.7	1.3

3.4.3 Enzyme Activities

Enzyme activities involved in pyruvate metabolism, hydrogen and ethanol production, and the "*malate shunt*" associated malate dehydrogenase and malic enzyme were assayed from mid-exponential phase (OD_{600nm} ~0.500) cell extracts of *C. thermocellum* ATCC 27405 grown under low N_2 (2 mL s^{-1}), high N_2 and H_2 (20 mL s^{-1}) gas sparging conditions. Activities displayed in Table 3.3 are reported in nmol min^{-1} mg $protein^{-1}$ and represent the average of at least three determinations.

Neither the activity of malic enzyme or malate dehydrogenase changed significantly when sparged with either N_2 or H_2 gas. Similarly, fructose-1,6-bisphosphate activated lactate dehydrogenase activity remained constant, despite the absence of observable lactate production, under each of the growth conditions tested. Of the enzymes catalyzing the oxidative decarboxylation of pyruvate, pyruvate:ferredoxin oxidoreductase activity was 10- to 20-fold greater than pyruvate:formate lyase activity under all conditions tested.

Table 3.3. Effect of gas sparging on activities of mid-exponential phase enzymes involved in conversion of pyruvate into end-products of *C. thermocellum* grown under carbon limited conditions (2 g L^{-1} cellobiose).

Enzyme	Substrate	Specific Activity (nmol min^{-1} mg $protein^{-1}$)		
		20 mL s^{-1} N_2	2 mL s^{-1} N_2	20 mL s^{-1} H_2
Malic enzyme	Malate: NADP+	78.4 (2)	ND	80.7 (2.2)
Malate dehydrogenase	Oxaloacetate: NADH	46.4 (2.3)	ND	40.7 (6.2)
	Oxaloacetate: NADPH	13.3 (0.7)	ND	20.7 (2.3)
Lactate dehydrogenase	Pyruvate: NADH: FDP	11 (3)	23 (2)	10 (4)
Pyruvate:formate lyase	Pyruvate: CoA	11 (3)	11 (1)	13 (2)
Pyruvate:ferredoxin oxidoreductase	Pyruvate: CoA: MV	218 (8)	327 (21)	256 (30)
Hydrogenase	H_2: MV	220 (6)	840 (55)	737 (73)
	H_2: Fd: metronidazole	ND	1 (< 1)	1 (< 1)
	H_2: NAD^+	1 (< 1)	1 (< 1)	3 (< 1)
	H_2: $NADP^+$	5 (< 1)	8 (1)	7 (1)
Alcohol dehydrogenase	Acetaldehyde: NADH	40 (14)	61 (5)	84 (29)
	Acetaldehyde: NADPH	15 (1)	15 (1)	18 (3)

ND: Not determined, MV: methylviologen, FDP: Fructose-1,6-bisphosphate

Of the enzymes involved in pyruvate catabolism, hydrogen metabolism, or ethanol production, only the activity of alcohol dehydrogenase appears to be influenced by sparging gas (H_2 or N_2). NADH-dependent alcohol dehydrogenase activity was 2.1-fold greater when sparged with H_2 than when sparged with N_2 (20 mL s^{-1}). Methylviologen dependent hydrogenase activity, which represents the sum total of hydrogenase activities, ranged from 220 to 737 units (nmol min^{-1} mg $protein^{-1}$) under the conditions tested. In contrast, the measured specific activity of Fd-dependent and NADH-dependent hydrogenase was very low. NADPH-dependent hydrogenase specific activity was approximately 2-3 fold greater than Fd/NADH-dependent hydrogenase activity.

3.4.4 Gene Expression

Expression of 17 genes encoding proteins involved in pyruvate catabolism and end-product synthesis was determined relative to 16S rRNA transcription by RT-qPCR (Figure 3.3). The expression of the 16S rRNA gene was typically between 2 to 4 orders of magnitude greater than the expression of the assayed genes (Supplementary Table 2). Gene expression ratios were determined relative to cultures sparged with 20 mL s^{-1} N_2 gas and are provided in Table 3.4. Most genes analyzed displayed very little (< 2 fold, *P* value < 0.05) significant changes in expression when grown in pH control or under high H_2 (20 mL s^{-1}) or low N_2 (2 mL s^{-1}) gas sparging conditions. This included the '*malate shunt*' genes (*mdh*, *malE*) and each of the hydrogenases assayed. No significant changes in expression were observed with either *pfl* or *ldh*, while a small increase was observed in two of the *pfor* genes grown under low N_2 sparging conditions. Only the genes putatively determined to participate in the conversion of acetyl-CoA to ethanol displayed significant

differential expression according to gas sparging. Expression of the bifunctional acetaldehyde/alcohol dehydrogenase, *adhE*, exhibited the most dramatic change of all genes assayed with a 10-fold increase observed when cultures were grown under conditions of high H_2 gas sparging (Figure 3.4). Alternatively, expression of *adhY* was 5-fold greater in cultures sparged with N_2 gas at 2 mL s^{-1}.

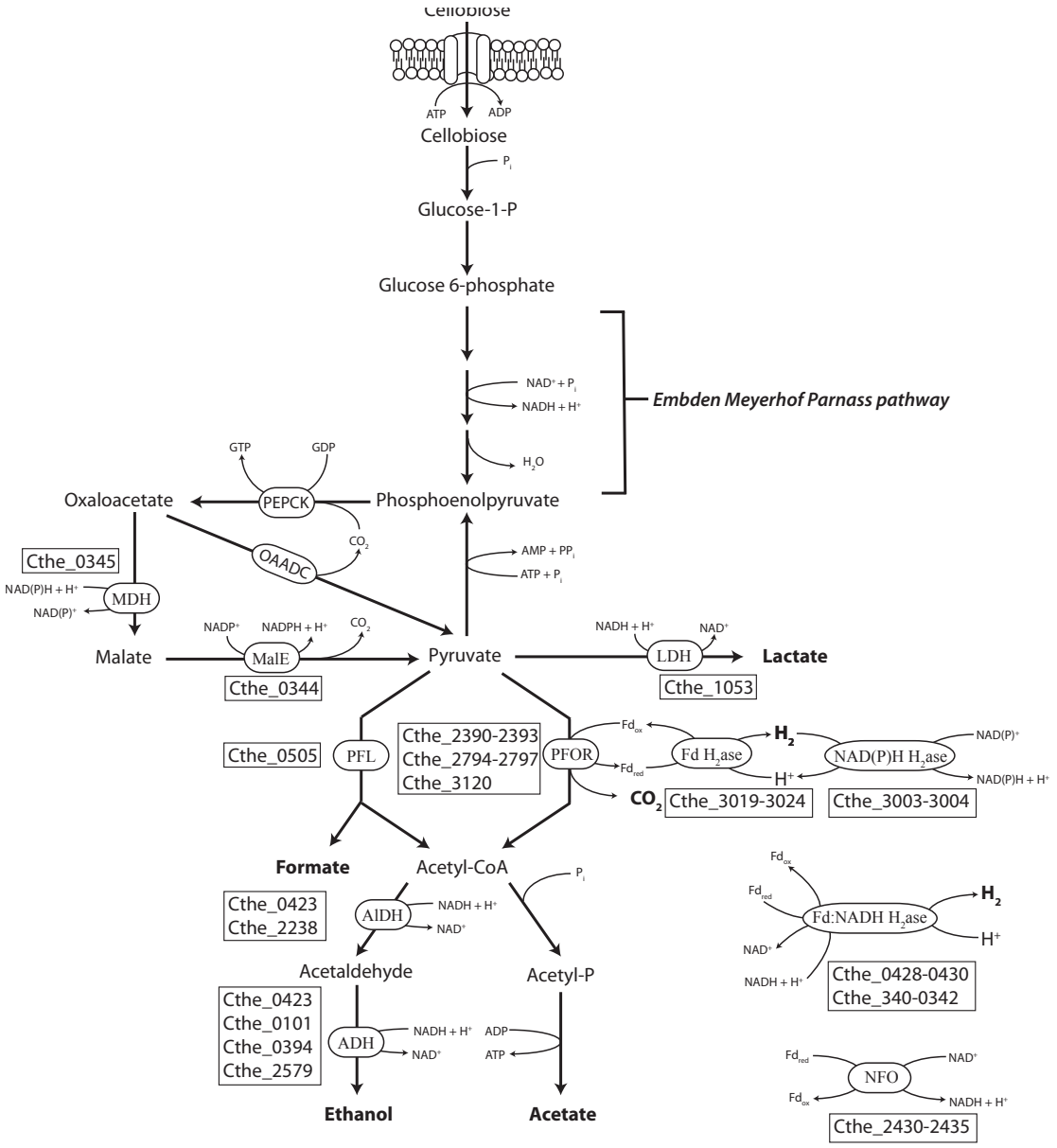


Figure 3.3. Metabolically deduced pathway of *C. thermocellum* ATCC 27405 illustrating the production of lactate, formate, acetate, ethanol and H_2 during cellobiose fermentation. LDH, lactate dehydrogenase; MDH, malate dehydrogenase; MalE, malic enzyme; PFL, pyruvate:formate lyase; PFOR, pyruvate:ferredoxin oxidoreductase; AlDH, acetaldehyde dehydrogenase; ADH, alcohol dehydrogenase; Fd H_2ase, ferredoxin dependent hydrogenase; NAD(P)H H_2ase, NAD(P)H dependent hydrogenase; Fd:NADH H_2ase, ferredoxin and NADH dependent bifurcating hydrogenase; NFO, NADH:ferredoxin oxidoreductase.

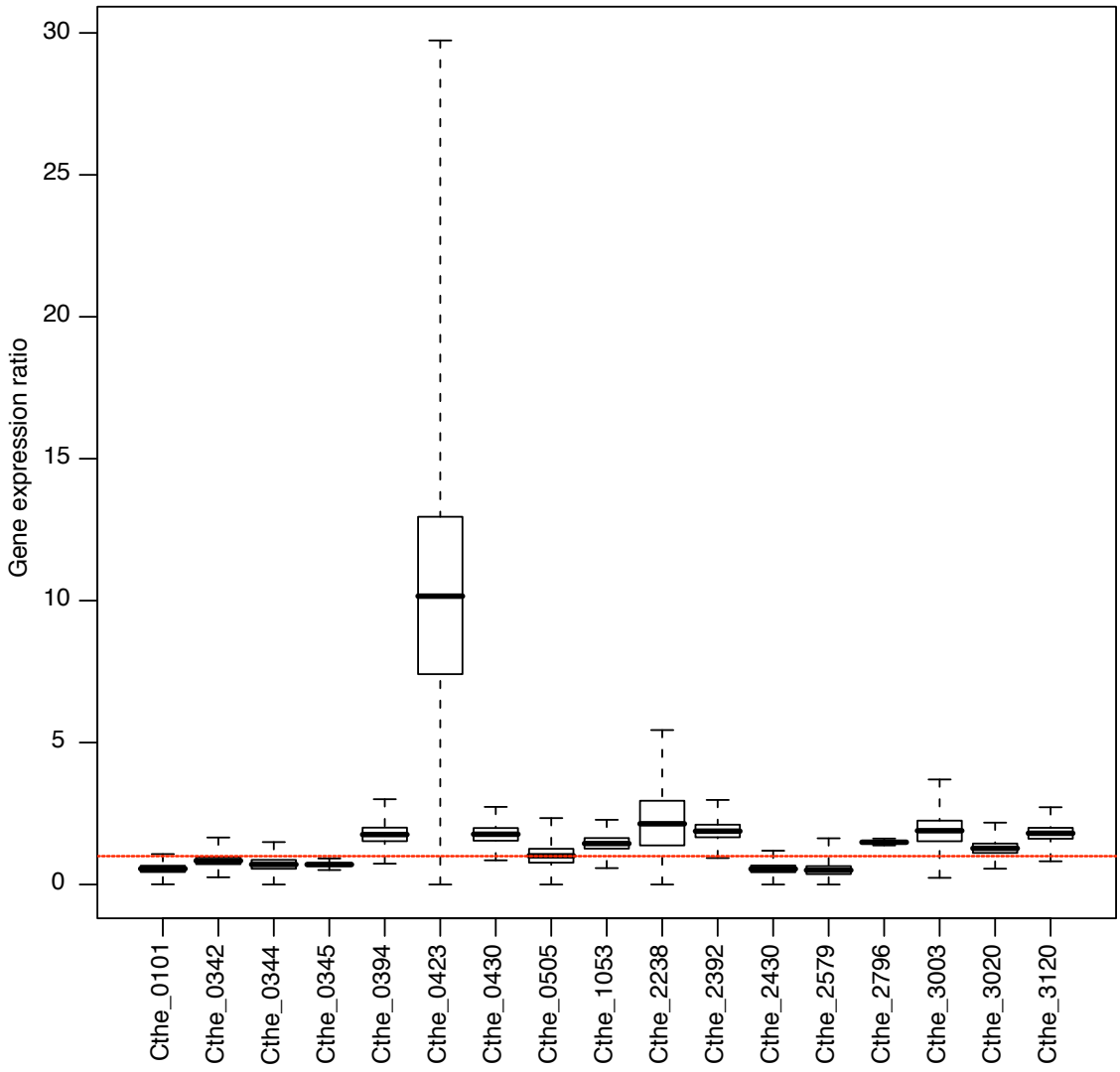


Figure 3.4. Relative expression of genes involved in pyruvate catabolism, ethanol and hydrogen production in *C. thermocellum* ATCC 27405 during mid-exponential phase of growth (OD_{600nm} ~ 0.500) under 20 mL s^{-1} H_2 gas sparging calibrated against 20 mL s^{-1} N_2 gas sparging. Locus tags and their associated gene annotations are listed in Table 3.4

.

Table 3.4. Relative expression of genes involved in pyruvate catabolism, ethanol and hydrogen production in *C. thermocellum* ATCC 27405 during mid-exponential phase of growth (OD_{600nm} ~ 0.600).

Gene	Locus	Gene expression ratio relative to 20 mL s^{-1} N_2:								
		20 mL s^{-1} H_2			2 mL s^{-1} N_2			pH		
		RE[a]	SE[b]	P[c]	RE[a]	SE[b]	P[c]	RE[a]	SE[b]	P[c]
Fe-adh	Cthe_0101	0.610	(0.102)	<0.01	0.555	(0.013)	<0.01	0.557	(0.056)	<0.01
[FeFe] H_2ase	Cthe_0342	0.778	(0.124)	NS	0.758	(0.086)	NS	0.787	(0.045)	<0.01
malE	Cthe_0344	0.645	(0.142)	NS	0.846	(0.187)	NS	0.882	(0.175)	NS
mdh	Cthe_0345	0.685	(0.040)	<0.01	0.544	(0.018)	<0.01	0.594	(0.017)	<0.01
adhY	Cthe_0394	1.714	(0.233)	<0.05	5.758	(0.343)	<0.01	0.937	(0.219)	NS
adhE	Cthe_0423	10.074	(2.753)	<0.01	4.245	(2.144)	NS	2.417	(0.627)	NS
[FeFe] H_2ase	Cthe_0430	1.681	(0.204)	<0.01	1.313	(0.307)	NS	1.465	(0.012)	<0.01
pfl	Cthe_0505	0.746	(0.206)	NS	0.774	(0.352)	NS	0.326	(0.098)	<0.01
ldh	Cthe_1053	1.456	(0.176)	<0.05	1.317	(0.131)	NS	1.212	(0.143)	<0.05
aldh	Cthe_2238	3.076	(0.719)	NS	1.233	(0.810)	<0.05	1.233	(0.415)	NS
pfor	Cthe_2392	1.747	(0.204)	<0.01	2.114	(0.059)	<0.01	1.613	(0.065)	<0.01

nfo	Cthe_2430	0.606	(0.108)	<0.01	0.749	(0.108)	<0.05	0.598	(0.351)	NS
adhZ	Cthe_2579	0.376	(0.134)	<0.05	0.522	(0.154)	NS	0.671	(0.119)	<0.05
pfor	Cthe_2796	1.499	0.019	<0.01	1.754	0.100	<0.01	0.608	0.0258	<0.01
[FeFe] H_2ase	Cthe_3003	1.817	0.332	<0.05	1.866	0.219	<0.01	1.468	0.228	NS
[NiFe] H_2ase	Cthe_3020	1.078	0.144	NS	1.442	0.227	NS	1.170	0.185	NS
pfor	Cthe_3120	1.899	0.186	<0.01	2.298	0.236	<0.01	0.981	0.049	NS

[a] RE: median observed relative expression ratio normalized to 16s rRNA expression and calibrated against 20 mL s^{-1} N_2 gas sparging cultures; values are determined from a minimum of three technical replicates each of biological duplicates

[b] SE: standard error of the observed gene expression ratio;

[c] P: Approximation to p-value for bootstrap sampling of mean ratio. If the median boot ratio is > 1 then P is the relative frequency of median ratio < 1. If median boot ratio is < 1 then P is the relative frequency of median boot ratio > 1; NS, not significant; P > 0.05

3.5 Discussion

The biofuel producing capabilities of *Clostridium thermocellum* have been actively investigated, following the species original isolation, for over fifty years [170]. While much effort has focused on the mechanisms underlying lignocellulose degradation [171], other studies have investigated how metabolic flux in *C. thermocellum* may be manipulated by stirring [125], carbon [62, 102, 167] and end-product loading [128], and enzyme inhibition [134]. In this study the influence on end-product distribution, when cultures of *C. thermocellum* were sparged with N_2 or H_2 gas at elevated rates was investigated. Additionally, changes in the expression of key genes identified to participate in hydrogen and ethanol metabolism was measured.

Fermentative bacteria generate ATP via substrate level phosphorylation during which cellular redox must be maintained in order for metabolism to proceed. In *C. thermocellum*, reducing equivalents NAD(P)H and ferredoxin (Fd) are oxidized via the catalyzed production of lactate, ethanol and H_2. When reducing equivalents are oxidized via hydrogenase, acetyl-CoA may be converted to acetate, yielding ATP. Hydrogenase mediated recovery of reducing equivalents, however, becomes thermodynamically unfavorable when H_2 is permitted to accumulate [125].

The shift in fermentation metabolites we observed towards ethanol production when cultures were sparged with H_2 is suggestive that the oxidation of NAD(P)H/Fd_{red}, via the suite of hydrogenases encoded by *C. thermocellum*, had become thermodynamically unfavorable. Our observed ethanol-to-acetate ratios for N_2 sparged cultures were consistent with reports of stirred *C. thermocellum* cultures grown on cellobiose under 2.5 atm N_2 [125]. Additionally, the shift towards ethanol production

(E:A ratio 1.3) that we observed when cultures were sparged with H_2 was similar to the findings of Lamed *et al.* when *C. thermocellum* strains YS, AS-39, and LQRI were grown under 2.5 atm H_2 with stirring [125]. In 2012, Li *et al* reported a 35 % decrease in acetate production and a 350 % increase in ethanol production when the sparge gas of steady-state cultures of *C. thermocellum* cultures was switched from N_2 to H_2 [134]. This corresponds to 14 mM ethanol, 19.5 mM acetate, and an ethanol-to-acetate ratio of 0.72.

In batch cultures of *C. thermocellum* grown on cellobiose, Rydzak *et al* [128] reported a modest 10 % increase in ethanol production (final yield 4.8 mM) and an ethanol-to-acetate ratio of 1.65 when the N_2 headspace was replaced with H_2 (100 %) gas. A 98 % increase in ethanol production was reported in cultures of *Caldanaerobacter subterraneus* subsp. *Tengcongensis* (formerly *Thermoanaerobacter tengcongensis*) in response to elevated H_2 partial pressure ($p(H_2)$) [172]. In this study the shift towards ethanol production was accompanied by elevated acetaldehyde dehydrogenase and alcohol dehydrogenase activities. Furthermore, reported NADH-dependent hydrogenase activity was 4-fold lower in the elevated $p(H_2)$ cultures. As with *C. subterraneus* subsp. *tengcongensis, C. thermocellum's* primary enzymatic mechanisms of NAD(P)H and ferredoxin oxidation are acetaldehyde dehydrogenase, alcohol dehydrogenase, and hydrogenase (Fd- and/or NAD(P)H-dependent). Additionally *C. thermocellum* expresses lactate dehydrogenase.

We observed that NADH-dependent alcohol dehydrogenase activity was 2.1-fold greater in cultures sparged with H_2 than when sparged with N_2 (20 mL s^{-1}). NADPH-dependent alcohol dehydrogenase activity did not change in response to gas sparging, and acetaldehyde dehydrogenase activity was not determined.

