Industrial Robustness: Understanding the mechanism of tolerance for the *Populus* hydrolysate tolerant strain of *Clostridium thermocellum*

A Dissertation Presented for

The Doctor of Philosophy Degree

The University of Tennessee

Jessica Leigh Linville

May 2013

Acknowledgements

I would like to thank Chris D. Cox and Jonathan R. Mielenz for their support and guidance during my tenure in graduate school. I would also like to thank Qiang He, Kevin G. Robinson and Gary S. Sayler for serving on my committee. Special thanks go to Miguel Rodriguez Jr., Kelsey Yee and Kelly Craven for their support, knowledge and encouragement during this process. I would also like to thank a number of people at Oak Ridge National Laboratory for their assistance in my research including: Tim Tschaplinski, and Nancy Engle with Populus hydrolysate analysis; Adam Guss and Donna Kridelbaugh with strain isolation and engineering; Steve Brown, Dawn Klingeman and Courtney Johnson with DNA and RNA isolation and the resulting genomic and transcriptomic analysis; and Loren Hauser, Miriam Land, Mustafa Syed and Tatiana Karpinets for help with Kbase and pathway tools.

I would also like to thank the U.S. Department of Energy funded Bioenergy Science Center and Institute for a Secure and Sustainable Environment at the University of Tennessee for providing funding. I would also like to thank the National Renewable Laboratory for providing the *Populus* hydrolysate and the Joint Genome Institute for analysis of DNA and RNA samples used in this project.

Abstract

An industrially robust microorganism that can efficiently degrade and convert lignocellulosic biomass into ethanol and next-generation fuels is required to economically produce future sustainable liquid transportation fuels. The anaerobic, thermophilic, cellulolytic bacterium *Clostridium thermocellum* is a candidate microorganism for such conversions but it, like many bacteria, is sensitive to potential toxic inhibitors developed in the hydrolysate produced during biomass processing. Microbial processes leading to tolerance of the inhibitory compounds found in the pretreated biomass hydrolysate are likely complex and involve multiple genes. In this study, a 17.5% v/v *Populus* hydrolysate tolerant mutant strain of *C. thermocellum* was developed by directed evolution.

The genome of the wild type strain, intermediate population samples and single colony isolates were sequenced to elucidate the mechanism of tolerance. Genetic mutations common to all isolates were matched with the observed phenotype through kinetic modeling and comparison of gene expression levels (RNA-seq) during fermentation by the wild type strain and mutant isolate #6 in various concentrations of *Populus* hydrolysate (0%, 10%, and 17.5% v/v). Inhibition was found to be a constant term based on initial hydrolysate concentrations. The mutant strain was found to have a faster growth rate and was less inhibited than the wild type. The mutant increased expression of genes encoding for energy production and conversion, and amino acid transport and metabolism, and decreased expression of genes encoding for the cell envelope and outer membrane, cell motility, cellulosome, inorganic ion transport and metabolism, sporulation and cell defense mechanisms when compared to the wild type in standard media. The wild type differently expressed twice as many genes when compared to the mutant in hydrolysate conditions. The mutant increased growth related genes where as the wild type increased cell defense mechanisms when placed in hydrolysate media.

The findings suggest that there are multiple mutations responsible for the *Populus* hydrolysate tolerant phenotype resulting in several simultaneous mechanisms of action. To date, this study provides the most comprehensive elucidation of the mechanism of tolerance to a pretreated biomass hydrolysate by *C. thermocellum*. These findings make important contributions to the development of industrially robust strains of consolidated bioprocessing microorganisms.

Table of Contents

List of Tables

List of Figures

List of Attachments

Supplemental File S1: Clostridium thermocellum genomic analysis

An excel file containing the full list of possible mutations received from JGI.

File Name: File S1 Clostridium thermocellum genomic analysis.xlx

Supplemental File S2: Differential Gene Expression Profiles

An excel file containing the significantly differentially expressed genes based of the ANOVA results.

File Name: File S2 Differential Gene Expression Profiles.xlx

Supplemental File S3: Cthe_2376-alignment

A text file containing the Clustawl alignment of the SNP in Cthe_2376 with the top 100 hits.

File Name: File S3 Cthe_2376-alignment.txt

Supplemental File S4: Cthe_2724-alignment

A text file containing the Clustawl alignment of the SNP in Cthe_2724 with the top 100 hits.

File Name: File S4 Cthe_2724-alignment.txt

Supplemental File S5: Cthe_2727-alignment

A text file containing the Clustawl alignment of the SNP in Cthe_2727 with the top 100 hits.

File Name: File S5 Cthe_2727-alighment.txt

Supplemental File S6: All statistically significant differentially expressed genes

An Excel file with all of the results of the ANOVA test for all 1,795 statistically significant differentially expressed genes based on all possible simple comparisons.

File Name: File S6 Significantly differentially expressed genes.xlx

Supplemental File S7: Significantly differentially expressed hypothetical proteins

An Excel file with the 492 genes that encode for hypothetical proteins, pseudo genes, and genes of unknown function.

File Name: File S7 Hypothetical Genes.xlx

Supplemental File S8: Significantly differentially expressed genes with category designation

An Excel file with the 1014 genes that were significantly differentially expressed along with the category designation assigned by this analysis.

File Name: File S8 Differentially Expressed Genes by Catergory.xlx

Introduction

With the depletion of the world's oil supply and rising concerns about the environmental impact associated with the use of oil, there is an increasing need for a more sustainable source of fuel. So far, 99% of the total biofuel consumption in the US comes from ethanol [1]. The conversion of lignocellulosic biomass (biomass) to ethanol is one option of creating biofuels. Biomass has the potential to meet most, if not all, of the energy demand for transportation needs [2] and has the ability to significantly reduce net carbon dioxide emission because it is part of a global carbon cycle [3]. The main technological roadblock to the widespread use of biomass for the production of fuels is the lack of technologies to overcome the recalcitrance of its cellulosic structure [4-6]. Recalcitrance is the resistance of cellulosic biomass to enzymatic conversion into sugars [5].

Current technology utilized at bio-refineries includes numerous complex, costly and energy-intensive steps with low efficiency and yields [2]. Before the conversion of biomass begins, it must be pretreated. The pretreatment step is usually a physical and/or chemical process, in which the polymeric sugars are made more accessible to conversion during the enzymatic process [6]. In addition to generating suitable substrates for conversion to biofuels, the pretreatment process of biomass typically produces a range of compounds that inhibit the organism used for fermentation [7].

Traditional biomass processing schemes involving enzymatic or microbial hydrolysis have four biological steps: the production of saccharolytic enzymes such as cellulases and hemicellulases; the hydrolysis of cellulosic polymer components in the pretreated biomass to sugars; the fermentation of hexose sugars such as glucose, mannose and galactose; and the fermentation of pentose sugars such as xylose and arabinose [4]. Recent advances in fermentation research have yielded microbial strains capable of simultaneous saccharification and fermentation (SSF) of biomass so that the hydrolysis of cellulose and fermentation of glucose are accomplished in a single step significantly reducing the cost of

biomass processing [2]. Further advances are needed to achieve the goal of a consolidated bioprocessing (CBP) scheme, which allows for the four biological steps to occur in a single processing unit, which will reduce the complexity, cost, and energy intensity of the process [2]. The simplified industrial CBP will allow for fundamental process cost savings necessary to have a competitive biofuels market [5].

However, CBP requires a highly-engineered microbe that is robust enough to efficiently breakdown the biomass into fermentable sugars and ferment the mixed- sugar hydrolysate [6], have balanced growth on hexose and pentose sugars [8] and have a higher rate of conversion and yield than current strains [2]. Furthermore, due to the variety of feedstocks and the diversity of pretreatment methods, the organism must be robust with regards to inhibitor tolerance [6].

This introduction will review the different inhibitory compounds produced from the pretreatment process and the effect of the inhibitory compounds have on the growth. The introduction will also examine gene expression of various bacteria used for the fermentation of biomass into ethanol, and methods for increasing and analyzing industrial robustness including genomic and transcriptomic analysis, and kinetic modeling.

Biomass pretreatment

Successful pretreatment methods are needed to help to overcome the recalcitrance of biomass and produce materials with acceptable enzymatic digestibility and fermentability for the production of biofuels [7]. The three main components of biomass are cellulose, hemicelluloses and lignin, which are assembled into an intricate composite that comprises plant cell walls [2]. Cellulose is long strands of D-

glucose subunits linked together by β-1,4 glycosidic bonds which are weakly bonded together through hydrogen bonding [9]. Hemicellulose is complex branches structure that consists of different polymers like pentoses (i.e. xylose and arabinose), hexoses (i.e. mannose, glucose and galactose), and sugar acids [9]. Lignin is an amorphous heteropolymer consisting of three different phenylpropane units (*p*-coumaryl, coniferyl and sinapyl alcohol) [9].

The various pretreatment methods are designed to make the sugars more available for subsequent hydrolysis and fermentation steps through the breakdown of the cell wall, and the degradation of cellulose, hemicelluloses and lignin matrix [7]. The different categories of pretreatment processes include: physical, base catalyzed, non-catalyzed (thermal), acid catalyzed, solvent based, chemical based or a combination of the above processes [7,9]. Table 1 lists the types of pretreatment methods available for each category. Although, there are many types of acid pretreatment, sulfuric acid pretreatment is the most commonly investigated [10].

The various pretreatment processes produce compounds that are inhibitory to the organisms used for fermentation. These inhibitory compounds come from the partial degradation of the different components of the biomass [11] and include carboxylic acids, primarily acetic acid, the sugar degradation products furfural and hydroxymethylfurfural (HMF), phenolic compounds, and inorganic salts [7]. When cellulose is degraded glucose is hydrolyzed and when hemicelluloses is degraded, xylose, mannose, acetic acid, galactose, and glucose are liberated [12]. The furan derivatives HMF and furfural are formed at high temperature and pressure as degradation products of hexoses and pentoses, respectively [1,12]. The phenolic compounds come from the degradation of lignin [1,12,13]. Vanillic acid and vanillin are formed from the degradation of the guaiacylpropane units of lignin and syringealdehydye and syringic acid are formed from the degradation of syringyl propane units [3,12,13].

Table 1: List of available pretreatment technologies for cellulosic biomass

Pretreatment category	Specific pretreatment technology
Physical	Milling
Base catalyzed	Ammonia fiber explosion/ fiber extrusion explosion, Ammonia recycle percolation, Lime
Non-catalyzed	Autohydrolysis – Steam, Steam explosion, Liquid hot water, Hot water pH neutral
Acid catalyzed	Carbonic acid, Nitric acid, Sulfuric acid
Solvent based	Organosolv
Chemical based	Wet oxidation (O_2, peroxide, ozone)

4-hydroxybenzene constitutes a large fraction of the lignin-derived compounds in hydrolysates from the hardwoods poplar and willow [12,13]. Different biomass feedstocks and pretreatment methods produce different combinations of toxic compounds [7].

The inhibitory compounds produced from corn stover and different pretreatment processes have been analyzed by several methods. Varga et al. (2004) measured the inhibitory compounds produced from corn stover by wet oxidation (WO) in both acidic and alkaline conditions [3]. Chen et al. (2005) measured the inhibitory compounds produced from corn stover pretreated in 1% (v/v) sulfuric acid at a solids concentration of 1000 g/L [14]. Du et al. (2010) recently did a comprehensive investigation of the different inhibitors produced from corn stover, poplar and pine feedstocks that were pretreated with eight different chemical conditions, which are representative of leading pretreatment processes. Pretreatment processes include: 0.7% H_2SO_4, 0.07% H_2SO_4, liquid hot water, neutral buffer solutions, aqueous ammonia, lime, lime with oxygen pressurization, and wet oxidation [11]. Joh et al (2010) pretreated Switchgrass using four different methods: water only at 55^{o}C, 60% methanol at 25^{o}C, 6% NaOH at 55^{o}C, and 1 M NaOH at 25^{o}C [15]. In general, it was observed that the types and concentration of degradation products produced during the various pretreatment processes depended more upon the feedstock than on the type of pretreatment used [11]. The range of possible inhibitory compounds identified for each of the feedstocks discussed for the various pretreatment methods is summarized in Table 2.

Currently, detoxification methods and strategies are employed to remove the inhibitory compounds in order to increase the efficiency of the fermentation process. Detoxification methods include biological, physical and chemical strategies [16]. These methods can include separation of hemicellulose hydrolysate, washing of fibers, and reducing toxicity by overliming, charcoal, or ion

Table 2: The range of concentrations of inhibitory compounds found in the leading biomass feedstocks subjected to various pretreatment methods

	Biomass Feedstock Hydrolysate			
Inhibitory compound [a,b]	**Corn Stover**	**Hybrid Poplar**	**Pine**	**Switch grass**
aliphatic acids				
malonic acid	0.32 – 2.6	0.15-2.7	0.21-3.1	
lactic acid	5.5 - 45	1.8-29	3.7-39.6	
maleic acid	0.80 – 3.1	0.25-3.7	0.11-2.2	
cis-aconitic acid	0.055 - 4.7	0.075-0.90	0.048-1.1	
methylmalonic acid	0.007 - 0.070	0.005 - 0.045	0.11 - 0.054	
succinic acid	1.7 - 11	0.93 - 6.4	0.34-4.4	
fumaric acid	1.8 - 5.9	0.11 - 1.7	0.054 - 1.5	
trans-aconitic acid	0.036 - 1.0	0.018 - 0.16	0 - 0.081	
levulinic acid	0.48 - 41	0.29 - 45	0.40-30	
glutaric acid	0.23 – 1.2	0.23 - 1.1	0.16-1.2	
itaconic acid	1.2 - 7.2	0.088-0.55	0.032 - 0.65	
2-hydroxy-2-methylbutyric acid	0.006 - 0.033	0-0.050	0.009 - 0.50	
gallic acid	0.003 – 0.70	0.002-0.052	0.005 - 0.68	
adipic acid	0.11 – 0.25	0.048-0.64	0.054 - 0.22	
acetic acid	34 - 240	57-310	14 - 120	
formic acid	43 - 250	52-210	42 - 210	
aromatic acids				
2-furoic acid	0.77-7[c]	0.30-3.1	0.23 - 1.1	
3,4-dihydroxybenzoic acid	0.031 – 0.73	0.018-1.7	0.021 - 0.82	0.162[d]
3,5-dihydroxybenzoic acid	0.006-0.048	0-0.022	0 - 0.040	
2,5-dihydroxybenzoic acid	0.015-0.23	0.007-0.070	0.005 - 0. 035	
salicylic acid	0.78-2.8	19-44	0.24 - 0.98	
vanillic acid	1.5-57[c]	2.5-5.9	2.3 - 8.7	0.632[d] - 36.22[d]
homovanillic acid	0.11-28[c]	0.049-0.18	0.090 - 0.31	
caffeic acid	0.038-0.20	0.003-0.053	0.005 - 0.028	

a) concentrations in mg/L unless stated otherwise, b) data taken from various pretreatment methods found in [11] unless stated otherwise, c) data taken from [3], d) data taken from [15] concentrations converted from ug/g dry switchgrass based on pretreatment method used.

Table 2: The range of concentrations of inhibitory compounds found in the leading biomass feedstocks subjected to various pretreatment methods cont.

	Biomass Feedstock Hydrolysate			
Inhibitory compound [a,b]	**Corn Stover**	**Hybrid Poplar**	**Pine**	**Switch grass**
aromatic acids cont.				
syringic acid	1.4-42[C]	1.6-3.5	0.40-1.0	0.676[d] - 13.32[d]
4-hydroxybenzoic acid	0.028-54[C]	0.016-0.12	0-0.033	
benzoic acid	0.16-1.7	0.78 - 2.6	1.7 -3.8	
4-hydroxycoumaric acid	5.6-17	0.018-0.93	0.020 - 0.33	
ferulic acid	0.76-9[C]	0.070-0.46	0.083 - 0.31	7.55[d] - 251.38[d]
3-hydroxy-4-methoxycinnamic acid	0.019 – 0.22	0.005 - 0.040	0.003 - 0.030	
ortho-toluic acid	0.016 - 0.030	0.013-0.041	0.015 - 0.031	
para- toluic aicd	0.36 – 0.98	0.37-0.93	0.43 - 0.90	
coumaric acid	32[C]-35[C]			
trans - p - Courmaic acid				8.70[d] - 245.44[d]
aldehydes and ketones				
5-hydroxymethylfurfural (5-HMF)	0.89-44	0.079 -64	0.16 - 170	
Furfural	0.40-220	0.50-220	0.65 - 190	
4-hydroxybenzaldehyde	1.5-102[C]	0.093-0.49	0.24 - 0.70	
3,4-dihydroxybenzaldehyde	0.057-0.082	0.038-0.50	0.041 - 0.66	0.124[d]
4 - hydroxyacetophenone	0.28-8[C]	0.015 - 0.70	0.020 - 0.090	
Vanillin	1.7-74[C]	2.6 -9.1	2.4 - 7.1	1.37[d] -41.17[d]
syringaldehyde	0.084-38[C]	0.93 - 7.4	0.20 - 1.1	
4-hydroxycoumarin	0.006 - 0.035	0.001 - 0.20	0.002 - 0.020	
Acetovanillone	13[C]-58[C]			
Acetosyringone	0[C]-18[C]			
Phenols				
Phenol	4.0[C]-6.0[C]			
Guaiacol	12.0[C]-13.0[C]			
Syringol	5[C]			

a) concentrations in mg/L unless stated otherwise, b) data taken from various pretreatment methods found in [11] unless stated otherwise, c) data taken from [3], d) data taken from [15] concentrations converted from ug/g dry switchgrass based on pretreatment method used

exchange [10]. However, this increases the production cost by adding equipment, materials and complexity [2,10]. Other research is focusing on the formation and modification of the biomass in order to develop varieties which are easier to breakdown into fermentable sugars that would require little or no pretreatment [2]. However, little is currently known about how cellulose and hemicelluloses are synthesized; distributed within cell walls; and attached to each other, to lignin, or to cell-wall proteins [5] making modifications to the biomass structure difficult.

Organism Inhibitory Compound Tolerance

Inhibitory compounds have been shown to reduce the rate of ethanol production and the overall yield [2,7]. Testing of model compounds, either individually or in mixtures, can indicate the resistance of the bacteria strains towards the inhibitory compounds. However, the different fermentation strains have different levels of resistance to the hydrolysate inhibitory compounds [7]. The fermentation performance of biomass hydrolysates also depend on biomass feedstocks and pretreatment conditions [17]. Several assays have been used to evaluate and compare toxicity on microbial fermentation, including cell growth, ethanol yield and ethanol productivity [17]. Due to the variety in the experimental design (methods used to for toxicity measurement, fermentation organisms, inoculum levels, biomass feedstock, pretreatment conditions, and choice of model compounds), the analysis of inhibition by hydrolysates and model compounds in the literature is often difficult to compare [17].

Saccharomyces cerevisiae is used extensively in the ethanol industry due to its high ethanol tolerance, high fermentation rate and robustness. However, *S. cerevisiae* cannot ferment xylose to ethanol and many of the yeast strains, including industrial strains, are susceptible to the multiple

inhibitory compounds present. Geng et al. (2010) compared the cell growth levels of six different yeast strains when exposed to stresses such as pH, ethanol, xylose, temperature and inhibitors. The cells were then exposed to various concentrations of a stock inhibitor cocktail with multiple inhibitory compounds [18]. Barber et al. (2000) examined inhibition for three yeast strains *S. cerevisiae*, *Saccharomyces pombe*, and *Saccharomyces roseus*, and three bacteria strains: *Bacillus subtilis Niger, Escherichia coli*, and *Pseudomonas syringe* [19]. Zaldivar et al. (1999) examined the toxicity of six aromatic alcohols and furfuryl alcohol on the growth and fermentation of *E. coli* LY01 singly and in binary combinations [20]. The binary combinations of inhibitory compounds were roughly additive in toxicity, although some were more toxic (60%-80% inhibition) and some combinations less than additive (25% -40% inhibition) [20]. Geddes et al. (2010) described the development of an ethanogenic *E. coli* (strain MM 160) with increased resistance to phosphoric acid hydrolysates inhibitors of sugarcane bagasse [10]. Franden et al. (2009) measured the effect of furfural, HMF, acetate, ethanol and dilute acid pretreated corn stover hydrolysate on *Zymomonas mobilis* recombinant strain 8b [17].

Only a few studies have been conducted on the effects of inhibitory compounds using cellulolytic bacteria. Joh et al (2010), in addition to quantifying the inhibitors present in Switchgrass, determined the growth rate of *Acidothermus cellulolyticus* when grown with known amounts of phenolic compounds from several classes (cinnamic acid derivatives, hydroybenzoic acids, benzaldehydes, and condensed tannin monomers) [15]. Another study on *A. cellulolyticus* by VanderGheynst et al. (2009) investigated conditions that supported growth on switchgrass in solid-state fermentation [21]. The extraction ratio had the most effect on growth inhibition [21]. Ezeji et al. (2007) observed at the inhibition of *Clostridium beijerinckii* BA101, a butanol producer, on inhibitors over the range of 0.3 to 3.0 g/L [22]. Growth of *C. beijerinckii* BA101was stimulated by some compounds where others were highly toxic [22]. Rani et al. (1999) observed the ethanol tolerance of isolated strains of *C.*

thermocellum and the effect of different solvents and acetic acid of the growth of the two ethanol tolerant strains [23]. The growth of each strain was tested with different concentrations of ethanol at different temperatures. The optimal growth temperature when grown without ethanol remained the same. However, the optimal growth temperature was reduced for one strain in the presence of ethanol.

In general these studies found that the cell density decreased with increasing inhibitor concentrations. Furthermore, different compounds inhibited the different organisms to various degrees. The range of concentrations of inhibitory compounds used and the amount that each concentration inhibited the various bacteria is summarized in Table 17 which can be found in Appendix A1: Amount inhibited for concentration ranges of inhibitory compounds found in pretreated biomass for various microorganisms.

Increasing Industrial Robustness through genomic mutations

In order to develop a CBP scheme for biofuel production, more robust microorganisms are needed with higher rates of conversion and yield with ability to withstand inhibitory compounds present in the pretreated hydrolysate. Unfortunately, very little is known about how to design robustness into dense, highly interconnected and highly stochastic systems of the biological cell [24]. In general, organisms with broad substrate ranges and cellulolytic and/or hemicellulolytic abilities generally suffer from poor growth characteristics or poor yield, titer, and rate, or produce mixtures of desirable and undesirable products [6]. Organisms with desirable product-producing qualities often suffer from limited substrate range including lack of cellulolytic ability, poor fermentation qualities, and sensitivity to inhibitory compounds present in pretreated biomass [6].

However, microbes are well suited to manipulation for a number of reasons: fast generation times, repeatability, the ease of maintaining large population sizes, and the ability to store populations for later examination [25]. Mutation studies seek to observe microbes under controlled scenarios in which mutations are expected to occur. Mutations can occur by either direct or indirect manipulation of the genome. Strain engineering directly manipulates the genome by targeting specific genes in order to coerce a cell to change its phenotype. However, it is necessary to know which genes to target. In contrast, directed evolution applies a selection pressure to induce mutations in the genome [25]. Under stressful conditions, higher rates of mutations can be selected [26]. Individuals with high mutation rates can have a competitive advantage in changing environments [26].

The development of a single robust organism for the production of cellulosic biofuels has focused on one of two strategies. The recombinant cellulolytic strategy involves engineering non-cellulolytic organisms with high product yields so that they express a heterologus cellulase system to enable cellulose utilization [6]. This includes eukaryotic organisms such as *S. cerevisiae*, *Kluyveromyces marxianus*, *Pichia stipitis* and *Hyphomonas polymorpha*; and prokaryotic organisms such as *E. coli*, *Z. mobilis*, *Clostridium acetobutylicum*, lactic acid bacteria and corynebacteria [6]. The native cellulolytic strategy involves engineering naturally cellulolytic microorganisms to improve product and related properties, such as yield [6]. Cellulolytic organisms include several clostridia species including *C. thermocellum* [6]. Clostridia are Gram- positive ($Gram^{+}$) organisms that are strict anaerobes and superb producers of primary metabolites for applications in biofuels and chemical production [27].

Recent advances in DNA sequencing and high-throughput technologies have enabled empirical studies that directly characterize the molecular and genomic bases of evolution [25]. Genomic sequencing can determine the complete set of mutations responsible for an evolved phenotype. When

genome sequencing is applied in a longitudinal manner, in which the genome is sequenced at various time points, the rate of mutation appearance may increase in the later phase of the adaption due to the development of a mutator phenotype [25]. However, the genome sequence that is obtained represents a population average rather than the sequence of any individual bacteria [28]. Therefore, it is not possible to characterize a species from a single genome sequence [28]. The best approximation to describe a species could be made by using the concept of the pan-genome [28]. The pan-genome can be divided into three elements: a core genome that is shared by all strains; a set of dispensable genes that are shared by some but not all isolates; and a set of strain-specific genes that are unique to each isolate [28]. The pan-genome reflects the selective pressure to generate new adaptive combinations by recombining and constantly restructuring gene variants (alleles) in the population [28]. When the longitudinal analysis and pan-genomic analysis are combined a fourth element of the pan-genome arises, a set of discarded mutations that are in the population samples but not in the isolate samples.

Statistical analysis of mutations detected by genome sequencing can provide further insight into the genomic differences between strains and often reveals non-random patterns [25,26]. Mutation hotspots found in genomes are often associated with insertions and deletions (indel) [26]. Other types of mutations include single nucleotide substitutions (SNP) that can either be synonymous or non-synonymous and multiple substitutions [26]. Synonymous SNPs do not cause a change in the amino acid where as non-synonymous SNPs do change the amino acid and therefore, typically have a greater effect. Zhu et al (2009) found that the nucleotide diversity is negatively correlated with the distance from an indel, and that nucleotide substitution rates are significantly higher in lineages with indel events than in lineages without derived indels [26]. Therefore, a conserved region with low mutation rate can maintain a stable inheritance, whereas a less conserved region can evolve more quickly which is expected because indels can disrupt the reading frame in the coding sequence [26].Indels that are 3x bp-long

(e.g., 3, 6, 9 bp long, and so on) have greater stability in the sequence and stay in coding sequences for longer periods because they do not cause frame-shifts [26]. Frame-shifts can disrupt functionality over longer spans of the genome.

Genomic manipulation has been used to increase resistance to inhibitory compounds found in hydrolysates. Improvement of *S. cerevisiae* tolerance to lignocellulosic hydrolysates has been achieved by overexpressing homologus or heterologus genes encoding enzymes that confer resistance towards specific inhibitors [29]. An NADPH-dependant alcohol dehydrogenase (ADH6p) was identified as one of the enzymes responsible for HMF and furfural reduction in *S. cerevisiae.* Furthermore, gene deletion in the pentose phosphate pathway (PPP) genes exhibited growth deficiency in the presence of furfural indicating that *S. cerevisiae* tolerance to furfural was associated with the activity of PPP. Increased PPP level might help in protecting and repairing furfural induced damage [29]. HMF and furfural tolerant *S. cerevisiae* strains were also improved by directed evolution by increasing concentrations of the inhibitors in the media. The mutant strains showed enhanced ability to reduce the inhibitors, and metabolized glucose and grew faster than the control strain [29]. A furfural-tolerant mutant of ethanologenic *E. coli* LY180 was produced and shown to have also acquired tolerance to HMF [30]. The increased tolerance to the inhibitory compounds results from the silencing of two NADPH-dependent oxidoreductases which compete with biosynthesis for this limiting cofactor [30]. The growth of *Z. mobilis* is reduced in the presence of acetate, vanillin, furfural, or HMF; however, a *hfq* mutant strain of *Z. mobilis* had increased resistance to those inhibitors showing that the *hfq* gene plays a role in inhibitor tolerance [31] The design of a genetically engineered *S. cerevisiae* strain tolerant to phenolic derivatives has also been explored where a phenylacrylic acid decarboxylase was overexpressed allowing the strain to convert ferulic and cinnamic acid at a faster rate [29]. Laccase leads to the removal of low-molecular-mass phenolic compounds. When a recombinant *S. cerevisiae* carrying a laccase gene was designed, the

mutant strain was able to grow in medium containing coniferyl aldehyde at a concentration that completely inhibited the control strain [29]. Additional resistance to the inhibitory compounds may be achieved through detoxification. An important example would be the case where a strain has been endowed by selection or metabolic engineering to convert the impurities to non-toxic chemicals thereby removing the toxicity [1]. *Bacilli* can degrade an array of chemicals, including phenols and catechols, and possess catabolic pathways for degradation of aromatics [1]. The metabolism of furfural and HMF by *C. acetobutylicum* produces furfuryl alcohol and HMF alcohol, compounds which are less inhibitory to microorganisms [32]. Furthermore, the growth was stimulated with the addition of low concentrations of furfural and HMF [32]. Cofactors (NADH and NADPH) are involved in the detoxification of furans and it is possible that the furans may enhance glycolysis via regeneration of NAD^+ since NADH may be involved in the reduction of those furans to alcohol [32].