The genome of *C. thermocellum* encodes multiple iron-containing alcohol dehydrogenases including Cthe_0394 (*adhY*), Cthe_2579 (*adhZ*), Cthe_0101 (*adh*) and a bifunctional acetaldehdye/alcohol dehydrogenase (Cthe_0423; *adhE*). In addition to *adhE*, the conversion of acetyl-CoA to acetaldehdye may be catalyzed by an NADH-dependent aldehyde dehydrogenase (Cthe_2238; *aldH*). While crude cell free extract displayed both NADH- and NADPH-dependent alcohol dehydrogenase activities, identifying cofactor specificity on the basis of sequence analysis is problematic. Proteomic analysis of the core catabolic enzymes of *C. thermocellum*, however, has shown AdhE to be the most abundant ADH expressed. The alcohol dehydrogenases encoded by Cthe_0394, Cthe_0101, and Cthe_2579 were reported at 78 %, 24 % and 9 % abundance relative to the level of AdhE expression, respectively [173].

The relative abundance of AdhE and the low levels of reported AldH expression is suggestive that AdhE is the primary enzyme involved in the reduction of acetyl-CoA to ethanol. The contribution of *adhE* to ethanol metabolism is further supported by reports of increased ethanol tolerance correlating with two non-synonymous mutations in *adhE* (Pro-704-Leu; His-734-Arg). In this mutant strain of *C. thermocellum*, the loss of NADH specificity occurred concomitantly with the acquisition of NADPH-dependent activity [142]. The authors proposed that the acquisition of NADPH dependence likely affected electron flow in the mutant strain but did not propose a mechanism for increased ethanol tolerance.

Interestingly, an additional two non-synonymous mutations were identified immediately upstream of *adhE* in a gene annotated as a '*Rex*'-type redox-sensing transcriptional repressor (Cthe_0422) [142]. These redox-sensing proteins have been

implicated in regulating the expression of genes important for fermentative growth in many Gram-positive bacteria [174, 175]. In *Bacillus subtilis* Rex senses NADH/NAD^+ and represses gene expression by binding to 20-bp AT-rich semi-palindromic operator region downstream of the *yjlC-ndh* promoter. Repression is greatest under low NADH:NAD^+ ratios with de-repression occurring as NADH:NAD^+ ratios increase [175]. The 11.5 kbp region of *C. thermocellum's* genome that encompasses Cthe_0422 to Cthe_0431 includes genes that encode the Rex-type redox sensing protein (Cthe_0422), AdhE (Cthe_0423), and both a PAS-fold containing sensory hydrogenase and a heterotrimeric bifurcating hydrogenase (Cthe_0426, Cthe_0428-0430; discussed below; Figure 3.5). It is tempting to imagine how this gene cassette is poised to regulate ethanol and hydrogen metabolism in response to cellular redox.

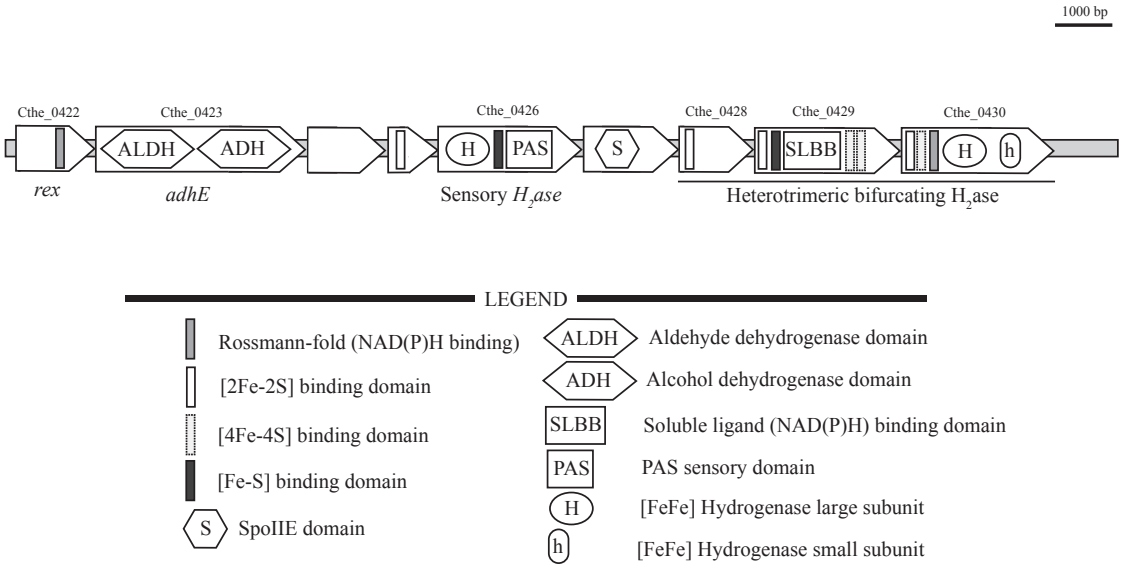


Figure 3.5. Illustration of 11.5 kbp genomic region of *Clostridium thermocellum* ATCC 27405 that includes ORFs Cthe_0422-Cthe_0430. Included in this gene cassette is a redox sensing transcription regulator (*rex*, Cthe_0422), a bifunctional acetaldehyde/alcohol dehydrogenase (*adhE*, Cthe_0423), and both a PAS-fold containing sensory hydrogenase and a heterotrimeric bifurcating hydrogenase (Cthe_0426 and Cthe_0428-0430, respectively).

The increase in transcription that was observed in response to H_2 gas sparging supports the hypothesis that *adhE* may be regulated in response to cellular redox. Transcriptional regulation of *adhE* has previously been reported in *Escherichia coli* in response to anaerobic conditions [176] and in the solventogenic species *Clostridium acetobutylicum* (here annotated as a bifunctional butyraldehdye/butanol dehydrogenase) [177]. In *C. acetobutylicum*, several ADHs appear to dictate solvent/alcohol production; including a second *adhE* encoded on the pSOL megaplasmid. In *C. thermocellum*, the regulation of alcohol metabolism appears to be very complex. Transcriptomic analysis of *C. thermocellum* grown on cellulose displayed increased expression of *aldH* (Cthe_2238) during stationary phase while Cthe_0101 displayed reduced expression in stationary phase cultures [171]. Neither *adhE* nor *adhY* expression changed significantly throughout growth. Stevenson and Weimer [39], alternatively, found (via RT-pPCR analysis) that *adhY* displayed differential expression in response to growth substrate and growth rate. The expression of *adhZ* did not change and *adhE* expression was not reported [167]. We observed that the expression of *adhY* was 5-fold greater in cultures sparged with N_2 gas at 2 mL s^{-1} compared to 20 mL s^{-1} N_2 sparged cultures. Given AdhY is highly expressed at the proteomic level [173], it is likely this enzyme plays an important role in ethanol metabolism that merits further investigation.

Unlike in *C. subterraneus* subsp. *Tengcongensis* [172], neither NADH- nor Fd-dependent hydrogenase activities changed considerably in *C. thermocellum* cultures in response to H_2 gas sparging. NADPH-dependent hydrogenase activity was 5- to 8-fold greater than Fd- and NADH-dependent activities, which is consistent with previous findings [66, 128]. Finally, methyl viologen dependent hydrogenase activity was high

under each condition tested but was reduced nearly 4-fold in 20 mL s-1 N_2 sparged cultures.

The *C. thermocellum* genome encodes at least five [FeFe] type hydrogenases and a single [NiFe] hydrogenase [59]. This includes a putatively identified dimeric NADPH-dependent enzyme (Cthe_3003-3004), two heterotrimeric hydrogenases that resemble the NADH:Fd bifurcating enzyme of *Thermotoga maritima* [155] (Cthe_0428-0430; Cthe_340-342), and a dimeric PAS-domain containing sensory hydrogenase (Cthe_0425-0426). Additionally, Cthe_0355 encodes a monomeric [FeFe] hydrogenase that is immediately downstream from a histidine kinase, while Cthe_3019-3024 encodes a [NiFe] hydrogenase that clusters with '*energy conserving*' type membrane bound enzymes similar to the Fd-dependent '*ech*' hydrogenase of *C. subterraneus* subsp. *tengcongensis* [172].

Given the bidirectional nature of hydrogenases and varied cofactor requirements, determinations of enzyme function *in vivo* can be difficult. The low levels of Fd- and NADH-dependent hydrogenase activity observed (compared to methyl viologen-dependent H_2ase activity), however, suggest the principle method of H_2 production *in C. thermocellum* may be via a bifurcating system. These enzymes utilize the reducing power of Fd to overcome the thermodynamic barriers associated with NADH oxidation to H_2 [155]. Both identified bifurcating hydrogenases have been reported at high levels at the proteomic level during cellobiose fermentation [173].

Transcription of the hydrogenase large subunit (Cthe_0430; Cthe_342) was high, but did not change significantly in response to gas sparging. Similarly, transcription of Cthe_3003 and Cthe_3020 did not display significant differential expression in response

to gas sparging or pH control. Consequently, it appears that the hydrogenases of *C. thermocellum* are not transcriptionally regulated and the observed metabolic shifts in response to gas sparging are likely dictated by thermodynamic considerations and the regulation of alternative redox pathways.

Two additional mechanisms that may function in the maintenance of cellular redox should be briefly addressed. A putatively identified NADH:Fd oxidoreductase has been identified in the *C. thermocellum* genome (NFO; Cthe_2430-2435). If functionally expressed, NFO could catalyze the direct transfer of electrons between Fd and NAD(P)H via this membrane-bound enzyme complex that may be additionally capable of generating a sodium motive force derived from the energy difference between reduced Fd and NADH. Although we detected transcripts of NFO subunit C (Cthe_2430), proteomic analysis, revealed low, or no, expression of several NFO subunits (specifically, only subunits B, C and D were detected), suggesting it does not significantly influence electron exchange between Fd and NADH [173].

Additionally, the genome of *C. thermocellum* encodes genes (Cthe_0373-0372) orthologous to *NfnA* and *NfnB* of *Clostridium kluyveri*. NfnAB encodes a cytoplasmic complex with electron bifurcating NADH-dependent reduced ferredoxin (Fd_{red}):$NADP^+$ oxidoreductase activity. The NfnAB complex of *C. kluyveri* and its' ortholog in *Morella thermoacetica* are proposed to participate in redox maintenance (NADH/NAD^+ and Fd_{ox}/Fd_{red}) via NADPH production during fermentation [178, 179]. We did not test for electron bifurcating NADH-dependent Fd_{red}: $NADP^+$ oxidoreductase activity or transcription in *C. thermocellum*. However, peptides corresponding to Cthe_0373 and Cthe_0372 are reported a high levels in the proteome [173].

Finally, it merits brief mention that the use of H_2 sparging for industrial bioethanol production from lignocellulosic substrates does not represent a cost-effective strategy for increasing yields. The findings reported here are demonstrative of how the H_2- and ethanol- producing pathways of *C. thermocellum* compete for reducing equivalents. Although sparging with H_2 is an effective strategy for directing flux towards ethanol production, metabolic engineering, via the targeted mutation or knockout of hydrogenases, represents a far more realistic approach for achieving the same effect.

3.6 Conclusions

C. thermocellum possesses numerous and seemingly redundant mechanisms to maintain cellular redox that ultimately results in lactate, ethanol, H_2 or biomass production. The diverse suite of hydrogenases encoded in the *C. thermocellum* genome, that includes Fd-, NAD(P)H-dependent and bifurcating enzymes suggests complicated systems are in place to ensure metabolism may proceed in response to changing environmental conditions. Furthermore, the genomic evidence is suggestive that electron transfer may occur directly between NADH, NADPH, or Fd. The shift towards ethanol production that was observed in response to H_2 gas sparging is consistent with previous reports, but this report is the first to investigate this change while considering the expression/activity of key metabolic pathways.

Recent metabolic engineering strategies aiming at improving ethanol yields (via deletion of the phosphotransacetlyase gene and, hence, acetate production) in *C. thermocellum* are illustrative that metabolic system perturbations continue to manifest in unexpected ways [159]. Future investigations at the proteomic, transcriptomic and

metabolomic scale offer the possibility to further resolve how *C. thermocellum* adapts in response to changing environmental conditions.

Chapter 4: Advances in the Development of a Plasmid Transfer System Suitable for Cellulolytic *Clostridial* Species

4.1 Abstract

Enhancing biofuel (H_2, ethanol) yields during the fermentation of lignocellulosic substrates by anaerobes such as *Clostridium thermocellum* is of paramount importance if production costs are to become economically competitive. Currently, a lack of efficient transformation strategies has limited the metabolic engineering of cellulolytic Clostridia to only a handful of studies. The following outlines the electro-transformation and conjugation of *C. thermocellum* 27405 and *C. termtidis* CT1112. Our findings show that the efficiency of transformation for *C. thermocellum* strain ATCC 27405 is consistently low with values of approximately 4.3 x 10^2 cfu/ug DNA. It was determined that the pBC1 replicon was able to maintain plasmid pNW33Nmob at ~ 5 copies $cell^{-1}$. Perhaps as a consequence of reduced host restriction endonuclease activity, transformation frequencies were approximately 100-fold higher in *C. termitidis*. Observed frequencies of plasmid transfer, via conjugation, were similar in both *C. thermocellum* and *C. termitidis* suggesting the transfer of single stranded DNA may circumvent aggressive restriction methylation systems. While transformants containing the replicative plasmid pNW33mob may be subcultured for multiple passages, plasmid loss remains a concern. Transformation efficiency appears to be related to the presence of restriction methylation (RM) systems encoded in the genomes of *C. thermocellum* (strains ATCC 27405, DSM 4150, DSM 2360 and DSM 1313), *C. phytofermentans*, *T. thermosaccharolyticum*, *C. cellulolyticum* H10, and *C. termitidis,* and suggests screening for RMs may be

advantageous prior to attempting genetic modification. Given the difficulties associated with isolating recombinant mutants using suicide vectors in Clostridial species, alternative strategies that utilize replicative plasmids are now being developed. Low frequency of transformation, however, continues to hinder attempts at genetic manipulation. Findings illustrate that although transformation must be evaluated on a 'case by case' basis, conjugative transfer of plasmid pNW33Nmob is possible into a range of recipient species.

4.2 Introduction

One necessary prerequisite to improve biofuel yield is the development of suitable microorgansims that are capable of efficiently fermenting a variety of sugars while simultaneously displaying tolerance to high end-product concentrations. Metabolic engineering of bacterial and yeast species has demonstrated that the construction of novel strains using recombinant DNA technologies can confer advantageous traits in regards to biofuels production.

The Firmicute Phylum includes an ever-increasing number of bacterial species that are noted for their ability to both efficiently degrade lignocellulose and ferment oligosaccharides into ethanol or other fuel chemicals. *Clostridium thermocellum*, a gram-positive thermophile, has served as a model for microbial cellulose hydrolysis because of its characteristic cellulosome, an extracellular multi-enzyme complex that facilitates degradation [33]. Other thermophilic species, (e.g. *Clostridium stercorarium*, *Thermoanaerobacter pseudethanolicus*, *Caldicellulosiruptor bescii*) and mesophiles (e.g. *Clostridium cellulolyticum*, *Clostridium phytofermentans, Clostridium termitidis*) are likewise being investigated as candidates for biofuels production.

A report of natural competence in *Thermoanaerobacter* and *Thermoanerobacterium* species [180] contrasts starkly with other studies illustrating the recalcitrance of Clostridial species to genetic manipulations [181-183]. Although the successful genetic manipulation of some cellulose degrading species has been recently reported, the genetic tools available remain largely under-developed [159, 181, 182, 184-187]. The nascent state of available technologies represents a significant obstacle to our understanding of cellulose hydrolysis, microbial metabolism and the development of

industrially relevant microorganisms. Here findings relating to the electroporation and conjugative plasmid transfer into, *Clostridium thermocellum* ATCC 27405, *Thermoanaerobacter pseudethanolicus* 39E and *Clostridium termitidis* CT1112 are reported. These findings represent both the first description of conjugative plasmid transfer into thermophilic Clostridia and the first report of transformation of *C. termitidis*.

4.3 Methods

4.3.1 Bacterial Strains and Plasmids Used

The various bacterial strains and plasmids used in conjugation and electroporation experiments are summarized in Table 4.1.

4.3.2 Culture Media and Bacterial Growth

Escherichia coli strains were maintained on Luria-Bertani medium at 37 °C [188]. *Clostridium thermocellum* 27405, *Thermoanerobacter pseudethanolicus* 39E and *Clostridium termitidis* CT1112 were both obtained from the American Type Culture Collection (ATCC) and used for all growth experiments. Fresh cultures were maintained by routinely transferring 10 % (v/v) into anaerobically prepared 1191 media containing (pH 7.2, per 1000 mL distilled deionized water) KH_2PO_4, 1.5 g; Na_2HPO_4-$12H_2O$, 4.2 g; NH_4Cl, 0.5 g; $MgCl_2$-$6H_2O$, 0.18 g; yeast extract (BD 212750), 2.0 g; L-cysteine-HCl, 1.0 g; 0.25 mg/ml resazurin, 1.0 mL; vitamin solution, 0.5 mL; mineral solution, 1 mL. Vitamin solution contained the following (per 1000 mL): biotin, 20 mg; *p*-aminobenzoic acid, 50 mg; folic acid, 20 mg; nicotinic acid, 50 mg; thiamine, 50 mg; riboflavin, 50 mg; lipoic acid, 50 mg; cyanocobalamin, 10 mg. Mineral solution contained (per 1000 mL): trisodium nitrilacetate, 20.2 g; $FeCl_3$-$6H_2O$, 2.1 g; $CoCl_2$-$6H_2O$, 2.0 g; $MnCl_2$-$4H_20$, 1.0;

$ZnCl_2$, 1.0 g; $NiCl_2$-$6H_2O$, 1.0 g; $CaCl_2$-$2H_2O$, 0.5 g; $CuSO_4$-$2H_2O$, 0.5 g; Na_2MoO_4-$2H_2O$, 0.5 g.

Balch tubes or serum bottles (Bellco Glass Co., New Jersey, USA) were prepared with 10 mL or 50 mL 1191 media respectively, sealed with butyl rubber stoppers, aluminum seals and gassed/degassed (1:4 min) for 3 cycles prior to autoclaving. After cooling, sterile cellobiose was added to a final concentration of 2 g L^{-1}. For plating experiments, 1191 media (as described above) in addition to 1.0 % Agar was prepared aerobically and autoclaved. Cellobiose was added to a final concentration of 2 g L^{-1} (in addition to appropriate antibiotics) and 25 mL plates were poured and allowed to solidify aerobically before being transferred into a GasPak EZ Gas Generating Container System (BD Diagnostics, Franklin Lakes, NJ, USA) for overnight reduction.

Table 4.1. Bacterial strains and plasmids used for conjugation/electroporation experiments.

Strain or plasmids	Relevant characteristics	Source/reference
Bacterial strains:		
Clostridium thermocellum ATCC 27405	Wild-type	ATCC
Thermoanaerobacter pseudethanolicus 39E	Wild-type	ATCC
Bacillus cereus	Nx^R	This study
Clostridium termitidis CT1112	Wild-type	ATCC
Escherichia coli DH5α	Dam^+ Dcm^+	Invitrogen
Escherichia coli BL21	Dam^+ Dcm^-	New England Biolabs
Escherichia coli S17-1	recA pro hsd^R RP4-2-Tc::Mu-Km::Tn7 tmp^R, spc^R, str^R	[189]
Plasmids:		
pNW33N	3.85 kb, pBC1, pBR322,CmR, Ap^R	BGSC No. ECE136
pIkm1	6.19 kb, ColE1, pIM13 ApR, Kan^R, MLS^R	[190]
pNW33Nmob	7.785 kb, pBC1, pBR322, Cm^R, Ap^R	This study
pIkm1mob	9.6 kb, pIM13, pBR322, Kan^R, MLS^R	This study
pSUP202	7.83 kb, mob^+, Cm^R, Tc^R, Ap^R	[189]
pRK2013	48kb, ColE1, tra^+, mob^+, Kan^R	[191]
pHV33	7.21 kb, ColE1, Tc^R, Ap^R	
pMTL007C-E2	9.03 kb, pCB102, ColE1, Cm^R, Em^R	[192]

4.3.3 Plasmid Construction

Strains and plasmids used in this study and their relevant features are provided in Table 4.1. Dr. Juergen Wiegel from the University of Georgia, USA, generously provided plasmid pIkm1. Plasmid pNW33N was obtained from the Bacillus Genetic Stock Center (http://www.bgsc.org/) and plasmid pMTL007C was obtained from DNA 2.0 (Menlo Park, CA) with the permission of Dr. Nigel Minton (University of Nottingham, UK).

Mobilizable plasmids pIkm1mob and pNW33Nmob were constructed by cloning a 3.4 kb *BamHI-EcoRI* fragment and a 4.2 kb *PstI-HindIII* fragment from plasmid pSUP202, containing the mobilization region (*mob*), into plasmids pIkm1 and pNW33N respectively. Vectors were constructed using standard molecular cloning techniques [188]. Plasmid DNA was isolated from *E. coli* DH5α cells using a QIAprep Miniprep kit (QIAGEN, Valencia, CA).

4.3.4 Plasmid Copy Number Determination

Plasmid copy number was estimated using real-time quantitative polymerase chain reaction (RT qPCR). Total DNA was extracted from stationary-phase clonal transformants using a Wizard Genomic DNA purification kit (Promega, Madison, WI) following the pretreatment protocol recommended for DNA isolation from Gram-positive bacteria. Approximately 10 ng of extracted DNA was used as a template for subsequent qPCR experiments. Reactions were run on a Stratagene Mx3005p thermocycler using iTaq SYBR Green Supermix with Rox (BioRad, Mississauga, ON) and primers designed to amplify the chloramphenicol acetyltransferase (*cat*) gene of plasmid pNW33Nmob and 16S rDNA of *C. thermocellum* ATCC 27405 (Supplementary Table 3). Specific reaction

conditions are provided in section 3.3.8. Relative ratios (pNW33Nmob:16S rDNA) were calculated from pNW33Nmob and 16S rDNA-pGEMt standard curves and used to estimate plasmid copy number.

4.3.5 Conjugation Methodology

The optimized filter mating conjugation protocol was adapted from reports of conjugation in *Clostridium cellulolyticum* [182] and other various saccharolytic clostridial species [192]. Donor (and helper) strains of *E. coli* harboring different plasmids were grown aerobically grown overnight, at 37 °C, to stationary phase in LB media supplemented with antibiotics appropriate for the selection of conjugative plasmids (50 μg mL^{-1} chloramphenicol, for plasmids pMTL007C and pNW33Nmob; 50 μg mL^{-1} kanamycin for plasmid pRK2013). The following morning stationary phase donor and helper cultures were used to inoculated fresh tubes (10 mL at 10 % v/v inoculum) and regrown to an OD_{600nm} 0.5-0.7 (~ 7.23 x 10^6 cfu ml^{-1}). Donor and helper strains were washed twice in 1 mL pre-warmed LB to remove residual antibiotics before mating.

Recipient cultures were grown anaerobically at 58 °C for *C. thermocellum* and *T. pseudethanolicus,* and 37 °C for *C. termitidis,* for 18-36 h in 1191 media to mid-exponential phase (OD_{600nm} 0.5-0.7). All manipulations of the recipient species were performed in a Thermo Forma Anaerobic Chamber (Model 1025, Thermo Fisher Scientific, Waltham, MA, USA) with an atmosphere of 10 % hydrogen, 90 % nitrogen. Donor and recipient cells were mixed at 10:1, 1:1, 1:5 and 1:10 ratios for biparental mating experiments. For triparental mating experiments, a helper:donor:recipient ratio of 10:10:1 was used. Mating mixtures were concentrated by centrifugation at 4500 x g for 10 min. The pellets were resuspended in 100 μL 1191 media (containing 2 g L^{-1}

cellobiose) and added, drop wise, onto a nitrocellulose filter (0.45 μm, 25 mm diameter, Millipore, Billerica, MA, USA) overlaid on an 1191 Agar plate supplemented with 2 g L^{-1} cellobiose. Mating mixtures were incubated in a GasPak EZ Gas Generating Container System for 6-24 h at 37 °C, bacteria uppermost, before being harvested by vortex mixing for 60 s in 1 mL 1191. Alternatively, liquid mating was performed in a 3 mL sterile, anaerobic glass vial to which 100 ul 10 % glucose was added. Mating mixtures were serially diluted with PBS, in order to prevent overwhelming, prior to plating in the presence of antibiotics appropriate for plasmid selection. Cycloserine (250 μg ml^{-1}) was used to counter-select for donor/helper *E. coli* strains. All transfer frequencies reported represent the mean of at least three determinations.

For triparental mating experiments, donor *E. coli* strains (DH5α, BL21), harboring conjugative plasmids, were mated with recipient cultures using *E. coli* DH5α (pRK2013) as a helper. For biparental mating experiments, recipient species were mated with donor *E. coli* S17-1 a strain with chromosomally integrated *tra* genes of plasmid RP4 [189].

4.3.6 Electro-Transformation Methodology

Electro-transformation protocols were modified from Guss *et al.,* [6] and Tripathi *et al.,* [11]. Four hundred (400) mL of cells were harvested for electroporation following overnight growth to an OD_{600} between 0.4 and 0.6, by centrifugation in a Sorvall RC6 centrifuge with a SH-3000 rotor at 4500 x g for 12 minutes at 4 °C. Following the removal of supernatant, cell pellets were incubated on ice for 20 min before the addition of 2 mL ice-cold electroporation buffer (EPB; 280 mM Sucrose, 100 mM $NaPO_4$, pH 7.0). The cells were washed and centrifuged a second time, as described above, before

finally being resuspended in 1 mL EPB and kept on ice until use. All cell manipulations were performed anaerobically within a Thermo Forma Anaerobic System Model 1025. Plasmid DNA (4 uL at 400-1200 ng μL^{-1}) was added to 20 µL washed cells in a pre-chilled 1 mm electroporation cuvette and mixed by gentle tapping. The Cell/Plasmid DNA mixture was then allowed to incubate on ice before being subjected to 10 kV cm^{-1} pulse (500 Ω resistance, 25 µF capacitance) using a BTX ECM 630 Exponential decay wave electroporation system. The resulting pulse duration was 1.5-2.5 ms. Immediately following pulse application, cells were suspended with 1 mL chilled 1191 growth medium containing 2 g L^{-1} cellobiose, incubated on ice for 10 min and then transferred into 15 mL conical tubes containing an additional 2 mL pre-warmed media and allowed to recover for 6 h at 37/58 °C.

Following the recovery period, 100 µL of the transformation mixture was plated onto plates that had been prepared 24 h previously (see *Culture media and bacterial growth*) and supplemented with appropriate antibiotics. Serial dilution of the transformation mixture (with PBS) was often required to ensure reliable colony counts and to avoid 'overwhelming' selection plates and the occurrence of false-positive growth. Plates were incubated at 37/58 °C for 5-8 days or until colony formation was observed. Successful transformation/conjugation was confirmed by re-isolation of plasmid DNA. From 3 mL cultures derived from antibiotic resistant colonies, either total DNA was isolated using the Wizard Genomic DNA purification kit (Promega, Madison, WI) following the pretreatment protocol recommended for DNA isolation from Gram-positive bacteria or, alternatively, plasmid DNA was isolated using a QIAprep Miniprep kit (QIAGEN, Valencia, CA). Isolated DNA was then transformed into *E. coli* DH5α cells,

re-isolated, and subjected to restriction enzyme digestion as a diagnostic to confirm plasmid identity.

4.3.7 Analysis of Host Restriction Methylation Systems

The genomes of *Clostridium thermocellum* (strains ATCC 27405, DSM 4150, DSM 2360 and DSM 1313), *Clostridium phytofermentans* ISDg, *Thermoanaerobacter thermosaccharolyticum* M0795, *Clostridium cellulolyticum* H10, and *Clostridium termitidis* were screened *in silico* for restriction methylation (RM) systems using the Joint Genome Institute's Integrated Microbial Genomes (IMG) database [67]. The gene annotations presented here reflect the numbering of the final assembly or most recent draft available (July, 2012). *In silico* screening for restriction endonucleases and DNA methylases were conducted using three annotation platforms independently: i) Clusters of Orthologs Groups (COGs)[68], ii) KEGG Orthology assignments (KO) [69], and (iii) TIGRFAMs [70]. Genes identified using a single annotation platform were cross-referenced against the other two platforms to determine a consensus functional annotation. Functional annotations not supported by all three platforms were evaluated on a case-by-case basis and decisions were made using a combination of manual analysis of genomic context, literature searches, and functional prediction through RPS-BLAST using the Conserved Domain Database website [71].

In tandem with the genomic investigation of *C. termitidis*, total proteome analysis of *C. termitidis* was performed to screen for expressed proteins that resembled restriction endonucleases. Compositional proteomic studies were performed on α–cellulose and cellobiose cultures of *C. termitidis* at mid-exponential and stationary phase to determine

the relative protein expression levels using 2D RP HPLC-MS/MS [Our unpublished data].

4.4 Results

4.4.1 Electro-Transformation of *Clostridium thermocellum* ATCC 27405

In preliminary experiments, *C. thermocellum* ATCC 27405 was electro-transformed with plasmids pIkm1 and pNW33Nmob. Liquid selection prevented an estimation of transformation efficiency, however, growth was consistently observed within 40 h of incubation at 58 °C (Figure 4.1). Application of a field pulse strength 10 kV cm^{-1} or greater resulted in positive outgrowth and, hence, transformation. Experiments with an applied field pulse of 7 kV cm^{-1} or less failed to yield growth in the presence of selection (Supplementary Table 4). No growth was observed within 72 h incubation in negative control tubes. A erythromycin resistant phenotype was retained in subsequent passages of pIkm1 transformation experiments (data not shown). In these early electroporation experiments, PCR amplification of the macrolide-lincosamide-streptogramin (*mls*) gene was used to confirm the presence of plasmid pIkm1 using total DNA preparations as template and thus be diagnostic of transformation (Supplementary Figure 3).

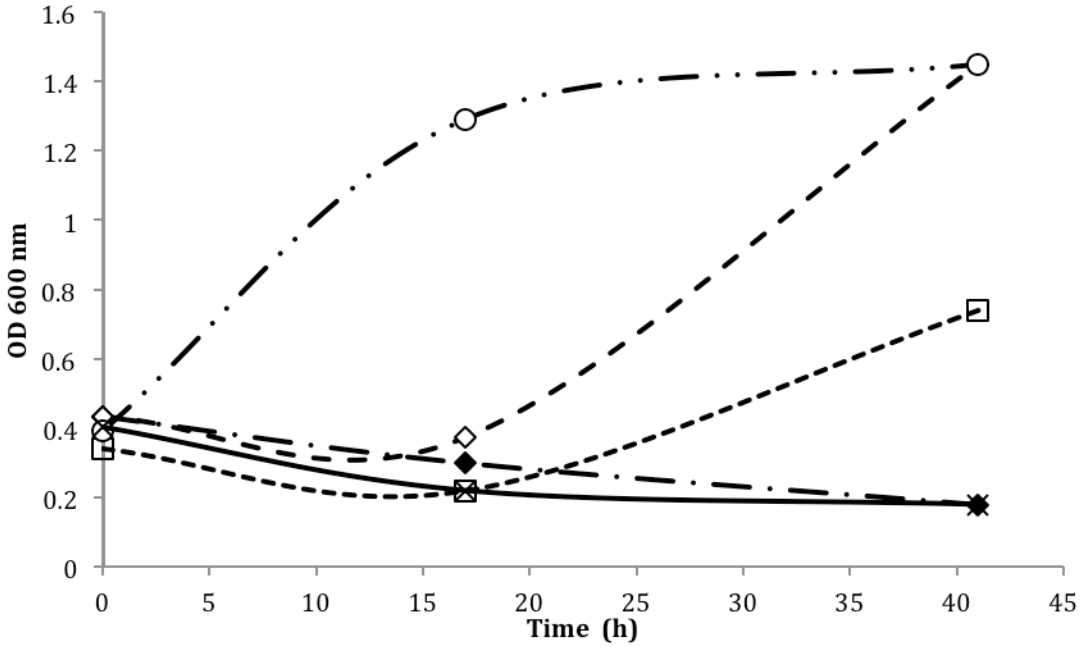


Figure 4.1. Growth of *Clostridium thermocellum* ATCC 27405 following electroporation with plasmid pIkm1. Liquid selection was performed with erythromycin at 20 μg mL^{-1}. Field pulses of 7 kV cm^{-1} (◆), 10 kV cm^{-1} (□), 12.5 kV cm^{-1} (◇), positive control (average of 7 kV cm^{-1}, 10 kV cm^{-1}, and 12.5 kV cm^{-1} field pulses with no selection following recovery; ○) and negative control (no plasmid, erythromycin at 20 μg mL^{-1}; ×) results are shown.

Electro-transformation efficiency of *Clostridium thermocellum* ATCC 27405 with plasmid pNW33Nmob, calculated from colony counts, was consistently low with values averaging 3.3 x 10^2 cfu ug^{-1} DNA (± 1.0 x 10^2 cfu ug^{-1} DNA, average of minimum 3 independent experiments). Neither alternative electroporation buffers (10 % glycerol, 0.2 M cellobiose, or 5 mM HEPES 270 mM sucrose at pH 7.4), nor the addition of isoniacin (20 ug mL^{-1}) to early exponential phase cultures was found to significantly affect transformation efficiency (data not shown).

Following electroporation, plasmid pNW33Nmob was stably maintained for five passages (10 % v/v) in the presence of selection. While growth following electroporation in the presence of selection was consistently observed following 40 h incubation, subculture growth was seen within 20 h of inoculation at 58 °C (Figure 4.2). Plasmid pNW33Nmob was recovered from each subculture and confirmed by back transformation and recovery from *E. coli* DH5α (Figure 4.3).

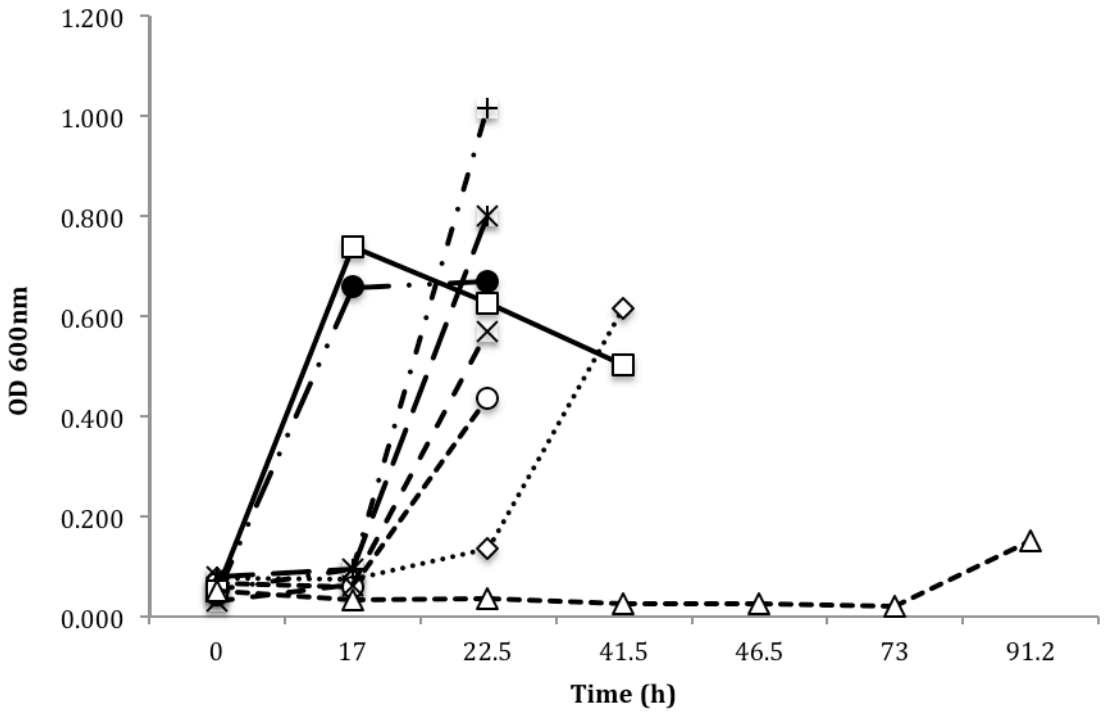


Figure 4.2. Electroporation and subsequent subculturing (10 % v/v) of *Clostridium thermocellum* ATCC 27405 transformants with plasmid pNW33N (liquid selection, chloramphenicol 50 μg ml^{-1}). Electroporation (◇), first subculture (○), second subculture (×), third subculture (+), fourth subculture (∗), fifth subculture (●), positive control (□), negative control (△). Data points represent the average of experimental triplicates.

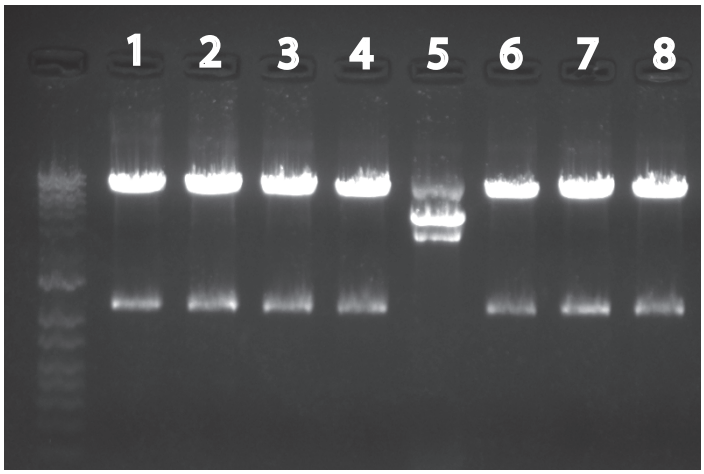


Figure 4.3. *EcoR*I restriction endonuclease digest of plasmid pNW33Nmob recovered from *E. coli* back transformed from *C. thermocellum* ATCC electroporations and subsequent passages. Lanes 1 and 2, electroporation; lane 3, subculture 1; lanes 4 and 5, subculture 2; lane 6, subculture 3; lane 7, subculture 4; lane 8, subculture 5.

Plasmid copy number was estimated with quantitative PCR by comparing pNW33Nmob:16S rDNA ratios from total DNA extractions of transformed *C. thermocellum*. Amplification of the chloramphenicol acetyltransferase gene of pNW33Nmob suggests that the plasmid is maintained at a low copy number in *C. thermocellum* ($\leq$ 5 copies cell^{-1}). Digestion of total DNA recovered from *C. thermocellum* transformants with enzyme *PstI* (which recognizes a single sequence in plasmid pNW33Nmob) and subsequent ligation prior to back transformation into *E. coli* was used to test for plasmid integration into the genome. Characterization of recovered pNW33Nmob plasmid from back transformation experiments by *EcoRI* digest continued to display the expected 7.2 kb and 1.2 kb fragments and suggests plasmid pNW33Nmob remains episomal (Figure 4.4).

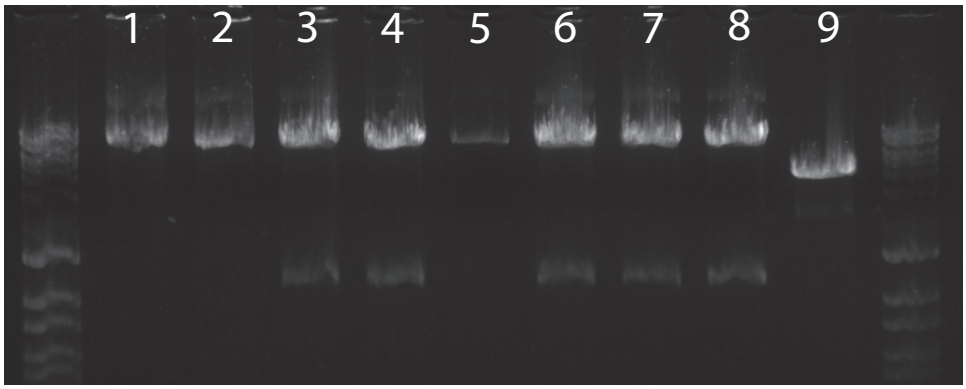


Figure 4.4. *EcoR*I restriction endonuclease digest of plasmid pNW33Nmob recovered from *E. coli* back transformed from *C. thermocellum* ATCC 27405 electroporations. Lanes 1-3, recovered total DNA from *C. thermocellum* transformants was used to transform *E. coli*; lanes 4-6, recovered total DNA from *C. thermocellum* transformants was PstI digested and re-ligated prior to transformation into *E. coli*; lanes 7, recovered plasmid DNA (via miniprep) from *C. thermocellum* transformants was used to transform *E. coli*; lane 8, pNW33Nmob; lane 9, pNW33N.

Preparations of plasmid pNW33Nmob, methylated *in vivo* and treated with the crude cell lysate of *C. thermocellum* ATCC 27405, indicated *dam* methylation ($G^{m}ATC$) is sufficient to maintain plasmid integrity and protect plasmid pNW33Nmob from host endonuclease activity. In contrast, plasmid DNA prepared from dam^{+}/dcm^{-} and dam^{-}/dcm^{-} *E. coli* strains displayed non-discrete degradation when incubated in the presence of *C. thermocellum* ATCC 27405 lysate.

4.4.2 Conjugation of *Clostridium thermocellum* ATCC 27405 and *Thermoanaerobacter pseudethanolicus* 39E

Plasmids pNW33Nmob and pIkm1mob were tested for their ability to be transferred by conjugation into both mesophilic and thermophilic recipient species. An overview of the experiments performed is outlined in Table 4.2. In these preliminary experiments, growth in the presence of selection (relative to controls) was deemed to be a sufficient indicator of transformation.