In *C. thermocellum*, the wild type strains produce ethanol as well as organic acids, but growth is inhibited by relatively low ethanol concentrations (<10g/L or <2% v/v) [33,34]. However, the mechanisms contributing to inhibition are many, complex, and still not completely understood but a major cause could be related to disturbance and disruption of membrane permeability [34]. Since ethanol tolerance is complex, both Brown et al. (2011) and Williams et al. (2007) adapted an ethanol-tolerant strain of *C. thermocellum* through sequential transfers with increasing concentrations of ethanol. Williams et al. (2007) adapted *C. thermocellum* to 5% v/v ethanol concentrations and Brown et al. (2011) adapted *C. thermocellum* to 80 g/L ethanol. However, both ethanol-tolerance strains of *C. thermocellum* had a lower growth rate and cell yield compared with the wild type strain in incubations not containing ethanol [33-35]. Furthermore, Williams et al. (2007) found that approximately 60% of proteins identified from purified membrane fractions were observed to be differently expressed in the wild type and ethanol-tolerant strains. A majority (73%) of differently expressed proteins were down-

regulated in the ethanol-adapted strain with a significant proportion involved with carbohydrate transport and metabolism [34]. Brown et al. (2011) resequenced the genomes of both the ethanol-tolerant and wild type strains of *C. thermocellum* and found that single nucleotide substitutions are the dominant type of mutation that occurs in the ethanol-tolerant strain and non-synonymous substitutions were approximately twice as abundant as synonymous substitutions [33]. Multiple substitutions, however, mainly targeted coding sequences of the genome, and indels were overrepresented in noncoding sequences [33]. Furthermore, 7 of the 16 mutation hotspots were in genes related to cellulose degradation, consistent with observed poor growth phenotype for the ethanol-tolerant mutant [33]. The ethanol-tolerant phenotype is primarily due to a mutated bifunctional acetaldehyde-CoA/ alcohol dehydrogenase gene (adhE) which contains two non-synonymous mutations resulting in predicted amino acid changes [33].

Stress response due to inhibitory compounds can be measured on a genetic level

The ability to investigate the functions of each individual gene and the highly complex regulatory networks inside the living is cell is now possible due to the availability of genetic data [36]. Significant clues to the mechanisms involved in adaptation to new environments, such as would be found in a CPB production scheme, have come from studies of gene expression in response to specific stresses [37]. The response of cells to environmental changes can provide clues to the molecular apparatuses that enable cells to adapt to new environments and the molecular mechanisms that have evolved to regulate the remodeling of gene expression that occurs in new environments [37].

Traditional methods for detecting gene expression changes include Northern and Southern blotting, and differential RNA display; however, these methods can only analyze one of a few genes at a time and offer a limited insight into the complex nature of cellular processes. DNA microarray data is a much more powerful tool in studying genome-wide gene expression, which will eliminate the bias associated with a preselected subset of genes thought to be involved in a certain cellular event [36]. Study of gene expression using DNA microarrays focuses primarily on identifying genes differentially expressed across different experimental conditions and recognizing and classifying common expression patterns by groups of genes. [38] Analysis of large microarray datasets requires application of normalization methods and arbitrary cut-off values that reflect 'significant' change in expression values [36]. Genomic approaches like DNA microarray analysis make it possible to identify novel cellular programs affected by stress environments [27]. However, there is very limited literature on microarray studies of Gram-positive prokaryotes and especially strict anaerobes, such as Clostridia [27].

Both Gasch et al. (2000) and Causton et al. (2001) examined the change in gene expression of *S. cerevisiae* in response to environmental changes such as heat shock, osmotic shock, hydrogen peroxide, amino acid starvation and other conditions using DNA microarrays [37,39]. The transcriptional responses demonstrated that a much larger fraction of the genome was involved in responses to environmental stress (~ 10%) than was previously known [37]. Gasch et al. (2000) reported gene expression changes in more than 14% of the genome. Clustering of the gene expression data revealed two groups of genes that tended to respond to the various stresses in a consistent manner: one group tended to be up regulated, while the other group tended to be down regulated [39].

The repressed genes could then be broken down further into two clusters. The first cluster of repressed genes consists of genes involved in growth-related processes, translation and protein

synthesis [37,39]. The transient translational arrest occurs in order to redirect resources towards synthesis of stress proteins. The second cluster of repressed genes, which had a delay in repression, consists almost entirely of genes encoding ribosomal proteins [39]. Causton et al. (2000) found 106 repressed genes that had unknown functions [37]. In general, the stress-inhibited genes can be categorized as general growth related genes.

Of the induced genes found by Gasch et al. (2000) approximately 60% of them were uncharacterized at the time and thought to be likely to provide clues as to how the cells survive in the stress environment [39]. The genes induced with known function include carbohydrate metabolism, cell stress, and the generation of energy [37]. Furthermore, the "cell stress" genes include heat shock genes and defenses against oxidizing agents [37]. The overall stress response can be characterized by a reallocation of cell resources away from growth and towards production of resources necessary to survive the stressful condition.

Stern et al. (2007) found that the global transcriptional response of *S. cerevisiae* is sensitive to the applied pressure of inhibitors and that the observed response is largely nonspecific with repeated experiments having a low reproducibility of their transcriptional states [40]. A large fraction of the responding genes, although allowed the cell to adapt to the stress environment, were nonspecific towards the challenge [40]. Genes that belonged to the same functional module, even those active within a linear chain of a biosynthesis pathway, might respond in opposite directions; some are highly induced whereas the others were highly repressed [40]. There was also a striking symmetry between the sets of genes that were induced and repressed which suggests that a global conservation principle, possibly competition for cellular resources, is involved in the dynamics of the genetic regulatory system [40].

Hirasawa et al. (2010) reviewed the DNA microarrays analysis for study of *S. cerevisiae* and other strains of brewing yeast [38]. In a comparison of the transcriptional responses of the laboratory strain and an industrial sake-brewing strain that was known as the stress-tolerant strain under high osmotic pressure, it was found that the overexpression of the osmotic stress genes were greater in the sake-brewing strain than in the laboratory strain [38]. It was also found that the genes required for growth under high ethanol concentrations are included in the genes whose expression is upregulated under normal growth conditions rather than high ethanol conditions [38]. This suggests that simple selection of upregulated or downregulated genes in a stress environment as genetically engineered candidates is not effective for improving the phenotype of the cells. However, the genes that were required for growth under high ethanol concentrations are among those whose expression patterns are different between the laboratory strain and sake-brewing yeast, suggesting that comparing the gene expression patterns of the strains having different phenotypes is an effective way to identify the important genes that are responsible in producing the intended phenotype [38].

In comparison to yeast very few studies of gene response have been conducted for cellulolytic bacteria. The closest related organism that has been studied is *B. subtilis*; however, it is not a strict anaerobe and is only distantly related to the Clostridia [41]. Heat- shock is the best-studied stress response in *B. subtilis*. The exposure of *B. subtilis* to heat shock activates the transcription of well over 100 genes [42]. Many of these genes are members of a general stress response regulon controlled by a single factor and were not known members of the heat shock regulon [42]. Additional investigations of the transcriptional response of *B. subtilis* during the early to middle stages of sporulation have been conducted (e.g. Fawcett et al. (2000)). The Spo0A protein is a master regulator of gene expression and induces the σ^F promoter in the sporulating cells of bacilli and clostridia [43]. Comparison of wild type, and Spo0A and σ^F deficient mutant strains of *B. subtilis* revealed three categories of gene expression:

genes whose expression was higher in the wild type than in the Spo0A mutant (283 genes), genes with lower expression in the wild type than in the Spo0A mutant (242 genes) and genes whose expression was higher in the wild type than in both the Spo0A mutant and the mutant σ^F (66 genes) [44]. This analysis was done in order to determine which genes were under the control of Spo0A but not σ^F and identify genes with unknown functions that may be involved in sporulations [44].

C. acetobutylicum and *C. beijerinckii* are among the prominent solventogenic species capable of acetone and butanol formation by fermentation [43]. None of the other closely or distantly related clostridia are understood genetically any better than *C. acetobutylicum* [41]. Tomas et al. (2003) compared the gene-expression patterns to specific phenotypes and investigated the gene expression differences between the wildtype, and Spo0A and σ^F mutant strain of *C. acetobutylicum* and also compared the similarities and differences from *B. subtilis* [41]. This analysis was done in an attempt to assign functions to groups of or individual *C. acetobutylicum* genes. Many of the genes known to be directly or indirectly influenced by Spo0A in *B. subtilis* are similarly regulated in the Spo0A mutant of *C. acetobutylicum*. Also, many similarities were observed between the two *C. acetobutylicum* mutants that have very similar phenotypes; however, several unique classes of genes were noted as being differently expressed in the σ^F mutant but not the Spo0A mutant [41]. Another study done by Tomas et al. (2006) investigated the transcriptional butanol stress tolerance in a *C. acetobutylicum* recombinant strain, which had increased butanol production and tolerance [45]. The recombinant bacteria were exposed to two concentrations of butanol while a plasmid control strain was exposed to a single high concentration of butanol. The stress-response and solvent-formation genes had elevated expression levels in all butanol-challenged cultures [45]. Tomas et al. (2004) concluded that genes with the same expression pattern for all three experiments are likely part of a general butanol stress response. Genes whose expression pattern are opposite for the recombinant and control bacteria are likely to represent

candidate genes that have a role in the increased tolerance of the recombinant bacteria to butanol [45]. Overexpression of the solvent formation genes in the butanol-challenged cultures was unexpected and counterintuitive; however, it was not associated with increased solvent production [45]. The increased expression of stress response genes, solvent formation genes, and many other genes appears to be butanol dose dependent as well, which suggest that *C. acetobutylicum* has a mechanism for sensing varying levels of butanol and altering gene expression accordingly [45].

Elucidations of system-level responses are likely to be revealed by examining gene expression responses over multiple stressful environments. By looking at multiple stressors there is an ability to classify genes and programs as general verses stress-specific responses [27]. To date only a few such comparative analyses have been reported in literature [27]. Alsaker et al. (2010) took on this task by examining the gene expression changes in *C. acetobutylicum* for three metabolite (butanol, butyrate, and acetate) stresses [27]. Stress by all three metabolites resulted in upregulation of genes related to post-translational modifications and chaperone activity, and down regulation of the translational machinery [27]. Alsaker et al. believe it is possible that the list of upregulated genes in response to the three stresses include several genes whose overexpression would likely convey tolerance and a valuable source for the development of tolerant recombinant strains [27].

Shi and Blaschek (2008) used DNA microarray analysis to examine gene expression in a parent strain and hyper-butanol-producing mutant strain of *C. beijerinckii* during the shift from acidogenesis to solventogensis [43]. The comparison analysis suggested differences in the regulation of primary metabolic genes in the two strains with lower induction of sporulation genes in the hyper-butanol-producing mutant strain resulting in lower levels of endospore formation than the parent strain [43].

The hyper-butanol-producing mutant strain also exhibited elevated expression of several primary metabolic genes and chemotaxis/motility genes [43].

Early studies on *C. thermocellum* focused on the expression levels of genes involved in the production of the extracellular multi-enzyme complex, termed the cellulosome, using traditional techniques such as Western blot analysis [46,47]. It wasn't until 2007, when Brown et al. developed a whole-genome microarray for *C. thermocellum* ATCC 27405 [35] that a more in depth look at gene expression levels in *C. thermocellum* under various growth conditions could be conducted. Raman et al. (2009) investigated the change in cellulosome composition of *C. thermocellum* during growth on a wide variety of growth substrates using DNA-microarrays in order to determine if *C. thermocellum* down regulates the expression of the energy-intensive cellulases in the presence of alternate readily metabolized substrates such as cellobiose [48]. The results suggest a coordinated substrate-specific regulation of cellulosomal subunit composition to better suit the organism's needs for growth under different conditions [35]. In 2011, Raman et al. again used microarray technology to investigate the temporal changes in gene expression associated with fermentation of crystalline cellulose by *C. thermocellum*. The results suggest that under low substrate availability, growth slows due to decreased metabolic potential and *C. thermocellum* alters its gene expression to coordinate its cellular strategy to increase its chances of survival in natural ecosystems [49]. Riederer et al. (2011) used microarrays to study the gene expression changes of *C. thermocellum* during chemostat growth on insoluble cellulose or soluble cellobiose and found 348 genes whose changes in expression patterns were growth rate and/or substrate dependent. The pattern of expression indicates that response to growth rate is the dominant global mechanism used for control of gene expression in *C. thermocellum* [50].

DNA microarrays have been the technology of choice for large-scale studies of gene expression levels. The arrays have the ability to simultaneously interrogate thousands of transcripts; however, array technologies have several limitations [51]. For example, background levels of hybridization limit the accuracy of expression measurements (particularly for transcripts in low abundance), probes differ in hybridization properties and arrays are limited to interrogating transcripts with relevant probes on the array [51]. More recently, sequence-based technologies provide more efficient and accurate measurements on a genome-wide scale [52] RNA Sequencing (RNA-seq) now rivals microarray-based gene expression for efficiency, cost, and accuracy and can be performed when microarrays cannot. Microarrays are simply not available for many non-model organisms whereas genome and sequence information are readily available for thousands of species [52]. Moreover, even when genomes are not available, RNA- seq can still be performed [53]. RNA-seq data gives unprecedented detail about transcriptional features that arrays cannot, such as novel transcribed regions, allele-specific expression, RNA editing and a comprehensive capability to capture alternative splicing [53].

Sequencing-based approaches to measure gene expression levels have the potential to overcome these limitations; however, as with any high-throughput technology, analysis methodology is critical to interpreting the data [53]. These technologies produce millions of short sequence reads and have to be mapped to unique positions on the reference genome [53]. Complications come from reads that can align to multiple locations or not matching perfectly with the reference [53]. The next task is to summarize the reads in a meaningful way such as over the whole length of a gene. The data has to be normalized to enable accurate comparison of expression between and within samples. A common method for normalization within a sample is to divide the summarized counts by the length of the gene [53]. Between-sample normalization is still essential for comparing counts from different libraries and can be done by adjusting by the total number of reads in the library [53]. As with microarray studies, the

goal of a differential expression analysis is to identify genes with significant changes in abundance across experimental conditions which means performing statistical testing between the normalized sequence libraries [53]. Although microarray intensities are typically log-transformed and analyzed as normally distributed random variables, there are statistical drawbacks [52]. When using log transformations with low or zero expression levels the calculation is numerically unstable and leads to extreme estimates of missing values [52]. Bergemann and Wilson (2011) found that by using a proportion statistic it is possible to detect differentially expressed genes that would be typically classified as missing data [52]. Furthermore, the Poisson distribution forms the basis for modeling RNA-seq count data over the normal distribution assumption in array data analysis [51,53].

Marioni et al. (2008) compared mRNA expression levels with microarray expression levels on the same samples while also assessing the variability across technical replicates for a single sample [51]. This study found that the variation across technical replicates of RNA-seq data can be captured using a Poisson model with only a small proportion of genes having clear deviation from this model and that a single technical replicate is comparable to that of a single array in enabling identification of differentially expressed genes [51]. Furthermore, the analysis suggested that the RNA-seq data allowed for better detection of low-expression genes with a large proportion of genes called differentially expressed from the RNA-seq data but not from the array being true positives [51].

In understanding tolerant phenotypes, comparison of genomic and transcriptomic characterizations of the wild type and mutant strains can be useful for inferring molecular scale mechanisms of increased tolerance [1]. While not every mutation or differentially expressed gene would be expected to provide tolerance upon engineered expression, much of the information may provide

useful clues as to the desirable targets for strain engineering. However, engineering transcription or sigma factors affecting these genes would be more likely to provide positive outcomes [1].

Transcription Factors

In bacteria, alterations in gene expression are often controlled at the transcriptional level through changes in association between the catalytic core of RNA polymerase and the different sigma factors present in a bacterial cell [54]. When bound together, the complex recognizes conserved DNA motifs called promoter regions (or sites) that precede gene sequences [54]. As the set of genes controlled by a single sigma factor can number in the hundreds, sigma factors provide effective mechanisms for simultaneously regulating large numbers of genes [54]. If cells contain multiple sigma proteins with unique promoter preferences, sigma factors substitution could be a potent vehicle for gene control [55].

Sigma factors are classified into two structurally unrelated families: σ^{54} and σ^{70} families [54]. The sigma factors below belong to the σ^{70} families. Sigma factor σ^{A} is the principle sigma factors present in vegetatively growing *B. subtilis* and other gram-positive bacteria [55] and is also referred to as the "housekeeping" sigma factor, as they direct transcription of genes important to metabolism [54]. σ^{A} is the most abundant sigma factor in purified RNAP from vegetatively growing *B. subtilis* [55].

Sigma factor σ^{B} was the first alternate sigma factor identified in bacteria [55] and has been assigned as the general stress responsive alternative sigma factor in Gram-positive bacteria [54]. The *B. subtilis* σ^{B} – dependent stress regulon contains over 200 genes that are expressed following bacterial

exposure to heat, acid, ethanol, salt stress, entry into stationary phase, or starvation for glucose, oxygen, or phosphate [54]. Numerous studies have shown that the activation of σ^B leads to an increased resistance to stress [56]. These studies placed *B. subtilis* and other bacteria in various stress environments in order to determine which genes are controlled by the σ^B factor. Examples of these stress environments include heat shock [42,56], ethanol stress [57], and salt stress [58]. Van Schaik et al (2007) found that genes from the σ^B regulon are generally conserved among four gram – positive bacteria to the same extent as all the genes in the genome [56]. However, even though σ^B – dependent genes from one organism are generally well conserved in other organisms, their σ^B dependency generally is not [56].

Other important sigma factors include σ^E, σ^F, σ^G, and σ^K which are the four-know sporulation sigma factors in *B. subtilis* [55]. In *B. subtilis*, σ^E is the mother cell-specific sigma factor and is also involved in the synthesis of σ^K, the late acting mother cell sigma factor [55]. Furthermore, σ^F – dependent transcription appears to be limited to the early expression of forespore-specific genes and σ^G appears to encode products that are synthesized within the forespore compartment during the later stages of sporulation to enhance spore survival and facilitate germination [55]. Fawcett et al. (2000) compared the gene expression of a wild type and σ^F deficient mutant of *B. subtilis* in early to middle sporulation to determine which genes were controlled by the sigma factor, σ^F [44]. In addition to the sporulation sigma factors, it has been proposed that the σ^D in *B. subtilis* controls flagellin production and possibly has a role in the expression of the methyl-accepting chemotaxis proteins [55]. The role of σ^H in *B. subtilis* transcription has not been fully defined. Although σ^H is essential for sporulation, it also is present during growth, when it may direct the transcription of a subset of vegetative genes [55].

Another important transcriptional unit is the anti-σ factors which negatively regulates the corresponding σ factor. The anti-σ factors have extracytoplasmic sensory domains and intracellular inhibitory domains [59]. For example, Kahel-Raifer et al. (2010) found that previously annotated 'hypothetical proteins' and 'membrane-associated proteins' in *C. thermocellum* have possible homology to membrane-associated anti-σ factors [59]. Further analysis of *C. thermocellum* genome found homologus genes in *B. subtilis* which encode for σ^{I} and its negative membrane-associated regulator RsgI [59,60]. These proteins were predicted to be involved in regulatory mechanisms that control carbohydrate degradation processes including formation and function of the cellulosome complex [60]. These observations suggest a novel carbohydrate-sensing mechanism in C. thermocellum, whereby the presence of polysaccharide biomass components are detected extracellularly and the signal is transmitted intracellularly, resulting in the activation of appropriate genes encoding proteins involved in cellulose utilization [59]. Bahari et al. (2011) showed how the anti-σ like RsgI glycoside hydrolases interacted with the crystalline cellulose but showed no hydrolytic activity [60]. The results of this study further suggest that the anti-σ like RsgI glycoside hydrolases function as polysaccharide binding agents rather than enzymatic components and thus serve as extracellular carbohydrate sensors [60].

Mathematical Modeling

Mathematical models that can accurately predict physical phenomena are essential since models provide the basis for the optimization of process systems [61]. The objective of the mathematical model is to obtain expressions that quantitatively describes the behavior of the process under consideration [61]. It is clear that good models should be valid and have explanatory and accurate predictive power [61]. However, there are no clear guidelines for what constitutes a good model [61]. The complexity of biological phenomena requires the use of nonlinear mathematical models to identify

growth parameters [62]. The basic growth phases can be characterized by the growth time course: lag phase, the exponential growth phase, the stationary phase and the exponential decay phase [62].

Various kinetic models have been proposed for the analysis of biochemical reactions. To describe a microbial process, two kinds of models can be developed, structured and unstructured. Structured models take into account some basic aspects of cell structure, their function and composition, but they can be complex [62]. In an unstructured model, only total cellular concentrations are considered; therefore, they do not involve any physiological characterization of the cells [62]. The unstructured model demonstrates the cell as an entity in the solution which interacts with the environment [63].

Basic bacterial growth of an unstructured model can be described as follows:

$$\frac{dX}{dt} = \mu * X$$

Where X is the concentration of biomass, μ is the specific growth rate and t is time [62]. This model can be expanded on in a number of ways. One such adaptation takes into account cell death in the growth kinetics model and considers an exponential decay for the decline phase:

$$\frac{dX}{dt} = \mu * X - k_d * X$$

Where k_d is the specific death rate [62].

The specific growth rate is typically modeled after the carbon substrate limiting Monod model:

$$\mu = \mu_{max} \frac{S}{S + k_s}$$

Where μ_{max} is the maximum specific growth rate, and S and k_s are the substrate concentration and substrate saturation constant, respectively [62,64,65]. The Haldane's equation added to the Monod model a term for inhibition by the carbon substrate:

$$\mu = \mu_{max} \frac{S}{S + k_s + {S^2}/{k_i}}$$

Where k_i is the substrate inhibition constant [62].

Various relationships have been proposed for ethanol product inhibition including: linear, exponential, quadratic, and hyperbolic functions. The most widely used relationships are linear and exponential [65]. Inhibition terms can also be modeled as either being competitive, uncompetitive, or noncompetitive [66]. In enzyme kinetics, competitive inhibitors are usually substrate analogs and compete with substrate for the active site of the enzyme [67]. Noncompetitive inhibitors are not substrate analogs. These inhibitors bind on sites other than the active site and reduce enzyme affinity to the substrate [67]. Uncompetitive inhibitors bind to the enzyme-substrate complex only and have no affinity for the enzyme itself [67]. A common form of the Monod equation that takes into account product inhibition uses a noncompetitive inhibition:

$$\mu = \mu_{max} \frac{S}{S + k_s} * \frac{k_p}{k_p + P}$$

Where P and k_p are the product concentration and product inhibition constant, respectively [61]. Depending on the system this specific growth rate equation can be expanded in a number of other ways to include other substrate and product inhibition terms, pH inhibitory effects, nitrogen limitations, liquid- to- gas mass transfer limitations, chemical equilibriums and others [62,68].

Product formation kinetics can also be modeled for a biological system. Product kinetics can either be growth or non-growth associated [62]. The Luedeking and Pirest growth associated model is the most widely used and the expression is:

$$q_p = \frac{dP}{Xdt} = A * \mu + B$$

In this relation, q_p, is the specific productivity rate and A and B are the growth and non-growth associated production constants [61,62]. Again this equation can be expanded to include other limitations depending on the system. The balanced systems of equations employed for a biological system can be solved by any number of global/local optimization methods.

These models can be applied to a myriad of biological systems. Gnanapragasam et al. (2011) used the Monod and Haldane models to model the color and chemical oxygen demand (COD) removal from a textile dye wastewater using methanogenic granular sludge [69]. Becker and Markl (2000) modeled olive oil degradation and oleic acid inhibition of *Bacillus thermoleovorans* IHI-91 by calculated biomass growth, substrate degradation, and lipid formation during chemostat and batch cultivation [64].

Hidaka et al. (2010) used the reaction rate and mass balance equations to determine the inhibitory effects of substrate, lactate (product) and, sodium chloride (NaCl) on bacterial growth in order to evaluate the lactate fermentation characteristics of *Bacillus coagulans* under non-sterile conditions [66]. Ljunggren et al. (2011) modeled batch fermentations of *caldicellulosiruptor saccharolyticus* on glucose using Monod kinetics and the Luedeking-Piret equation for growth and non-growth associated products to determine inhibition of growth and hydrogen production based osmolarity and hydrogen liquid-to-gas transfer [68]. Boyer et al. (1992) used a Monod model that took into effect product and furfural inhibition to determine the effects of furfural on ethanol production by *S. cerevisiae* in batch culture [65]. Huang and Wang (2010) used three different models: the S-system which uses a set of nonlinear ordinary differential equations, the rate-law equations modeled as a linear sum of nonlinear logarithmic concentration terms (lin-log model) and the Monod model to analyze the growth dynamics of *Sccharomyces disastaticus* LORRE 316 utilizing mixed-sugar to produce ethanol and glycerol [61]. Of the three systems, the S-system had the most predictive power based on comparisons to experimental data [61].

Alternative models to the classic Monod model include the logistic equation which has the advantage of direct biological meaning of its parameters:

$$\frac{dX}{dt} = \mu_m * \left(\frac{X_m - X}{X_m}\right) * X$$

Where X is the biomass (with X_m as asymptotic maximum), t is time and μ_m the maximum specific growth rate [70]. Rial et al. (2011) used the logistic equation coupled with a dose-response analysis to determine the effects of three heavy metals on the bacteria growth kinetics [70]. Nasrollahzadeh et al. (2010) compared a Monod and logistic model for the biodegradation of phenanthrene in anaerobic

batch reactors by a mixed consortia of microorganisms and found that the logistic model had better predictability of the system [63].

Structured models can take a closer look at the effects that inhibitors have on the cells. Desvaux et al. (2001) modeled the carbon flux distribution and kinetics of cellulose fermentation of *Clostridium cellulolyticum* on a defined media in order to optimize the percentage of carbon that was distributed towards acetate and ethanol production. The model took into account the ATP/ADP ratio, the total pool of ATP and ADP and the $NADH/NAD^+$ ratio [71]. Palmqvist et al. (1998) studied the furfural inhibition of *S. cerevisiae* in batch culture and found that furfural inhibition is a function of furfural concentration, cell density and aeration [72]. The model used nonlinear regression which took into account ATP/ADP recirculation, glucose limitation and furfural reduction [72]. Modig et al. (2002) went a step further and modeled the furfural and 5-HMF inhibition effects specifically on alcohol dehydrogenase, aldehyde dehydrogenase and pyruvate dehydrogenase to determine the mechanism of inhibition on *S. cerevisiae* [73]. Although, structured models give better predictability as to what is happening inside the cell, they can easily become complex and difficult to solve.

Proposed Research

An industrial robust microorganism capable of degrading lignocellulosic biomass in the presence of inhibitory compounds produced from the pretreatment process in a consolidated bioprocessing scheme is needed. One organism of interest is *Clostridium thermocellum* because it has the ability to break down the biomass sugars and convert them into fermentation end products; however, *C. thermocellum* lacks industrial robustness and suffers from poor growth, is inhibited by the inhibitory

pretreatment compounds and has a low yield and titer of fermentation end products. Improvements in the industrial robustness of the organism used will reduce the cost of ethanol based biofuels making them economically feasible.