Plasmid pNW33Nmob was transferred to recipient species *C. thermocellum* (strains ATCC 27405 and DSM 1313) and *Thermoanaerobacter pseudethanolicus* 39E by conjugation. As with the electroporation of *C. thermocellum*, outgrowth following a 6-24 h conjugation period, in the presence of liquid selection, was observed within 40 h of incubation at 58 °C (Figure 4.5). Wildtype *E. coli* S17-1 and a dcm^{-} mutant strain both functioned as suitable donor strains for the conjugative transfer of plasmid pNW33Nmob (data not shown). As with the electroporation of *C. thermocellum*, subculturing of exoconjugants maintained chloramphenicol resistance. We estimated conjugation frequency into *C. thermocellum* ATCC 27405 to be $7.61 \times 10^{-6} \pm 1.67 \times 10^{-6}$ exoconjugants $donor^{-1}$. Exoconjugant colonies, however, frequently would develop no

further than pinprick sized colonies and liquid selection proved to be far more reliable. Nevertheless, the presence of plasmid pNW33Nmob was confirmed by both PCR amplification (Supplementary Figure 4) and back-transformation into *E. coli* followed by restriction endonuclease digestion.

Table 4.2. An overview of conjugation experiments performed. Outgrowth in the presence of liquid selection (relative to controls) was used as evidence of successful transformation. Nx, naladixic acid; Cm, chloramphenicol; Cyc, cycloserine; Kan, kanamycin.

Donor	Plasmid	Recipient	Selection condition	Outgrowth ? (Y/N)
E. coli S17-1	pNW33Nmob	*E. coli* Nx^R	37 °C aerobic Nx_{50} Cm_{20}	Yes
	pIkm1mob		37 °C aerobic Nx_{50} Kan_{50}	Yes
	pNW33Nmob	*B. cereus* Nx^R	37 °C aerobic Nx_{50} Cm_{20}	Yes
	pIkm1mob		37 °C aerobic Nx_{50} Kan_{50}	Yes
	pNW33Nmob	*G. debilis*	60 °C anerobic Cm_{20}	Yes
	pIkm1mob		48 °C aerobic Kan_{50}	Yes
	pNW33Nmob	*G. sulfurreducens*	34 °C anerobic Cm_{50}	Yes
	pNW33Nmob	*T. pseudethanolicus*	58 °C anerobic Cm_{50}	Yes
	pIkm1mob		58 °C anerobic Kan_{10}	No
	pNW33Nmob	*C. thermocellum*	58 °C anerobic Cm_{100}	Yes
	pIkm1mob		58 °C anerobic Kan_{10}	No
	pMTL007C-E2	*C. termitidis*	37 °C anerobic Cm_{50} Cyc_{250}	Yes
	pNW33Nmob			Yes

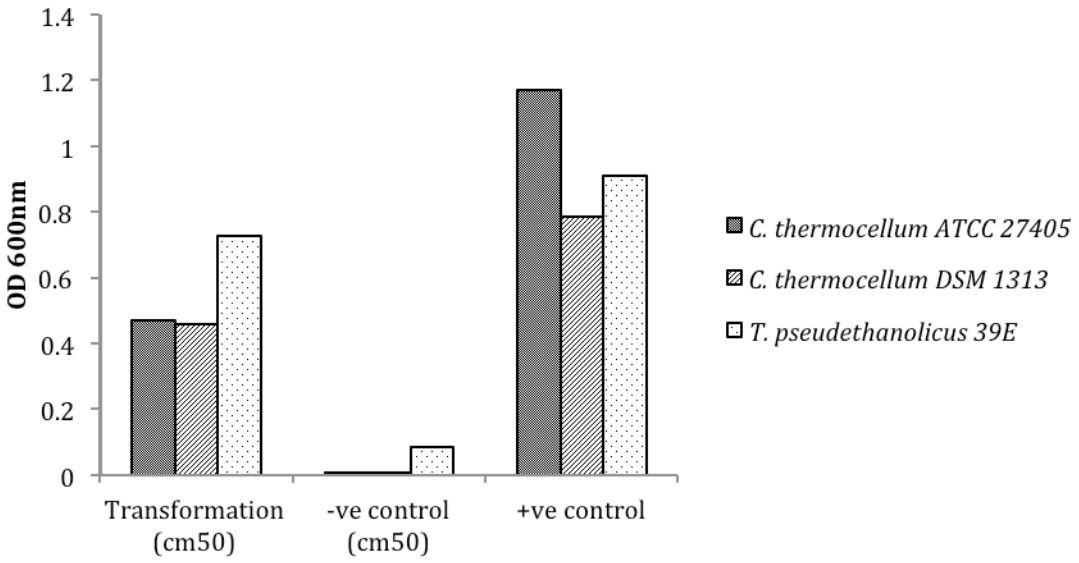


Figure 4.5. Growth of recipient species *Clostridium thermocellum* (strains ATCC 27405, DSM 1313) and *Thermoanaerobacter pseudethanolicus* 39E following conjugation with *E. coli* S17-1 harboring the mobilizable plasmid pNW33Nmob. Growth was evident within 48 h of incubation at 58 °C.

4.4.3 Conjugation and Electro-Transformation of *Clostridium termitidis*

The mesophile, *Clostridum termitidis* was successfully transformed via conjugation and electroporation with both plasmid pMTL007C and plasmid pNW33mob. Growth of electro-transformed cells in liquid selection was observed within 40 h (Figure 4.6). *C. termitidis* electro-transformation efficiency was calculated to be 1.88×10^4 cfu ug^{-1} DNA ($\pm 1.10 \times 10^4$ cfu ug^{-1} DNA), approximately 10^2 fold more frequent compared to both the transformation of *C. thermocellum* ATCC 27405 or to reported transformation frequencies of *Clostridium cellulolyticum* [182]. The appearance of colonies in the presence of selection became evident within 3 to 5 days of incubation at 37 °C.

Both biparental and triparental *C. termitidis* mating experiments were performed. We observed transfer frequencies from triparental matings into *C. termitidis* from *E. coli* DH5α of 7.23×10^{-6} ($\pm 2.21 \times 10^{-6}$) exoconjugants $donor^{-1}$ and from *E. coli* BL21 1.76×10^{-6} ($\pm 9.01 \times 10^{-7}$) exoconjugants $donor^{-1}$. Conjugation frequency does not appear to be significantly influenced by the donor strain used (DH5α, BL21 or S17-1) and consequentially the *dcm* methylation state of plasmid DNA. It is possible that potential effects may have been masked by the poor plating efficiencies of *C. termitidis*, by difficulty in recovering cells or as a result of the extended mating period. Predictably, at elevated chloramphenicol concentrations the number of exoconjugants was reduced.

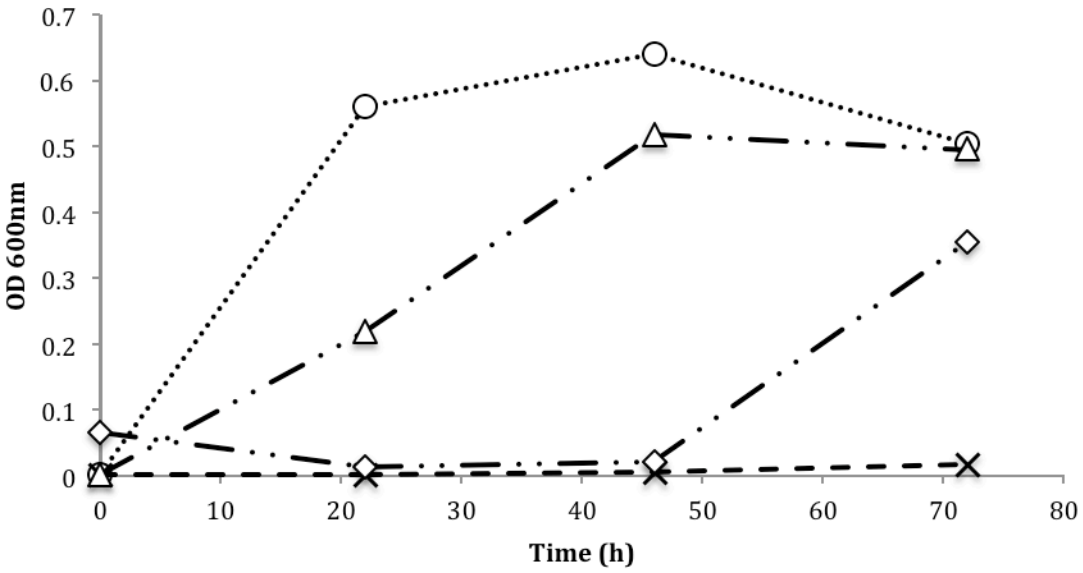


Figure 4.6. Electroporation of *Clostridium termitidis* CT1112 with plasmids pNW33Nmob (◇) and pMTL007C-E2 (△) (liquid selection, Thiamphenicol 50 μg ml^{-1}). Positive control (○), negative control (×). Data points represent the average of technical triplicates.

The effect of donor:recipient ratio on conjugation efficiency was investigated and is shown in Figure 4.7. A modest 3.5-fold increase in transfer frequency was observed in 1:10 donor:recipient mating mixtures compared to 1:1 mixtures. The appearance of colonies on selective plates consistently followed 7-11 days of incubation and, unlike the exoconjugnt colonies of *C. thermocellum*, *C. termitidis* exoconjugant colonies developed fully. In all instances no colonies were observed on either the *C. termitidis* or *E. coli* negative-control plates. Conjugation was successfully confirmed by recovering plasmid pNW33Nmob DNA from *C. termitidis* colony growth and "back-transforming" *E. coli* DH5α. A HindIII/PstI restriction digest of recovered pNW33Nmob DNA was used verify plasmid identity with expected band sizes of 4.21 and 3.58 kb (Figure 4.8).

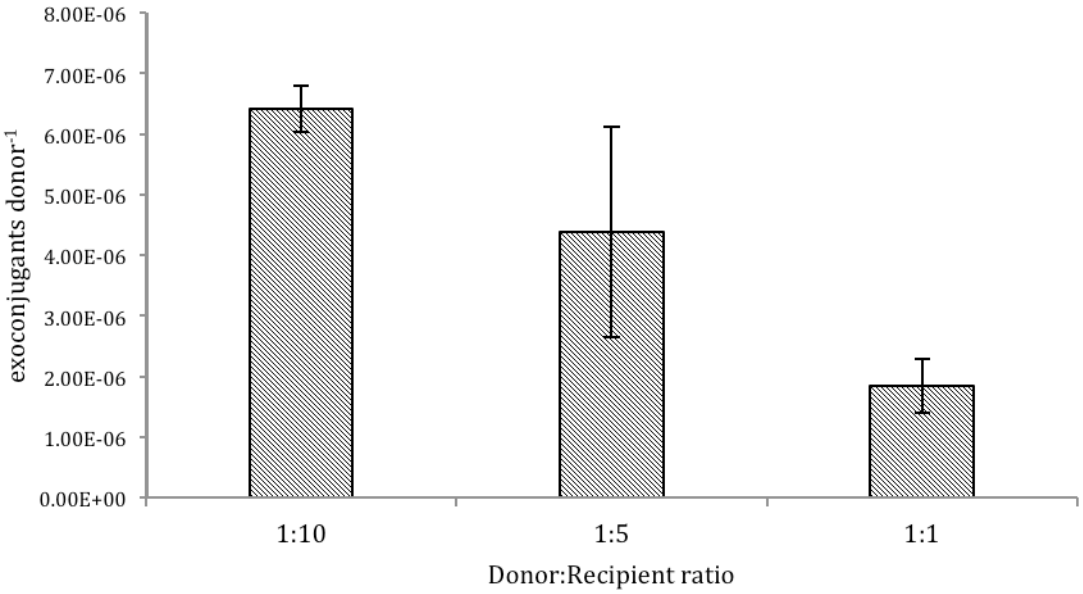


Figure 4.7. Biparental conjugal plasmid transfer into *Clostridium termitidis* at 1:10, 1:5 and 1:1 Donor:Recipient ratios. Displayed values represent the mean of at least three conjugal matings with plasmid pNW33Nmob. Error bars represent the standard deviation.

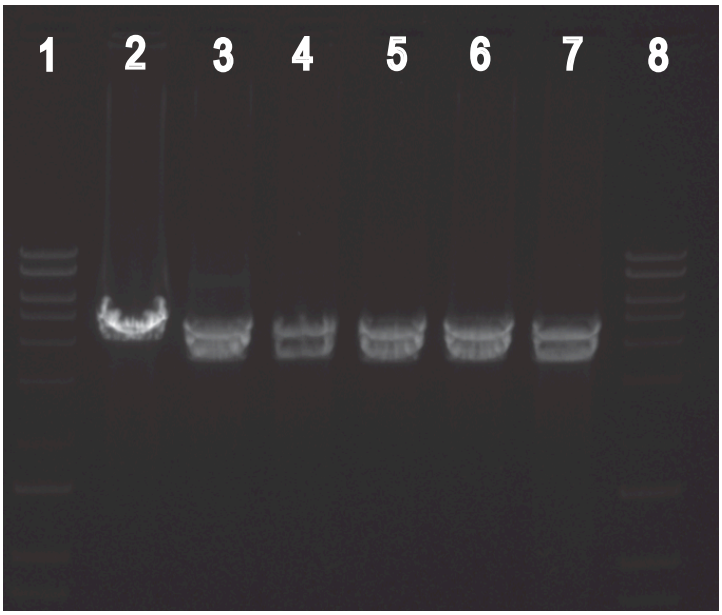


Figure 4.8. HindIII/PstI restriction digest confirmation of plasmid pNW33Nmob (expected band sizes: 4.21 and 3.58 kb). Plasmid pNW33Nmob was recovered from *C. termitidis* exoconjugants and used to "back-transform" *E. coli* DH5α. Lanes 1 and 8: 10 kb ladder; lanes 2-6: plasmid pNW33Nmob recovered from *C. termitidis* exoconjugants; lane 7: pNW33Nmob positive control.

4.4.4 Host Encoded Restriction Methylation Systems

Identified restriction methylation systems are provided in Table 4.3. Type I restriction modification systems were identified in *C. thermocellum* (ATCC 27405, Cthe_1144-1145; DSM 4150, Clo4150_0423-0434) and *C. cellulolyticum* (Ccel_2724-2726). Type II restriction methylation (RM) systems were identified in the genomes of all *C. thermocellum* strains analyzed (ATCC 27405, Cthe_1511-1513; DSM 2360 Clo2360_1474-1475; DSM 1313, Clo1313_2273-2275) but not in *C. termitidis*, *C. cellulolyticum*, C. *phytofermentans* or *T. thermosaccharolyticum*. *C. thermocellum* strain 27405 was the only genome analyzed to encode a Type III RM system (Cthe_0518-0519), and Type IV restriction endonucleases, encoded by a single gene, were identified in the genomes of *T. thermosaccharolyticum* (Thethe_00825), *C. termitidis* (Ct_00002530, Ct_00021470), *C. cellulolyticum* (Ccel_1925) and *C. phytofermentans* (Cphy_1303).

Analysis of *C. termitidis* total proteome revealed peptides specific for Ct_00021470, a Type IV endonuclease in log and stationary phase cultures grown on either α-cellulose or cellobiose. No peptides corresponding to Ct_00002530 were detected in the proteome. Additionally, peptides specific for a cryptic Type IV endonuclease (Ct_00014740) that was not identified during genome analysis were detected.

Table 4.3. Restriction modification systems within the genomes of select thermophilic and mesophilic Firmicutes and reported transformation efficiencies.

Species	RM type	Gene locus	Reported transformation efficiency (cfu/ µg DNA)	Reference:
Clostridium cellulolyticum H10				
	Type I	Ccel_2724-2726	2.0×10^2	[182]
	Type IV	Ccel_1925		
Clostridium phytofermentans ISDg				
	Type IV	Cphy_1303	ND	ND
Clostridium termitidis DSM 5398				
	Type IV	Ct_00002530[A]	$1.8 \times 10^{4\dagger}$	This study
	Type IV	Ct_00021470[B]		
	Type IV	Ct_00014740[B]		
Clostridium thermocellum ATCC 27405				
	Type I	Cthe_1144-1145	$3.3 \times 10^{2\dagger}$	This study
	Type II	Cthe_1511-1513	$5.0 \times 10^{4\dagger}$	[186]
	Type III	Cthe_0518-0519		
	Type IV	Cthe_0850		
Clostridium thermocellum DSM 2360				
	Type II	Clo2360_1474-1475	ND	ND
	Type IV	Clo2360_1185		
Clostridium thermocellum DSM 4150				
	Type I	Clo4150_0432-0434	1.0×10^3	[186]
	Type II	Clo4150_0255-0257		
	Type IV	Clo4150_2682		
Clostridium thermocellum DSM 1313				
	Type II	Clo1313_2273-2275	$5.7 \times 10^{2\dagger}$	[181]
	Type IV	Clo1313_1370	$4.4 \times 10^{3\P}$	
			$2.2 \times 10^{5\dagger}$	[186]
Thermoanaerobacter thermosaccharolyticum M0795				
	Type IV	Thethe_00825	2.5×10^5	[180]

†. Plasmid Dam+/Dcm+ methylated; ¶. Plasmid Dam+/Dcm- methylated; ND. Not determined

A: peptides corresponding to this locus were not detected during proteome analysis.

B: Detected in total proteome analysis

4.5 Discussion

Several published reports of plasmid transfer into other clostridia have involved the circumvention of host restriction/methylation systems or have proposed that poor efficiencies of transformation were a consequence of host RM activity. Typically this has involved the characterization of host RM systems by the systematic digestion of plasmid DNA with crude cell extracts and a determination of endonuclease recognition sequence on the basis of observed electrophoretic banding patterns. Characterization of restriction endonuclease activity in *C. thermocellum* ATCC 27405 previously identified an *Mbo*I-like enzyme (GATC) that was sensitive to Dam methylation [193]. Although not observed by Klapatch [193], despite the presence of appropriate recognition sequences in their DNA templates, isoschizomers of *BcII* (TGATCA) and *EcoRII* (CC(A/T)GG) have additionally been reported in *C. thermocellum* ATCC 27405 [194].

Investigations using cell extracts of *C. thermocellum* prepared by high-pressure disruption, confirmed that plasmid DNA (pNW33Nmob) prepared from dam^+ strains of *E. coli* were protected from nuclease digestion. Plasmid pNW33Nmob harbors 0 *Bcl*I sites, 11 *EcoR*II sites, and 30 *Mbo*I sites. While Klapatch *et al* [193] reported specific endonuclease activity in Dam^- DNA, non-specific patterns of digestion in both our dam^-/dcm^- and dam^-/dcm^+ methylated DNA was observed in these studies (data not shown). Considering the genome of *C. thermocellum* ATCC 27405 encodes Type I, Type III, and Type IV RM systems (Table 4.3), in addition to the aforementioned *Mbo*I-like type II endonuclease, non-discrete degradation is plausible. Nuclease activity may go unobserved for a variety of reasons, and Klapatch cites [193] both cell-disruption

methodology and the ionic strength of buffers used during digestion as possible factors influencing both observed restriction endonuclease and non-specific exonuclease activity.

Following confirmation that *dam* methylation was sufficient to protect foreign DNA from a suite of endogenously encoded nucleases; the transformation of *C. thermocellum* ATCC 27405 via both electro-transformation and conjugation was attempted. Initial electroporation experiments focused on reproducing the methodologies of Tyurin *et al*. [186] using plasmids pIkm1 and pNW33Nmob. The addition of isoniacin (20 μg mL^{-1}) was not found to significantly affect transformation efficiency despite reports to the contrary [186]. Similar to Tyurin *et al*, growth in the presence of antibiotic selection only following the application of a pulse ≥ 10 kV cm^{-1} was observed. Transformation efficiencies were 100-fold lower than the values reported by Tyurin *et al* [186] (3.3×10^2 cfu ug^{-1} DNA compared to 5.0×10^4) however values were comparable with more recent reports of electrotransformation in *C. thermocellum* strain DSM 1313 [181] and in *Clostridium cellulolyticum* H10 [182].

Clostridium termitidis, a mesophilic cellulose degrader known to ferment both the hexose [195] and pentose sugars [Our unpublished data] released during lignocellulose hydrolysis, was selected as a candidate for transformation largely due the absence of identifiable type I, type II or type III RM systems within its genome (Table 4.3). Perhaps as a consequence of reduced host endonuclease activity, it was observed that *C. termitidis* was transformed with plasmid pNW33Nmob approximately 100-fold more frequently than *C. thermocellum* ATCC 27405. Recently, it was reported that dcm-methylation was detrimental to the transformation of *Clostridium thermocellum* strain DSM 1313. It has been hypothesized that some Clostridial restriction systems may target methylated DNA

at recognition sequences that overlap with dcm-methylation sites [181]. Observations from electroporation experiments using plasmid pNW33Nmob (11 *dcm* recognition sequences), methylated *in vivo* in *dcm*$^+$ and dcm$^-$ *E. coli* strains, however, suggest that transformation efficiencies in *C. termitidis* are unaffected by *dcm* methylation. These findings are consistent with the apparent lack of identifiable methylation dependent restriction endonucleases in the *C. termitidis* draft genome. Although many factors complicate direct comparisons of transformation frequency, it is tempting to reason that an *in silico* analysis of host encoded RM systems should be considered, in future studies, prior to attempting genetic manipulations.