Current research suggests that comparing the change in phenotype between a wildtype strain and a mutant strain of bacteria can give the greatest insight into the genes which are essential to the change in phenotype between the two strains. One desired phenotype with regard to *C. thermocellum* would be an increased resistance to the inhibitory compounds found in pretreated *Populus* hydrolysate. Genomic sequencing can determine the complete set of mutations. Some of these mutations may be responsible for the evolved phenotype. Genomic sequencing can be applied at different time points in a mutating population in a longitudinal manner in order to correlate certain mutations with the onset of particular phenotypes. Alternatively, sequencing of multiple isolates from a single evolved population may be applied in order to find the core genes shared by all isolates, genes shared by some but not all isolates and a set of isolate-specific genes. Understanding of the transcriptional stress response of the microorganism on a genetic level will also provide insight to the increased robustness. Mutations may result in gene expression levels that are tuned to increase the probability of survival when placed in a stress environment such as the presence of inhibitory compounds found in the pretreated biomass. Sequence – based approaches such as RNA-seq can also be used to measure gene expression changes in order to determine how the gene mutations affect the highly complex regulatory networks inside the living is cell. Kinetic based modeling of a wild type and mutant strain is also helpful in quantifying the extent of the tolerance to the inhibitory compounds. The increased industrial robustness of *C. thermocellum* on *Populus* hydrolysate has been the focus of this work and has been performed under the following hypotheses:

Hypothesis #1: Phenotypic changes in *C. thermocellum* driven by directed evolution on increasing concentrations of *Populus* hydrolysate will result in a mutant strain with improved industrial robustness through increased tolerance to the inhibitory compounds from the pretreatment process when compared to the wild type strain of *C. thermocellum*

In order to investigate the causes of increased robustness in *C. thermocellum*, first it is necessary to obtain a mutant strain of the organism with a desired phenotype. One of the model biomass feedstocks being considered for consolidated bioprocessing is a hardwood tree, *Populus*. A *Populus* hydrolysate-tolerant mutant strain of *C. thermocellum* with increased resistance to the inhibitory compounds would be one such desired phenotype. The mutant can be developed by conducting daily serial dilutions in liquid cultures with a low concentration of *Populus* hydrolysate from a wildtype strain of *C. thermocellum* ACCT 27405. The *Populus* hydrolysate will be from a dilute acid pretreatment process at elevated temperatures since it is the most common process used [10]. Once the mutant had the ability to grow as well as the wild type bacteria in a certain concentration of *Populus* hydrolysate, the concentration of *Populus* hydrolysate will be increased and the process repeated until the mutant has the ability to grow in a sufficiently high concentration of *Populus* hydrolysate. The target concentration is 17.5% v/v *Populus* hydrolysate. A stock culture of the mutant strain will then obtained from a single cell isolate that grows similarly to the population average.

Since there is more variety in the types of compounds that the *Populus* hydrolysate-tolerant mutant strain of *C. thermocellum* will be adapted to, it is possible that mutations in more than one gene will be responsible for the increased tolerance compared to previously reported increases in tolerance to a single inhibitor coming from a mutation in a single gene (e.g., Brown et al. [31]). In order to test the mechanism of tolerance, the genomic sequencing of the wild type and *Populus* hydrolysate-tolerant

mutant strain will be conducted. The genomic sequence of five intermediate *Populus* hydrolysate-tolerant mutant strains of *C. thermocellum* will be obtained in order to determine the sequence of mutations over time. Since these sequences have not come from a single cell isolate and therefore represent a population average, they will be sequenced to a depth of 150x coverage. The genomic sequence of seven single cell isolates from the final *Populus* hydrolysate-tolerant mutant *C. thermocellum* culture will be used to determine the core genes shared by all of the isolates, the genes shared by some of the isolates, gene changes to unique isolates and genes in the population samples but discarded from the isolates. The genomic sequence of the frozen stock wild type *C. thermocellum* culture will be used as baseline for mutation comparisons. The single cell isolates of the final *Populus* hydrolysate-tolerant mutant *C. thermocellum* culture and wild type stock culture of *C. thermocellum* will be sequences to a 15x coverage since there should be less variability in the genome.

A draft genome sequence can be generated for each of the mutants using a combination of shotgun and paired-end library reads with the Genome Sequencer FLX System. Differences from the wild type strain can then be detected using the GSMapper software. A statistical analysis of the detected mutations by pyrosequencing will then be conducting in order to give further insights into the genomic differences in the wild type and *Populus* hydrolysate-tolerant mutant strain by revealing non-random mutation patterns. Analysis will include detection of SNPs, insertions and deletions with confidence level position in genome, coding and amino acid information. Further computational analysis of the mutations and literature searches will be conducted in order to determine the effect of the mutations on the tolerant phenotype. There are two possible mechanisms responsible for the tolerant phenotype, which is discussed below.

Hypothesis #2: Increased tolerance to *Populus* hydrolysate by the *Populus* hydrolysate-tolerant mutant strain of *C. thermocellum* is due to mutations of specific genes capable of detoxifying specific compounds in the hydrolysate.

The first possible mechanism responsible for the increased tolerance could be caused by changes to specific genes capable of detoxifying specific compounds in the *Populus* hydrolysate. *C. thermocellum* has the ability to degrade four of the known inhibitors in the *Populus* hydrolysate: furfural, 5-hydroxymethylfurfural, syringic acid and vanillin. Further evidence of the mechanism can be shown by comparing the degradation rate of furfural and 5-hydroxymethylfurfural in the wild type and *Populus* hydrolysate-tolerant mutant strain of *C. thermocellum*. The extent of degradation is comparable between the two strains; however, the rate of degradation is faster in the *Populus* hydrolysate-tolerant mutant strain of *C. thermocellum*. This increased rate of inhibitor degradation in the *Populus* hydrolysate-tolerant mutant strain of *C. thermocellum* may be due to the increased growth as described in Hypothesis #3.

Further studies comparing the growth of the both the wild type and *Populus* hydrolysate-tolerant mutant strain of *C. thermocellum* on detoxified *Populus* hydrolysate may determine if this increased rate of degradation confers the increased tolerance or if the *Populus* hydrolysate-tolerant mutant strain of *C. thermocellum* has obtained tolerance to an inhibitory compound that it cannot degrade. Identifying mutations to specific detoxification genes in the *Populus* hydrolysate-tolerant mutant strain of *C. thermocellum* from resequencing project will give additional evidence to this hypothesis. The mutations would make the enzymes responsible for the detoxification of these compounds more efficient. If there are significant mutations causing specific detoxifying enzymes to be more efficient, it is likely that there will be a lower expression of these enzymes in the *Populus*

hydrolysate-tolerant mutant strain of *C. thermocellum* allowing the strain to devote more energy into cell growth.

Hypothesis # 3: Increased tolerance to Populus hydrolysate by the *Populus* hydrolysate-tolerant mutant strain of *C. thermocellum* is due to mutations that modify the overall metabolism thereby increasing the growth rate of the mutant strain.

The second mechanism of increased tolerance could be related to a change in the overall metabolism resulting in an increase in the growth rate of the *Populus* hydrolysate-tolerant mutant strain of *C. thermocellum* compared to the wild type. A faster growth rate may allow the mutant to detoxify the hydrolysate faster due more active cell mass. This hypothesis comes from the increased growth of the *Populus* hydrolysate-tolerant mutant strain of *C. thermocellum* when grown in standard media compared to the wild type strain. Additional evidence for this mutation may come from genetic mutations scattered over the genome. Wide spread changes in gene expression in the *Populus* hydrolysate-tolerant mutant strain of *C. thermocellum* compared to the wild type strain in standard media (0% v/v *Populus* hydrolysate) could be further evidence of a an overall change in metabolic profile resulting in the observed increase in growth rate and modification of the metabolic products. This hypothesis can be further developed and tested based on the specific mutations and changes in gene expression observed from the genomic and RNA sequencing projects.

Hypothesis #2 and #3 will be evaluated in two ways. Both the wild type and *Populus* hydrolysate-tolerant strain of *C. thermocellum* will be grown in 0%, 10% and 17.5% v/v *Populus* hydrolysate. First, the increased industrial robustness of the *Populus* hydrolysate-tolerant mutant strain of *C. thermocellum* will be verified by comparing various metrics of the wild type strain and one single-

cell isolate of the *Populus* hydrolysate-tolerant mutant strain of *C. thermocellum* with the highest cell density. These metrics include growth rate, maximum cell density, sugar utilization rate, and fermentation end product yield and titer using mathematical modeling based on the Monod equation. Second, two samples will be taken during the fermentation to determine the temporal changes in gene expression as a function of hydrolysate concentration and *C. thermocellum* strain. The high-throughput technology of RNA-seq will then be used to determine the gene expression levels. The gene expression level will be normalized and a statistical analysis will be applied to identify genes that are differentially expressed between the wild type and *Populus* hydrolysate-tolerant mutant strain of *C. thermocellum.* This analysis should provide clues to the structural changes in enzymes that enable cells to adapt to new environments and/or the molecular mechanisms that have evolved to regulate the remodeling of gene expression that occurs in new environments. Of particular interest will be the change in gene expression for the genes identified as having been mutated in the *Populus* hydrolysate-tolerant mutant strain of *C. thermocellum.* Furthermore, organisms typically decrease their expression of growth-related genes and increase their expression stress response genes when placed in a stress environment [37,39]. Therefore, identifying differentially expressed genes between *Populus* hydrolysate concentrations may lead to insight of genes with unknown function.

Chapter 1 Industrial Robustness: Understanding the mechanism of tolerance for the Populus hydrolysate tolerant strain of Clostridium thermocellum

This chapter is revised based on the partial revised manuscript submitted to publication by Linville et al. The initial manuscript was accepted with major revisions. The revisions will be completed and submitted to Plos One in April, 2013.

Jessica L. Linville, Miguel Rodriguez, Jr., Miriam Land, Mustafa H. Syed, Nancy L. Engle, Timothy J. Tschaplinski, Jonathan R. Mielenz, Chris D. Cox. Industrial Robustness: Understanding the mechanism of tolerance for the Populus hydrolysate tolerant strain of Clostridium thermocellum. Submitted to Plos One (February 2013).

My primary contributions to this paper include: (i) design and conduction of experiments, (ii) interpretation and analysis of all experimental data, (iii) literature searches and (iv) majority of the writing of this paper.

1.1 Abstract

An industrially robust microorganism that can efficiently degrade and convert lignocellulosic biomass into ethanol and next-generation fuels is required to economically produce future sustainable liquid transportation fuels. The anaerobic, thermophilic, cellulolytic bacterium *Clostridium thermocellum* is a candidate microorganism for such conversions but it, like many bacteria, is sensitive to potential toxic inhibitors developed in the liquid hydrolysate produced during biomass processing. Microbial processes leading to tolerance of these inhibitory compounds found in the pretreated biomass hydrolysate are likely complex and involve multiple genes.

In this study, we developed a 17.5% v/v *Populus* hydrolysate tolerant mutant strain of *C. thermocellum* by directed evolution. The genome of the wild type strain, six intermediate population samples and seven single colony isolates were sequenced to elucidate the mechanism of tolerance. Analysis of the 224 putative mutations revealed 73 high confidence mutations. A longitudinal analysis of the intermediate population samples, a pan-genomic analysis of the isolates, and a hotspot analysis revealed 24 core genes common to all seven isolates and 8 hotspots. Genetic mutations were matched with the observed phenotype through comparison of RNA expression levels during fermentation by the wild type strain and mutant isolate #6 in various concentrations of *Populus* hydrolysate (0%, 10%, and 17.5% v/v).

The findings suggest that there are multiple mutations responsible for the *Populus* hydrolysate tolerant phenotype resulting in several simultaneous mechanisms of action, including increases in cellular repair, more efficient molecular pumps, and altered energy metabolism. To date, this study provides the most comprehensive elucidation of the mechanism of tolerance to a pretreated biomass hydrolysate by *C. thermocellum*. These findings make important contributions to the development of industrially robust strains of consolidated bioprocessing microorganisms.

1.2 Introduction

Organic fuels, chemicals, and materials developed from plant biomass are among the leading options to meet sustainability requirements of the future [33,74]. Cellulosic biomass is an attractive feed stock for production of renewable fuels because of its widespread availability and relatively low cost [74]. The conversion process of cellulosic biomass to fuels utilizes biological fermentations. The most economic process combines the cellulase production, biomass hydrolysis, and sugar fermentation into a

single step by a cellulose-fermenting microorganism in a consolidated bioprocessing (CBP) scheme. CBP eliminates the cost of exogenous cellulase addition and consolidates capital equipment [74].

The plant cell walls of cellulosic biomass consist of several intertwined heterogeneous polymers, primarily cellulose, hemicelluloses (e.g., substituted xylans, mannans, ect.), pectin and lignin [60]. Various pretreatment methods are designed to make the sugars more available for subsequent hydrolysis and fermentation steps through the breakdown of the cell wall and the degradation of cellulose, hemicelluloses, and lignin matrix [7]. Most pretreatment processes produce compounds that are inhibitory to the organism used for fermentation. These inhibitory compounds come from the partial degradation of biomass components [11] and include carboxylic acids (primarily acetic acid), furfural, 5-hydroxymethylfurfural (HMF), phenolic compounds, and inorganic salts [7]. The furan derivatives HMF and furfural are formed at high temperature and pressure as degradation products of hexoses and pentoses, respectively [1]. The phenolic compounds originate from the degradation of lignin [1]. Different biomass feedstocks and pretreatment methods produce different combinations of inhibitory compounds that stress the microorganisms [7]. Acetic acid, furfural and HMF are the most studied inhibitory compounds [31].

A variety of cellulolytic microorganisms are under development for biofuel production, but these organisms are typically poorly characterized [33,75]. An exception is *Clostridium thermocellum* which has been the subject of significant research for decades. *C. thermocellum* is a Gram–positive, anaerobic, thermophilic, cellulolytic bacterium that can rapidly solubilize biomass and use cellulose as a carbon and energy source [33,50,76]. High efficiency cellulose hydrolysis is aided by the cell surface attached multi-enzyme protein complex termed the cellulosome [48,50]. *C. thermocellum*'s cellulolytic ability gives it an advantage over organisms that are currently used for bioethanol production (e.g.,

yeast and *Zymomonas*), which can only ferment nonpolymeric carbohydrates, and has the potential to be a model organism for CBP [34,75]. *C. thermocellum* produces a number of industrially important fermentation products in addition to ethanol, including acetic acid, formic acid, lactic acid, and hydrogen[75]. Inhibitory compounds have been shown to reduce the rate of ethanol production and the overall yield [2,7]. Improved tolerance to inhibitory compounds found in pretreated biomass hydrolysate should improve the fermentation process and increase economic feasibility of consolidated bioprocessing.

C. thermocellum 27405 is among the rapidly growing number of microorganisms whose genome has been sequenced and annotated, and serves as a baseline for comparison to strains improved through engineering or evolution. Thus genomic sequencing can assist in determine the complete set of mutations responsible for an evolved phenotype. When genome sequencing is applied in a longitudinal manner, in which the genome is sequenced at various time points during its evolution, the mutation rate can be determined. Studies have shown that the mutation rate increases in the later phase of the adaption due to the development of a mutator phenotype [25]. However, the genome sequence that is obtained represents a population average rather than the sequence of any individual bacteria [28]. Therefore, it is not possible to characterize a species from a single genome sequence [28]. The best approximation to describe a species uses the concept of the pan-genome from multiple single cell isolates [28]. The pan-genome can be divided into three elements: a core genome that is shared by all single cell isolates; a set of shared mutations that are shared by some but not all single cell isolates; and a set of isolate-specific mutations that are unique to each single cell isolate [28]. When the pan-genome is combined with population samples, a fourth category of discarded mutations can be identified which occur in the population samples but not in the single cell isolates. The pan-genome reflects the selective

pressure to generate new adaptive combinations by recombining and constantly restructuring gene variants (alleles) in the population [28].

RNA Sequencing (RNA-seq) is an emerging technology that is being used for expression studies and it offers several advantages over DNA microarrays such as better detection of genes expressed at low levels [51]. RNA-seq analysis is particularly relevant for controlled experiments comparing expression in wild type and mutant strains of the same microorganism [53].

To date, the majority of genetic regulation studies for *C. thermocellum* have focused on the cellulosome [46,47,59,60,76-78] or ethanol tolerance [23,33,34,74]. Only a few studies have looked at gene regulation of *C. thermocellum* on a global level [49,50,75]. At this time, there is no known research that investigates the effects of hydrolysates from pretreated biomass on *C. thermocellum* gene expression. Current research suggests that comparing the change in phenotype between a wild type strain and a mutant strain of bacteria can give the greatest insight into the genes which are essential to the change in phenotype between the two strains. Therefore, in this study, we created a mutant population of *C. thermocellum* tolerant to 17.5% v/v *Populus* hydrolysate from which we isolated seven single cell colonies. We have sought to understand the mechanism of tolerance by (1) sequencing the genomes of the wild type strain (WT), intermediate population samples, and final isolate samples; (2) conducting comparative fermentative growth studies with the wild type and a final mutant isolate stain; and (3) comparing gene expression between the two strains during fermentative growth studies.

1.3 Material and Methods

1.3.1 Strain and Culture Conditions

Clostridium thermocellum ATCC 27405 was obtained from Prof. Herb Strobel, University of Kentucky collection and denoted as the wild type strain (WT). In all experiments the cells were grown in media for thermophilic clostridia (MTC) with a substrate loading of 5 g/L cellobiose. The medium was composed of 0.336 g/L potassium chloride [KCl], 0.25 g/L ammonium chloride [NH_4Cl], 1.00 g/L magnesium sulfate heptahydrate [$MgSO_4*7H_2O$], 1.70 g/L potassium phosphate [KH_2PO_4], 0.50 g/L MOPS [$C_7H_{14}NO_4S$], 0.15 g/L calcium chloride dehydrate [$CaCl_2*2H_2O$], 1.75 g/L Trisodium citrate dehydrate [$Na_3C_6O_7*2H_2O$], 0.6 g/L Urea [CH_4N_2O], 1.00 g/L L-cysteine HCL, 0.30 mg/L resazurin, 2.0 mL of 1000x MTC minerals and 1.25 mL of 50x MTC vitamins [79,80]. All chemicals were reagent grade and obtained from Sigma (St. Louis, MO), unless otherwise indicated. The cells were grown in 25 mL Balch tubes (Bellco Glass, Inc., Vineland, NJ) and were incubated in the dark in an orbital shaker at 58 ^oC and 100 rpm.

1.3.2 *Populus* hydrolysate preparation and analysis

Milled *Populus trichocarpa* hydrolysate was pretreated at the National Renewable Energy Lab (NREL) using a 20% w/w solid loading and dilute concentrations of H_2SO_4 at temperatures of 165-195 ^oC [81]. Solids were removed by filtration. The *Populus* hydrolysate was adjusted to a pH of 7.0 using 50% w/w NaOH and filter-sterilized before being added to the MTC medium at various concentrations. HPLC analysis used a LaChrom Elite system (Hitachi High Technologies America, Inc. Pleasanton, CA) equipped

with a refractive index detector (Model L-2490) and UV lamp (Model L-242OU). Metabolites were separated at a flow rate of 0.5 mL/min in 5 mM H_2SO_4 using an Aminex HPX-87H column (Bio-Rad Laboratories, Inc, Hercules, CA). GC-MS analysis used sorbitol (10 µl of a 1 mg/mL aqueous solution) added to 10 µl of sample as an internal standard to correct for differences in derivatization efficiency and changes in sample volume during heating. The sample was dried in a nitrogen stream and then dissolved in 500 µL of silylation–grade acetonitrile, followed by the addition of 500 mL N-methyl-N-trimethylsilyltrifluoroacetamide (MSTFA) with 1% trimethylchlorosilane (TMCS) (Thermo Scientific, Bellefonte, PA), and then heated for 1 h at 70 °C to generate trimethylsilyl (TMS) derivatives [82]. After 2days, 0.5 to 1-mL aliquots were injected into an Agilent Technologies Inc. (Santa Clara, CA) 5975C inert XL gas chromatograph-mass spectrometer (GCMS), fitted with an Rtx-5MS with Integra-guard (5% diphenyl/95% dimethyl polysiloxane) 30 m x 250 µm x 0.25 µm film thickness capillary column. The standard quadrupole GC-MS was operated in the electron impact (70 eV) ionization mode; gas (helium) flow is set at 1.2 mL per minute with the injection port configured in the splitless mode. The injection port, MS Source, and MS Quad temperatures are set to 250 ºC, 230 ºC, and 150 ºC, respectively. The initial oven temperature is held at 50 ºC for 2 min and is programmed to increase at 20 ºC per min to 325 ºC and held for another 11 min, before cycling back to the initial conditions.

1.3.3 Adaptation of *C. thermocellum* to *Populus* hydrolysate

Adaption of *C. thermocellum* was performed by serial transfers using crimp-sealed 25 mL Balch tubes containing fresh anaerobic medium at 24-hour intervals with increasing concentration of *Populus* hydrolysate [83]. The tubes were sealed empty and purged with N_2 and sterilized by autoclaving at 121 ^{o}C. The tubes were then injected with 9.5mL MTC medium containing cellobiose and a known concentration of *Populus* hydrolysate. Each inoculation/transfer was 5% volume (0.5 mL). The optical

density (OD_{600}) was measured directly (without sampling) using a Spectronic 21D (Milton Roy, Ivyland, PA). The OD was referenced to an uninoculated sample of the *Populus* hydrolysate medium incubated with the samples. The OD was used to determine the increase in cell growth over time. A given concentration of *Populus* hydrolysate was maintained for each transfer until the OD reached that of the WT strain at 24 hours. When the benchmark OD was obtained at 24 hours, the culture was transferred to a higher *Populus* hydrolysate concentration. The final adaptation culture after 117 transfers is tolerant to 17.5% v/v *Populus* hydrolysate. Various viable culture samples were stabilized with 33% glycerol and stored at -80 ^{o}C during the course of the mutation process.[74].

1.3.4 Isolation of single colonies

Agar (Fisher Scientific, Pittsburg, PA, USA) solution (15 g/L) was prepared with 5g/L cellobiose MTC medium. The agar-containing bottles were then transferred into an anaerobic chamber (Coy Laboratory Products, Grass Lake, MI, USA) where the following manipulation occurred. The final adaption culture was plated with serial dilutions of (100 µL, 10 µL, 1 µL, and 0.1 µL) on quad- Petri plates (BD Biosciences, Bedford, MA, USA) and mixed with the agar solution. The plates were allowed to sit for 30 min to solidify the agar and then incubated at 55 ^{o}C. Colonies were picked using a needle after 32-48 hours incubation. Seven petri plates were filled with the agar solution and let dry in the anaerobic chamber for 2-3 days. A picked colony was streaked onto the pre-poured plates and incubated for an additional 32-48 hours. A picked colony was then transferred into a microcentrifuge tube with 1 mL sterilized DI water, which was mixed and injected into a crimp-sealed Balch tube with 9.5 mL of cellobiose MTC medium. The tube was incubated at 58 ^{o}C and 100 rpm. After about 24 hours, stock culture was prepared with 33% glycerol and stored at -80 ^{o}C [74]. These cultures were named isolates 1-7.

1.3.5 Genome resequencing and analysis

Genomic DNA for the WT strain, six intermediate mutant populations, and seven 17.5% v/v *Populus* hydrolysate tolerant mutant isolates were extracted using a Qiagen DNeasy kit (Qiagen, Valencia, CA). The DNA samples were shipped in dry ice to the DOE Joint Genome Institute (JGI, Walnut Creek, CA). The samples were sequenced using JGI's whole-genome shotgun sequencing method, which began with the creation of DNA libraries [84]. Sequencing was performed from both sides of the library insert, producing paired ends, typically resulting in approximately 10X depth coverage for the single colony isolates and 50X depth coverage for the intermediate populations. Sequenced reads were aligned using MAQ [85]. A report with Single Nucleotide Polymorphisms (SNPs) and statistical analysis was returned. Further analysis was conducted on the mutations using the Genome Resequencing Toolkit available from BESC Knowledgebase [86]. The "Function change predictor" tool available from the toolkit was used to evaluate potential changes in the protein function, namely a loss or a gain of a protein family (Pfam) domain. This tool calculates the Pfam score of the non-synonymous amino acid substitution within a Pfam domain. The Pfam score is calculated from the Position Specific Scoring Matrix (PSSM) scores of the Pfam model, taking into consideration the position of the substitution and the amino acid in the mutant and in the reference sequence. The resulting Pfam score of the mutation can be positive, which indicates a potential gain of function, or negative, which points to a potential loss of the protein function. Detailed documentation of these tools is available at http://cricket.ornl.gov/html/download/resequencing/ResequencingToolkitDocumentation_16Dec2011.pdf [86-88]. Multiple sequence alignment was conducted on the top 100 Blastp hits using Clustalw. Amino acids were termed 'very conserved' upon visual inspection showing 100% agreement or 'non-conserved' if there was no conservation at that position. Genes with unknown function or hypothetical proteins where analyzed with BLAST to identify homologus genes with known functions.

1.3.6 Fermentation

Batch fermentations were conducted in triplicate in 1.5 L Q-plus jacketed glass fermentors (Sartorius Stedim Biotech, Bohemia, NY) using a 1 L working volume of MTC medium at 58^{o}C and 300 rpm, with pH controlled to 7.0 using 3N NaOH. Fermentors containing only 5 g/L cellobiose were sparged with filtered 20% CO_2/80% N_2 gas mixture and vigorously agitated overnight, followed by addition of the remaining medium components and *Populus* hydrolysate. *Populus* hydrolysate was added to the fermentors at 0%, 10%, or 17.5% v/v concentration. The fermentors were then sparged for an additional 4 hours with a 20% CO_2/80% N_2 gas mixture. The inoculum was grown overnight in 125 mL serum bottles (11-13 hrs) in 5g/L cellobiose, diluted to an OD_{600} of 0.200 $\pm$ 0.013, and added as 10% v/v to inoculate the fermentors. Isolate #6 was chosen as the *Populus* mutant (PM) strain.

1.3.7 Cell Growth and metabolite analysis

Well-mixed 5 mL aliquots of culture were harvested at regular intervals. Cell growth was monitored based on an increase in the optical density. The OD_{600} of the culture was measured in triplicate by a Spectramax Plus 384 (Molecular Devices, Sunnyvale, CA). Metabolite analysis was performed using HPLC as previously described.

1.3.8 RNA isolation

Fermentation samples for RNA isolations were harvested by centrifugation of an ~30 mL culture in 50 mL conical tubes at 8000 rpm and 4 ^{o}C for 10-15 min. The solid pellet fraction containing cells was flash frozen in liquid nitrogen and stored at -80 ^{o}C until further use. Total RNA was extracted from the

cell pellets as follows. Briefly, the frozen cell solution was thawed on ice and an equal volume of TRIzol (Invitrogen, Carlsbad, CA) was added. The cell solution was divided into ~1 mL aliquots and added to 2 mL tubes containing 1 mL of 0.1 mm glass beads (Biospec Products, Bartesville, OK) ashed at 250 ^{o}C overnight. Cells were lysed by rapid agitation of the tubes at 6500 rpm for 1 min in three 20s-ON/20s-OFF cycles using the Precellys® bead beater (Bertin Technologies, France). Subsequently, the cell lysate (~0.8 mL) in TRIzol was phase separated by addition of 200 µL chloroform and the RNA was precipitated by addition of 500 µL isopropanol. The precipitated RNA pellet was washed with 1 mL of 75% ethanol and resuspended in 100 µL of RNase-free water. Any contaminating DNA was digested by in-solution DNase-I (Qiagen, Valencia, CA) treatment and the RNA sample was purified using the RNeasy mini kit (Qiagen, Valencia, CA) as per the manufacturer's instructions.

Methods for fermentation experiments through RNA isolation techniques were adapted from Raman et. al (2011) [49].