Plasmids pIkm1/pIkm1mob utilize a gram-positive replicon from the *Bacillus subtilis* plasmid pIM13 (rolling circle, ssDNA plasmid) and confers resistances to kanamycin, macrolide-lincosamide-streptogramin (MLS) and ampicillin. Although *C. thermocellum* ATCC 27405 could be transformed with plasmid pIkm1, reproducibility was poor and experiments sometimes failed to yield transformants. Vectors pHV33 and pSUP202, neither containing an apparent gram-positive origin of replication predictably failed to transform *C. thermocellum* ATCC 27405. Plasmid pNW33Nmob contains a gram-positive origin of replication derived from the cryptic plasmid replicon pBC1, isolated from *Bacillus coagulans*. Previous studies have shown that plasmids expressing the pBC1 replicon may be introduced and stably maintained in a variety of gram-positive species including *Bacillus subtilis*, *Bacillus amyloliquefaciens*, *Geobacillus* sp., *Staphylococcus aureus*, *Staphylococcus carnosus, Lactobacillus reuteri* [196] and *C. thermocellum* DSM 1313 [181]. These studies have expanded the range in which pBC1 may function to include *Bacillus cereus*, *Geobacillus debilis, Geobacter sulfurreducens,*

Thermoanaerobacter pseudethanolicus 39E, *Clostridium thermocellum* ATCC 27405 and *Clostridium termitidis* DSM 5398. While De Rossi *et al.* determined the pBC1 replicon was maintained at 60 copies cell^{-1} in *B. subtilis*, plasmid pNW33Nmob is estimated to be maintained at very low copy number (≤ 5 copies cell^{-1}) in *C. thermocellum* ATCC 27405.

During intergeneric conjugation, the transfer of DNA into recipient cells is believed to occur in the single-stranded form [197]. Plasmid DNA transferred into recipient cells may therefore avoid degradation by endogenously encoded endonucleases. While reports suggest this single-stranded plasmid transfer may help circumvent host restriction barriers in *Neisseria gonorrhoeae* [198], others cite a reduction in conjugation frequency as a consequence of recipient RMs [199]. Reports of conjugative plasmid transfer into clostridia are numerous and include cellulose degraders such as *C. cellulolyticum* [182], solventogenic species such as *C. acetobutylicum* and C. *beijerinkii* [200], and human pathogens including *C. dificile* [183]. Conjugal plasmid transfer into thermophilic firmicutes, however, has not yet been reported extensively (if at all).

Results presented here demonstrate the successful introduction of plasmid pNW33Nmob into *C. thermocellum* (strains ATCC 27405 and DSM 1313) and *Thermoanaerobacter pseudethanolicus* 39E via conjugation from *E. coli* S17-1. Observed transfer frequencies were comparable to *C. termitidis* and to reports of conjugative plasmid transfer into *C. dificile* (between 10^{-6}-10^{-7} exoconjugants donor^{-1}) following a 6-24 h mating period, with colonies sometimes failing to develop further than pinprick sized.

Although the apparent size of exoconjugant colonies is suggestive of aggressive host restriction activity or an incompatibility of plasmid replication machinery, reliable

outgrowth in liquid selection throughout multiple passages demonstrated that pNW33Nmob was able to replicate stably. Poor growth kinetics on solid media and consequentially a low plating efficiency is likely a contributing factor to both the observed low conjugation/transformation frequencies and poor colony development of *C. thermocellum* ATCC 27405 on selective media.

Similar to reports of conjugation into other mesophilic Clostridia, a 1:10 ratio of donor to recipient was found to be optimal for plasmid transfer into *C. termitidis*. Transfer frequencies were nearly 3.5 times greater in 1:10 mating cultures compared to 1:1 ratios. The *dcm*-methylation state of transferred ssDNA did not significantly affect transfer frequency. Consistent with both electrotransformation findings and genome/proteome analysis, it seems evident that unlike *C. thermocellum* or *C. dificile*, *C. termitidis* does not express a restriction system the is sensitive to *dcm* methylation. These findings may help streamline the process of genetic manipulation of *C. termitidis* in future studies.

4.6 Conclusions

The results reported here demonstrate conjugal plasmid transfer into *C. thermocellum, T. pseudethanolicus* and *C. termitidis*. Although low plating efficiency and poor growth kinetics on solid media likely depressed estimations of transfer frequency and colony formation, reliable growth in liquid media suggest that plasmid pNW33Nmob may be stably maintained in *C. thermocellum* despite the presence of several host encoded RM systems. The mesophile, *C. termitidis*, alternatively, is an attractive candidate for genetic manipulation as evidenced by plasmid transfer frequencies that support a genome/proteome analysis suggestive of a less-extensive host restriction

methylation system. Reports of genetic manipulation in Clostridia are increasing in both number and sophistication as the factors influencing stable DNA uptake continue to be resolved. Given the highly variable results often observed between strains, it seems evident that the tools for genetic manipulation must continue to be developed on a case-by-case basis. The findings presented here illustrate that conjugative transfer of plasmid pNW33Nmob is possible into a range of recipient species. This includes, the first report of conjugation into thermophiles *C. thermocellum* and *T. pseudethanolicus* as well as the mesophile *C. termitidis*.

Summary and General Discussion

5.1 Revisiting Clostridial Metabolism

Cellulolytic species of clostridia have been receiving attention because of their potential to convert lignocellulosic biomass into industrially valuable chemicals. Both mesophilic and thermophilic species have been isolated and are characterized as strictly anaerobic spore-formers with many expressing multienzyme extracellular complexes, known as a cellulosome, that rapidly and efficiently hydrolyzes cellulose [6, 7]. It is important to remain mindful that the physiological conditions under which these species have been characterized does not necessarily reflect the environments in which these bacteria have evolved. Rather, their metabolism has evolved to simultaneously exploit a narrow ecological niche and to function as part of a diverse microbial ecosystem. Guedon *et al* [41] contend that cellulolytic species such as *C. cellulolyticum* are not adapted to utilize carbon sources and other nutrients in excess, as many natural ecosystems rarely contain all the nutrients required in saturating quantities. *C. thermocellum* is incapable of utilizing the products of hemicellulose or lignin hydrolysis and instead is restricted to catabolizing the oligosaccharides released during cellulose decomposition. Co-cultures of *C. thermocellum* with saccharolytic bacteria such as *Clostidium thermosaccharolyticum* [201], *Clostridium thermohydrosulfurican* [202, 203] and *Thermoanaerobacter ethanolicus* have reported increased rates of cellulose hydrolysis and higher ethanol titres than observed in mono-cultures. Indeed, the isolation of *C.*

thermocellum was fraught with challenges resultant from a proclivity to exist in stable, synergistic relationships with other species [203]. Naturally occurring co-cultures of cellulose fermenting mesophiles have also been described [204]. A comprehensive understanding of lingnocellulose fermentation may require, in future studies, the utilization of holistic approaches that involve meta- genomic, proteomic and transcriptomic methodologies. Nevertheless, understanding the role and behaviour of the individual participants in lignocellulosic hydrolysis/fermentation remains a necessary prerequisite.

5.2 The genomics of biofuels production

Recall that in chapter 1 an ideal biofuel-producing microorganism was described as being capable of: (i) producing high yields of the desired product, (ii) simultaneous utilization of sugars released from lignocellulose hydrolysis, (iii) growth at elevated temperatures, and (iv) low product inhibition. In the second chapter of this thesis, a comparative genomics approach was used to examine the link between reported biofuel chemical yields with genome content and reaction thermodynamics within select Firmicutes, Euryarchaeota, and Thermotogae. While the cellulolytic and, to a lesser extent, H_2 and ethanol producing capabilities of cellulolytic bacteria have been reviewed [64, 65, 100], no studies had yet attempted to describe end-product distribution patterns in relation to genome content and reaction thermodynamics. The hypothesis, that reported differences in product yields would be reflected in the complement of genes encoding enzymes that participate in pyruvate metabolism and reductive end-product synthesis, was confirmed. A number of potential biomarkers that could be diagnostic of the H_2 and/or ethanol producing capabilities of sequenced and novel isolates were

identified. Perhaps the most surprising finding from these analyses was the observation that within the genomes of species that produce mixed fermentation products, numerous, seemingly redundant, mechanisms capable of balancing cellular redox were identified. These included multiple catabolic routes leading to the production of reduced products including ethanol, lactate, and H_2 in addition to acetate, formate and CO_2. In *C. thermocellum*, for example, three putative class IV Fe-containing ADHs in addition to an NADH-dependent bifunctional acetaldehdye/alcohol dehydrogenase (*adhE*) and an aldehyde dehydrogenase were identified (Table 2.5). Alcohol metabolism in *C. thermocellum* (discussed in section 3.5) appears to be very complex and may be regulated, at least in part, transcriptionally. The numerous hydrogenases, a putative NFO and an apparent ortholog to the recently described *NfnAB*, an electron bifurcating NADH-dependent Fd_{red}: $NADP^+$ oxidoreductase, of *C. kluyveri* [178, 179] futher illustrates how *C. thermocellum* appears poised to respond to changing environmental conditions in order to ensure continued metabolism. The 11.5 kbp region that encompasses Cthe_0422 to Cthe_0431 and includes genes that encode a Rex-type redox sensing protein (Cthe_0422), AdhE (Cthe_0423), and both a PAS-fold containing sensory hydrogenase and a heterotrimeric bifurcating hydrogenase (Cthe_0426, Cthe_0428-0430) was identified as unique to *C. thermocellum* and it's potential role in regulating cellular redox was considered (chapter 3) and merits further investigation. Future studies, perhaps involving the systematic knockout or heterologous expression of the identified regulatory elements in this gene cassette may offer further insight into how *C. thermocellum* responds to changing environmental conditions.

5.3 Closing the "production gap"

It is both tempting and fallacious to imbue metabolic flux with purposefulness, such as an industrial application (ie cellulose degradation, biofuels production etc.), that is not solely directed towards persistent survival in a natural environment. It is important, therefore, to remain mindful to develop hypotheses concerning metabolism within this context. In order to achieve maximal biofuels production yields, intelligent species/strain selection and both process and genetic engineering approaches will likely all be required. The "ideal biofuel-producing microorganism" that was described in chapter 1 remains elusive. In chapter 3 of this thesis how carbon and electron flux was influenced by environmental changes during fermentation of cellobiose by *C. thermocellum* was investigated. The hypothesis, that gas sparging, alternatively with N_2 or H_2, would significantly affect fermentation product yields was confirmed. Observed ethanol-to-acetate ratios for N_2 sparged cultures that were consistent with reports of stirred *C. thermocellum* cultures grown on cellobiose under 2.5 atm N_2 [125]. The shift towards ethanol production was observed in response to H_2 gas sparging was consistent with the findings of Lamed *et al.* [125], Li *et al* [134] and Rydzak *et al* [128]. Similar findings in cultures of *Caldanaerobacter subterraneus* subsp. *tengcongensis* have been reported in response to elevated H_2 partial pressure [172]. The hypothesis, that the expression of key genes involved in biofuels production would exhibit regulation in response to gas sparging, appears, at least, to be partially confirmed. Simlar to reports of cellobiose fermentation by *C. subterraneus* subsp. *Tengcongensis*, it was observed that alcohol dehydrogenase activity was greater in cultures of *C. thermocellum* sparged with H_2 gas

than within cultures sparged with N_2 gas. Additionally, the transcription of *adhE* may be regulated in response to cellular redox and is discussed in chapter 3.

Variations in substrate availability and/or dilution rate [102, 109-111, 114-117], substrate composition [110, 118-123], media composition [111], pH [124], gas partial pressure [90, 98, 125, 126], growth phase [113], and end-product accumulation [103, 118, 125, 127, 128] are all known to dramatically influence metabolic flux. Consequently, the highly branched catabolic pathways present in most cellulose degrading Clostridia continue to stymie the realisation of homogeneous product yields.

Genetic modification of these species offers, perhaps, the greatest potential for meeting the criteria outlined for an ideal biofuel-producing microorganism. In section 1.7.1 the heterologous expression of *pdc* and *adh* from *Zymomonas mobilis* within *C. cellulolyticum* was discussed [54]. Recent metabolic engineering strategies aiming at improving ethanol yields (via deletion of *pta* and, hence targeting acetate production) in *C. thermocellum* DSM 1313 are illustrative that metabolic system perturbations continue to confound prediction [159]. The major obstacle to the further genetic modification of these species appears to be the recalcitrance of Clostridia to transformation. In chapter 4, the lack of highly efficient transformation strategies was discussed and the development of both electrotransformation and conjugative plasmid transfer into recipient thermophilic species *C. thermocellum*, *T. pseudethanolicus* 39E and the mesophile *C. termitidis* CT1112 was reported. The findings presented in Chapter 4 have expanded the known range in which the pBC1 replicon may function to include *Bacillus cereus*, *Geobacillus debilis, Geobacter sulfurreducens, Thermoanaerobacter pseudethanolicus* 39E, *Clostridium thermocellum* ATCC 27405 and *Clostridium termitidis*. In spite of these developments, transformation efficiencies, particularly for *C. thermocellum* strain 27405,

are poor. The role of host restriction methylation systems is discussed in chapter 4 and represents a likely obstacle to achieveing high transformation efficiencies/frequencices. In the future, candidate species for genetic modification should be evaluated, in part, based on the compliment of nucleases encoded in their genomes.

5.4 Conclusions and future perspectives

The development of "third generation" biofuels depends heavily on a detailed understanding of the metabolism of cellulolytic bacteria. Both recombinant strategies that aim to introduce/improve biosynthetic pathways into naturally cellulolytic species [54, 142, 159, 187] and those strategies that endeavour to confer cellulolytic ability to microorganisms that display favourable product formation properties have made considerable progress [205]. Nevertheless, recombinant strains far from achieve the titers and productivities that are required for the consolidated bioprocessing of lignocellulose to be economically viable. The emergence of genomics and "next-generation" DNA sequencing technologies now permits the rapid *in silico* characterization of novel isolates and can inform rational metabolic engineering strategies. It is likely, however, that the successful development of "third generation" biofuels will require the confluence of intelligent species/strain selection and both process and genetic engineering approaches.

Literature Cited

1. EIA: **Annual energy outlook 1999, with projections to 2020**. In: *DOE/EIA-0383.* vol. 99. Washington, DC: Energy Information Administration, USA Department of Energy.; 1998.

2. Monserrate E, Leschine, S.B., Canale-Parola, E.: ***Clostridium hungatei* sp. Nov., a mesophillic, N2-fixing cellulolytic bacterium isolated from soil**. *International Journal of Systematics, Evolution and Microbiology* 2001, **51**:123-132.

3. Levin DB, Pitt, L., Love, M.: **Biohydrogen production; prospects and limitations to practical application**. *International Jounral of Hydrogen Energy* 2004, **29**(2):173-185.

4. Levin DB, Sparling, R., Islam, R., Cicek, N.: **Hydrogen production by *Clostridium thermocellum* 27405 from cellulosic biomass substrates**. *International Journal of Hydrogen Energy* 2006, **31**(11):1496-1503.

5. Lynd LR, van Zyl, W.H., McBride, L.E., Laser, M.: **Consolidated bioprocessing of cellulosic biomass: an update**. *Current Opinion in Biotechnology* 2005, **16**(5):577-583.

6. Lynd LR, Weimer, P.J., van Zyl, W.H., Pretorius, I.S.: **Microbial cellulose utilization: Fundamentals and biotechnology**. *Microbiology and Molecular Biology Reviews* 2002, **66**:506-577.

7. Demain AL, Newcomb, M., Wu, J.H.D.: **Cellulase, clostridia, and ethanol**. *Microbiology and Molecular Biology Reviews* 2005, **69**:124-154.

8. OFD: **Review of the research strategy for biomass-derived transportation fuels.** In: *Committee to review the R and D strategy for biomass- derived ethanol and biodeisel transportation fuels.* Washington, DC: National Academy 1999: 1-48.

9. Brown MA, Levine, M.D., Romm, JPRAH, Koomey, J.H.: **Engineering-economic studies of energy technologies to reduce greenhouse gas emissions: opportunites and challenges**. *Annual review of energy environment* 1998, **23**:31-39.

10. Carmack R: **Redox enzymes: Splitting molecular hydrogen.** *Nature* 1995, **373**:556-557.

11. Hallenbeck PC, Benemann, J.R.: **Biological hydrogen production; fundamentals and limiting processes**. *International Journal of Hydrogen Energy* 2002, **27**:1123-1505.

12. Hallenbeck PC: **Fundamentals of the fermentative production of hydrogen**. *Water Science and Technology* 2005, **52**:21-29.

13. Kim S, Dale, B.: **Allocation procedure in ethanol production system from corn grain: I. system expansion**. *International Journal of Life Cycle Assessment* 2002, **7**(4):237-243.

14. Shapouri H, Duffield J.A., Wang, M.: **The energy balance of corn ethanol: an update. Agricultural Economic Report**. In., vol. 813. Washington, DC: US Department of Agriculture; 2002.

15. Shapouri H, Duffield, J.A., Graboski, M.S.: **Estimating the net energy balance of corn ethanol**. In: *Agricultural Economic Report.* vol. 721. Washington, DC: US Department of Agriculture; 1995.

16. Wheals AE, Basso, Luiz C., Denise, M., Alves, G., Amorim, Henrique.: **Fuel ethanol after 25 years**. *Trends in Biotechnology* 1999, **17**:482-487.

17. Zaldivar J, Nielson, J., Olsson, l.: **Fuel ethanol production from lignocellulose: a challenge for metabolic engineering and process integration**. *Applied Microbiology and Biotechnology* 2001, **56**:17-34.

18. Administration USEI: **Biofuels Issues and Trends**. In.; 2012.

19. U.S. Department of Energy. Center for Transportation Research. Argonne National Laboratory. **GREET 1.5: transportation fuel-cycle model, Volume 1: methodology, development, use and results.** *The University of Chicago* 1999.

20. Mizuno O, Dinsdale R, Hawkes FR, Hawkes DL, Noike T: **Enhancement of hydrogen production from glucose by nitrogen gas sparging**. *Bioresource Technology* 2000, **73**(1):59-65.

21. Chang J, Chen, W., Shih, S., Yu, S., Lay, J., Wen, F., Huang, C.: **Molecular detection of the clostridia in an ananerobic biohydrogen fermentation system**

by hydrogenase mRNA-targeted reverse transcription-PCR. *Applied Microbiology and Biotechnology* 2005.

22. Collet C, Adler, N., Schwitzguebel, J., Peringer, P: **Hydrogen production by *Clostridium thermolacticum* during continuous fermentation of lactose**. *International Journal of Hydrogen Energy* 2004, **29**:1479-1485.

23. Council IG: **Five year global supply and deman projections to 2017/18**. In.; 2012.

24. Desvaux M: ***Clostridium cellulolyticum*: model organism of mesophillic cellulolytic clostridia**. *FEMS Microbiology Reviews* 2005, **29**:741-764.

25. Boisset C, Chanzy, H., Henrissat, B., Lamed, R., Shoham, Y., Bayer, E.A: **Digestion of crystalline cellulose substrates by the *Clostridium thermocellum* cellulosome: Structural and morphological aspects**. *Biochemical Journal* 1999, **340**:829-835.

26. Bergquist PL, Gibbs, M.D., Morris, D.D., Te'O, V.S., Saul, D.J., Morgan, H.W.: **Molecular diversity of thermophillic cellulolytic and hemicellulolytic bacteria**. *FEMS Microbiology Ecology* 1999, **28**:99-110.

27. Rampersad K, Goldstone, L.A., Tivchev, G.N.: **Study of methods for the cultivation of anaerobic cellulose-degrading bacteria**. *Water SA* 1998, **24**:343-346.

28. O'sullivan CA, Burrell, P.C., Clarke, W.P., Blackall, L.L.: **Structure of a cellulose degrading bacterial community during anaerobic digestion**. *Biotechnology and Bioengineering* 2005, **92**(7):871-878.

29. Weimer PJ, Zeikus, G.: **Fermentation of cellulose and cellobiose by *Clostridium thermocellum* in the absence and presence of *Methanobacterium thermoautotrophicum***. *Applied and Environmental Microbiology* 1977, **33**:289-297.

30. Lamed R, Setter, E., Bayer, E.A.: **Characterization of a cellulose-binding, cellulose-containing complex in *Clostridium thermocellum***. *Journal of Bacteriology* 1983, **156**:828-836.

31. Felix RC, Ljungdahl, L.G.: **The cellulosome: the exocellular organelle of clostridium**. *Annual Reviews of Microbiology* 1993, **47**:791-819.

32. Walter L, Staudenbauer, Wolfgang, H., Schwarz.: **A model of the *Clostridium thermocellum* cellulosome**. *Fachgebiet Mikrobielle Biotechnologie* 2007:http://www.wzw.tum.de/mbiotec/celostruct.htm.