1.3.9 RNA sequencing

The RNA samples were shipped in dry ice to DOE Joint Genome Institute (JGI, Walnut Creek, CA). JGI prepared an indexed, directional, amplified Illumina RNA_seq library for each sample using the RiboZero kit for rRNA removal [84]. Triplicate fermentation samples were pooled into a single channel and sequenced using 10 2x100 channels. The reads were mapped to the reference transcriptome. A report from JGI was returned with a correlation analysis between replicate samples and contained counts (raw and normalized to the gene length and total number of counts for the sample) for known gene models. Further analysis of the gene expression level was conducted using JMP Genomics 10. The raw count data was log-2 transformed and normalized by Upper Quartile Scaling [89]. An analysis of variance (ANOVA) test was conducted using a false discovery rate of 5% and the WT in 0% hydrolysate

as the baseline. A gene was considered statistically significantly differentially expressed if the log-2 difference was greater than 1 and the –log10(p) was greater than 1.714. The differentially expressed genes were then clustered using hierarchal clustering.

1.4 Results

1.4.1 *Populus* hydrolysate analysis

The *Populus* hydrolysate used in these experiments was analyzed by HPLC and GCMS. Table 3 shows the concentration of metabolites found in the hydrolysate. Analysis of the effects of various feedstocks and pretreatment combinations on the formation and the accumulation of potentially inhibitory degradation products in biomass hydrolysate have been investigated [11]. Additional literature discusses inhibition by products formed during biomass pretreatment [3,11-13,15,16]. Many of the non-carbohydrate compounds are believed to be inhibitory in nature [10,15,18,19,22].

2.4.2 Mutant development

The mutant development process can be seen in Figure 1, with the mutant development steps listed in Figure 1A. In order to determine the inhibitory effect of the *Populus* hydrolysate, the wild type strain (WT) of *C. thermocellum* was cultured with low concentrations of *Populus* hydrolysate (0.05% - 10% v/v *Populus* hydrolysate) in Figure 1B. During these early selection experiments the medium was prepared with only 1/3 of the vitamin solution compared to later experiments which explains the decreased growth of the WT in Figure 1B. Late in the fermentation (approximately 200 hours) the

Table 3: Results of Populus hydrolysate metabolite analysis by HPLC and GC-MS.

Sugar	**g/L**	**Aliphatic acids**	**mg/L**
Cellobiose	0.68	lactic acid	137
Glucose	22.74	maleic acid	11
Xylose	42.66	succinic acid	76
Arabinose	1.84	fumaric acid	74
Mannose	6.34	levulinic acid	1082
Aromatic acids	**mg/L**	itaconic acid	7
2-furoic acid	64	adipic acid	29
3,4-dihydroxybenzoic acid	55	acetic acid	13971
salicylic acid	3	arabinonic acid	709
syringic acid	127	**Aldehydes and Ketones**	**mg/L**
4-hydroxybenzoic acid	8047	5-hydroxymethylfurfural	425
benzoic acid	39	Furfural	960
hydroquinone	4	4-hydroxybenzaldehyde	3
catechol	1	syringaldehyde	226
glucuronic acid	429	vanillin	86
galacturonic acid	10535	acetosyringone	8

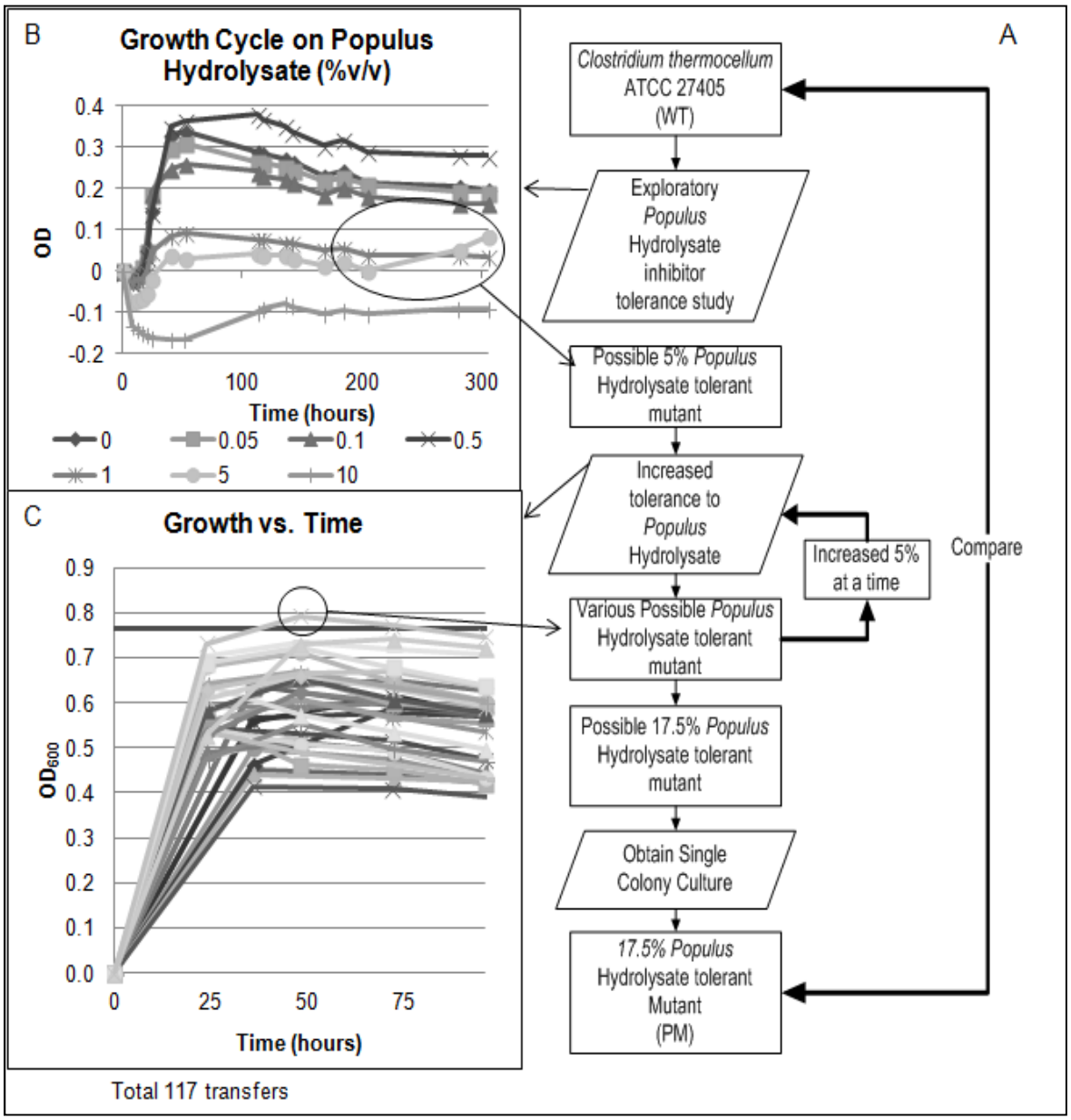


Figure 1: Flow chart for mutant development.

(A) The flow chart shows the mutant development process from the initial WT strain through the final mutant. Single colony isolate #6 was chosen as the final 17.5% v/v *Populus* hydrolysate tolerant mutant strain (PM). (B) Initial growth of WT culture on various concentrations of *Populus* hydrolysate to determine inhibitory effect. The culture in 5% v/v *Populus* hydrolysate started growing late in the fermentation. (C) An example of serial transfers showing increased growth over time in 10% v/v Populus hydrolysate. Illustrative curves show each daily serial transfer.

culture with 5% v/v *Populus* hydrolysate started to grow (Figure 1B). When cultured in fresh medium containing 5% v/v *Populus* hydrolysate this culture grew as well as the WT strain in standard medium (without hydrolysate) and was denoted as the final 5% v/v *Populus* hydrolysate tolerant mutant strain of *C. thermocellum*. Selection of a resistant culture began using 10% v/v *Populus* hydrolysate. The growth rate of each serial dilution was tracked at 24 hour intervals for 96 hours and plotted as optical density (OD) versus Time (Figure 1C). The maximum OD of 0.0750 for the WT in standard medium was selected as the benchmark. The OD of the culture in 10% v/v *Populus* hydrolysate at 24 hours slowly increased with each serial dilution transfer until the OD reached the required benchmark. The concentration of *Populus* hydrolysate was then increased by 5% v/v and the process repeated. The target OD of 0.75 at the 24 hour sample point was not reached after 30 serial dilutions when the culture was grown in 20% v/v *Populus* hydrolysate. At that point, the concentration of hydrolysate was reduced to 17.5% v/v *Populus* hydrolysate. This subculture was able to reach the benchmark OD and was denoted as the final 17.5% v/v *Populus* hydrolysate culture. As described, culture samples were saved throughout the process. After a total of 117 transfers, the 5% v/v *Populus* hydrolysate tolerant population generated the final 17.5% v/v *Populus* hydrolysate tolerant mutants. Seven single colony isolates were obtained from the final 17.5% v/v *Populus* hydrolysate tolerant mutant population. Isolates 1 and 4 grew poorly compared to the other isolates (Figure 2). Isolate #6 grew most similarly to the population average; therefore, it was selected for further strain comparisons and denoted PM for "*Populus* Mutant" (Figure 1A). The PM has a similar growth rate on insoluble cellulose (Avicel) compared to the WT strain despite being selected with cellobiose as the only carbon source (Figure 3).

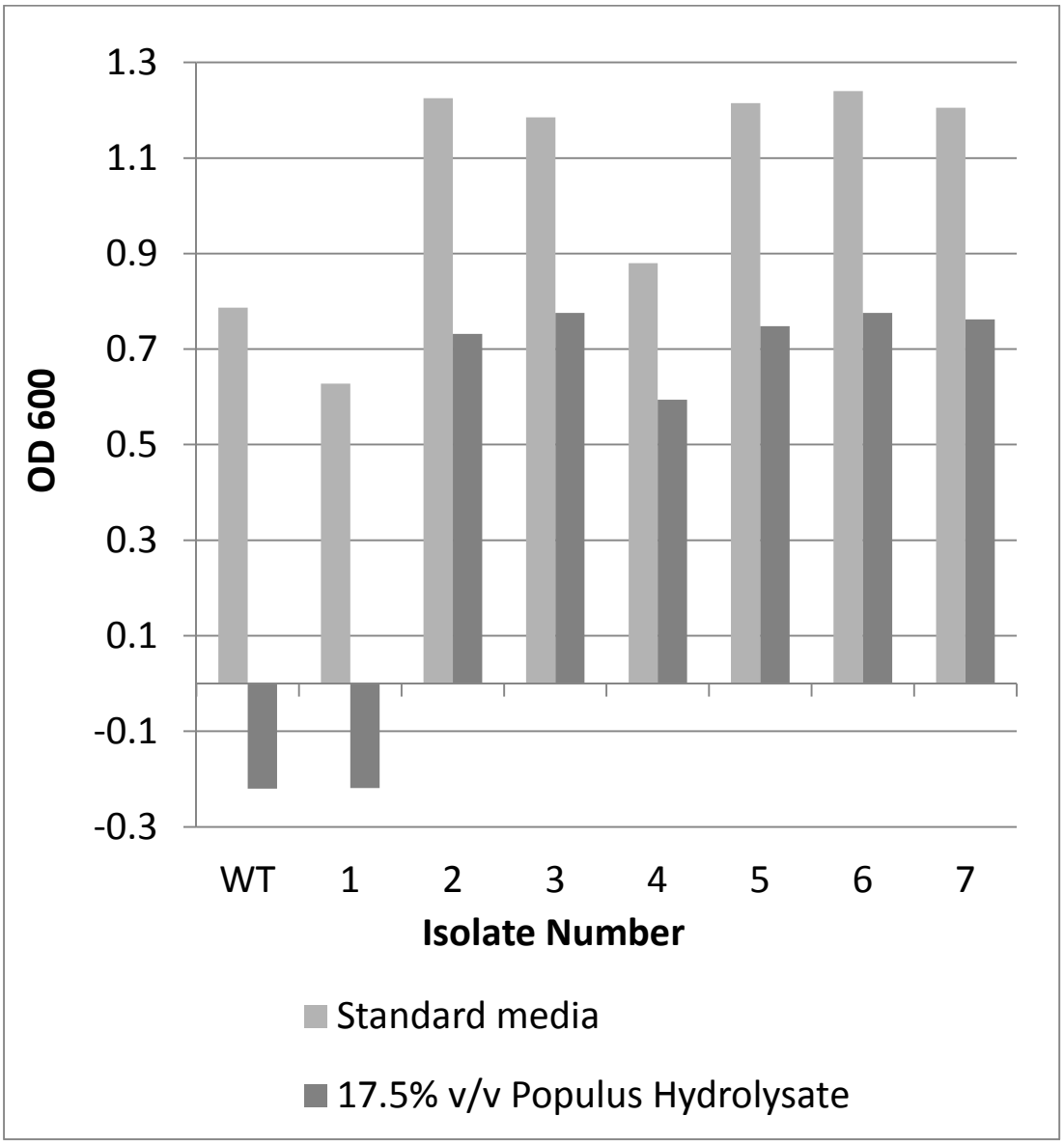


Figure 2: Growth comparison of WT and seven single colony isolates in standard media and 17.5% v/v Populus hydrolysate.

The OD was referenced to an uninoculated sample of medium incubated with the samples. The PM and WT have the ability to degrade two of the compounds in the *Populus* hydrolysate which lightens the color and leads to the negative OD.

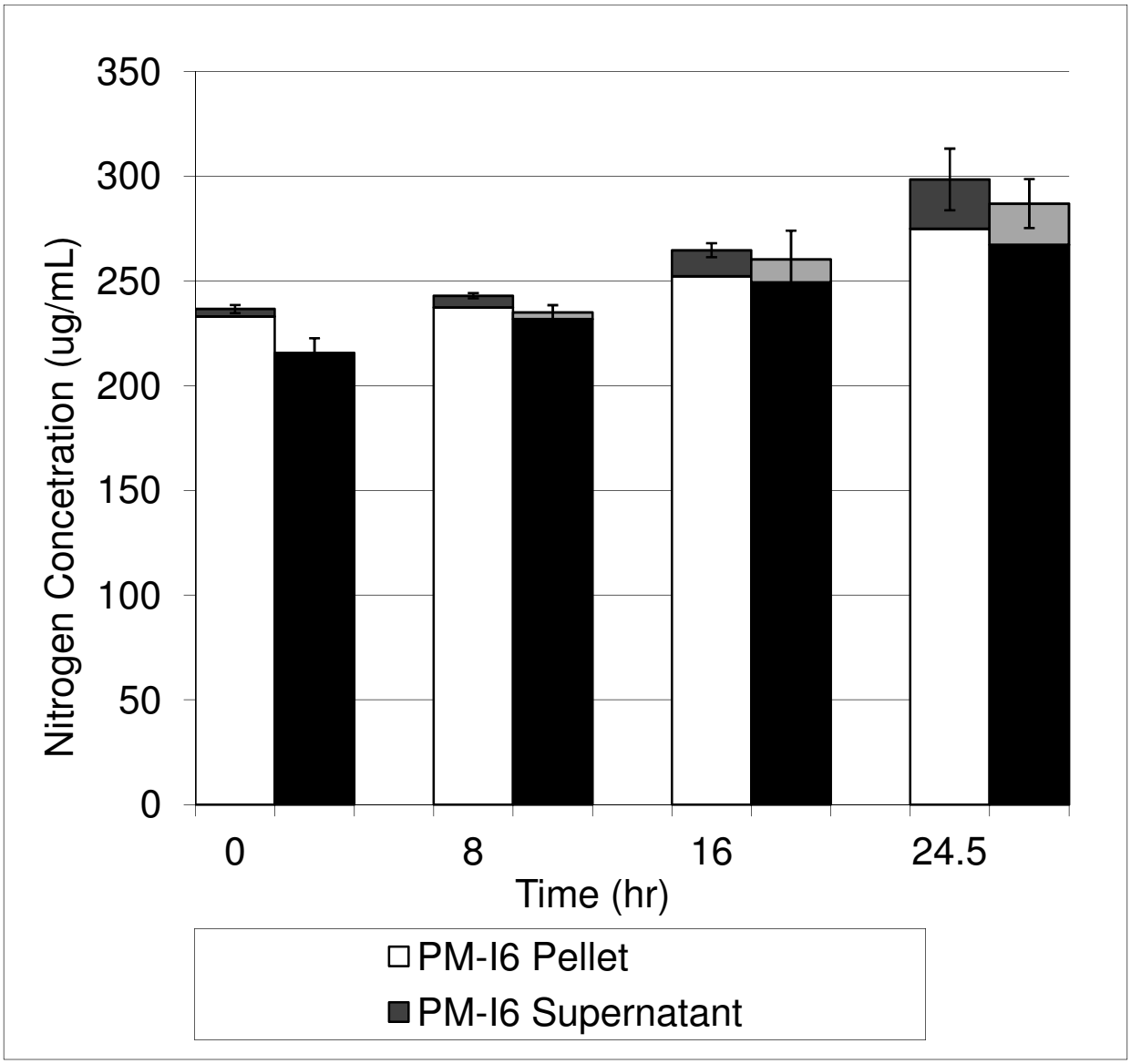


Figure 3: PM and WT growth comparison on insoluable substrate (Avicel) as a function of pellete and supernatant nitrogen concentration taken over time.

1.4.3 Genetic changes during the mutation process

To determine the genetic basis of the *C. thermocellum Populus* hydrolysate- tolerant phenotype, the genomes of the WT strain, six intermediate populations and the seven single colony isolates were resequenced by JGI using the whole-genome shotgun sequencing method. JGI reported 224 putative mutations across all 14 samples (Supporting File S1) [84]. There were 20 putative differences compared to the reference genome in all of the sample sequences including the WT which were removed from consideration. The remaining 204 putative differences were screened for 'high-confidence differences' by removing mutations listed as false positive (fp), Illumina Sequence-specific error leading to a false positive (ISE), and structural variants (SV), resulting in 73 high confidence differences.

A longitudinal analysis was conducted on the six intermediate populations to determine when in the evolutionary process the mutations occurred. The number of mutations increased at a rate of 5±3 mutations between samples (Figure 4). The greatest increase in the number of mutations, 9 differences, occurred between the final 15% v/v *Populus* hydrolysate population and the final 17.5% v/v *Populus* hydrolysate population samples. The second highest increase in differences, 8 differences, occurred between the WT and the final 5% v/v *Populus* hydrolysate population samples. Although, there was no difference in the total number of mutations between the final 10% population and the intermediate 15% v/v *Populus* hydrolysate population sample, the actual mutations in these two populations were different from one another.

A pan genomic analysis (Figure 5) was conducted on the seven single colony isolates to determine the distribution of differences among the isolates. Each of the isolates contained an average of 27±2 differences. Of the 73 differences, 33% of the differences were found in the population samples

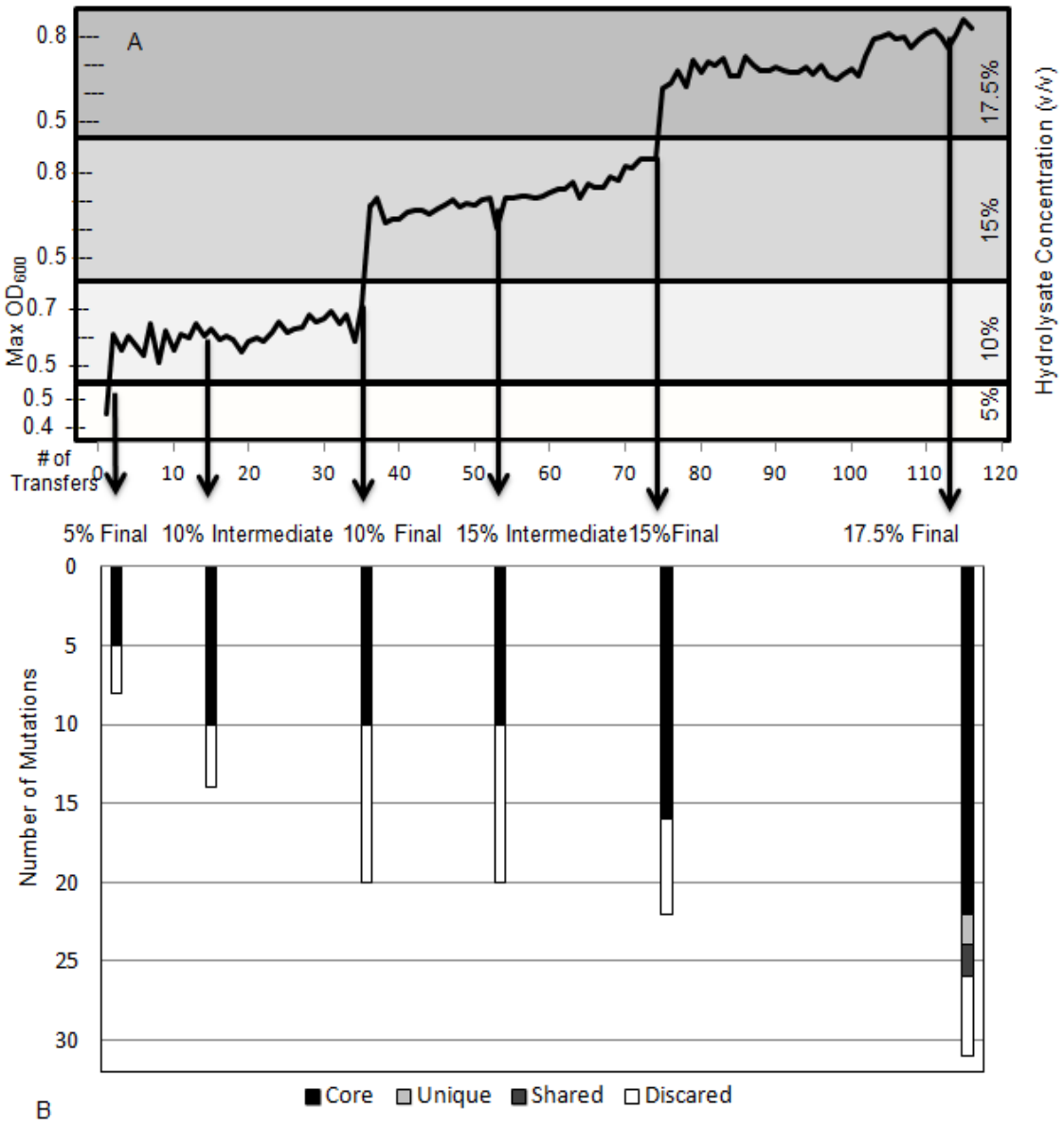


Figure 4: Longitudinal Genomic Analysis in which six population samples along the evolutionary process were sequenced to determine when the mutations occurred.

(A) Daily transfers over the entire evolutionary process in various hydrolysate concentrations. The arrows indicate where in the process the different population samples were taken for sequencing. Samples were considered either final mutations if taken at the end of a given concentrations evolutionary process or intermediate mutants if taken during the middle of the processes. (B) The total number of mutations that occurred for each population by catergory of the pan genomic analysis.

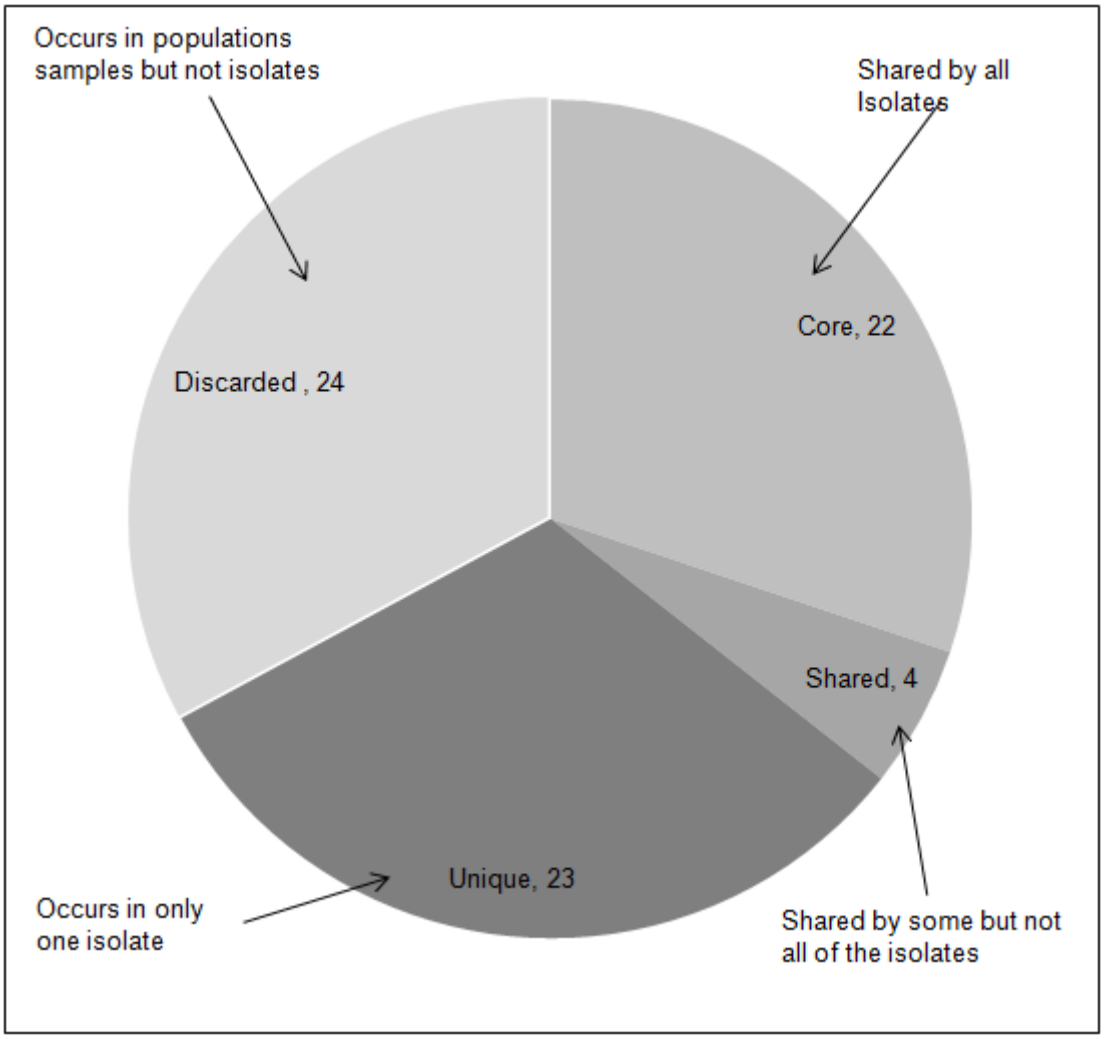


Figure 5: Pan Genomic Analysis of the seven single colony isolates from the final 17.5% v/v Populus hydrolysate tolerant mutant strain.

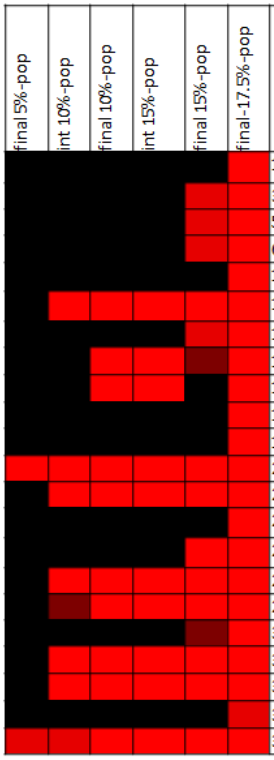
The core mutations are most likely responsible for the increased robustness since they are shared by all isolates.

but not in the isolate samples and were considered discarded mutations. Of the remaining differences, 32% were unique to one of the isolates, 5% were shared by multiple isolates, and 30% belonged to the core genome which is common to all of the isolates (designated as core mutations). Figure 6 lists the core mutations plus their annotated function, and when they occurred in the evolutionary process.

Table 4 lists the unique and shared mutations for the seven single colony isolates. There are a total of 27 unique and shared mutations among the isolates. Only four of these mutations were detected in the final 17.5% v/v *Populus* hydrolysate tolerant mutant population sample at a concentration of < 2% of the population. None of the mutations were detected in earlier population samples. These mutations may explain some of the variability seen in the growth of the different single colony isolates (Figure 2). However, only the six unique and shared mutations contained isolate #6 will be considered further since the fermentation and RNA-seq experiments used this isolate.