33. Lamed R, Bayer, E. A.: **The cellulosome of *Clostridium thermocellum***. *Advances in Applied Microbiology* 1988, **33**:1-46.

34. Tomme P, van Tilbeurgh, H., Peterson, G., van Damme, J., Vandekerckhove, J., Knowles, J., Teeri, T., Claeysens, M.: **Studies of the cellulolytic system of *Trichoderma reesei* QM 9414: Analysis of domain function in two**

cellobiohydrolases by limited proteolysis. *European Journal of Biochemistry* 1988, **170**:575-581.

35. Johnson EA, Bouchot, F., Demain, A.L.: **Regulation of cellulase formation in *Clostridium thermocellum***. *Journal of General Microbiology* 1985, **131**:2303-2308.

36. Zhang Y, Lynd, L.R.: **Quantification of cell and cellulase mass concentrations during anaerobic cellulose fermentation: Development of an enzyme-linked immunoabsorbent assay-based method with applications to *Clostridium thermocellum* batch cultures**. *Analytical Chemestry* 2003, **75**:219-227.

37. Zhang YP, Lynd, L.R.: **Regulation of cellulase synthesis in batch and continuous cultures of *Clostridium thermocellum***. *Journal of Bacteriology* 2005, **187**:99-106.

38. Gelhaye E, Gehin, A., Petitdemange, H.: **Colonization of crystalline cellulose by *Clostridium cellulolyticum* ATCC 35319**. *Applied and Environmental Microbiology* 1993, **59**:3154-3156.

39. Strobel HJ, Caldwell, F.C., Dawson, K.A.: **Carbohydrate transport by the anaerobic thermophile *Clostridium thermocellum* LQR1**. *Applied and Environmental Microbiology* 1995, **61**:4012-4015.

40. Desvaux M, Guedon, E., Petitdemange, H.: **Metabolic flux in cellulose-fed continuous cultures of *Clostridium cellulolyiticum* in response to acidic environment**. *Microbiology* 2001, **147**:1461-1471.

41. Guedon E, Payot, S., Desvaux, M., Petitdemange, H.: **Carbon and electron flow in *Clostridium cellulolyticum* grown in chemostat culture on synthetic medium**. *Journal of Bacteriology* 1999, **181**:3262-3269.

42. Payot S, Guedon, E., Caillez, C., Gelhaye, E., Petitdemange, H.: **Metabolism of cellobiose by *Clostridium cellulolyticum* growing in continuous culture: Evidence for decreased NADH reoxidation as a factor limiting growth**. *Microbiology* 1998, **144**:375-384.

43. Tomme P, Warren, R.A.J., Gilkes, N.R.: **Cellulose hydrolysis by bacteria and fungi**. *Advances in Microbial Physiology* 1995, **37**:1-81.

44. Lynd LR, Grethlein, H.G., Wolkin, R.H.: **Fermentation of cellulose substrates in batch and continuous culture by *Clostridium thermocellum***. *Applied and Environmental Microbiology* 1989, **55**:3131-3139.

45. Lynd LR, Grethlein, H.G., Wolkin, R.H.: **Hydrolysis of dilute acid pretreated hardwood and purified microcyrstalline cellulose by cell-free broth from *Clostridium thermocellum***. *Biotechnology and Bioengineering* 1987, **29**:92-100.

46. Islam R, Cicek, N., Sparling, R., Levin, D.B.: **Effect of substrate loading on hydrogen production during anaerobic fermentation by *Clostridium thermocellum* 27405**. *Applied Microbiology and Biotechnology* 2006, **72**(3):576-583.

47. Sparling R, Islam, R., Cicek, N., Carere, C., Chow, H., Levin, D.B.: **Formate synthesis by *Clostridium thermocellum* during anaerobic fermentation**. *Canadian Journal of Microbiology* 2006, **52**:681-688.

48. Rydzak T, Levin, D.B., Cicek, N., Sparling, R. : **Optimization of hydrogen production by *Clostridium thermocellum* 27405 via manipulation of growth condition**. In: *Proceedings of the 2nd International Hydrogen Energy Congress & Exhibition: July 13-15 2007; Istanbul, Turkey*; 2007.

49. Rani KS, Swamy, M.V., Seenayya, G.: **Increased ethanol production by metabolic modulation of cellulose fermentation in *Clostridium thermocellum***. *Biotechnology Letters* 1997, **19**:819-823.

50. Bothun GD, Knitson, B.L., Berberich, J.A., Strobel, H.J., Nokes, S.E.: **Metabolic selectivity and growth of *Clostridium thermocellum* in continuous culture under elevated hydrostatic pressure**. *Applied Microbiology and Biotechnology* 2004, **65**:149-157.

51. Bailey E, James: **Toward a Science of Metabolic Engineering**. *Science, New Series* 1991, **252**(5013):1668-1675.

52. Chung KT: **Inhibitory effects of H_2 on growth of *Clostridium cellobioparum***. *Applied and Environmental Microbiology* 1976, **31**:342-348.

53. Dabrock B, Bahl, H., Gottschalk.: **Parameters affecting solvent produciton in *Clostridium pasteurianum***. *Applied and Environmental Microbiology* 1992, **58**(4):1233-1239.

54. Guedon E, Desvaux, M., Petitdemange, H.: **Improvement of cellulolytic properties of *Clostridium cellulolyticum* by metabolic enginering**. *Applied Environmental Engineering* 2002, **68**:53-58.

55. Desai RP, Papoustakis, E.T.: **Antisense RNA strategies for metabolic engineering of Clostridium acetobutylicum**. *Applied and Environmental Microbiology* 1999, **65**(3):936-945.

56. Coleman JP, Green, J., Inouye, M.: **The use of RNAs complementary to specific mRNAs to regulate the expression of individual bacterial genes**. *Cell* 1984, **37**(429-436).

57. Van den Berg WA, Walter, M.A., van Dongen, M., Veeger, C.: **Reduction of the amount of periplasmic hydrogenase in *Desulfovibrio vulgaris* with antisense RNA: Direct evidence for an important role of this hydrogenase in lactate metabolism.** *Journal of Bacteriology* 1991, **173**(12):3688-3694.

58. Carere CR, Sparling R, Cicek N, Levin DB: **Third Generation Biofuels via Direct Cellulose Fermentation**. *International Journal of Molecular Science* 2008, **9**(7):1342-1360.

59. Carere C, Kalia V, Sparling R, Cicek N, Levin D: **Pyruvate catabolism and hydrogen synthesis pathway genes of *Clostridium thermocellum* ATCC 27405**. *Indian Journal of Microbiology* 2008, **48**(2):252-266.

60. Levin DB, Pitt L, Love M: **Biohydrogen production: prospects and limitations to practical application**. *International Journal of Hydrogen Energy* 2004, **29**(2):173-185.

61. Lynd LR, van Zyl WH, McBride JE, Laser M: **Consolidated bioprocessing of cellulosic biomass: an update**. *Current Opinion in Biotechnology* 2005, **16**(5):577-583.

62. Islam R, Cicek N, Sparling R, Levin D: **Influence of initial cellulose concentration on the carbon flow distribution during batch fermentation by *Clostridium thermocellum* ATCC 27405**. *Applied Microbiology and Biotechnology* 2009, **82**(1):141-148.

63. Yang SJ, Kataeva I, Hamilton-Brehm SD, Engle NL, Tschaplinski TJ, Doeppke C, Davis M, Westpheling J, Adams MWW: **Efficient Degradation of Lignocellulosic Plant Biomass, without Pretreatment, by the Thermophilic Anaerobe "*Anaerocellum thermophilum*" DSM 6725**. *Applied and Environmental Microbiology* 2009, **75**(14):4762-4769.

64. Bruggemann H, Gottschalk G: **Comparative genomics of clostridia: link between the ecological niche and cell surface properties**. *Annals of the New York Academy of Sciences* 2008, **1125**:73-81.

65. Desvaux M: **Unravelling carbon metabolism in anaerobic cellulolytic bacteria**. *Biotechnology Progress* 2006, **22**(5):1229-1238.

66. Rydzak T, Levin DB, Cicek N, Sparling R: **Growth phase-dependant enzyme profile of pyruvate catabolism and end-product formation in *Clostridium thermocellum* ATCC 27405**. *Journal of Biotechnology* 2009, **140**(3-4):169-175.

67. Markowitz VM, Korzeniewski F, Palaniappan K, Szeto E, Werner G, Padki A, Zhao X, Dubchak I, Hugenholtz P, Anderson I *et al*: **The integrated microbial genomes (IMG) system**. *Nucleic Acids Research* 2006, **34**(Database issue):D344-348.

68. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN *et al*: **The COG database: an updated version includes eukaryotes**. *BMC Bioinformatics* 2003, **4**:41.

69. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T *et al*: **KEGG for linking genomes to life and the environment**. *Nucleic Acids Research* 2008, **36**(Database issue):D480-484.

70. Haft DH, Loftus BJ, Richardson DL, Yang F, Eisen JA, Paulsen IT, White O: **TIGRFAMs: a protein family resource for the functional identification of proteins**. *Nucleic Acids Research* 2001, **29**(1):41-43.

71. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool**. *Journal of Molecular Biology* 1990, **215**(3):403-410.

72. Calusinska M, Happe T, Joris B, Wilmotte A: **The surprising diversity of clostridial hydrogenases: a comparative genomic perspective**. *Microbiology* 2010, **156**(Pt 6):1575-1588.

73. Saitou N, Nei M: **The neighbor-joining method: a new method for reconstructing phylogenetic trees**. *Molecular Biology and Evolution* 1987, **4**(4):406-425.

74. Felsenstein J: **Confidence Limits on Phylogenies: An Approach Using the Bootstrap**. *Evolution* 1985, **39**(4):783-791.

75. Zuckerkandl E, Pauling L: **Evolutionary Divergence and Convergence in Proteins**. In: *Evolving Genes and Proteins.* Edited by Bryson B, Vogel H. New York: New York; 1965: 97–166.

76. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods**. *Molecular Biology and Evolution*, **28**(10):2731-2739.

77. Thauer RK, Jungermann K, Decker K: **Energy conservation in chemotrophic anaerobic bacteria**. *Bacteriological Reviews* 1977, **41**(1):100-180.

78. Chemical Rubber Company.: **CRC handbook of chemistry and physics**. In. Cleveland, OH: CRC Press; 1977.

79. Hall TA: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT**. In: *Nucleic acids symposium series: 1999*; 1999: 95-98.

80. de Vrije T, Mars AE, Budde MAW, Lai MH, Dijkema C, de Waard P, Claassen PAM: **Glycolytic pathway and hydrogen yield studies of the extreme thermophile *Caldicellulosiruptor saccharolyticus***. *Applied Microbiology and Biotechnology* 2007, **74**(6):1358-1367.

81. Bredholt S, Sonne-Hansen J, Nielsen P, Mathrani IM, Ahring BK: ***Caldicellulosiruptor kristjanssonii* sp nov., a cellulolytic extremely thermophilic, anaerobic bacterium**. *International Journal of Systematic Bacteriology* 1999, **49**:991-996.

82. Kadar Z, De Vrijek T, van Noorden GE, Budde MAW, Szengyel Z, Reczey K, Claassen PAM: **Yields from glucose, xylose, and paper sludge hydrolysate during hydrogen production by the extreme thermophile *Caldicellulosiruptor saccharolyticus***. *Applied Biochemistry and Biotechnology* 2004, **113-16**:497-508.

83. Kataeva IA, Yang SJ, Dam P, Poole FL, Yin Y, Zhou FF, Chou WC, Xu Y, Goodwin L, Sims DR *et al*: **Genome Sequence of the Anaerobic, Thermophilic, and Cellulolytic Bacterium "*Anaerocellum thermophilum*" DSM 6725**. *Journal of Bacteriology* 2009, **191**(11):3760-3761.

84. Svetlichnyi VA, Svetlichnaya TP, Chernykh NA, Zavarzin GA: ***Anaerocellum-thermophilum* Gen-Nov Sp-Nov - an extremely thermophilic cellulolytic**

Eubacterium isolated from hot-Springs in the valley of Geysers. *Microbiology* 1990, **59**(5):598-604.

85. Chou CJ, Shockley KR, Conners SB, Lewis DL, Comfort DA, Adams MW, Kelly RM: **Impact of substrate glycoside linkage and elemental sulfur on bioenergetics of and hydrogen production by the hyperthermophilic archaeon *Pyrococcus furiosus***. *Applied and Environmental Microbiology* 2007, **73**(21):6842-6853.

86. Kengen SW, de Bok FA, van Loo ND, Dijkema C, Stams AJ, de Vos WM: **Evidence for the operation of a novel Embden-Meyerhof pathway that involves ADP-dependent kinases during sugar fermentation by *Pyrococcus furiosus***. *The Journal of Biological Chemistry* 1994, **269**(26):17537-17541.

87. Schicho RN, Ma K, Adams MW, Kelly RM: **Bioenergetics of sulfur reduction in the hyperthermophilic archaeon *Pyrococcus furiosus***. *Journal of Bacteriology* 1993, **175**(6):1823-1830.

88. Kanai T, Imanaka H, Nakajima A, Uwamori K, Omori Y, Fukui T, Atomi H, Imanaka T: **Continuous hydrogen production by the hyperthermophilic archaeon, *Thermococcus kodakaraensis* KOD1**. *Journal of Biotechnology* 2005, **116**(3):271-282.

89. Munro SA, Zinder SH, Walker LP: **The fermentation stoichiometry of *Thermotoga neapolitana* and influence of temperature, oxygen, and pH on hydrogen production**. *Biotechnology Progress* 2009, **25**(4):1035-1042.

90. Nguyen TA, Han SJ, Kim JP, Kim MS, Sim SJ: **Hydrogen production of the hyperthermophilic eubacterium, *Thermotoga neapolitana* under N2 sparging condition**. *Bioresource Technology* 2010, **101 Suppl 1**:S38-41.

91. Eriksen NT, Nielsen TM, Iversen N: **Hydrogen production in anaerobic and microaerobic *Thermotoga neapolitana***. *Biotechnology Letters* 2008, **30**(1):103-109.

92. Takahata Y, Nishijima M, Hoaki T, Maruyama T: ***Thermotoga petrophila* sp. nov. and *Thermotoga naphthophila* sp. nov., two hyperthermophilic bacteria from the Kubiki oil reservoir in Niigata, Japan**. *International Journal of Systematic Evolutionary Microbiology* 2001, **51**(Pt 5):1901-1909.

93. Nguyen TN, Borges KM, Romano AH, Noll KM: **Differential gene expression in *Thermotoga neapolitana* in response to growth substrate**. *FEMS Microbiology Letters* 2001, **195**(1):79-83.

94. Schröder C, Selig M, Schönheit P: **Glucose fermentation to acetate, CO_2 and H_2 in the anaerobic hyperthermophilic eubacterium *Thermotoga maritima*: involvement of the Embden-Meyerhof pathway**. *Archives of Microbiology* 1994, **161**(6):460-470.

95. Lakhal R, Auria R, Davidson S, Ollivier B, Dolla A, Hamdi M, Combet-Blanc Y: **Effect of oxygen and redox potential on glucose fermentation in *Thermotoga maritima* under controlled physicochemical conditions**. *International Journal of Microbiology* 2010, **2010**:896510.

96. Nguyen TAD, Pyo Kim J, Sun Kim M, Kwan Oh Y, Sim SJ: **Optimization of hydrogen production by hyperthermophilic eubacteria, *Thermotoga maritima* and *Thermotoga neapolitana* in batch fermentation**. *International Journal of Hydrogen Energy* 2008, **33**(5):1483-1488.

97. 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**. *International Journal of Systematic Evolutionary Microbiology* 2001, **51**(Pt 4):1335-1341.

98. Soboh B, Linder D, Hedderich R: **A multisubunit membrane-bound [NiFe] hydrogenase and an NADH-dependent Fe-only hydrogenase in the fermenting bacterium *Thermoanaerobacter tengcongensis***. *Microbiology* 2004, **150**(7):2451-2463.

99. Xing D, Ren N, Li Q, Lin M, Wang A, Zhao L: ***Ethanoligenens harbinense* gen. nov., sp. nov., isolated from molasses wastewater**. *International Journal of Systematic Evolutionary Microbiology* 2006, **56**(Pt 4):755-760.

100. Ren Z, Ward TE, Logan BE, Regan JM: **Characterization of the cellulolytic and hydrogen-producing activities of six mesophilic Clostridium species**. *Journal of Applied Microbiology* 2007, **103**(6):2258-2266.

101. Warnick TA, Methe BA, Leschine SB: ***Clostridium phytofermentans* sp. nov., a cellulolytic mesophile from forest soil**. *International Journal of Systematic Evolutionary Microbiology* 2002, **52**(Pt 4):1155-1160.

102. Islam R, Cicek N, Sparling R, Levin D: **Effect of substrate loading on hydrogen production during anaerobic fermentation by *Clostridium thermocellum* 27405**. *Applied Microbiology and Biotechnology* 2006, **72**(3):576-583.

103. Freier D, Mothershed CP, Wiegel J: **Characterization of *Clostridium thermocellum* JW20**. *Applied and Environmental Microbiology* 1988, **54**(1):204-211.

104. Lacis LS, Lawford HG: **Ethanol production from xylose by *Thermoanaerobacter ethanolicus* in batch and continuous culture**. *Archives of Microbiology* 1988, **150**(1):48-55.

105. Lacis LS, Lawford HG: ***Thermoanaerobacter ethanolicus* growth and product yield from elevated levels of xylose or glucose in continuous cultures**. *Applied and Environmental Microbiology* 1991, **57**(2):579-585.

106. Wiegel J, Ljungdahl LG: ***Thermoanaerobacter ethanolicus* gen. nov., spec. nov., a new, extreme thermophilic, anaerobic bacterium**. *Archives of Microbiology* 1981, **128**(4):343-348.

107. Ouhib-Jacobs O, Lindley ND, Schmitt P, Clavel T: **Fructose and glucose mediates enterotoxin production and anaerobic metabolism of Bacillus cereus ATCC14579(T)**. *Journal of Applied Microbiology* 2009, **107**(3):821-829.

108. Tang YJ, Sapra R, Joyner D, Hazen TC, Myers S, Reichmuth D, Blanch H, Keasling JD: **Analysis of metabolic pathways and fluxes in a newly discovered**

thermophilic and ethanol-tolerant *Geobacillus* strain. *Biotechnology and Bioengineering* 2009, **102**(5):1377-1386.

109. Stevenson DM, Weimer PJ: **Expression of 17 genes in *Clostridium thermocellum* ATCC 27405 during fermentation of cellulose or cellobiose in continuous culture**. *Applied and Environmental Microbiology* 2005, **71**(8):4672-4678.

110. Strobel HJ: **Growth of the thermophilic bacterium *Clostridium thermocellum* in continuous culture**. *Current Microbiology* 1995, **31**(4):210-214.

111. Guedon E, Payot S, Desvaux M, Petitdemange H: **Carbon and electron flow in *Clostridium cellulolyticum* grown in chemostat culture on synthetic medium**. *Journal of Bacteriology* 1999, **181**(10):3262-3269.

112. Özkan M, Ylmaz, E., Lynd, L.R., Özcengiz, G.: **Cloning and expression of the *Clostridium thermocellum* L-lactate dehydrogenase in *Escherichia coli* and enzyme characterization**. *Canadian Journal of Microbiology* 2004, **50**:845-851.

113. Willquist K, Zeidan AA, van Niel EW: **Physiological characteristics of the extreme thermophile *Caldicellulosiruptor saccharolyticus*: an efficient hydrogen cell factory**. *Microbial Cell Factories* 2010, **9**:89.

114. Desvaux M, Guedon E, Petitdemange H: **Metabolic flux in cellulose batch and cellulose-fed continuous cultures of *Clostridium cellulolyticum* in response to acidic environment**. *Microbiology* 2001, **147**(Pt 6):1461-1471.

115. Desvaux M, Petitdemange H: **Flux analysis of the metabolism of *Clostridium cellulolyticum* grown in cellulose-fed continuous culture on a chemically defined medium under ammonium-limited conditions**. *Applied and Environmental Microbiology* 2001, **67**(9):3846-3851.

116. Desvaux Ml, Guedon E, Petitdemange H: **Kinetics and metabolism of cellulose degradation at high substrate concentrations in steady-state continuous cultures of *Clostridium cellulolyticum* on a chemically defined medium**. *Applied and Environmental Microbiology* 2001, **67**(9):3837-3845.

117. Guedon E, Payot S, Desvaux M, Petitdemange H: **Relationships between cellobiose catabolism, enzyme levels, and metabolic intermediates in *Clostridium cellulolyticum* grown in a synthetic medium**. *Biotechnology and Bioengineering* 2000, **67**(3):327-335.

118. Ben-Bassat A, Lamed R, Zeikus JG: **Ethanol production by thermophilic bacteria: metabolic control of end product formation in *Thermoanaerobium brockii***. *Journal of Bacteriology* 1981, **146**(1):192-199.