There was a nonrandom distribution of mutations with only 11% of mutations occurring in the non-coding region. The number of insertions or deletions (INDELs), synonymous single nucleotide polymorphous (SNPs), non-synonymous SNPs and STOP codons in the coding parts of the genome is presented in Figure 7. SNPs are the dominant type of mutation that occurred, and non-synonymous substitutions were approximately four times as abundant as synonymous.

Further analysis of the distribution of mutations revealed an increase number of mutations at several locations termed hotpots. In order to be considered a hotspot, multiple mutations had to be located within a gene, in an operon, or along the same pathway and at least one of the genes had to be located in the core mutation genome (Figure 6). A manual analysis of the distribution of the mutations



final 5%-pop	int 10%-pop	final 10%-pop	int 15%-pop	final 15%-pop	final-17.5%-pop	pos	type	name	Name		type	Placement	strand
						199088	CDS	Cthe_0158	Rng G	ribobuclease, Rne/Rng family DEL:199088-199233	DEL		
						365961	CDS	Cthe_0291	yyaL*	Uncharacterized protein	Syn	L19L	-
						530675	NC	Cthe_0422	Rex	CoA-binding domain protein	NC		NC
						646838	CDS	Cthe_0531		hypothetical protein	Non-Syn	C30Y	+
						1135183	CDS	Cthe_0948	PyrDII	dihydroorotate dehydrogenase,	Non-Syn	D173N	-
						1137862	CDS	Cthe_0949	CarB	carbamoyl-phosphate synthase, large subunit	Non-Syn	P361H	-
						1221346	CDS	Cthe_1020	CbpB	extracellular solute-binding protein family 1	Non-Syn	A99V	-
						1432911	CDS	Cthe_1202	MFS_1	major facilitotor superfamily, MFS_1	Non-Syn	R343G	-
						1434019	NC	Cthe_1202	MFS_1	82 bases 5' of Cthe_1202	NC		NC
						1523834	CDS	Cthe_1256	BglX	glycoside hydrolase family 3 domain protein	Non-Syn	E458K	-
						1897805	CDS	Cthe_1569	metY	Cysteine synthase	STOP	E229STOP	+
						2088723	CDS	Cthe_1766	proB	glutamate 5-kinase	Non-Syn	A149T	-
						2179038	NC	NC		String of 9 As not close to anything	INDEL		
						2187817	CDS	Cthe_1842	metY	O-acetylhomoserine/O-acetylserine sulfhydrylase	Non-Syn	P29Q	-
						2212180	CDS	Cthe_1866	argD	Acetylornithine aminotransferase	Non-Syn	E55G	+
						2517570	CDS	Cthe_2119	RsgL	glycoside hydrolase family 10	Syn		-
						2838931	CDS	Cthe_2376	gyrB	DNA gyrase, B subunit	Non-Syn	R26S	+
						3073025	NC	Cthe_2602	ATPase	127 bases 5' of Cthe_2602 - ATP synthase operon	NC		NC
						3151358	NC	NC		5 bases 3' of an insertion sequence (transposase)	INDEL		
						3214087	CDS	Cthe_2724	rpoB	DNA-directed RNA polymerase, beta subunit	Non-Syn	E885K	+
						3219211	CDS	Cthe_2727	rpsL	30S ribosomal protein S12	Non-Syn	A14V	+
						3642539	CDS	Cthe_3087	spo0A	sporulation transcriptional activator SpoOA	STOP	Q76STOP	+

Figure 6: Core mutations common to all seven single colony isolates.

*Blast search of the uncharacterized protein returned yyaL as a homologus protein. Type of mutations: Syn: synonymous SNP, Non-Syn: Non-Synonymous SNP, NC: non-coding region, STOP: stop codon inserted, INDEL: insertion or deletion.

Table 4: Shared and unique mutations in the seven single colony isolates.

Isolate 1	solate 2	Isolate 3	Isolate 4	Isolate 5	Isolate 6	Isolate 7	Position	Name	Product	Type of Mutation	Placement	Strand
	X						212735	Cthe_0175	polysaccharide deacetylase	Non-Syn	S280G	+
	X						606073	Cthe_0492	CheC, ingibitor of MCP methylation	Non-Syn	E53K	+
						X	662249	Cthe_0541	peptidase M24	Syn	N286N	+
X			X				1187137	Cthe_0992	ribosomal protein L7Ae/L30e/S12e/Gadd45	Syn	T95T	-
			X				1221367	Cthe_1020	extracellular solute-binding protein family 1	Non-Syn	A92V	-
X							1230221	Cthe_1028	acetate kinase	INDEL		
X			X				1342025	Cthe_1127*	integrase family protein	Non-Syn	K5E	+
						X	1663282	NC	115 bp 5' end of Cthe_1368: S-layer domain-containing protein	INDEL		
					X		1702173	Cthe_1393	multi-sensor signal transduction histidine kinase	Non-Syn	K559R	-
				X			1717600	NC	1735 bp from 5' end of Cthe_1401	INDEL	Deletion of 1717552-1717617	
					X		1799942	Cthe_1480	hypothetical protein	INDEL		
						X	1880654	NC	606 bp from end of F Cthe_1550 and 868 bp from end of R Cthe_1552	NC		NC
					X		2153823	Cthe_1819	urea ABC transporter, ATP-binding protein UrtE	Non-Syn	N99I	-
						X	2270432	Cthe_1909	copper-amine oxidase-like domain-containng protein	Syn	V423V	-
				X			2421066	NC*	1519 bp from 5' end of Cthe_2036	NC		NC

'X' marks which isolate contained the mutation. * Indicates that the mutation was detected in the final-17.5% v/v *Populus* hydrolysate mutant populuation sample at less than 2%. The rest of the mutations were not detected in that population sample. None of the mutations were detected in earlier population samples. Type of mutations: Syn: synonymous SNP, Non-Syn: Non-Synonymous SNP, NC: non-coding region, STOP: stop codon inserted, INDEL: insertion or deletion.

Table 4: Shared and unique mutations in the seven single colony isolates.

Isolate 1	solate 2	Isolate 3	Isolate 4	Isolate 5	Isolate 6	Isolate 7	Position	Name	Product	Type of Mutation	Placement	Strand
				X			2614779	Cthe_2193	carbohydrate binding family 6	Non-Syn	I601V	-
				X			2699028	Cthe_2270	ABC transporter related protein	Syn	A535A	-
					X		2732545	NC	3524 bp from 3' end of Cthe_2295 and 8703 bp from 3' end of Cthe_2297	INDEL	Deletion of 2732545-2732612	
						X	2997777	NC	98 bp from 5' end of Cthe_2529: delta-aminolevulinic aicd dehydratase	DEL		
X	X	X	X	X	X		3073931	Cthe_2603*	ATP synthase F0, C subunit (operon), signal peptide, amino acid has several substitutions	Non-Syn	I4M	+
						X	3073969	Cthe_2603	ATP synthase F0, C subunit	Non-Syn	V17A	+
	X						3086626	Cthe_2611	S-layer domain-containing protein	Non-Syn	M1473T	+
X			X				3136682	NC	167 bp from 3' end of Cthe_2655: AbrB family transcriptional regulator	INDEL	Insertion: TTTTTTT	
						X	3224162	Cthe_2731	RNA polymerase, sigma-24 subunit, ECF subfamily	Non-Syn	R42G	+
				X			3233583	Cthe_2741	ATP-dependent Clp protease, ATP-binding subunit Clpx	Non-Syn	E20G	+
				X			3525977	Cthe_3003*	hydrogenase, Fe-only	Non-Syn	V329I	-
					X		3628267	Cthe_3078	cellulosome anchoring protein cohesin region	Syn	V1013V	+

'X' marks which isolate contained the mutation. * Indicates that the mutation was detected in the final-17.5% v/v *Populus* hydrolysate mutant populuation sample at less than 2%. The rest of the mutations were not detected in that population sample. None of the mutations were detected in earlier population samples. Type of mutations: Syn: synonymous SNP, Non-Syn: Non-Synonymous SNP, NC: non-coding region, STOP: stop codon inserted, INDEL: insertion or deletion.

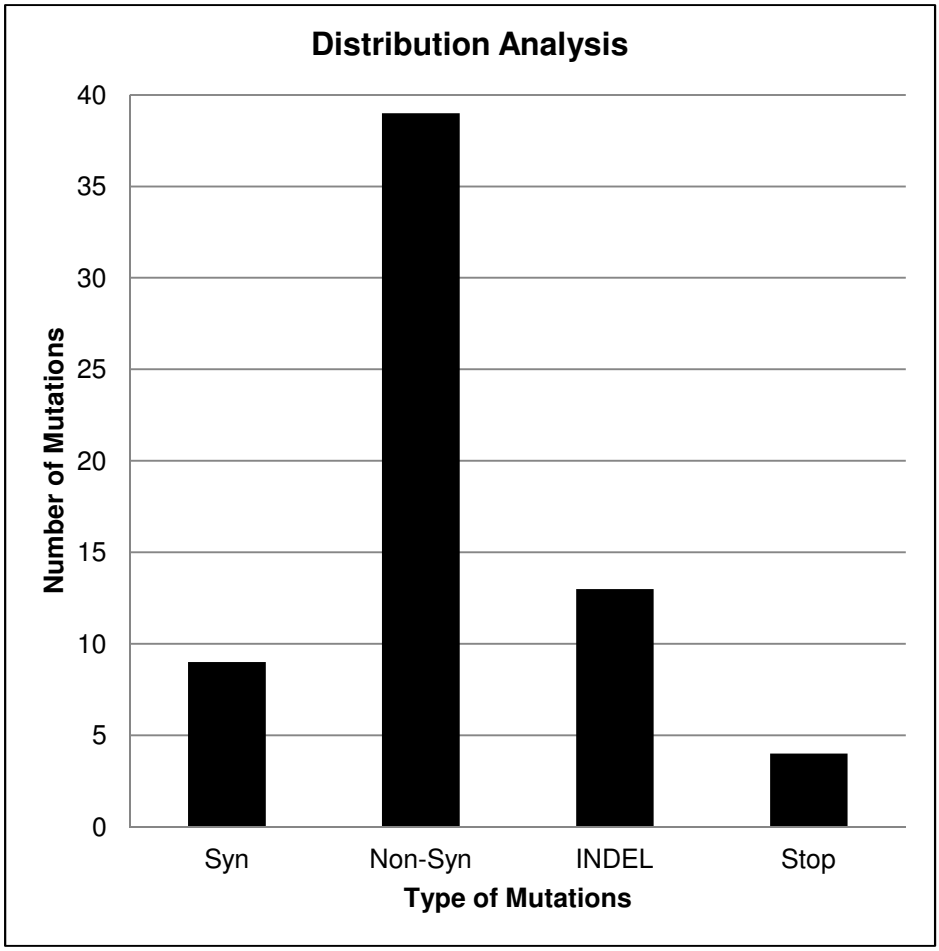


Figure 7: Distribution analysis of mutation data.

identified 8 putative hot spots (Table 5). Genes linked to the hot spots included 24 out of the 73 putative mutations and 14 of the 22 core mutations.

1.4.4 Genes and mutations with unknown function

There are four mutations in the core genome with unknown function. Two of the mutations are INDELS located in non-coding regions and are located at positions 2179038 and 3151358. The other two of the mutations are in Cthe_0291: uncharacterized protein and Cthe_0531: hypothetical protein. In an attempt to determine functionality of these genes a BLAST similarity search was run between the unknown genes and genes from public databases. The BLAST search for Cthe_0531 returned only genes with hypothetical protein annotations. However, the BLAST search for Cthe_0291 returned a number of genes annotated as the thioredoxin, yyaL, protein.

1.4.5 Fermentative Growth

Batch fermentations were conducted in triplicate in Q-plus fermentors to further evaluate the performance of the PM versus WT *C. thermocellum* strains in different levels of hydrolysate. The PM was grown in three test conditions: standard medium (0% v/v *Populus* hydrolysate), 10% v/v and 17.5% v/v *Populus* hydrolysate. The WT was grown in two test conditions: standard medium and 10% v/v *Populus* hydrolysate. The WT grown in the presence of 17.5% v/v *Populus* hydrolysate severely inhibited growth in an unpredictable manner (data not shown). Samples were taken at regular intervals from each fermentation unit based on their growth rate. Samples were analyzed for OD and metabolite

Table 5: Mutational Hot Spots identified in the sequenced genome

Hot Spot ID	Number of Mutations	Genes (# mutations, location)	Products
1	5	**Cthe_0948 (1, CDS), Cthe_0949 (1, CDS), Cthe_1766 (1, CDS), Cthe_1866 (1, CDS),** Cthe_2529 (1, CDS)	Production of various amino acids
2	5	**Cthe_2602 (2, NC)***, Cthe_2603 (2, CDS), Cthe_2607 (1, CDS)	Single transcription unit Cthe_2602-9: Subunits for F-type H+-transporting ATPase
3	4	**Cthe_0422 (1 NC and 1 CDS) ***, Cthe_1028 (1 CDS), Cthe_1029 (1 CDS)	Redox-sensing transcriptional repressor Rex and acetate formation from acetyl-CoA I
4	2	**Cthe_1569 (1 CDS), Cthe_1842 (1 CDS)**	Homocysteine biosynthesis
5	2	**Cthe_1202 (1 NC and 1 CDS)**	Major facilitotor superfamily, MFS_1
6	2	**Cthe_1020 (2 CDS) ***	Extracellular solute-binding protein family 1, PBP2_LTTR_substrate superfamily
7	2	**Cthe_3087 (2 CDS) ***	Sporulation transcriptional activator Spo0A
8	2	**Cthe_1256 (1 CDS), Cthe_2119 (1 CDS)**	glycoside hydrolases

CDS, coding region, NC, non-coding region, bold genes belong to the core mutational genome, *only one of the mutations are in the core mutational genome

concentration by HPLC. Figure 8 shows the average OD (Figure 8A) and concentrations of ethanol (Figure 8B) and acetic acid (Figure 8C) for each of the test conditions. End products are calculated as a function of the amount of glucan utilized from the fermentation broth since the hydrolysate contained additional cellobiose and glucose. In brief, the PM had an increased growth rate when compared to the WT in the same condition. The PM also produced more ethanol and the same amount of acetic acid compared to the WT under the same test conditions.

1.4.6 RNA-seq Analysis

The samples for RNA analysis were harvested from the fermentors during the mid-log phase based on OD (see Figure 8). Resulting RNA-seq data was used to elucidate the role of the genomic changes that helped confer the hydrolysate tolerant phenotype. Gene expression data were analyzed for genes that: (1) were mutated, (2) were located downstream or in an operon with a mutated gene, (3) were another copy of a mutated gene, or (4) were under regulatory control of a mutated gene. An ANOVA analysis was conducted on the resulting subset of 41 genes using a false discovery rate of 5% and the expression of the WT in 0% v/v *Populus* hydrolysate as the baseline. Of the 41 genes 25 of them were differentially expressed in at least one of the comparisons. The 25 differentially expressed genes were clustered using hierarchal clustering as shown in Figure 9. The differential expression profiles for the subset of 41 genes can be found in Supporting File S2.

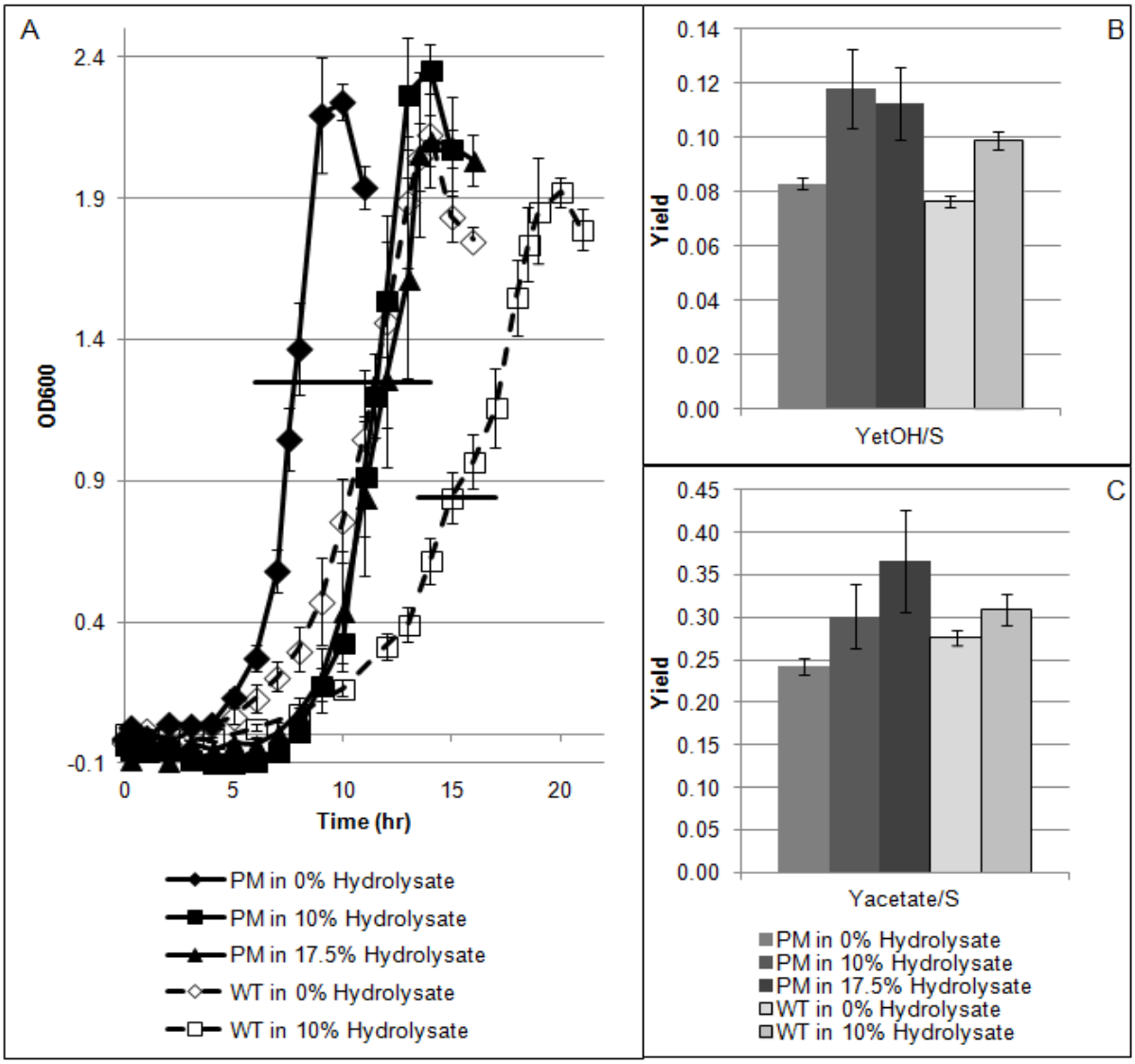


Figure 8: Growth comparison for WT and PM in various concentrations of Populus hydrolysate.

(A) Optical density versus time. The horizontal bars indicate sample point for RNA-seq analysis. (B) Yield of ethanol produced (g/L)/glucan utilized (g/L). (C) Yield of acetic acid produced (g/L)/ glucan utilized (g/L).

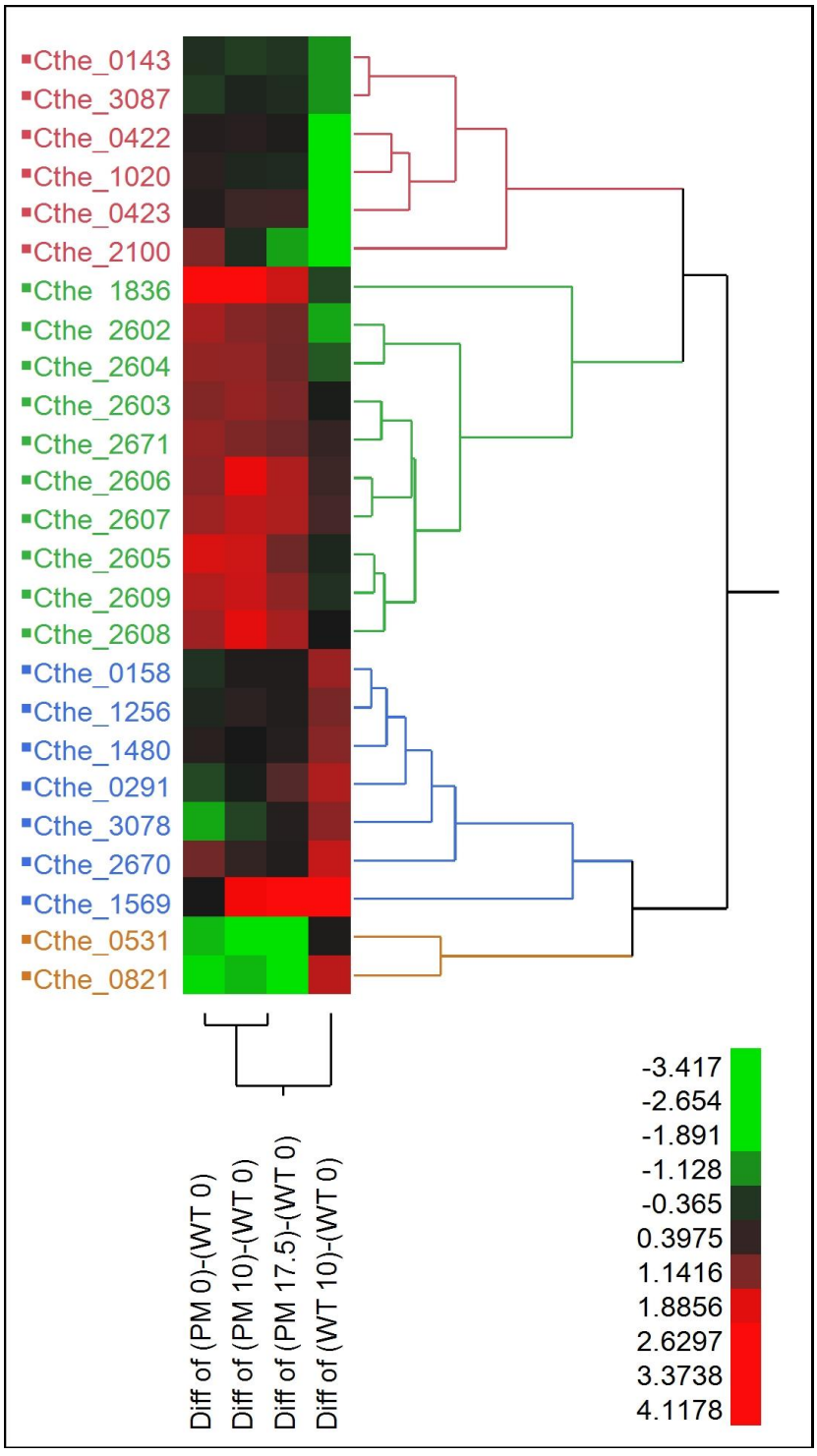


Figure 9: Hierarchical clustering of differentially expressed genes.

The scale represents the log-2 transformed difference between the two conditions; therefore a change greater than one represents a >2-fold change in gene expression.

1.5 Discussion

Further analysis was conducted on the core genes (Figure 6) with insight from the hot spot analysis (Table 5). The mutations were broken down into three basic categories: mutations beneficial to the tolerant phenotype, extraneous or detrimental mutations, and mutations of unknown function.

1.5.1 Mutations beneficial to the tolerant phenotype

Inhibitory compounds may affect the cell by damaging and denaturing biological molecules resulting in adverse outcomes including the improper unfolding of proteins, DNA damage, improper RNA unfolding and degradation, and the impairment of biophysical changes to cell membranes necessary for energy generation and the proper functioning of molecular pumps [1,32]. Figure 10 provides an overall schematic of the mechanisms of tolerance for the PM. The tolerant phenotype is the result of several simultaneous mechanisms of action, including increases in cellular repair, and altered energy metabolism. These and related mechanisms are apparently independent from each other and can involve genes or gene clusters widely dispersed on the chromosome [1].

1.5.1.1 DNA repair and mRNA and protein turnover

Cells contain multiple lines of defense against protein aggregation, DNA and RNA damage, and membrane and organelle damage [1]. The PM has one mutation affecting a gene encoding for a protein related to DNA repair (Cthe_2376) and three mutations affecting genes related to transcription (Cthe_2724), translations (Cthe_2727) and RNA degradation (Cthe_0158). The mutations in Cthe_2376,

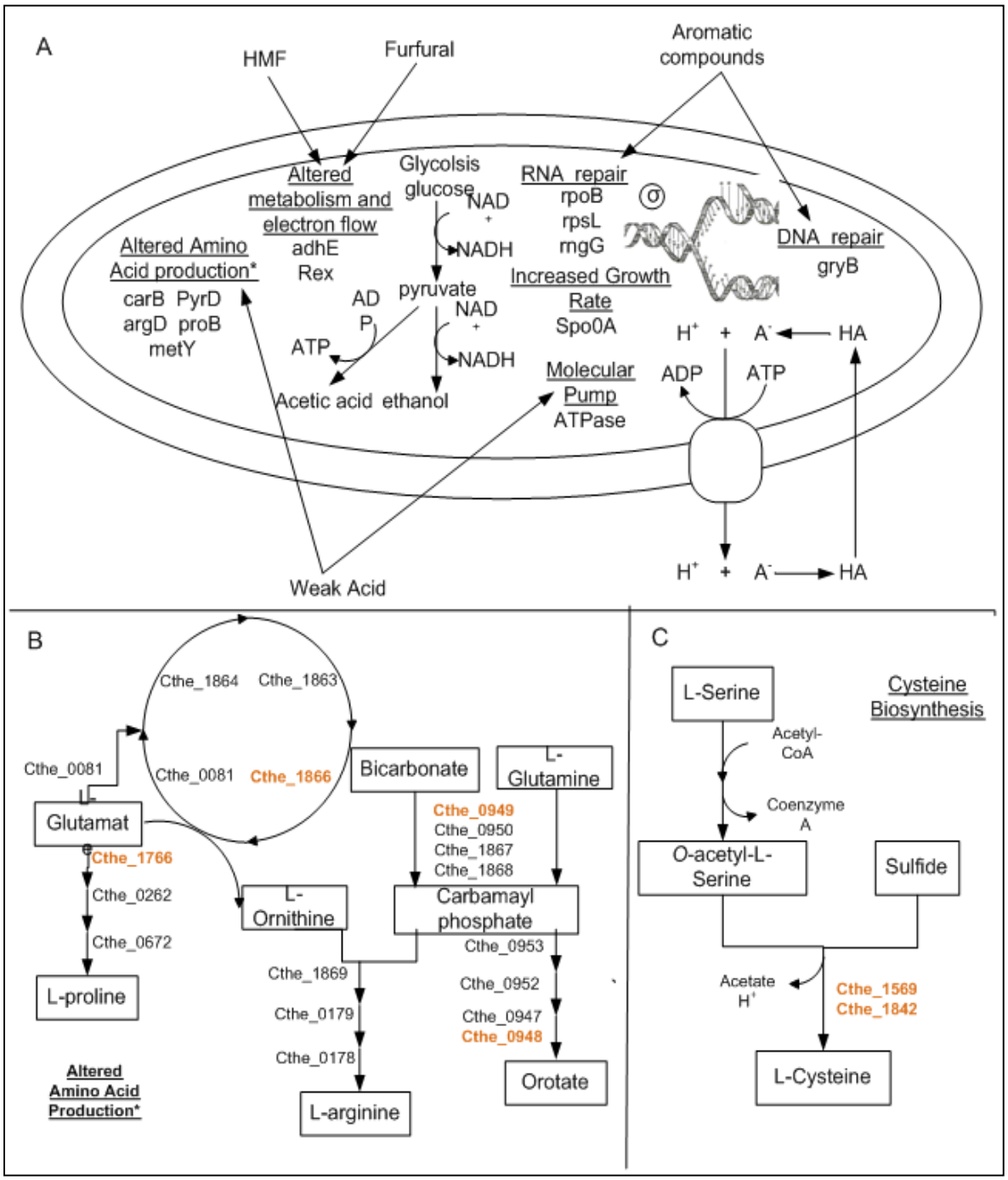


Figure 10: Mechanisms of tolerance.

A) Graphical representation for the mechanism of tolerance in PM. In depth look at the mechanism of tolerance from alter amino acid production (B) and in homocysteine biosynthesis (C) (Taken in part from [29,90]).

Cthe_2724 and Cthe_2727 are all non-synonymous SNPS and Cthe_0158 has a deletion near the 3' end.

The *gyrB* gene (Cthe_2376) encodes an ATP-dependent DNA gyrase enzyme which cleaves and reseals double stranded DNA thereby introducing negative supercoils into DNA [91]. This activity is essential for DNA replication, transcription and recombination [91]. The mutation in *gyrB* (R26S) occurs in a very conserved amino acid based on the Clustlw alignment (Supporting File S3). *C. thermocellum* contains a second copy of the *gyrB* gene (Cthe_0305) which is expressed at approximately one tenth of the level of the Cthe_2376 in the WT under standard conditions. Neither copy of the *gyrB* gene is differentially expressed across the different conditions (data not shown). It is likely that the Cthe_2376 is the dominant copy of the *gyrB* gene in *C. thermocellum*. However, the potential change in specificity, selectivity or stability resulting from this mutation is unknown.

The *rpoB* gene (Cthe_2724) encodes for the β-subunit of RNA polymerase. The RNA polymerase controls the flow of information from genotype to phenotype and is composed of an essential catalytic core enzyme and one of several alternative sigma (σ) factors [92]. *C. thermocellum* contains a single copy of the *rpoB* gene and the PM strain has a mutation (E885K) in a strongly conserved amino acid based on the Clustalw alignment (Supporting File S4). This mutation potentially changes the specificity, activity and/or stability of the RNA polymerase. Since, the RNA polymerase contacts every promoter in the genome, a single amino acid substitutions in the critical portions of the enzyme may lead to global changes in gene expression and have significant impact of cellular functions [92]. The *rpsL* gene (Cthe_2727) produces the ribosomal protein S12 [93] which is part of the 30S ribosomal subunit. Alterations in *rpsL* are known to confer resistance to the error-inducing antibiotic streptomycin [94]. The *rpsL* mutation (A14V) is in a location that can yield many different amino acids based on the Clustalw alignment showing no conservation (Supporting File S5). Together, mutations in the subunits of RNA

polymerase and ribosomes have significant potential to change the overall protein expression profile in the cell. However, these mutations do not significantly change the gene expression level in any of the test conditions (data not shown). On the other hand, mutations in both *rpoB* and *rpsL* have been shown to impart increased tolerance to aromatic compounds such as 4-hydroxybenzoate and a variety of other organic chemicals and organic solvents by blocking the uptake of these compounds by the membrane transport system [93].

RNase G (Cthe_0158) endoribonuclease has been shown to be required for the normal decay of several transcripts including the functional form of enolase (*eno*) and acetaldehyde dehydrogenase (*adhE*) mRNA, both are involved with glycolysis [95]. When the *rng* protien is disrupted, the *eno* and *adhE* proteins are known to be over produced as a consequence of mRNA stabilization and the resulting increase of their half-lives [95]. Although the mutation does not occur in a known domain, it is possible that the deletion of codons 438 through 486 of 498 may disrupt the functionality of the *rng* protein and slow down the rate of RNA degradation by the PM. A reduction in mRNA degradation can either increase the activity of mRNA or maintain mRNA activity with reduced transcriptional effort. The *rng* gene is significantly upregulated in the WT in the presence of hydrolysate (Figure 9) which would result in more rapid mRNA turnover. Upregulation of *rng* does not occur in PM, which, when combined with the potential disruption of functionality would result in significant reduction of mRNA turnover. Consistent with the differences in *rng* activity, the WT significantly downregulates *adhE* and *eno* in hydrolysate medium whereas the PM has constant expression levels (See Figure 9). Since DNA is particularly sensitive to damage during transcription, a reduction in mRNA turnover rates would also lead to reduced DNA damage by the hydrolysate. A reduction in mRNA turnover could also conserve energy and increase metabolic efficiency in stable environments that do not require frequent changes in gene expression profile. Furthermore, constant expression of the *eno* and *adhE* genes could cause a

greater flux through the glycolysis pathway. Together, the benefits of reduced mRNA turnover may play a significant role in the increased tolerance of the PM.

1.5.1.2 Altered Energy Metabolism

The engagement of energy-dependent processes for dealing with inhibitory effects, such as cellular repair would suggest that energy generation is a crucial process in dealing with chemical stress [1]. Energetic processes that may help counteract the increased demand include genes involved in glucose catabolism, cellular redox, amino acid production, and growth rate. An increased growth rate yields higher volumetric conversion rates which enhances the fermentability of hydrolysates [29].

Furfural and HMF are reduced to their corresponding alcohols which directly inhibits *adhE*, resulting in NAD(P)H depletion affecting glycolysis and TCA fluxes [29,32]. The PM had a mutation in the non-coding region upstream of the Cthe_0422-3 operon which encodes the *Rex* (redox) repressor and *adhE*. The mutation resulted in a slightly higher expression level (<2-fold) of both genes in all test conditions for the PM whereas the WT had significant decreased expression of these genes in hydrolysate medium (See Figure 9). Rex represses transcription of multiple target genes involved in fermentation pathways changing the product distribution to meet the needs of the cell [96]. The redox stoichiometry has an essential impact on the product distribution because increased NADH availability plays a key role in the alcohol production [96]. The inhibition of glycolysis by the WT would lead to a decrease in reducing power (in the form of NADH) being available for downstream electron transport and ethanol production. The decrease in ethanol production and reduced electron flux may generate insufficient NAD+ to ensure cellular metabolism for the WT which is not the case in the PM. Furthermore, insufficient energy may be available for many NADPH-dependent enzymes that are

responsible for cellular defense mechanisms against various stressed [29]. As to be expected with the increased expression of *adhE*, the PM has an increase in ethanol production and a decrease in acetic acid production when compared to the WT strain. This mutation combined with the deletion in RNase G (above) could result in greater tolerance to the hydrolysate.

The PM has a mutation in the non-coding region 127 bp upstream of the first gene in the operon (Cthe_2602-09) coding for the ATPase complex. The PM has significantly increased expression levels for the entire ATPase operon compared to the WT in all conditions tested (Figure 9). Of the single colony isolates, isolates 1-6 also have a mutation in Cthe_2603 (I4M) which encodes for the C subunit of the ATP synthase (Table 4); however, the effect of this mutation is unknown at this time.

The ATPase shuffle chemicals (H^+/NA^+) across the cellular membrane. If the transport is against the concentration gradient of the cell the ATPase operates as an antiport mechanism at the expense of ATP [1]. The ATPase may be used as a mechanism of tolerance for the undissociated form of weak acid in the hydrolysate which can diffuse from the fermentation medium across the plasma membrane and dissociate due to higher intracellular pH (see Figure 10A), thus decreasing the intracellular pH. The decrease in intracellular pH is compensated by the ATPase, which pumps protons out of the cell at the expense of ATP hydrolysis [29,90]. Alternatively, the ATPase might also work in conjunction with the membrane bound Ferredoxin-dependent Ech-type NiFe-hydrogenase (Cthe3013-3024) for the production of molecular hydrogen to dispose of excess reducing equivalents generated during carbohydrate catabolism [49]. The Ech hydrogenase complex reoxidizes the ferredoxin reduced during the conversion of pyruvate to acetyl-CoA by pumping H^+/Na^+ ions across the cell membrane and creates proton gradients for powering ATP synthesis by the ATPase [49]. The oxidation of electron carriers (NADH and/or reduced ferredoxins) is required for maintaining glycolytic flux and leads to the ultimate production of reduced products (ethanol, lactate, and H_2) [97]. Given the fact that the PM both in the

presence and absence of hydrolysate grows faster than the WT strain, the ATPase is most likely operating as a mechanism to produce ATP.

Cells respond to the effect of inhibitory compounds by alterations in the biosynthetic programs to produce metabolites, such as amino acids, that counter toxic effects [1]. The PM contained several mutations involved in amino acid production which are similar to those seen by an ethanol tolerant mutant under similar conditions [74]. The PM has two mutations involved with glutamate catabolism. A mutation to a gene in the arginine pathway, *argD* (Cthe_1866 E55G) resulted in a pfam score of +1180. This is a common mutation, including an occurrence in *C. thermocellum* DSM 1313 strain [74]. Although, there is no change in gene expression under any of the test conditions, the positive pfam score suggests a gain in function. Another mutation to a gene in the proline pathway, *proB* (Cthe_1766 A149T) resulted in a pfam score of -1766 suggesting a loss in function. This mutation is found in *Clostridium phytofermentans*. There was not a significant change in gene expression among any of the test conditions. These two mutations suggest a shift from proline production to arginine production in PM which seems to be a beneficial mutation that commonly occurs (see Figure 10B). An increase in amino acid production can also help overcome weak acid stress. The decarboxylation and antiport of the amino acids glutamate and arginine generates a net efflux of intracellular protons thus increasing intracellular pH [1,90]. Decarboxylases operate by combining an internalized anion acid (arginine or glutamate) with a proton and exchanging the resultant product (agmatine, γ-amiobutyrate) for another amino acid substrate. Thus, an extracellular amino acid is converted to an extracellular product, but the consumption of an intercellular proton results in an increase in intracellular pH [90].

There are two copies (Cthe_1569 and Cthe_1842) of the *metY* gene in *C. thermocellum*, which produces L-cysteine, both of which have mutations in the PM that seem to inactivate the proteins

(Figure 10C). Cthe_1569 has a stop sequence inserted at codon 229 and the non-synonymous SNP in a highly conserved region of Cthe_1842 (P29Q) results in a pfam score of -1198. These two mutations essentially cut off the conversion of serine and hydrogen sulfide to L-cysteine. It has been shown the *metY* is not essential for *E. coli* cell growth [98]; the mutation may be an attempt to reduce acetate and H^+ formation (see Figure 10C). There is no significant change in gene expression under any of the test conditions for Cthe_1842. However, the WT and PM significantly increase the expression of Cthe_1569in hydrolysate medium compared to the WT in standard medium.

The *spo0A* transcription factor is responsible for the initiation of sporulation [99,100]. *Spo0A* has been shown to repress transcription of the transcriptional pleiotropic regulator of transition state genes (*abrB)* and to activate operons encoding stage II sporulation genes [44,99]. The competitive advantage that the *abrB* product provides the cell is to minimize the expression of extraneous gene products that would interfere with the maximal rate of growth of the cell under nutrient-excess conditions [100]. *Spo0A* mutants which are unable to go into sporulation are thought to be locked in exponential growth since they continue to grow under nutritional conditions that would normally induce sporulation. Essentially they appear to maintain growth until the nutrients are exhausted, whereupon cell lysis occurs [100]. This is consistent with the PM growth kinetics and other *spo0A* defective mutants [41,43,44].

There are two copies of *spo0A* in *C. thermocellum,* Cthe_0812 and a distantly related homologue Cthe_3087. The PM has a stop codon placed early in the coding region of Cthe_3087 which should disrupt the protein function. As would be expected, the gene expression of Cthe_3087 is not significantly different in any test condition. Cthe_0812 is significantly downregulated by an unknown mechanism for the PM under all test conditions (Figure 9). Since there is a disruption in expression of

both *spo0A* proteins in the PM, the PM maintains the expression level of the *abrB* (Cthe_2100) under all test conditions (Figure 9). In comparison, the WT apparently significantly down regulates the expression of *abrB* in the hydrolysate medium.

1.5.2 Genes and mutations with unknown function

Of the three mutations in non-coding regions with unknown function, the first mutation (located at position 2179038) is an insertion of a string of nine adenine nucleotides 388 bp upstream of Cthe_1836, a hypothetical protein. The second mutation (located at position 3151358) is an INDEL of unknown size and 245 bp downstream of the 3' end of Cthe_2670, a hypothetical protein and 221 bp upstream of the 3' end of Cthe_2671, mutator type transposase. The third mutation (Located at position 2732545) is unique to isolate six and is a deletion of 67 bp in an intergenic region (Table 4) more than 3000 bp from the closest gene. Of the mutations in the coding regions, Cthe_0531 was mutated at the C30Y position. However, it is uncertain how this mutation affects the cell since the BLAST search returned no similar genes with known function. The BLAST search of Cthe_0291 returned a number of genes annotated as the thioredoxin, *yyaL*, gene. The mutation in the gene is a synonymous SNP located at codon 19. Thioredoxins are responsible for maintaining a cellular reducing environment and fulfill an important role in DNA synthesis and protein repair [101]. If Cthe_0291 is a thioredoxin homologue, it is possible that this mutation also is involved in the tolerant phenotype. The third mutation is an INDEL in the hypothetical protein Cthe_1480. The BLAST search returned a number of genes annotated as RND family transporter and *mmpL* domain-containing proteins. The pfam score of -492 suggest that this mutation is a loss of function. Although there are differences in the gene expression levels for these genes (Figure 9), the potential impacts of these mutations are unknown at this time.

1.5.3 Extraneous or detrimental mutations

The PM has three mutations affecting the cellulosome. There are two mutations in different glycoside hydrolases (GH); one each in a family 10 GH and in a family 3 GH. Cthe_2119 (*RsgI6*), is an anti-σ factor extracellular module with an apparent xylanase GH10 catalytic module. Rsgl6 is involved in regulating the expression of cellulosomal genes in the presence of xylans and cellulose [60,77,102]. Since the mutation is a synonymous SNP in codon 415, which is outside of the catalytic domain, it is unlikely that the mutation causes a significant change in function. There is no significant change in expression level between the WT and PM in any of the conditions. The other mutation is in *bglX* (Cthe_1256) which encodes a *beta-D-glucoside glucohydrolase* which breaks the β-1,4 glucosidic bonds of cellobiose to produce glucose monomers [103]. Insufficient β-glucosidase activity causes an accumulation of cellobiose with the consequence that both endoglucanase and exoglucanase activities are severely inhibited [103]. The mutation (E458K) is next to a common variant. GH3 β-glucosidases only have a few structures solved [103] so the effect of the E458K mutation is unknown; however, the -539 pfam score suggests a loss of function. Furthermore, there is no significant change in gene expression level for the PM in any the test conditions whereas the WT significantly increases expression in the presence of hydrolysate. Therefore, if compared with the WT, the mutation seems to effect the expression of GH3 at higher hydrolysate concentrations which could be detrimental in the presence of solid biomass. Isolate 6 also has a unique mutation in cellulosome anchoring protein Cthe_3078. The mutation is a synonymous SNP in codon 1013, which is outside of a known domain. The PM downregulates the gene expression compared to the WT in standard medium and keeps the expression constant in hydrolysate medium, whereas, the WT increases the expression in hydrolysate medium.

The PM also had mutations in two different types of sugar transport systems. The first is in an ATP-binding cassette (ABC) transporter and the second is in a Major Facilitator Superfamily 1 (MFS_1) transporter. ABC transporters utilize ATP energy to transport inorganic ions, amino acids, hydrocarbons, polypeptides or hydrophobic compounds [1]. *C. thermocellum* prefers to grow on long cellodextrins whose transport has been suggested to be mediated by ABC transporters [76]. The Cthe_1018-1020 system might serve as the major transporters in cellodextrin assimilation [76]. In some Gram-positive organisms, the ATP-binding subunit of an ABC system is not part of a specific transporter complex; instead, it is shared by multiple transporters [76] increasing the efficiency of the cell. Cthe_1020 (*cbpB*) is the extracellular-binding protein in the complex. The WT in hydrolysate significantly down regulated *cbpB* (see Figure 9); however, the expression levels in the other two components remain near that of the WT in standard medium (data not shown). The PM has a mutation (A99V) in *cbpB* which has a pfam of -227 but a near normal expression levels for all three genes in all test conditions. Although, it is not clear the role of this mutation, the negative pfam suggests that it might be detrimental. However, the increased expression of *cbpB* may be the result of a more global change in gene expression.

MFS_1 transporters differ in structure from ABC transporters and utilize the proton motive force as their energy source [1,104]. Solutes transported by the MFS_1 family include galactose, arabinose, xylose, and glucose in bacteria [104]. The PM has two mutations in the MFS_1 gene (Cthe_1202); one is in the non-coding region 82 bp upstream of the 5' end and the other is R343G. There is no significant change in gene expression between the WT and PM so it is unlikely that the mutation in the non-coding region affects gene expression (data not shown); however, the mutation in the coding region may change the selectivity or specificity of this protein. Since the PM was evolved only on soluble cellobiose, it is possible that these mutations are a result of an attempt to conserve energy and may be detrimental in an industrial setting.

The PM also has two mutations along the pyrimidine pathway (Cthe_0947-0953). The gene *carB* (Cthe_0949 P361H) encodes for a subunit of the carbamoyl-P synthase enzyme which catalyzes its synthesis from ammonia and ATP [2]. The other mutation in *pyrDII* (Cthe_0948 D173H) is located on the pathway from carbamoyl-P to orotate. These two mutations do not have a significant change in the pfam score and are similarly upregulated between the WT and PM in the same test condition (data not shown); therefore, these mutations are unlikely to contribute significantly to the tolerant phenotype (see Figure 10B).

Isolate #6 also has two extraneous unique mutations. The first mutation (K559R) is in a multi-sensor signal transduction histidine kinase (Cthe_1393). The pfam score of 55 suggest that this mutation does not affect the functionality of the protein greatly. There is also no significant change in expression of this gene. The other mutation is in the urea ABC transporter ATP-binding protein *urtE*(Cthe_1819). The mutation (N99I) is outside of a known domain and does not result in a significant change in gene expression.

Future work includes a global analysis of gene expression and kinetic modeling of growth for the PM and WT strains to further understanding the mechanisms of *Populus* hydrolysate tolerance. Additional analysis of these mutations could be pursued using system biology methods, genetic tools, biochemical assays and strain engineering.

1.6 Conclusion

This study is the first investigation of the mechanism of tolerance for *C. thermocellum* to inhibition by a complex hydrolysate generated during pretreatment, in this case using *Populus* hydrolysate. Genome sequencing used in the study proved to be a useful tool to reveal cellular evolution to stress such as that provided by the composition of *Populus* hydrolysate. Analysis of the mutations included a longitudinal analysis, a pan genomic analysis, and a hotspot analysis. Additional systems biology tools including pfam scores and RNA_seq analysis led to the identification of the multiple mutations most likely to confer tolerance to the *Populus* hydrolysate based of change in function and gene expression. These mutations are located in several simultaneous mechanisms of action, including increases in cellular repair, more efficient molecular pumps, and altered energy metabolism. This study may provide the building blocks for the construction of an industrially robust organism for economical ethanol production.

Chapter 2 Kinetic modeling of batch fermentation for *Populus* hydrolysate tolerant mutant and wild type strains of *Clostridium thermocellum*

This chapter is revised based on a near final draft of a journal article to be submitted to publication by Linville et al.

Jessica L. Linville, Miguel Rodriguez, Jr., Jonathan R. Mielenz, Chris D. Cox. Kinetic modeling of batch fermentation for Populus hydrolysate tolerant mutant and wild type strains of Clostridium thermocellum. Submitted to Bioresource Technology (March 2013).

My primary contributions to this paper include: (i) design and conduction of experiments, (ii) interpretation and analysis of all experimental data, (iii) literature searches and (iv) majority of the writing of this paper.

2.1 Abstract

Lignocellulosic biomass is an abundant and renewable feedstock for liquid fuels; however, its pretreatment produces a large number of inhibitory compounds. A Monod-based model of wild type (WT) and *Populus* hydrolysate tolerant mutant (PM) strains of the cellulolytic bacterium *Clostridium thermocellum* was developed to quantify growth kinetics in standard media and the extent of inhibition to a *Populus* hydrolysate. The PM was characterized by a higher growth rate (μ_{max} = 0.705 vs. 0.642) and less inhibition ($K_{I,gen}$ = 0.877 vs. 0.744) in 10% v/v *Populus* hydrolysate compared to the WT. In 17.5 % v/v hydrolysate inhibition of PM increased slightly ($K_{I,gen}$ = 0.823), whereas the WT was strongly inhibited and did not grow in a reproducible manner. Of the individual inhibitors tested, 4-hydroxybenzoic acid was the most inhibitory, followed by galacturonic acid. The PM did not have a greater ability to detoxify the hydrolysate than the WT.

2.2 Introduction

Lignocellulosic biomass provides an abundant and renewable energy source. Lignocellulosic biomass contains sugars polymerized in the form of cellulose and hemicelluloses, which can be liberated by hydrolysis, and subsequently fermented to ethanol by microorganisms, such as *Clostridium thermocellum* [16]. *C. thermocellum* is a Gram-positive, anaerobic, thermophilic, cellulolytic bacterium that can rapidly solubilize biomass, use cellulose as the sole carbon and energy source, and is under development for biofuel production [33,50,76,105]. *C. thermocellum*'s cellulolytic ability gives it an advantage over organisms that are currently used for bioethanol production (e.g., yeast and *Zymomonas*), which can only ferment nonpolymeric carbohydrates [34,75]. *C. thermocellum* produces a number of industrially important fermentation products in addition to ethanol, including acetic acid, formic acid, lactic acid, and hydrogen [75].

However, rapid and efficient fermentation of the lignocellulosic biomass hydrolysate is a not yet achievable on an industrial scale. Among the limitations is inhibition caused by a range of toxic compounds generated during steam pretreatment and hydrolysis [16]. Inhibitory compounds can be classified according to three main functional groups: weak acids, furan derivatives, and phenolic compounds[16]. Xylose, mannose, acetic acid, galactose, and glucose are liberated from hemicelluloses, and glucose is liberated from cellulose [12]. Furans are the primary aromatic inhibitor formed from sugar degradation during pretreatment [13]. Xylose is further degraded to 2-furfural and hexose is degraded to 5-hydroxymenthyl furfural (HMF) [12]. Galacturonic acid is formed from the degradation of galactose. Phenolic compounds are generated from the partial breakdown of lignin. Vanillin and syringic acid are formed by the degradation of the guaiacylpropane units of lignin and syringyl propane units,

respectively. In addition 4-hydroxybenzoic acid constitutes a large fraction of the lignin-derived compounds in hydrolysates from hardwood poplar [12].

These compounds have potential inhibitory effects, which decreases the ethanol yield and growth rate of the microorganism [12,16]. In order to obtain an economically feasible conversion process, reduction in the inhibitory effect of the toxic compounds is necessary [13]. Current detoxification methods include biological, physical, and chemical methods [16]. Detoxification of pretreated hydrolysates has been shown to improve their fermentability; however, processing options that minimize inhibitor formation coupled with tolerant fermenting microorganisms will likely be more cost effective.

Mathematical modeling is a proven tool in the quantitative analysis of complicated processes such as fermentative growth [61]. Mathematical models that accurately predict biochemical phenomena are essential since the model provides the basis for design, control, optimization and scale-up of process systems [61]. Various approximate kinetic formulations have proven useful to predict the performance of biochemical conversion processes [69]. A kinetic Monod-based model has the ability to accurately describe cell growth and product formation rates [68,69]. Such models can also be extended to include product and toxic compound inhibition [61,62,65].

Improved tolerance to inhibitory compounds found in pretreated biomass hydrolysate should improve the fermentation process and increase economic feasibility of consolidated bioprocessing. Inhibitory compounds have been shown to reduce the rate of ethanol production and the overall yield in *C. thermocellum* [2,7]. The purpose of this study was to quantify the extent of inhibition caused by a *Populus* hydrolysate with respect to growth, cellobiose utilization, and ethanol production of a wild type

(WT) and *Populus* hydrolysate tolerant mutant (PM) strain of *C. thermocellum* in batch culture using a Monod-based model. Studies of the growth on various concentrations of the full hydrolysate and individual inhibitors in the hydrolysate were also included to determine the relative inhibitory effect of various compounds in the hydrolysate. Also, a study was conducted to determine if the increased growth of the PM strain of *C. thermocellum* was due to an improved ability to detoxify the hydrolysate.

2.3 Materials and Methods

2.3.1 Strain and Culture Conditions

Clostridium thermocellum ATCC 27405 was obtained from Prof. Herb Strobel, University of Kentucky collection and denoted as the wild type (WT) strain. A *Populus* hydrolysate tolerant strain (PM) was developed from the WT strain as described in [105]. In all experiments, the cells were grown in media for thermophilic clostridia (MTC) with a substrate loading of 5 g/L cellobiose. The media was composed of 0.336 g/L potassium chloride [KCl], 0.25 g/L ammonium chloride [NH_4Cl], 1.00 g/L magnesium sulfate heptahydrate [$MgSO_4 \cdot 7H_2O$], 1.70 g/L potassium phosphate [KH_2PO_4], 0.50 g/L MOPS [$C_7H_{14}NO_4S$], 0.15 g/L calcium chloride dehydrate [$CaCl_2 \cdot 2H_2O$], 1.75 g/L Trisodium citrate dehydrate [$Na_3C_6O_7 \cdot 2H_2O$], 0.6 g/L urea [CH_4N_2O], 1.00 g/L L-cysteine HCL, 0.30 mg/L resazurin, 2.0 mL of 1000x MTC minerals and 1.25 mL of 50x MTC vitamins [79,80]. All chemicals were reagent grade and obtained from Sigma (St. Louis, MO, USA), unless otherwise indicated. The inoculum culture was grown in Balch (Bellco Glass, Inc., Vineland, NJ) tubes containing 9.5 mL of 5 g/L cellobiose MTC and 0.5 mL of the frozen stock culture. Cultures were allowed to reach exponential growth phase by incubation at 58^{o}C and 100 rpm, diluted to an optical density (OD_{600}) of 0.250 using sterile, anaerobic Milli-Q water and

injected into the test conditions with an inoculum size of 10% by volume [105]. Milled *Populus trichocarpa* hydrolysate was pretreated at the National Renewable Energy Lab (NREL) using a 20% w/w solid loading and dilute concentrations of H_2SO_4 at temperatures of 165-195^{o}C [81]. Solids were removed by filtration. The *Populus* hydrolysate was adjusted to a pH of 7.0 using 50% w/w NaOH and filter sterilized before being added to the MTC media.

2.3.2 Fermentation

Batch fermentations were conducted in triplicate in 1.5 L Q-plus jacketed glass fermentors (Sartorius Stedim Biotech, Bohemia, NY) using a 1 L working volume of MTC medium at 58^{o}C and stirred at a rate of 300 rpm, with pH controlled to 7.0 using 3N NaOH. Fermentors containing only 5g/L cellobiose were sparged with a filtered 20% CO_2/80% N_2 gas mixture and vigorously agitated overnight, followed by addition of the remaining medium components and *Populus* hydrolysate. *Populus* hydrolysate was added to the fermentors at 0%, 10% or 17.5% v/v concentration. The fermentors were then sparged for an additional 4 hours with a 20% CO_2/80% N_2 gas mixture. The inoculum was a bottle culture grown overnight (11-13 hrs) in 5g/L cellobiose, diluted to an OD_{600} of 0.200 $\pm$ 0.013, and added as 10% v/v to inoculate the fermenters. Well-mixed 5mL aliquots of culture were harvested at regular intervals. Cell growth was monitored based on an increase in the OD_{600} of the culture blanked against a sterile sample of the same concentration of Populus hydrolysate containing media that was stored in an incubator at 58 ^{o}C. The OD_{600} of the culture was measured in triplicate by a Spectramax Plus 384 spectrophotometer (Molecular Devices, Sunnyvale, CA). The corresponding dry cell weight was obtained from a calibration curve. Metabolite analysis was performed using HPLC. Metabolites were separated at a flow rate of 0.5 mL/min in 5 mM H_2SO_4 using an Aminex HPX-87H column (Bio-Rad Laboratories, Inc, Hercules, CA).

2.3.3 Parameter Estimation

The lag time was determined from a plot of log(biomass) as a function of time and removed from the experimental data before parameter optimization. The lag time was the same for the WT and PM strains; 4 hours in standard media and 6 hours in hydrolysate media. Model simulations were conducted in Matlab 7.10.0 (R2010a) (The Mathworks, Inc., Natick MA). The model parameters were optimized using simulated annealing to minimize the weighted least square errors for the biomass, substrate and fermentation products at each time point of experiment in standard media. In hydrolysate media the weighted least square errors for furfural and 5-HMF was also minimized to determine values of the inhibitor biodegradation parameters. The Matlab files used for these calculations can be found in Appendix B.