119. Levin DB, Islam R, Cicek N, Sparling R: **Hydrogen production by *Clostridium thermocellum* 27405 from cellulosic biomass substrates**. *International Journal of Hydrogen Energy* 2006, **31**(11):1496-1503.

120. Strobel HJ, Caldwell FC, Dawson KA: **Carbohydrate transport by the anaerobic thermophile *Clostridium thermocellum* LQRI**. *Applied and Environmental Microbiology* 1995, **61**(11):4012-4015.

121. Zhang YH, Lynd LR: **Regulation of cellulase synthesis in batch and continuous cultures of *Clostridium thermocellum***. *Journal of Bacteriology* 2005, **187**(1):99-106.

122. Girbal L, Soucaille P: **Regulation of Clostridium acetobutylicum metabolism as revealed by mixed-substrate steady-state continuous cultures: role of NADH/NAD ratio and ATP pool**. *Journal of Bacteriology* 1994, **176**(21):6433-6438.

123. Vasconcelos I, Girbal L, Soucaille P: **Regulation of carbon and electron flow in Clostridium acetobutylicum grown in chemostat culture at neutral pH on mixtures of glucose and glycerol**. *Journal of Bacteriology* 1994, **176**(5):1443-1450.

124. Desvaux Ml, Guedon E, Petitdemange H: **Metabolic flux in cellulose batch and cellulose-fed continuous cultures of Clostridium cellulolyticum in response to acidic environment**. *Microbiology* 2001, **147**(6):1461-1471.

125. Lamed RJ, Lobos JH, Su TM: **Effects of Stirring and Hydrogen on Fermentation Products of Clostridium thermocellum**. *Applied and Environmental Microbiology* 1988, **54**(5):1216-1221.

126. Bothun GD, Knutson BL, Berberich JA, Strobel HJ, Nokes SE: **Metabolic selectivity and growth of Clostridium thermocellum in continuous culture under elevated hydrostatic pressure**. *Applied Microbiology and Biotechnology* 2004, **65**(2):149-157.

127. Lamed R, Zeikus JG: **Ethanol production by thermophilic bacteria: relationship between fermentation product yields of and catabolic enzyme activities in *Clostridium thermocellum* and *Thermoanaerobium brockii***. *Journal of Bacteriology* 1980, **144**(2):569-578.

128. Rydzak T, Levin DB, Cicek N, Sparling R: **End-product induced metabolic shifts in Clostridium thermocellum ATCC 27405**. *Applied Microbiology and Biotechnology* 2011, **92**(1):199-209.

129. Sauer U, Eikmanns BJ: **The PEP-pyruvate-oxaloacetate node as the switch point for carbon flux distribution in bacteria**. *FEMS Microbiology Reviews* 2005, **29**(4):765-794.

130. Bapteste E, Moreira D, Philippe H: **Rampant horizontal gene transfer and phospho-donor change in the evolution of the phosphofructokinase**. *Gene* 2003, **318**(0):185-191.

131. Bielen AAM, Willquist K, Engman J, Van Der Oost J, Van Niel EWJ, Kengen SWM: **Pyrophosphate as a central energy carrier in the hydrogen-producing extremely thermophilic *Caldicellulosiruptor saccharolyticus***. *FEMS Microbiology Letters* 2010, **307**(1):48-54.

132. Mukund S, Adams MW: **Glyceraldehyde-3-phosphate ferredoxin oxidoreductase, a novel tungsten-containing enzyme with a potential glycolytic role in the hyperthermophilic archaeon *Pyrococcus furiosus***. *The Journal of Biological Chemistry* 1995, **270**(15):8389-8392.

133. Gowen CM, Fong SS: **Genome-scale metabolic model integrated with RNAseq data to identify metabolic states of *Clostridium thermocellum***. *Biotechnology Journal* 2010, **5**(7):759-767.

134. Li HF, Knutson BL, Nokes SE, Lynn BC, Flythe MD: **Metabolic control of *Clostridium thermocellum* via inhibition of hydrogenase activity and the glucose transport rate**. *Applied Microbiology and Biotechnology* 2012, **93**(4):1777-1784.

135. Axley MJ, Grahame, D. A., Stadtman, T. C.: ***Escherichia coli* formate-hydrogen lyase. Purification and properties of the selenium-dependent formate dehydrogenase component**. *Journal of Biological Chemistry* 1990, **265**(30):18213-18218.

136. Garvie EI: **Bacterial lactate dehydrogenases**. *Microbiological Reviews* 1980, **44**(1):106-139.

137. van de Werken HJ, Verhaart MR, VanFossen AL, Willquist K, Lewis DL, Nichols JD, Goorissen HP, Mongodin EF, Nelson KE, van Niel EW *et al*: **Hydrogenomics of the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus***. *Applied and Environmental Microbiology* 2008, **74**(21):6720-6729.

138. Membrillo-Hernandez J, Echave P, Cabiscol E, Tamarit J, Ros J, Lin EC: **Evolution of the *adhE* gene product of *Escherichia coli* from a functional**

reductase to a dehydrogenase. Genetic and biochemical studies of the mutant proteins. *The Journal of Biological Chemistry* 2000, **275**(43):33869-33875.

139. Zhu J, Shimizu K: **Effect of a single-gene knockout on the metabolic regulation in *Escherichia coli* for D-lactate production under microaerobic condition**. *Metabolic Engineering* 2005, **7**(2):104-115.

140. Asanuma N, Hino T: **Effects of pH and energy supply on activity and amount of pyruvate formate-lyase in *Streptococcus bovis***. *Applied and Environmental Microbiology* 2000, **66**(9):3773-3777.

141. Asanuma N, Yoshii T, Hino T: **Molecular characteristics and transcription of the gene encoding a multifunctional alcohol dehydrogenase in relation to the deactivation of pyruvate formate-lyase in the ruminal bacterium *Streptococcus bovis***. *Archives of Microbiology* 2004, **181**(2):122-128.

142. Brown SD, Guss AM, Karpinets TV, Parks JM, Smolin N, Yang S, Land ML, Klingeman DM, Bhandiwad A, Rodriguez M, Jr. *et al*: **Mutant alcohol dehydrogenase leads to improved ethanol tolerance in *Clostridium thermocellum***. *Proceedings of the National Academy of Sciences U S A* 2011, **108**(33):13752-13757.

143. Trinh CT, Li J, Blanch HW, Clark DS: **Redesigning *Escherichia coli* metabolism for anaerobic production of isobutanol**. *Applied and Environmental Microbiology* 2011, **77**(14):4894-4904.

144. Liu X, Dong Y, Zhang J, Zhang A, Wang L, Feng L: **Two novel metal-independent long-chain alkyl alcohol dehydrogenases from *Geobacillus thermodenitrificans* NG80-2**. *Microbiology* 2009, **155**(Pt 6):2078-2085.

145. Pei J, Zhou Q, Jiang Y, Le Y, Li H, Shao W, Wiegel J: ***Thermoanaerobacter* spp. control ethanol pathway via transcriptional regulation and versatility of key enzymes**. *Metabolic Engineering* 2010, **12**(5):420-428.

146. Burdette D, Zeikus JG: **Purification of acetaldehyde dehydrogenase and alcohol dehydrogenases from *Thermoanaerobacter ethanolicus* 39E and characterization of the secondary-alcohol dehydrogenase (2 degrees Adh) as a bifunctional alcohol dehydrogenase--acetyl-CoA reductive thioesterase**. *Biochemical Journal* 1994, **302 (Pt 1)**:163-170.

147. Lovitt RW, Shen GJ, Zeikus JG: **Ethanol production by thermophilic bacteria: biochemical basis for ethanol and hydrogen tolerance in *Clostridium thermohydrosulfuricum***. *Journal of Bacteriology* 1988, **170**(6):2809-2815.

148. Bernard N, Johnsen K, Holbrook JJ, Delcour J: **D175 Discriminates Between NADH and NADPH in the Coenzyme Binding Site of *Lactobacillus delbrueckii* subsp. *bulgaricus* D-Lactate Dehydrogenase**. *Biochemical and Biophysical Research Communications* 1995, **208**(3):895-900.

149. Nair RV, Bennett GN, Papoutsakis ET: **Molecular characterization of an aldehyde/alcohol dehydrogenase gene from *Clostridium acetobutylicum* ATCC 824**. *Journal of Bacteriology* 1994, **176**(3):871-885.

150. Hamilton-Brehm SD, Mosher JJ, Vishnivetskaya T, Podar M, Carroll S, Allman S, Phelps TJ, Keller M, Elkins JG: ***Caldicellulosiruptor obsidiansis*** **sp. nov., an anaerobic, extremely thermophilic, cellulolytic bacterium isolated from Obsidian Pool, Yellowstone National Park**. *Applied and Environmental Microbiology*, **76**(4):1014-1020.

151. Vignais PM, Billoud, B., Meyer, J.: **Classification and phylogeny of hydrogenases**. *FEMS Microbiology Reviews* 2001, **25**:455-501.

152. Vignais PM: **Hydrogenases and H(+)-reduction in primary energy conservation**. *Results and Problems in Cell Differentiation* 2008, **45**:223-252.

153. Buhrke T, Lenz O, Porthun A, Friedrich B: **The H2-sensing complex of *Ralstonia eutropha*: interaction between a regulatory [NiFe] hydrogenase and a histidine protein kinase**. *Molecular Microbiology* 2004, **51**(6):1677-1689.

154. Angenent LT, Karim K, Al-Dahhan MH, Wrenn BA, Domiguez-Espinosa R: **Production of bioenergy and biochemicals from industrial and agricultural wastewater**. *Trends in Biotechnology* 2004, **22**(9):477-485.

155. Schut GJ, Adams MW: **The iron-hydrogenase of *Thermotoga maritima* utilizes ferredoxin and NADH synergistically: a new perspective on anaerobic hydrogen production**. *Journal of Bacteriology* 2009, **191**(13):4451-4457.

156. Rydzak T, McQueen PD, Krokhin OV, Spicer V, Ezzati P, Dwivedi RC, Shamshurin D, Levin DB, Wilkins JA, Sparling R: **Proteomic analysis of *Clostridium thermocellum* core metabolism: Relative protein expression**

profiles and growth phase-dependent changes in protein expression. *BMC Microbiology* 2012, **12**(1):214.

157. Chang T, Yao S: **Thermophilic, lignocellulolytic bacteria for ethanol production: current state and perspectives**. *Applied Microbiology and Biotechnology* 2011, **92**(1):13-27.

158. Guedon E, Desvaux M, Petitdemange H: **Improvement of cellulolytic properties of *Clostridium cellulolyticum* by metabolic engineering**. *Applied Environmental Microbiology* 2002, **68**(1):53-58.

159. Tripathi SA, Olson DG, Argyros DA, Miller BB, Barrett TF, Murphy DM, McCool JD, Warner AK, Rajgarhia VB, Lynd LR *et al*: **Development of pyrF-based genetic system for targeted gene deletion in *Clostridium thermocellum* and creation of a pta mutant**. *Applied and Environmental Microbiology* 2010, **76**(19):6591-6599.

160. Bradford MM: **A rapid and sensitive method for the estimation of microgram quantities of protein utilizing the principle of protein-dye binding.** *Analytical Biochemestry* 1976, **72**:248-254.

161. Guedon E, Payot, S., Desvaux, M. Petitdemanger, H.: **Carbon and electron flow in *Clostridium cellulolyticum* grown in chemostat culture on synthetic medium**. *Journal of Bacteriology* 1999, **181**(10):3262-3269.

162. Sanders R: **Compilation of Henry's law constants for inorganic and organic species of potential importance in environmental chemistry.** *Air Chemistry Department* 1999, **Max-Planck Institute of Chemistry**(Mainz, Germany).

163. Vella F: **Biochemistry: By R H Garrett and C M Grisham. pp 1154. Saunders College Publishing: Harcourt Brace, Orlando, FL. 1995. £19.95**. *Biochemical Education* 1995, **23**(2):108-108.

164. Sridhar J, Eiteman MA, Wiegel JW: **Elucidation of enzymes in fermentation pathways used by *Clostridium thermosuccinogenes* growing on inulin**. *Applied and Environmental Microbiology* 2000, **66**(1):246-251.

165. Lamed R, Zeikus JG: **Thermostable, ammonium-activated malic enzyme of *Clostridium thermocellum***. *Biochimica et Biophysica Acta* 1981, **660**(2):251-255.

166. Sparling R, Islam R, Cicek N, Carere C, Chow H, Levin DB: **Formate synthesis by *Clostridium thermocellum* during anaerobic fermentation**. *Canadian Journal of Microbiology* 2006, **52**(7):681-688.

167. Stevenson DM, Weimer PJ: **Expression of 17 genes in *Clostridium thermocellum* ATCC 27405 during fermentation of cellulose or cellobiose in continuous culture**. *Applied and Environmental Microbiology* 2005, **71**(8):4672-4678.

168. Pfaffl MW: **A new mathematical model for relative quantification in real-time RT-PCR**. *Nucleic Acids Research* 2001, **29**(9):e45.

169. Cleries R, Galvez J, Espino M, Ribes J, Nunes V, de Heredia ML: **BootstRatio: A web-based statistical analysis of fold-change in qPCR and RT-qPCR data using resampling methods**. *Computers in Biology and Medicine* 2012, **42**(4):438-445.

170. Mc BR: **The characteristics of *Clostridium thermocellum***. *Journal of Bacteriology* 1954, **67**(4):505-506.

171. Raman B, McKeown CK, Rodriguez M, Jr., Brown SD, Mielenz JR: **Transcriptomic analysis of *Clostridium thermocellum* ATCC 27405 cellulose fermentation**. *BMC Microbiology* 2011, **11**:134.

172. Soboh B, Linder D, Hedderich R: **A multisubunit membrane-bound [NiFe] hydrogenase and an NADH-dependent Fe-only hydrogenase in the fermenting bacterium *Thermoanaerobacter tengcongensis***. *Microbiology* 2004, **150**(Pt 7):2451-2463.

173. Rydzak T, McQueen PD, Krokhin OV, Spicer V, Ezzati P, Dwivedi RC, Shamshurin D, Levin DB, Wilkins JA, Sparling R: **Proteomic analysis of *Clostridium thermocellum* core metabolism: relative protein expression profiles and growth phase-dependent changes in protein expression**. *BMC Microbiology* 2012, **12**:214.

174. Brekasis D, Paget MS: **A novel sensor of NADH/NAD+ redox poise in Streptomyces coelicolor A3(2)**. *EMBO J* 2003, **22**(18):4856-4865.

175. Gyan S, Shiohira Y, Sato I, Takeuchi M, Sato T: **Regulatory loop between redox sensing of the NADH/NAD(+) ratio by Rex (YdiH) and oxidation of NADH by NADH dehydrogenase Ndh in *Bacillus subtilis***. *Journal of Bacteriology* 2006, **188**(20):7062-7071.

176. Leonardo MR, Cunningham PR, Clark DP: **Anaerobic regulation of the adhE gene, encoding the fermentative alcohol dehydrogenase of *Escherichia coli***. *Journal of Bacteriology* 1993, **175**(3):870-878.

177. Durre P, Fischer RJ, Kuhn A, Lorenz K, Schreiber W, Sturzenhofecker B, Ullmann S, Winzer K, Sauer U: **Solventogenic enzymes of *Clostridium acetobutylicum*: catalytic properties, genetic organization, and transcriptional regulation**. *FEMS Microbiology Reviews* 1995, **17**(3):251-262.

178. Huang H, Wang S, Moll J, Thauer RK: **Electron bifurcation involved in the energy metabolism of the acetogenic bacterium *Moorella thermoacetica* growing on glucose or H2 plus CO2**. *Journal of Bacteriology* 2012, **194**(14):3689-3699.

179. Wang S, Huang H, Moll J, Thauer RK: **NADP+ reduction with reduced ferredoxin and NADP+ reduction with NADH are coupled via an electron-bifurcating enzyme complex in *Clostridium kluyveri***. *Journal of Bacteriology* 2010, **192**(19):5115-5123.

180. Shaw AJ, Hogsett DA, Lynd LR: **Natural competence in Thermoanaerobacter and Thermoanaerobacterium species**. *Applied Environmental Microbiology* 2010, **76**(14):4713-4719.

181. Guss AM, Olson DG, Caiazza NC, Lynd LR: **Dcm methylation is detrimental to plasmid transformation in *Clostridium thermocellum***. *Biotechnology for Biofuels* 2012, **5**(1):30.

182. Jennert KC, Tardif C, Young DI, Young M: **Gene transfer to *Clostridium cellulolyticum* ATCC 35319**. *Microbiology* 2000, **146 Pt 12**:3071-3080.

183. Purdy D, O'Keeffe TA, Elmore M, Herbert M, McLeod A, Bokori-Brown M, Ostrowski A, Minton NP: **Conjugative transfer of clostridial shuttle vectors from *Escherichia coli* to *Clostridium difficile* through circumvention of the restriction barrier**. *Molecular Microbiology* 2002, **46**(2):439-452.

184. Li Y, Tschaplinski TJ, Engle NL, Hamilton CY, Rodriguez M, Jr., Liao JC, Schadt CW, Guss AM, Yang Y, Graham DE: **Combined inactivation of the *Clostridium cellulolyticum* lactate and malate dehydrogenase genes substantially increases ethanol yield from cellulose and switchgrass fermentations**. *Biotechnology for Biofuels* 2012, **5**(1):2.

185. Olson DG, Tripathi SA, Giannone RJ, Lo J, Caiazza NC, Hogsett DA, Hettich RL, Guss AM, Dubrovsky G, Lynd LR: **Deletion of the Cel48S cellulase from *Clostridium thermocellum***. *Proceedings for the National Academy of Sciences* 2010, **107**(41):17727-17732.

186. Tyurin MV, Desai SG, Lynd LR: **Electrotransformation of *Clostridium thermocellum***. *Applied and Environmental Microbiology* 2004, **70**(2):883-890.

187. Argyros DA, Tripathi SA, Barrett TF, Rogers SR, Feinberg LF, Olson DG, Foden JM, Miller BB, Lynd LR, Hogsett DA *et al*: **High ethanol titers from cellulose by using metabolically engineered thermophilic, anaerobic microbes**. *Applied Environmental Microbiology* 2011, **77**(23):8288-8294.

188. Sambrook J, Fritsch EF, Maniatis T: **Molecular cloning: a laboratory manual**, 2nd edn. Cold Springs Harbor, New York.: Cold Springs Harbor Laboratory Press; 1989.

189. Simon R, Priefer U, Puhler A: **A broad host range mobilization system for *in vivo* genetic engineering: Transposon mutagenesis in Gram negative Bacteria**. *Nature Biotechnology* 1983, **1**(9):784-791.

190. Mai V, Lorenz WW, Wiegel J: **Transformation of *Thermoanaerobacterium* sp. strain JW/SL-YS485 with plasmid pIKM1 conferring kanamycin resistance**. *FEMS Microbiology Letters* 1997, **148**(2):163-167.

191. Figurski DH, Helinski DR: **Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans**. *Proceedings for the National Academy of Sciences* 1979, **76**(4):1648-1652.

192. Heap JT, Pennington OJ, Cartman ST, Carter GP, Minton NP: **The ClosTron: a universal gene knock-out system for the genus Clostridium**. *Journal of Microbiological Methods* 2007, **70**(3):452-464.

193. Klapatch TR, Demain AL, Lynd LR: **Restriction endonuclease activity in *Clostridium thermocellum* and *Clostridium thermosaccharolyticum***. *Applied Microbiology Biotechnology* 1996, **45**(1-2):127-131.

194. Young M, Minton NP, Staudenbauer WL: **Recent advances in the genetics of the clostridia**. *FEMS Microbiology Reviews* 1989, **5**(4):301-325.

195. Ramachandran U, Wrana N, Cicek N, Sparling R, Levin DB: **Hydrogen production and end-product synthesis patterns by *Clostridium termitidis* strain CT1112 in batch fermentation cultures with cellobiose or alpha-cellulose**. *International Journal of Hydrogen Energy* 2008, **33**(23):7006-7012.

196. De Rossi E, Brigidi P, Rossi M, Matteuzzi D, Riccardi G: **Characterization of gram-positive broad host-range plasmids carrying a thermophilic replicon**. *Research in Microbiology* 1991, **142**(4):389-396.

197. Vielmetter W, Bonhoeffer F, Schutte A: **Genetic evidence for transfer of a single DNA strand during bacterial conjugation**. *Journal of Molecular Biology* 1968, **37**(1):81-86.

198. Piffaretti JC, Arini A, Frey J: **pUB307 mobilizes resistance plasmids from Escherichia coli into Neisseria gonorrhoeae**. *Molecular Genetics and Genomic* 1988, **212**(2):215-218.

199. Guiney DG, Jr.: **Promiscuous transfer of drug resistance in gram-negative bacteria**. *The Journal of Infectious Diseases* 1984, **149**(3):320-329.

200. Bertram J, Dürre P: **Conjugal transfer and expression of streptococcal transposons in *Clostridium acetobutylicum***. *Archives of Microbiology* 1989, **151**(6):551-557.

201. Wang DI, Biosic I, Fang Y, Wang JD: **Direct microbiological conversion of cellulosic biomass to ethanol.** *Proceedings of the 3rd annual biomass energy systems conference* 1979, **National Technical Information Service, Springfield Va.**:61-67.

202. Ng TK, Ben-Bassat A, Zeikus JG: **Ethanol production by thermophilic bacteria: Fermentation of cellulosic substrates by cocultures of *Clostridium thermocellum* and *Clostridium thermohydrosulfuricum***. *Applied Environmental Microbiology* 1981, **41**(6):1337-1343.

203. Mori Y: **Characterization of a symbiotic coculture of *Clostridium thermohydrosulfuricum* YM3 and *Clostridium thermocellum* YM4**. *Applied Environmental Microbiology* 1990, **56**(1):37-42.

204. Murray WD: **Symbiotic relationship of *Bacteroides cellulosolvens* and *Clostridium saccharolyticum* in cellulose fermentation**. *Applied Environmental Microbiology* 1986, **51**(4):710-714.

205. Fan LH, Zhang ZJ, Yu XY, Xue YX, Tan TW: **Self-surface assembly of cellulosomes with two miniscaffoldins on *Saccharomyces cerevisiae* for cellulosic ethanol production**. *Proceedings for the National Academy of Sciences* 2012, **109**(33):13260-13265.

Appendix

Supplementary Table 1. Cofactor specificity (ATP, ADP or PP_i) of phosphofructokinases based on sequence alignments. Alignments of key residues determining ATP, ADP or PP_i specificity, as determined by Bapteste *et al*. [130] and Bielen *et al*. [131], were performed using BioEdit v.7.0.9.0. The *P. furiosus* and *Th. kodakarensis* genes are very distinct (different COG and different KO) and are annotated as Archaeal phosphofructokinases.

6-Phosphofructokinase				
Organism	Locus Tag	AA #104	AA #124	ATP, ADP or PPi Dependence
Bacillus cereus ATCC 14579	BC4600	G	G	ATP
Caldanaerobacter subterraneus subsp. tengcongensis MB4	TTE1816	G	G	ATP
Ethanoligenens harbinense YUAN-3T, DSM 18485	Ethha_1347	G	G	ATP
Geobacillus thermoglucosidasius C56-YS93	Geoth_0897	G	G	ATP
P. furiosus DSM 3638	PF0312	S	E	ADP
	PF1784	L	F	ADP
Thermococcus kodakaraensis KOD1	TK0376	L	F	ADP
	TK1110	S	E	ADP
Thermotoga maritima MSB8	TM0209	G	A	ATP
	TM0289	D	K	PP_i
Thermotoga neapolitana DSM 4359	CTN_0395	D	K	PP_i
	CTN_0476	G	A	ATP
Thermotoga petrophila RKU-1	Tpet_0623	D	K	PP_i

	Tpet_0715	G	A	ATP
Caldicellulosiruptor bescii DSM 6725	Athe_1265	G	G	ATP
	Athe_1824	D	K	PP_i
Caldicellulosiruptor saccharolyticus DSM 8903	Csac_1830	G	G	ATP
	Csac_2366	D	K	PP_i
Clostridium cellulolyticum H10	Ccel_2223	D	K	PP_i
	Ccel_2612	G	G	ATP
Clostridium phytofermentans ISDg	Cphy_0336	G	G	ATP
	Cphy_3345	D	K	PP_i
Clostridium thermocellum ATCC 27405	Cthe_0347	D	K	PP_i
	Cthe_01261	G	G	ATP
Clostridium thermocellum DSM 4150	Cther_1670	G	G	ATP
Thermoanaerobacter pseudethanolicus 39E	Teth39_0494	D	K	PP_i
	Teth39_0683	G	G	ATP

Supplementary Table 2. Transcription of key genes involved in pyruvate catabolism, ethanol synthesis and hydrogen production in C. thermocellum ATCC 27405 during mid-exponential phase of growth (OD600nm ~ 0.500) under low N2 (2 mL s^{-1}), high N2 and high H2 (20 mL s^{-1}) sparging conditions. Values represent the expression ratio relative to 16s rRNA as determined from a minimum of technical triplicates of biological duplicates ($n \geq 6$) . Stadard deviations (in brackets) are provided

Gene	Annotation	Relative expression ratio (16s rRNA)		
		20 mL s^{-1} N_2	2 mL s^{-1} N_2	20 mL s^{-1} H_2
malE	Cthe_0344	2.13E-01 (0.124)	1.75E-01 (0.058)	1.30E-01 (0.013)
mdh	Cthe_0345	7.90E-03 (<0.001)	4.30E-03 (<0.001)	5.60E-03 (< 0.001)
ldh	Cthe_1053	4.70E-04 (<0.001)	5.98E-04 (<0.001)	6.63E-04 (<0.001)
pfor	Cthe_2796	8.25E-03 (0.006)	2.01E-02 (0.003)	1.41E-02 (0.006)
	Cthe_2392	4.06E-02 (0.002)	8.40E-02 (0.002)	7.67E-02 (0.014)
	Cthe_3120	5.23E-04 (< 0.001)	1.26E-03 (< 0.001)	9.37E-04 (< 0.001)
pfl	Cthe_0505	6.81E-02 (0.057)	6.20E-02 (0.016)	4.90E-02 (0.012)
aldh	Cthe_2238	1.13E-02 (0.002)	1.32E-02 (0.003)	1.41E-02 (0.003)
adhE	Cthe_0423	6.17E-03 (0.005)	1.74E-02 (0.008)	3.84E-02 (0.018)

Fe-adh	Cthe_2579	5.08E-02 (0.002)	3.01E-02 (< 0.001)	1.65E-02 (0.001)
	Cthe_0394	5.96E-02 (0.037)	9.57E-02 (0.041)	3.10E-02 (0.006)
	Cthe_0101	6.13E-03 (< 0.001)	3.56E-03 (< 0.001)	3.40E-03 (< 0.001)
[FeFe] H_2ase	Cthe_0430	1.21E-01 (0.012)	1.42E-01 (0.086)	2.14E-01 (0.035)
	Cthe_3003	1.15E-02 (0.002)	2.03E-02 (< 0.001)	2.10E-02 (0.003)
	Cthe_0342	1.85E-02 (0.008)	1.48E-02 (0.005)	1.63E-02 (0.010)
[NiFe] H_2ase	Cthe_3020	1.13E-02 (0.002)	1.32E-02 (0.003)	1.41E-02 (0.003)
rnf	Cthe_2430	9.74E-02 (0.035)	6.30E-02 (0.005)	4.84E-02 (0.004)

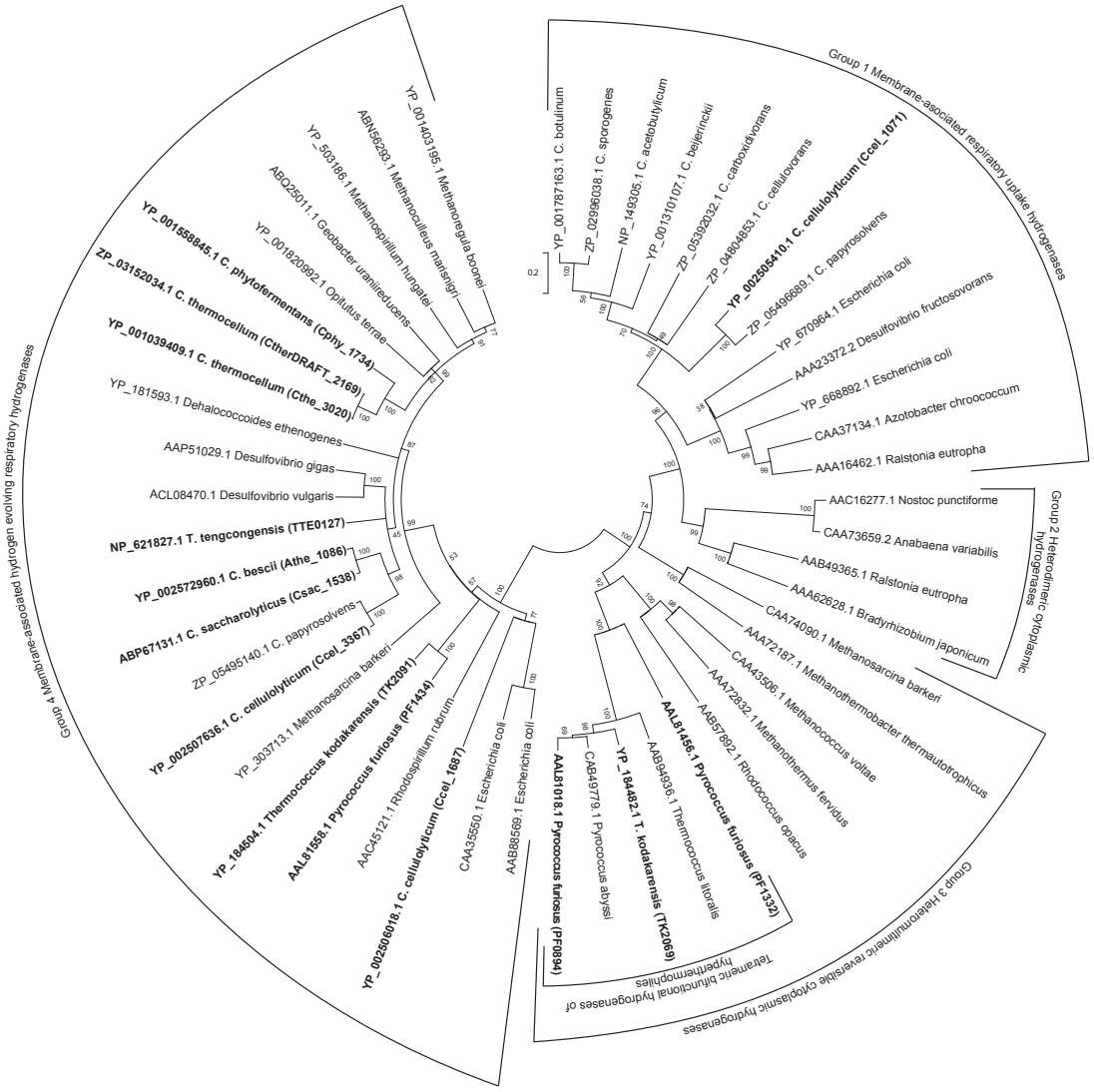
Supplementary Table 3. Primer sequences used for PCR amplification of macrolide-lincosamide-streptogramin (*mls)* region of plasmid pIkm1, the chloramphenicol acetyltransferase (*cat)* region from plasmid pNW33Nmob and 16s rDNA region of *Clostridium thermocellum* ATCC 27405.

Primer name	Nucleotide sequence
mls-fwd.	5'-AAGAGGTTATAATGAACGAGAAAA-3'
mls-rvs.	5'-AAAAATAGGTACACGAAAAACAAGTTA-3'
cat-fwd.	5'-GCGTGTTACGGTGAAAACCT-3'
cat-rvs.	5'-ATCACAAACGGCATGATGAA-3'
16s rDNA fwd.	5'-TGACGGGCGGTGTGTACAAGG-3'
16s rDNA rvs.	5'-GGTGGGGACGACGTCAAATCA-3'

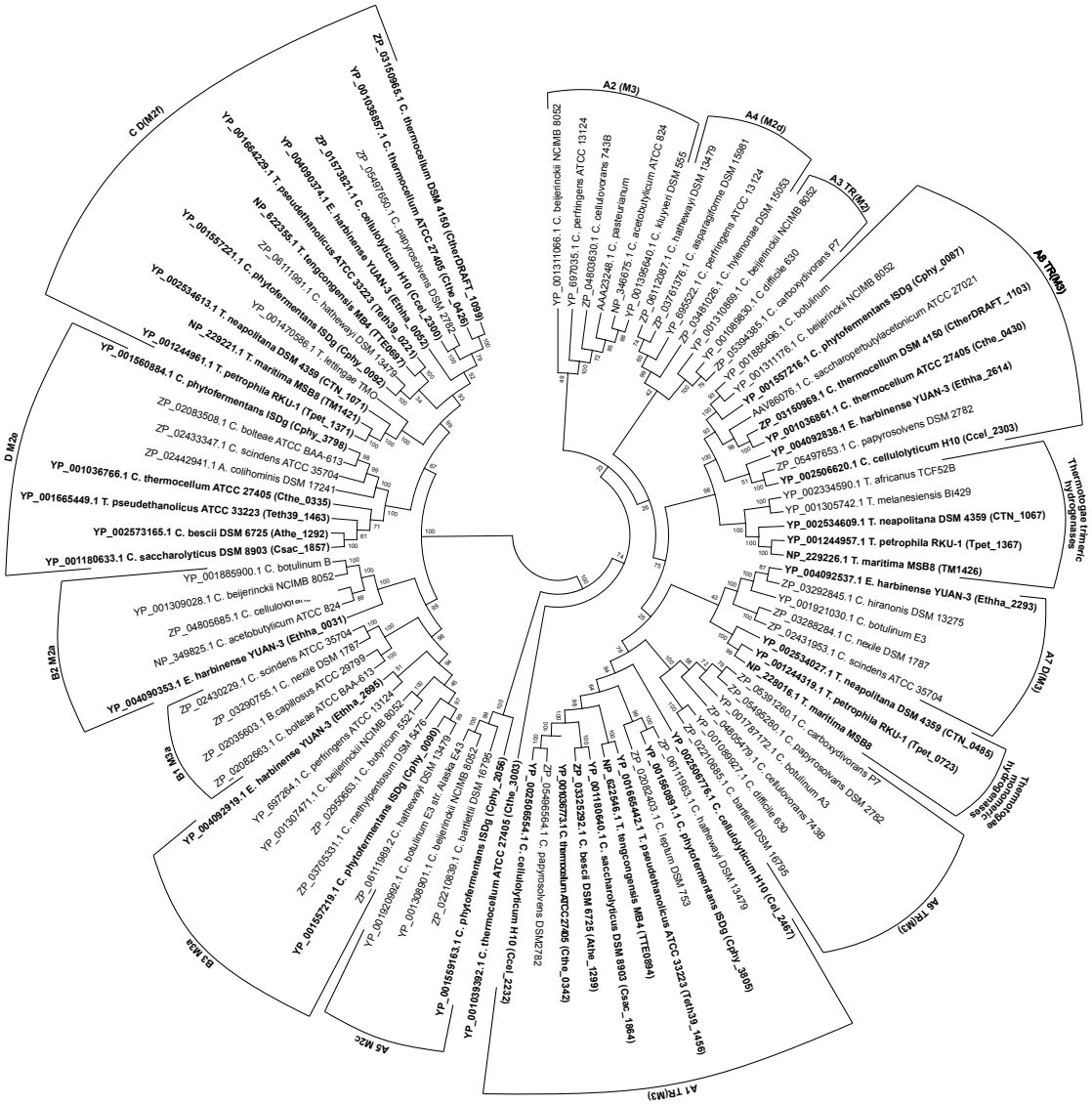
Supplementary Table 4. Electroporation of *Clostridium thermocellum* ATCC27405 with plasmid pIKm1.

Field Pulse (kV cm^{-1})	Resistance (Ω)	Pulse length (ms)	Capacitance (uF)	Viable (Y/N)	Transformed (Y/N)
7	500	8.6	50	Yes	No
10	500	3.2	50	Yes	Yes (Em^{R*})
12.5	500	1.4	50	Yes	Yes (Em^{R})

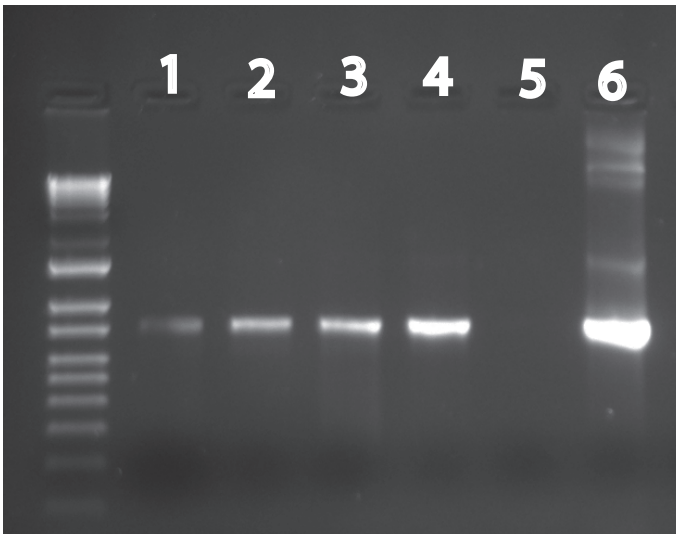
* cultures exhibited resistance to Erythromycin at 20 μg/ml



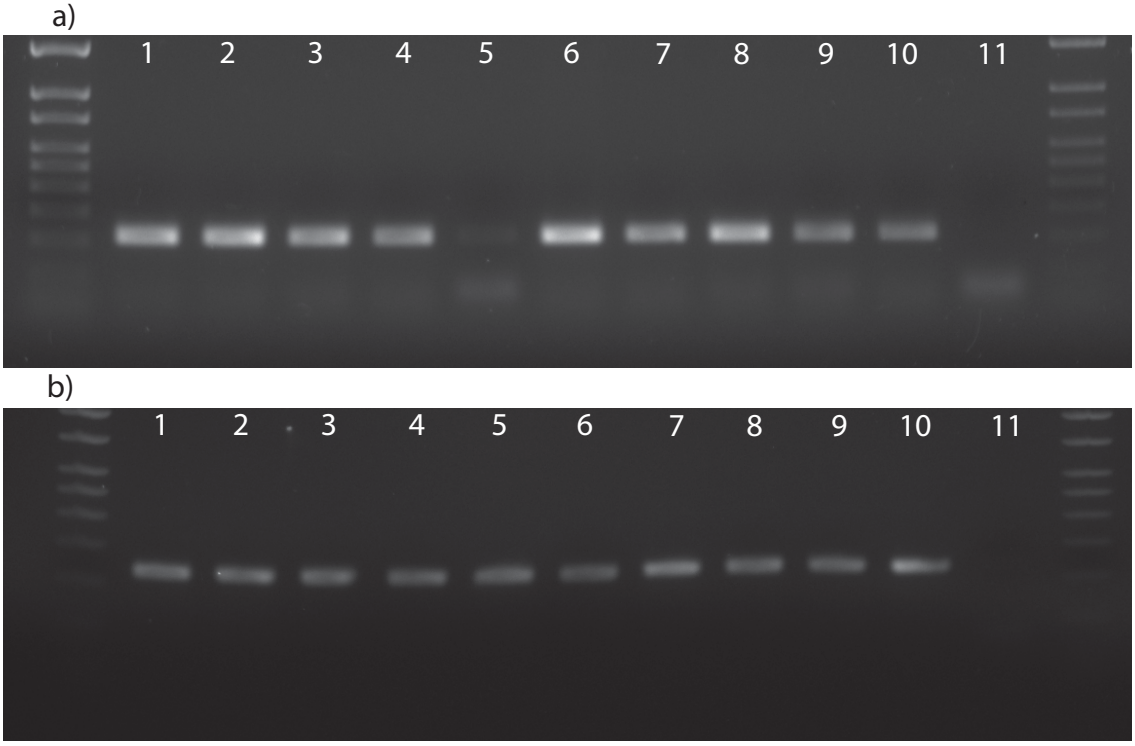
Supplementary Figure 1. Phylogenetic clustering of [NiFe] hydrogenases large (catalytic) subunits. Catalytic (large) subunits of [NiFe] H_2ases were identified based upon the modular signatures as described by Calusinska *et al*. [72]. Species considered in this manuscript are highlighted and corresponding H_2ase gene loci are provided.



Supplementary Figure 2. Phylogenetic clustering of [FeFe] hydrogenases large (catalytic) subunits. Catalytic (large) subunits of [FeFe] H_2ases were identified based upon the modular signatures as described by Calusinska *et al*. [72]. Species considered in this manuscript are highlighted and corresponding H_2ase gene loci are provided**.**



Supplementary Figure 3. PCR amplification of the *mls* cassette of plasmid pIkm1 recovered from total DNA extractions of *C. thermocellum* ATCC 27405 following electroporation. Lanes 1-4, *C. thermocellum* electroporated with plasmid pIkm1; lane 5, negative control; lane 6, positive control.



Supplementary Figure 4. PCR amplification of the a) chloramphenicol acetyl transferase gene of plasmid pNW33Nmob and b) 16s rDNA from total DNA extractions of *Clostridium thermocellum* ATCC 27405 and *Thermoanaerobacter pseudethanolicus* 39E exoconjugants. Lanes 1-4, *C. thermocellum* exoconjugant total DNA; lane 5, *C. thermocellum* genomic DNA; lanes 6-10, *T. pseudethanolicus* exoconjugant DNA; lane 11, no template control.