2.3.4 *Populus* hydrolysate tolerance

The WT and PM strains were cultured in cellobiose MTC media containing 0%, 10%, 20%, and 30% v/v *Populus* hydrolysate in 25 mL Balch tubes. The OD_{600} was measured directly (without sampling) using a Spectronic21D (Milton Roy, Ivyland, PA). The OD_{600} was blanked using a sterile sample of media with the same concentration of Populus hydrolysate stored in the incubator with the samples. The OD_{600} was used to determine the increase in cell growth over time.

2.3.5 Individual inhibitor tolerance

The WT and PM strains were cultured in cellobiose MTC media containing a single inhibitor at a concentration equivalent to its concentration in 17.5% and 35%v/v *Populus* hydrolysate. The six individual inhibitors chosen were furfural, 5-hydroxymethlyfurfural, vanillin, syringic acid, galacturonic acid, and 4-hydroxybenzoic acid (Sigma-Aldrich Co., St. Louis, MO). In addition, the six inhibitors were combined at the appropriate concentrations to make a 17.5% and 35% v/v model hydrolysate. The preparation of individual inhibitor media and model hydrolysate media was the same as the *Populus* hydrolysate media above except the pH was not adjusted for the individual inhibitor media. Samples were taken at 12, 24, and 48 hours for OD_{600} and product concentrations. Measurement of the OD_{600} of the culture and metabolite analysis was performed as described above.

2.3.6 Detoxification

The ability of the WT and PM strains to detoxify *Populus* hydrolysate was tested by conducting fermentation re-growth experiments in media obtained from previous fermentations. To produce the media for the re-growth experiments, the WT and PM strains were grown uninterrupted for 72 hours in 17.5% v/v *Populus* hydrolysate with 5 g/L cellobiose-MTC media. The concentration of the metabolites and inhibitors were measured before (T= 0 hr) and after (T= 72 hr) growth by HPLC. Cells were removed from the liquid media after the 72 hour growth period by 0.2 µm filtration. The concentration of sugars and inhibitors were adjusted for both the PM and WT detoxified liquids so that the concentrations were roughly the same in both media. Both the PM and WT were re-grown in each of these detoxified media. Biological triplicate experiments were conducted except for the WT-WT which was conducted in

duplicate because one of the series did not grow. Measurements of the OD_{600} of the culture and metabolite analysis were performed as described above.

2.4 Results and Discussion

2.4.1 Modeling of Fermentative Growth

Batch fermentations were conducted in fermentors to compare the performance of the PM and WT *C. thermocellum* strains in different concentrations of hydrolysate. The PM was grown in three test conditions: standard media (0% v/v *Populus* hydrolysate), 10% v/v and 17.5% v/v *Populus* hydrolysate. The WT was grown in two test conditions: standard media and 10% v/v *Populus* hydrolysate. Attempts to grow the WT strain in the presence of 17.5% v/v *Populus* hydrolysate resulted in a severely inhibited and non-reproducible growth pattern (data not shown). Samples were taken at regular intervals from each fermentation unit based on their growth rate.

Most models of biochemical processes are empirical and based on Monod's equations [61]. In this work, the Monod's model as formulated below was used to fit the experimental results:

$$\frac{dX}{dt} = (\mu - k_D) * X$$

$$\frac{dS}{dt} = -\frac{\mu * X}{Y_{X/S}}$$

$$\frac{dP_E}{dt} = \frac{Y_{E/S} * \mu * X}{Y_{X/S}}$$

$$\frac{dP_A}{dt} = \frac{Y_{A/S} * \mu * X}{Y_{X/S}}$$

where X is the cell biomass concentration; μ is the specific growth rate; k_D is the first-order cell decay constant; S is the substrate concentration; P_E and P_A are the product concentrations of ethanol and acetate, respectively; and $Y_{X/S}$, $Y_{E/S}$ and $Y_{A/S}$ are the yields of biomass, ethanol, and acetate per unit substrate consumed, respectively. For growth in standard conditions, the specific growth rate is described by Monod's equation:

$$\mu = \mu_{max} * \left(\frac{S}{k_S + S}\right)$$

where μ_{max} is the maximum specific growth rate, and k_S is the half saturation concentrations for substrate [61,62]. Ethanol inhibition was not included in the model since it has been shown that ethanol inhibition is not significant at ethanol concentrations below low 1 g/L [106] and including this process did not improve the model.

The first-order cell decay constant, k_D, was estimated directly from the decay portion of the growth curves of each strain in standard conditions by:

$$k_D = \frac{\ln\left(\frac{X_{max}}{X_{final}}\right)}{\Delta t}$$

Where X_{max} is the maximum biomass concentration and X_{final} is the biomass concentration at the final time point and Δt is the difference in time between the two biomass concentrations (Figure 11). The cell decay constant was calculated as 0.12 h^{-1} for both strains. The decay constant is of the same order as another model of *C. thermocellum* grown on crystalline cellulose (Avicel) [107]. Initial guesses of the five model parameters (μ_{max}, k_s,$Y_{X/S}$,$Y_{E/S}$, and $Y_{A/S}$) were determined by manual adjustment of the parameters to obtain a reasonable visual fit to the data. The simulated annealing function in Matlab was used to optimize the parameters (P) by minimizing the weighted least square errors:

$$min\chi^2(P) = \sum_i \sum_j \left(\frac{\left(\overline{y_{i,j}} - f_j(\mathbf{Y_i}, \mathbf{P})\right)}{\sigma_{i.j}} \right)^2$$

Where $\bar{y}_{i,j}$ is the averaged value, $f_j(\mathbf{Y_i}, \mathbf{P})$ is the calculated value, and $\sigma_{i,j}$ is the sample standard deviation of variable j at time point i [108]. Values of $f_j(\mathbf{Y_i}, \mathbf{P})$ were calculated using the ode45 function in Matlab. The confidence interval and codependence of each variable were determined from the covariance matrix. Derivative terms in the covariance matrix were determined numerically.

The optimized parameters and the 95% confidence intervals for both strains in standard media are shown in Table 6. The optimal variables for the WT and PM strains can be seen in Figure 11. The model was able to accurately describe most of the variance found in the data as evidenced by the R^2 values of the entire system (Table 6). The model parameters suggest that the PM mutant has a twice specific growth rate, μ_{max}, which could be due to the mutation in the Spo0A homologue (Cthe_3087)

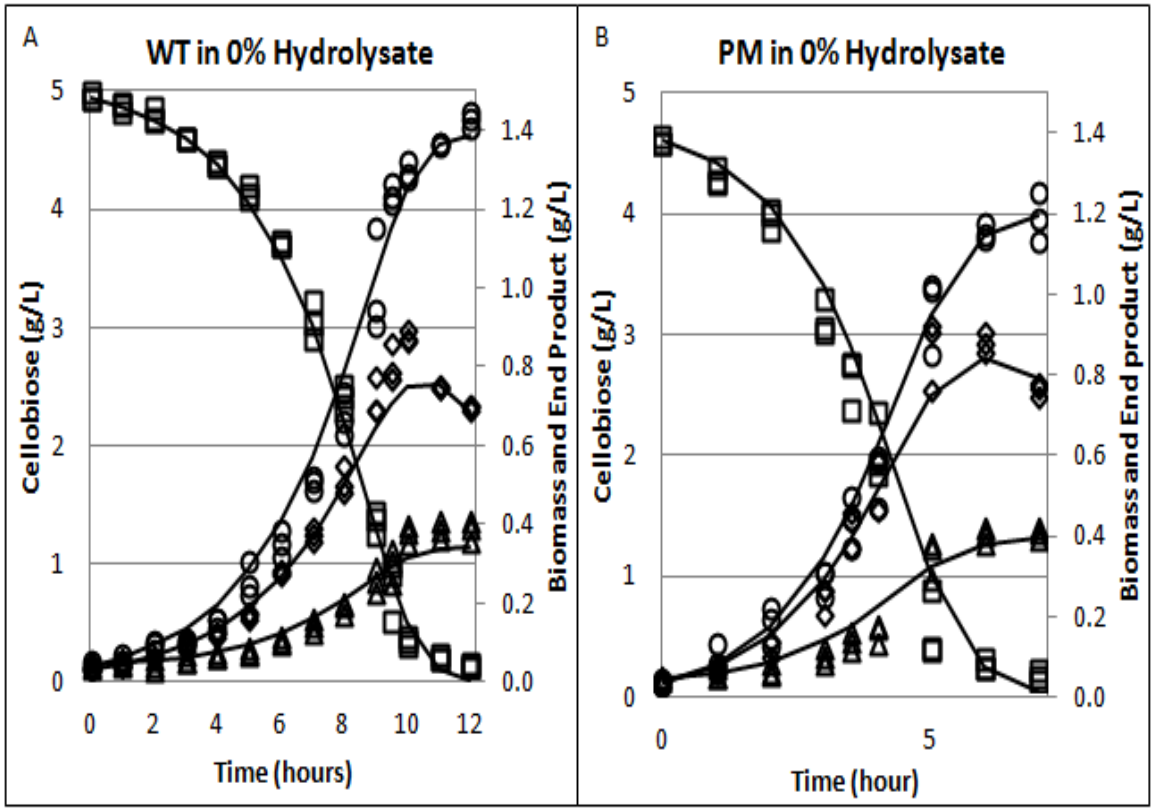


Figure 11: Comparison of model-calculated and experimental results.

(A) WT in 0% v/v Populus hydrolysate and (B) PM in 0% v/v Populus hydrolysate for cellobiose utilization([]), biomass formation(◊), and ethanol (Δ), and acetic acid (o) production. Experimental data are represented by points; model results are represented by solid lines.

Table 6: Optimal parameter estimates for Monod model in standard and hydrolysate conditions.

	μ_{max}	K_S	$Y_{X/S}$	$Y_{E/S}$	$Y_{A/S}$	$K_{I,gen}$	$K_{I,HMF}$	$K_{I,F}$	R^2
	(h^{-1})	(g/L)	(g/g)	(g/g)	(g/g)	(h^{-1})			
WT in 0% hydrolysate	0.569 ± 0.007	0.901 ± 0.044	0.234 ± 0.002	0.062 ± 0.004	0.273 ± 0.002	--	--	--	0.995
WT in 10% hydrolysate	0.569 ± 0.007	0.901 ± 0.044	0.234 ± 0.002	0.094 ± 0.007	0.330 ± 0.026	0.758 ± 0.003	2.013 ± 0.059	11.057 ± 1.072	0.969
PM in 0% hydrolysate	1.289 ± 0.040	2.499 ± 0.140	0.245 ± 0.008	0.077 ± 0.005	0.254 ± 0.009	--	--	--	0.974
PM in 10% hydrolysate	1.289 ± 0.040	2.499 ± 0.140	0.245 ± 0.008	0.102 ± 0.009	0.307 ± 0.055	0.985 ± 0.002	2.304 ± 0.154	9.248 ± 0.778	0.978
PM in 17.5 % hydrolysate	1.289 ± 0.040	2.499 ± 0.140	0.245 ± 0.008	0.115 ± 0.003	0.429 ± 0.013	0.881 ± 0.002	2.048 ± 0.026	10.416 ± 0.106	0.953

Spo0A mutants, which are unable to go into sporulation, are thought to be locked in exponential growth since they continue to grow under nutritional conditions that would normally induce sporulation. Essentially they appear to maintain growth until the nutrients are exhausted, whereupon cell lysis occurs [100]. The PM has a mutation in the non-coding region 127 bp upstream of the first gene in the operon (Cthe_2602-09) which codes for the ATPase complex [105]. The ATPase is believed to help with the electron flux in the cell by pumping protons across the cell membrane [49]. The biomass yield, $Y_{X/S}$, is similar between the two strains and similar to a recent model of *C. thermocellum* grown on crystalline cellulose (Avicel) [107]. This further suggests that the increased tolerance is due to a faster growth rate for the PM. The PM strain also has a number of mutations affecting a genes related to DNA repair (Cthe_2376), and RNA transcription (Cthe_2724), translations (Cthe_2727) and degradation (Cthe_0158) [105] . These mutations may allow the PM to conserve energy; therefore grow faster. The lower yield of acetate, $Y_{A/S}$, and higher yield of ethanol, $Y_{E/S}$, might be caused by the mutation in the non-coding region upstream of the Cthe_0422-3 operon which encodes the Rex (redox) repressor and adhE [96]. This may cause a switch from acetic acid production to ethanol production [105].

Experiments with hydrolysate could be modeled using the same Monod model adapted to include a general inhibition factor $k_{I,gen}$ ($0 \leq k_{I,gen} \leq 1$) as follows:

$$\mu = \mu_{max} * \left(\frac{S}{k_S + S}\right) * k_{I,gen}$$

Inhibition terms that were dependent on the concentration of the hydrolysate or of individual inhibitors were not found to describe the data well. The WT and PM strains have the ability to degrade two of the inhibitors in the hydrolysate that can be easily monitored during the growth period: furfural and HMF [65,72]. The furfural reduction rate has been shown to increase with increasing specific growth rate and

increasing furfural concentrations [12]. Therefore, degradation was modeled as a pseudo-second order rate equation by being a first order equation with respect to both inhibitor concentration and biomass as follows:

$$\frac{dI_{HMF}}{dt} = -k_{I,HMF} * I_{HMF} * X$$

$$\frac{dI_F}{dt} = -k_{I,F} * I_F * X$$

where I_{HMF} and I_F are the inhibitor concentrations of HMF and furfural, respectively; and $k_{I,HMF}$ and $k_{I,F}$ are the kinetic constants for each compound. The matrix from the standard condition optimized model was supplied to the hydrolysate model along with a new matrix of guessed values of inhibition. The parameters were optimized in a similar manner where the weighted least square errors of the six variables (biomass, substrate, ethanol, acetate, HMF, and furfural) were minimized. When the Monod parameters were held constant, the model consistently under predicted the formation of end products; therefore, the model was also adjusted to allow for possible changes in product yields. All other model parameters estimated from the 0% hydrolysate experiments (μ_{max}, k_S, and $Y_{X/S}$) were held constant.

The optimal parameters estimated for the hydrolysate conditions are shown in Table 6 and the model-calculated profiles for the WT and PM strains using these parameters are shown in Figure 12. The model accurately describes the experimental data based on the R^2 value for the entire system (Table 6). Less inhibition is indicated as values of $k_{I,gen}$, increase. As to be expected, the WT in 10% hydrolysate is

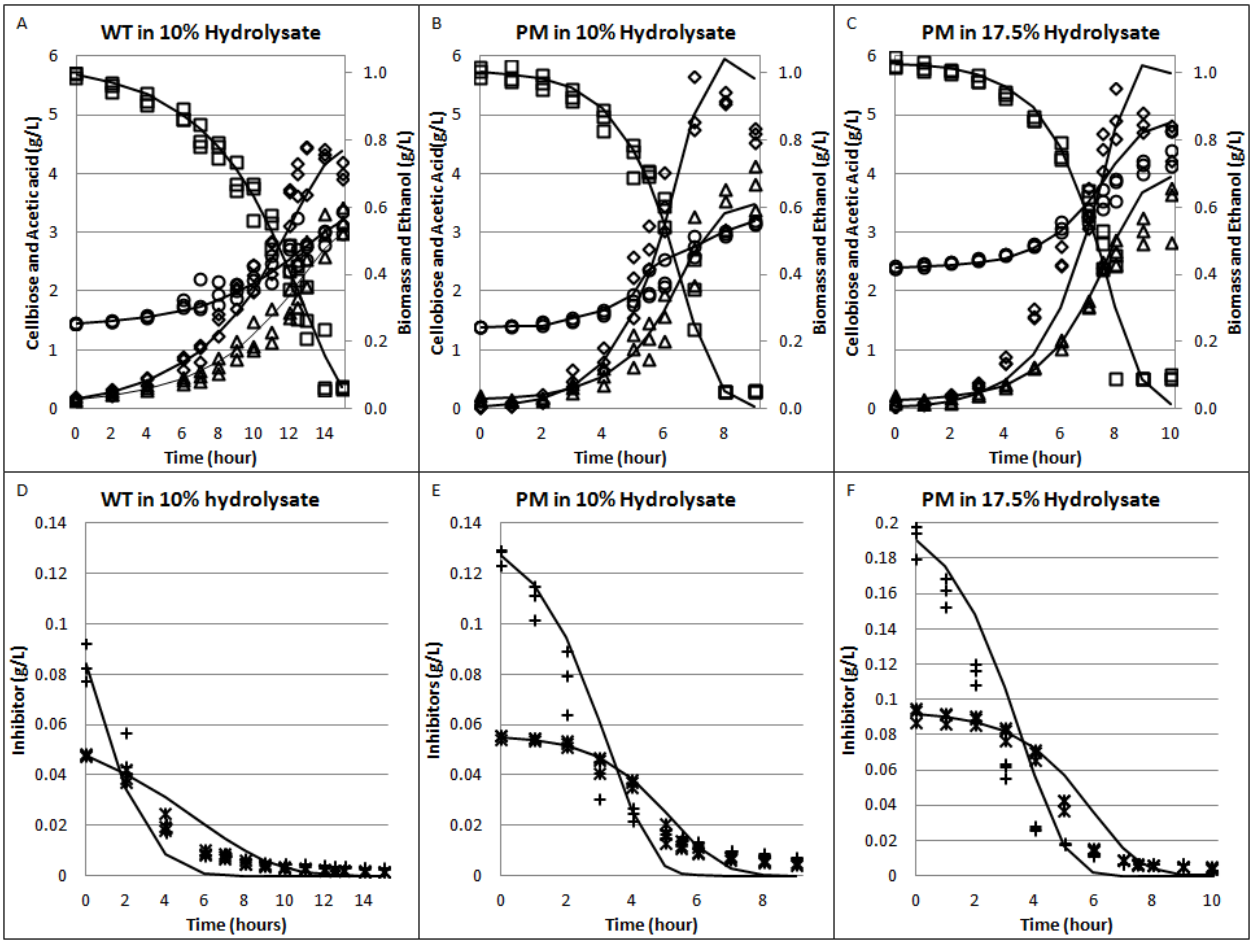


Figure 12: The model-calculated and experimental results for hydrolysate conditions.

(A) WT in 10% v/v *Populus* hydrolysate, (B) PM in 10% v/v *Populus* hydrolysate and (C) PM in 17.5% v/v *Populus* hydrolysate for cellobiose utilization([]), biomass formation(◊), and ethanol (Δ), and acetic acid (o) production. The estimated and experimental results for (D) WT in 10% v/v *Populus* hydrolysate, (E) PM in 10% v/v *Populus* hydrolysate, and (F) PM in 17.5% v/v *Populus* hydrolysate for furfural (+) and HMF (*). Experimental data are represented by points; model results are represented by solid lines.

the most inhibited, followed by the PM in 17.5% hydrolysate, and PM in 10% hydrolysate. The degradation rate constants $k_{I,HMF}$ and $k_{I,F}$ are very similar for the WT and PM (Table 6) which suggests that the PM does not have a greater ability to detoxify the hydrolysate. Comparison of the lag time in biomass growth versus the degradation of the inhibitors for the WT and PM further suggests that the WT dedicates more resources to the degradation of furfural and HMF than the PM does. Both the WT and PM reach the exponential growth phase 6 hours into the fermentation. At the end of the lag phase, the WT in 10% v/v *Populus* hydrolysate has degraded 40% of the furfural and 12% of the HMF. The PM in 10% v/v *Populus* hydrolysate only degraded 12% of the furfural and 5% of the HMF in the lag phase. In 17.5% v/v *Populus* hydrolysate, the PM degrades 23% of the furfural and 7% of the HMF by the end of the lag phase. Furthermore, the WT in 10% v/v *Populus* hydrolysate does not reach a biomass concentration of greater than 0.1 g/L until approximately 4.6 hours after it reaches the exponential growth phase; at that point 90% of the furfural and 71% of the HMF have been degraded. The PM reaches a biomass concentration of greater than 0.1 g/L in approximately 3.25 and 3.5 hours after it reaches the exponential growth phase in 10 and 17.5% v/v *Populus* hydrolysate, respectively. At approximately 3.25 hours into the fermentation, the PM has only degraded 68% of the furfural and 23% of the HMF in 10% v/v *Populus* hydrolysate, and 81% of the furfural and 27% of the HMF in 17.5% v/v *Populus* hydrolysate. Similarly, increases furfural tolerance in an *E. coli* EMFR9 strain was accompanied by a decreased in the rate of furfural reduction in vivo and a decrease in the NADPH-dependent furfural reductase activity in vivo [109]. Reducing the NADPH-dependent furfural reducase activity permitted increased growth in the presence of furfural by slowing the depletion of NADPH that is required for biosynthesis [109].

Furthermore, furfural and HMF directly inhibit adhE, resulting in NAD(P)H depletion due to their reduction to their corresponding alcohols thereby affecting glycolysis and TCA fluxes [29,32,110]. The

PM has a mutation in the non-coding region upstream of the Cthe_0422-3 operon which encodes the Rex (redox) repressor and adhE [105]. The mutation resulted in a slightly higher expression level of both genes in all test conditions for the PM whereas the WT had significant decreased expression of these genes in hydrolysate media [105]. Therefore, it is surprising that both strains similarly increased the ethanol production in the *Populus* hydrolysate conditions. Comparative proteomics indicate that the central carbon metabolism, levels of alcohol dehydrogenase, the redox balance, and other general stress responses are needed to tolerate furfural inhibition [110]. The increased ethanol production for the WT may be due to the need to balance the electron flow while converting the furfural and HMF into their corresponding alcohols. However, since the WT lacks the mutation of the PM, the increased ethanol may come at the expense of other cellular functions.

The codependence between the optimized parameters, as shown by their correlation, was determined from the covariance matrix. The correlation matrix for the WT in 0% and 10% v/v *Populus* hydrolysate can be seen in Table 7. The maximum specific growth rate, μ_{max}, the half saturation constant, k_S, and the yield of ethanol, $Y_{E/S}$ are all strongly correlated among one another. The parameters in the hydrolysate media calculations are less correlated. The furfural degradation constant, k_F, is correlated to the yield of both ethanol, $Y_{E/X}$, and acetate, $Y_{A/S}$. The PM has similar correlation between the parameters for the standard and hydrolysate calculations (data not shown).

2.4.2 *Populus* Hydrolysate tolerance

In order to further determine the inhibitory effects of the hydrolysate, both strains were grown in selected concentrations of *Populus* hydrolysate (0%, 10%, 20%, and 30%v/v). Figure 13A shows the

Table 7: Covariance matrix for optimized parameters of the WT in 0 and 10% hydrolysate media.

WT in 0% hydrolysate					
	μ_{max}	K_S	$Y_{X/S}$	$Y_{E/X}$	$Y_{A/X}$
μ_{max}	1.00	0.90	0.37	-0.93	-0.07
K_S	0.90	1.00	0.17	-0.98	0.14
$Y_{X/S}$	0.37	0.17	1.00	-0.32	0.01
$Y_{E/X}$	-0.93	-0.98	-0.32	1.00	-0.19
$Y_{A/X}$	-0.07	0.14	0.01	-0.19	1.00
WT in 10% hydrolysate					
	$K_{I,HMF}$	$K_{I,F}$	$K_{I,gen}$	$Y_{E/X}$	$Y_{A/X}$
$K_{I,HMF}$	1.00	-0.20	-0.30	-0.05	-0.06
$K_{I,F}$	-0.20	1.00	0.12	-0.58	-0.60
$K_{I,gen}$	-0.30	0.12	1.00	-0.05	-0.04
$Y_{E/X}$	-0.05	-0.58	-0.05	1.00	-0.25
$Y_{A/X}$	-0.06	-0.60	-0.04	-0.25	1.00

The covariance matrix shows strong correlation between a number of parameters. The PM has similar correlations between the optimized parameters (data not shown).

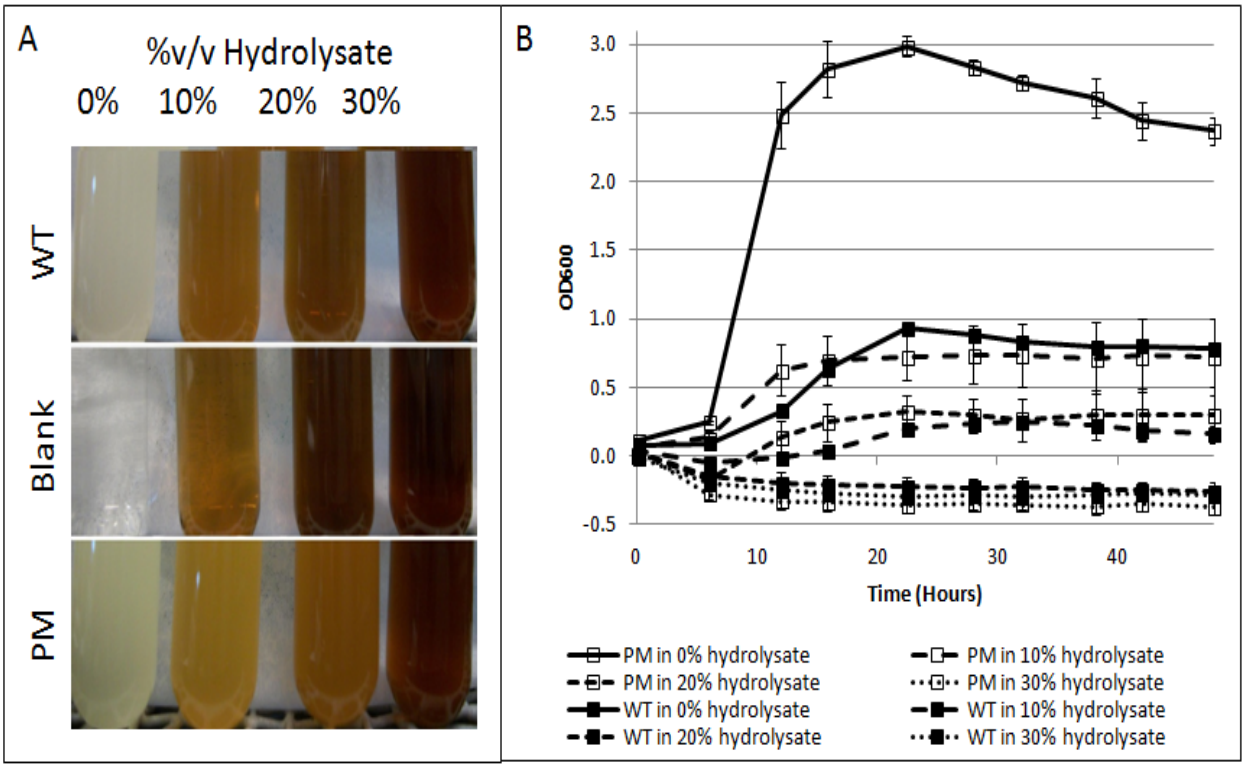


Figure 13: Growth in various concentrations of Populus hydrolysate.

(a) WT and PM strain test tube cultures at 24 hours in various concentrations of hydrolysate. (b) OD_{600} of the PM ([]) and WT (■). Solid line is 0% v/v *Populus* hydrolysate, long dash line is 10% v/v *Populus* hydrolysate, short dash line is 20% v/v *Populus* hydrolysate and dotted line is 30% v/v *Populus* hydrolysate. Error bars represent standard deviation of triplicate results.

change in culture appearance of both strains following 24 hours of fermentation in the various concentrations of *Populus* hydrolysate while Figure 13B shows the OD_{600} over time. The *Populus* hydrolysate is a dark brown color whose intensity increased with concentration. As the strain grows in the presence of hydrolysate, the color lightens due to degradation of some of the compounds and becomes milky as the biomass increases. This color change accounts for the negative OD_{600} seen at the higher concentrations of hydrolysate. For any given concentration of hydrolysate, the PM has a higher OD_{600} than the WT strain. The largest difference is in standard media (0% v/v hydrolysate) with the PM having 3-times greater OD_{600} than the WT strain. The WT strain is completely inhibited at 20% v/v *Populus* hydrolysate whereas the PM strain is completely inhibited at 30% v/v *Populus* hydrolysate.

2.4.3 Individual Inhibitors

In order to determine the inhibition contributed to the hydrolysate by the individual compounds, both the WT and PM strain of *C. thermocellum* were grown in six individual inhibitors (furfural, 5-HMF, syringic acid, vanillin, galacturonic acid, and hydroxybenzoic acid) at concentrations equal to those found in a media containing 17.5% and 35% v/v *Populus* hydrolysate. The six inhibitors were combined to make a model hydrolysate representative of the 17.5% v/v and 35% v/v concentrations. Table 8 lists the concentrations of individual and combined inhibitors in the media. Samples were taken at 12, 24 and 48 hours after inoculation. The optical density, cellobiose consumption, and ethanol and acetic acid production at each time point are shown in Figure 14. Overall, the PM is less inhibited in all conditions than the WT strain as demonstrated by its greater OD_{600}, cellobiose consumption, and ethanol production. However, the PM strain produced roughly the same amount of acetic acid, which could be due to mutations in the PM strain in the non-coding region upstream of Cthe_0442-3 as discussed above [105]. There is no significant reduction (less than 10%) in

Table 8: Concentration of compounds for Individual Inhibitor Study.

	Cellobiose	Glucose	Furfural	5-HMF	Syringic Acid	Vanillin	Galacturonic Acid	Hydroxybenzoic Acid
	g/L	g/L	mg/L	mg/L	mg/L	mg/L	g/L	g/L
0% Populus Hydrolysate	4.97	0.04						
150 mg/L Furfural	4.66	0.02	108.6					
300 mg/L Furfural	4.67	0.02	233.8					
75 mg/L 5-HMF	4.81	0.02		67.6				
150 mg/L 5-HMF	4.75	0.03		136.0				
50 mg/L Syringic Acid	5.38	0.02			50.4			
75 mg/L Syringic Acid	4.90	0.02			73.0			
25 mg/L Vanillin	4.61	0.02				24.2		
50 mg/L Vanillin	4.77	0.02				50.4		
2 g/L Galacturonic Acid	5.23	0.03					2.2	
4 g/L Galacturonic Acid	5.55	0.06					4.3	
1.5 g/L Hydroxybenzoic Acid	5.14	0.03						1.7
3 g/L Hydrolxybenzoic Aicd	4.37	0.03						2.8
17.5% v/v Model hydrolysate	6.24	4.02	165.5	54.8	50.0	25.0	2.6	1.9
35% v/v Model hydrolysate	7.46	7.82	314.8	121.6	75.0	50.0	4.5	3.6
17.5% v/v *Populus* Hydrolysate	5.98	3.98	193.6	78.5	40.0	15.0	1.1	0.2

Concentrations of individual inhibitors representing 17.5% and 35% v/v *Populus* hydrolysate, combined inhibitor concentration for 17.5% and 35% v/v model hydrolysate and the actual concentration in 17.5% v/v *Populus* hydrolysate where used to determine the toxicity of the individual compounds

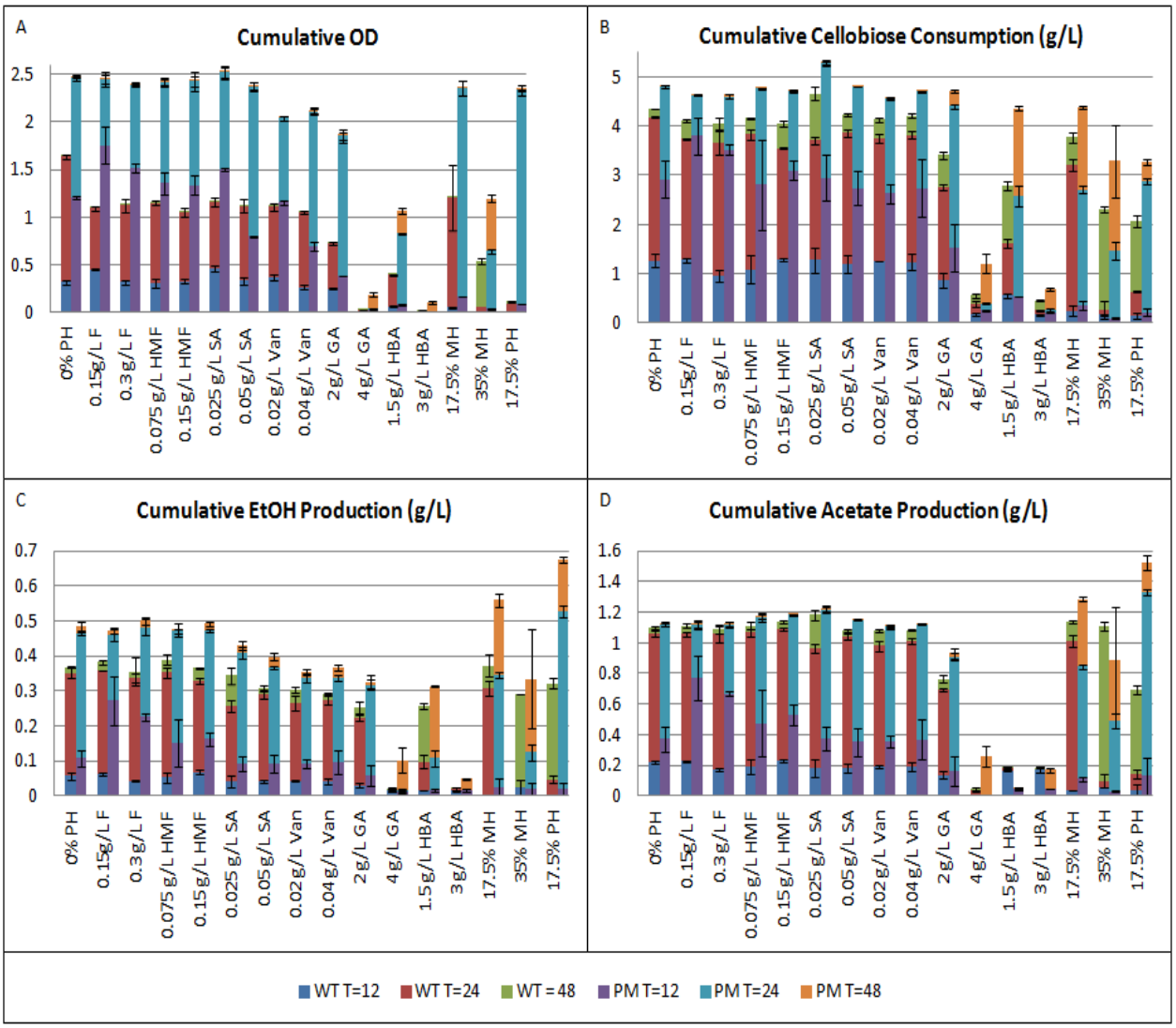


Figure 14: Effect of various inhibitors on fermentation.

For each inhibitor, stacked bars represent cumulative consumption or production at 12, 24, and 48 hours. (A) $OD_{600,}$ (B) cellobiose consumption, (C) ethanol production, (D) acetate production.

OD_{600} for the PM in the furfural, 5-HMF, or syringic acid conditions when compared to the OD_{600} of the PM in standard media based on the average maximum value over the time course. Also, there is no significant reduction in OD_{600} for the model or actual 17.5% v/v hydrolysate conditions. There is a slight reduction (10-25%) in OD_{600} in both vanillin and the lower concentration of galacturonic acid, moderate reduction (25%-75%) in OD_{600} for the lower concentration of hydroxybenzonic acid and the 35% v/v model hydrolysate, and sever reduction (greater than 75%) in the higher concentration of galacturonic acid and hydrolxybenzoic acid. For the WT strain, furfural, 5-HMF, syringic acid, vanillin and the 17.5% v/v model hydrolysate all caused similar moderate reductions in OD_{600}. There is also moderate reduction in OD_{600} for the lower concentrations of galacturonic acid. There is severe reduction in growth for both hydroxylbenzoic acid concentrations, the 35% v/v model hydrolysate, the higher concentration of galacturonic acid, and the 17.5% v/v *Populus* hydrolysate.

Both the WT and PM strain may be able to grow better in the model hydrolysate than the galacturonic acid and hydroxybenzoic acid conditions likely because the solution was neutralized to a pH of 7.0 in the model hydrolysate before the addition of the other media components. Furthermore, based of the growth of the WT in the 17.5% v/v model hydrolysate and actual hydrolysate, the six inhibitors chosen do not represent the full inhibitory effect of the hydrolysate. Similar trends were seen in the cellobiose consumption and ethanol production. The PM consumes the smallest amount of cellobiose per unit growth in the 17.5% v/v *Populus* hydrolysate. One possible explanation is that the *Populu*s hydrolysate triggers a stress response in which non-essential pathways are turned off to conserve energy. The highest amount of ethanol and acetic acid produced was in the 17.5% v/v *Populus* hydrolysate followed by the 17.5% v/v model hydrolysate concentrations. Although, the cellobiose consumption is low, the high ethanol and acetic production could be triggered by the need to balance the electron flow to increase the availability of ATP as the cell responds to the inhibitors in the

hydrolysate [105]. The WT strain lacks the ability to increase ethanol and acetic acid production under these conditions which may explain its greater level of inhibition.

2.4.4 Detoxification

Biotransformation of the inhibitory compounds in the hydrolysate (detoxification) was investigated as a mechanism of tolerance by re-growth experiments in which the PM and WT strains were grown in media reused from earlier fermentation experiments. The initial hydrolysate media (T=0) contained 5.9 g/L cellobiose, 4.1 g/L glucose and 5g/L acetic acid. Of the various inhibitors, the two that can be most readily measured by HPLC are furfural and 5-hydroxymethalfurfurl (5-HMF) with initial concentrations of 58 and 163 mg/L, respectively. During the 72 hour preparatory fermentation the PM strain utilized 3.7 g/L cellobiose and 2.4 g/L glucose and produced 0.14 g/L lactic acid, 0.76 g/L ethanol and 3.38 g/L acetic acid. However, the WT strain utilized only 1.9 g/L cellobiose and 0.9 g/L glucose and produced 0.1 g/L lactic acid, 0.25 g/L ethanol and 1.46 g/L acetic acid. Both strains degraded almost all of the furfural and 5-HMF present in the hydrolysate. The cells were removed via filter sterilization with a 0.2 μm filter and the metabolites were adjusted to roughly the same concentration; 3.75 g/L cellobiose, 2.86 g/L glucose, 0.09 g/L lactic acid, and 0.47 g/L ethanol. The acetic acid concentration was not adjusted. Both the WT and PM strains were grown in the adjusted media detoxified by either the WT or PM strain. Figure 15A - Figure 15D shows the cell growth, glucan utilization, ethanol production, and acetic acid production over time for each of the four regrowth experiments. The initial ethanol and acetic acid concentrations were subtracted from the final concentrations measured at the end of the experiment for better comparison. Neither strain produced significant concentrations of lactic acid. The fermentation profiles of both strains showed very little dependence on the type of media (WT-detoxified or PM-detoxified) used for the re-growth experiment. The PM had approximately twice the

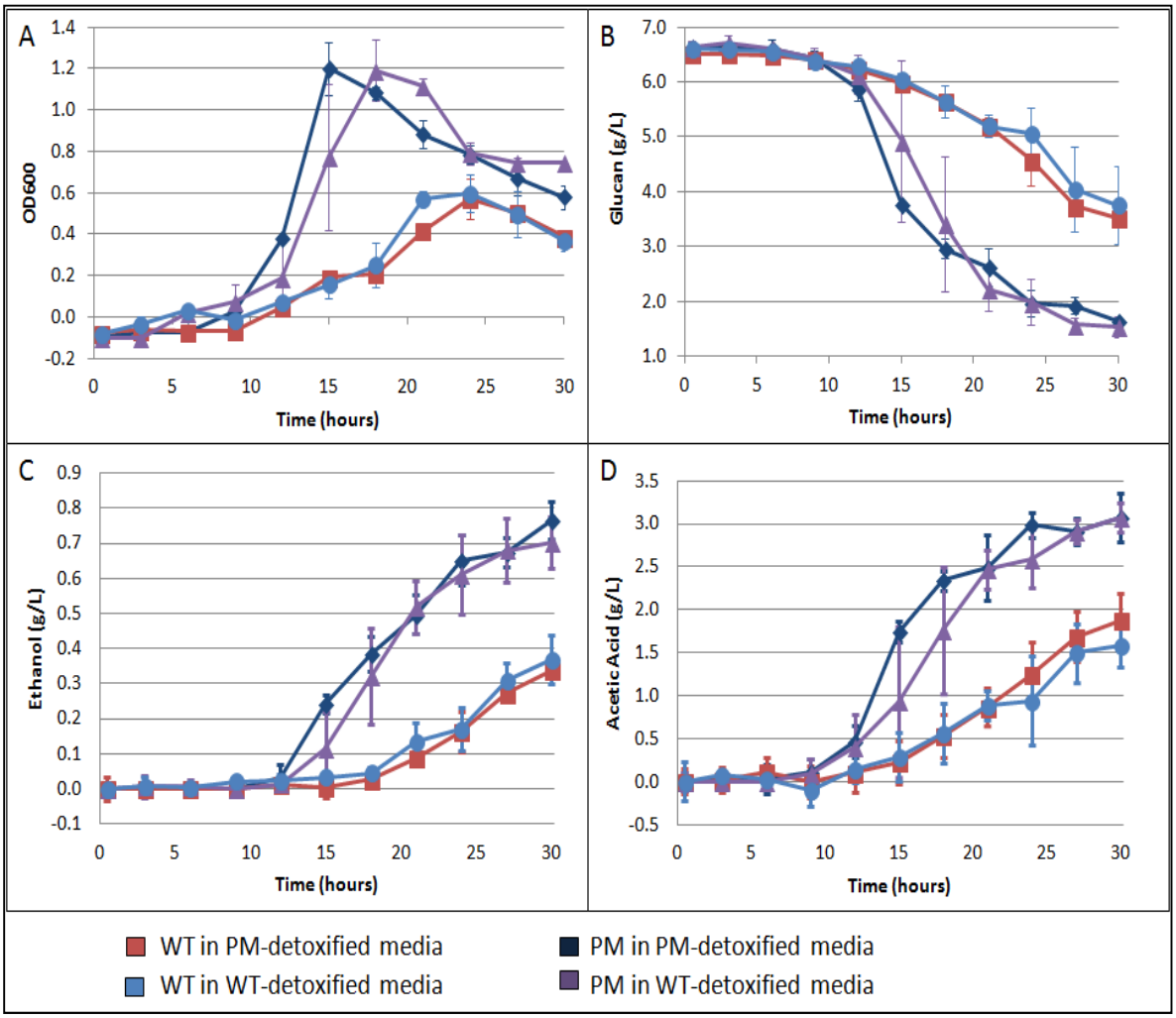


Figure 15: Re-growth detoxifixcation experiments.

A) biomass growth shown as OD_{600}, B) glucan utilization, C) ethanol and D) acetic acid production showing equal detoxification for the WT and PM strain.

maximum optical density of the WT strain did and reached it sooner (15 hours compared to 24 hours). Furthermore, in both conditions the PM utilized approximately 2.9 g/L cellobiose, 2.1 g/L glucose and produced 0.73 g/L ethanol and 3.0 g/L acetic acid. The WT strain utilized 2.7 g/L cellobiose but only 0.2 g/L glucose and produced 0.35 g/L ethanol and 1.8 g/L acetic acid. The WT strain performed better in the detoxified hydrolysate than in 17.5% v/v *Populus* hydrolysate; however, the PM did not (data not shown). This may be due to the higher acetic acid concentration in the PM detoxified media. The results show that the PM does not have a better ability to detoxify the hydrolysate and that this is not the primary mechanism of tolerance.

2.4. Conclusions

The results from batch fermentations of *C. thermocellum* on cellobiose were successfully simulated using Monod kinetics. The PM had a higher maximum growth rate and was less inhibited by ethanol and the hydrolysate than the WT strain; however, the PM is less inhibited by individual compounds than the WT strain. This study also found that the PM does not have a greater ability to detoxify the *Populus* hydrolysate than the WT. Therefore, the increased tolerance to the hydrolysate is apparently due to the combined effects of several mutations in the PM that increased its growth rate [105].

Chapter 3 Transcriptomic analysis of *Clostridium thermocellum Populus* hydrolysate- tolerant mutant and wildtype strains

This chapter is revised based on a draft of a journal article to be submitted to publication by Linville et al.

Jessica L. Linville, Miguel Rodriguez, Jr., Steve D. Brown, Jonathan R. Mielenz, Chris D. Cox. Transcriptomic analysis of *Clostridium thermocellum Populus* hydrolysate- tolerant mutant and wildtype strains. Will be submitted to BCM Microbiology (May 2013).

My primary contributions to this paper include: (i) design and conduction of experiments, (ii) interpretation and analysis of all experimental data, (iii) literature searches and (iv) majority of the writing of this paper.

3.1 Abstract

Clostridium thermocellum is a Gram-positive, anaerobic, thermophilic, cellulolytic bacterium that can rapidly solubilize biomass, use cellulose as the sole carbon and energy source, and is under development for biofuel production. *C. thermocellum* is a candidate for consolidated bioprocessing (CBP) the produces ethanol and other fermentation products, including acetic acid, formic acid, lactic acid, and hydrogen. Biocatalyst growth and metabolism inhibition is caused by a range of toxic compounds generated during the pretreatment process and that accumulate during fermentation and product formation.

In this study, a transcriptomic analysis was conducted on a 17.5% v/v *Populus* hydrolysate tolerant mutant (PM) and wild type stain (WT) of *C. thermocellum.* The WT had increased expression of approximately twice as many genes in the standard media as the PM did. The PM increases the

expression of genes belonging to transcription, energy production and conversion, and amino acid transport and metabolism, specifically the Histidine metabolism pathway. The WT has an increases expression in cell division and sporulation, cell defense mechanisms, cell envelop and outer membrane, cell motility, the cellulosome, and inorganic ion transport and metabolism. The WT also had a greater number of differentially expressed genes in hydrolysate media than the PM did in hydrolysate media. The PM increased expression in energy production and conversion, cellulosome, and inorganic ion transport and metabolism and decreased expression in transcription, and cell defense mechanisms compared to standard media. The WT increased expression in cell defense mechanisms, cell motility, and the cellulosome, and decreases expression in cell envelope and outer membrane, amino acid transport and metabolism, inorganic transport and metabolism, and lipid metabolism.

This study further suggests that the mechanisms of tolerance for the 17.5% v/v *Populus* hydrolysate tolerant mutant strain of *C. thermocellum* is based on increased cellular efficiency by downregulated extraneous genes and increasing the expression of energy production and conversion rather than tolerance to specific hydrolysate components.

3.2 Introduction

Lignocellulosic biomass provides an abundant and renewable energy source for conversion into liquid transport fuels. Lignocellulosic biomass contains sugars polymerized in the form of cellulose and hemicelluloses, which can be liberated by hydrolysis, and subsequently fermented to ethanol by microorganisms, such as *Clostridium thermocellum* [16]. *C. thermocellum* is a Gram-positive, anaerobic, thermophilic, cellulolytic bacterium that can rapidly solubilize biomass, use cellulose as the sole carbon

and energy source, and is under development for biofuel production [33,50,76,105]. The cellulolytic capability of *C. thermocellum* gives it an advantage over organisms that are currently used for bioethanol production (e.g., yeast and *Zymomonas*), which can only ferment nonpolymeric carbohydrates [34,75]. *C. thermocellum* is a candidate for consolidated bioprocessing (CBP) the produces ethanol and other fermentation products, including acetic acid, formic acid, lactic acid, and hydrogen [75,111].

Compounds generated during pretreatment, hydrolysis and growth can have inhibitory effects on the fermenting microorganism, which decreases ethanol yields [12,16]. In order to obtain an economically feasible conversion process, reduction in the inhibitory effect of the toxic compounds is necessary [13]. Improved tolerance to inhibitory compounds found in pretreated biomass hydrolysate should improve the fermentation process and increase economic feasibility of CBP. Significant clues to the mechanisms involved in adaptation to new environments, such as would be found in a CPB production scheme, have come from studies of gene expression in response to specific stresses [37]. The response of cells to environmental changes can provide clues to the molecular apparatuses that enable cells to adapt to new environments and the molecular mechanisms that have evolved to regulate the remodeling of gene expression that occurs in new environments [37]. By understanding the genetic basis for mechanisms of improved tolerance to inhibitors there is a possibility to rationally engineer there traits in the future [33,38,52,112].

There have been a number of studies that have analyzed the effect of various stress associated with biofuel production and inhibitory compounds from pretreated biomass [49,110,111,113,114]. RNA Sequencing (RNA-seq) is an emerging technology that is being used for expression studies and it offers

several advantages over DNA microarrays such as better detection of genes expressed at low levels [51].

The development of a *Populus* hydrolysate tolerant strain of *C. thermocellum,* which can grow in 17.5% v/v *Populus* hydrolysate as well as the wild type (WT) can in standard media has been reported [105]. Genomic analysis of the mutant strain (termed PM for *Populus* mutant) reveled several mutations had occurred in the strain that may be responsible for its faster growth rate and *Populus* hydrolysate tolerance with selected mutations related to the transcriptional changes shown in Table 9 [105]. The extent of the growth, end product production and Populus hydrolysate tolerance was described by kinetic modeling [115]. In this study, the WT and PM strains were grown in various concentrations of *Populus* hydrolysate (0% or standard media, 10% and 17.5% v/v *Populus* hydrolysate) and a genome wide transcriptomic analysis was conducted at mid-log and late-log time points via RNA-seq. Two types of comparisons were used to further elucidate the mechanism of tolerance for the PM strain: a strain comparison in standard media and a hydrolysate response comparison for each strain in standard versus Populus hydrolysate containing media.

Table 9: Select mutations in the Populus mutant strain of Clostridium thermocellum.

Gene Name	Name	Description	Category	Type	Mutation
Cthe_2724	rpoB	DNA-directed RNA polymerase, beta subunit	1	Non-Syn	E885K
Cthe_0158	Rng G	ribobuclease, Rne/Rng family DEL:199088-199233	2	DEL	codon 438-486
Cthe_3087	spo0A	sporulation transcriptional activator SpoOA	4	STOP	Q76STOP
Cthe_2119	RsgL	glycoside hydrolase family 10	10	Syn	
Cthe_0422-3	Rex/adhE	CoA-binding domain protein/alcohol dehydrogenase	11	NC	
Cthe_2602-9	ATPase	127 bases 5' of Cthe_2602 -9 ATP synthase operon	11	NC	
Cthe_1766	proB	glutamate 5-kinase	13	Non-Syn	A149T
Cthe_1866	argD	Acetylornithine aminotransferase	13	Non-Syn	E55G
Cthe_1569	metY	Cysteine synthase	13	STOP	E229STOP
Cthe_1842	metY	O-acetylhomoserine/O-acetylserine sulfhydrylase	13	Non-Syn	P29Q

Table adapted from Linville et al. [105]. Non-Syn is non-synonymous SNP, Syn is synonymous SNP, DEL is a deletion, STOP is the insertion of a stop codon, NC is a mutation in the non-coding region.

3.3 Materials and Methods

3.3.1 Strain and Culture Conditions

C. thermocellum ATCC 27405 was obtained from Prof. Herb Strobel, University of Kentucky collection and denoted as the wild type (WT) strain. A *Populus* hydrolysate tolerant strain, referred to as the *Populus* Mutant (PM) strain was developed from the WT strain has been described [105]. In all experiments the cells were grown in media for thermophilic clostridia (MTC) with a substrate loading of 5 g/L cellobiose. Briefly, media was composed of 0.336 g/L KCl, 0.25 g/L NH_4Cl, 1.00 g/L $MgSO_4*7H_2$), 1.70 g/L KH_2PO_4, 0.50 g/L $C_7H_{14}NO_4S$, 0.15 g/L $CaCl_2*2H_2O$, 1.75 g/L $Na_3C_6O_7*2H_2O$, 0.6 g/L CH_4N_2O, 1.00 g/L L-cysteine HCL, 0.30 mg/L resazurin, MTC minerals and MTC vitamins [79,80,105]. All chemicals were reagent grade and obtained from Sigma (St. Louis, MO, USA), unless otherwise indicated. The inoculum culture was grown in Balch (Bellco Glass, Inc., Vineland, NJ) tubes containing 9.5 mL of 5 g/L cellobiose MTC and 0.5 mL of the frozen stock culture. Cultures were allowed to reach exponential growth phase by incubation at 58^{o}C and 100 rpm, diluted to an optical density (OD_{600}) of 0.250 using sterile, anaerobic Milli-Q water and injected into the test conditions with an inoculum size of 10% by volume [105]. Milled *Populus trichocarpa* hydrolysate was pretreated at the National Renewable Energy Laboratory (NREL) using a 20% w/w solid loading and dilute concentrations of H_2SO_4 at temperatures of 165-195^{o}C [81]. Solids were removed by filtration. The *Populus* hydrolysate was adjusted to a pH of 7.0 using 50% w/w NaOH and filter sterilized before being added to MTC media [105].

3.3.2 Fermentation

Methods for fermentation experiments through RNA isolation techniques were adapted from Raman et al. (2011) and previously described in Linville et al. [49,105]. Batch fermentations were conducted in triplicate in 1.5 L Q-plus jacketed glass fermentors (Sartorius Stedim Biotech, Bohemia, NY) using a 1 L working volume of MTC medium at 58^{o}C and stirred at a rate of 300 rpm, with pH controlled to 7.0 using 3N NaOH. Fermentors containing only 5g/L cellobiose were sparged with a filtered 20% CO_2/80% N_2 gas mixture and vigorously agitated overnight, followed by addition of the remaining medium components and *Populus* hydrolysate. *Populus* hydrolysate was added to the fermentors at 0%, 10% or 17.5% v/v concentration. The fermentors were then sparged for an additional 4 hours with a 20% CO_2/80% N_2 gas mixture. The inoculum was a bottle culture grown overnight (11-13 hrs) in 5g/L cellobiose, diluted to an OD_{600} of 0.200 $\pm$ 0.013, and added as 10% v/v to inoculate the fermenters. Well-mixed 5mL aliquots of culture were harvested at regular intervals. Cell growth was monitored based on an increase in the OD_{600} of the culture, measured in triplicate by a Spectramax Plus 384 spectrophotometer (Molecular Devices, Sunnyvale, CA). The corresponding dry cell weight was obtained from a calibration curve. Metabolite analysis was performed using HPLC analysis used a LaChrom Elite system (Hitachi High Technologies America, Inc. Pleasanton, CA) equipped with a refractive index detector (Model L-2490) and UV lamp (Model L-242OU). Metabolites were separated at a flow rate of 0.5 mL/min in 5 mM H_2SO_4 using an Aminex HPX-87H column (Bio-Rad Laboratories, Inc, Hercules, CA).

3.3.3 RNA isolation

Fermentation samples for RNA isolations were harvested by centrifugation of ~30mL culture in 50 mL conical tubes at 8000rpm and 4^{o}C for 10-15 min. The solid pellet fraction containing cells was flash frozen in liquid nitrogen and stored at -80^{o}C until further use. Total RNA was extracted from the cell pellets as follows. Briefly, the frozen cell solution was thawed on ice and an equal volume of TRIzol (Invitrogen, Carlsbad, CA) was added. The cell solution was divided into ~1mL aliquots and added to 2mL tubes containing 1 mL of 0.1 mm glass beads (Biospec Products, Bartesville, OK) ashed at 250^{o}C overnight. Cells were lysed by rapid agitation of the tubes at 6500 rpm for 1 min in three 20s-ON/20s-OFF cycles using the Precellys® bead beater (Bertin Technologies, France). Subsequently, the cell lysate (~0.8mL) in TRIzol was phase separated by addition of 200 µL chloroform and the RNA was precipitated by addition of 500 µL isopropanol. The precipitated RNA pellet was washed with 1mL of 75% ethanol and resuspended in 100 µL of RNase-free water. Any contaminating DNA was digested by in-solution DNase-I (Qiagen, Valencia, CA) treatment and the RNA sample was purified using the RNeasy mini kit (Qiagen, Valencia, CA) as per the manufacturer's instructions.

3.3.4 RNA sequencing and Analysis

RNA samples were shipped in dry ice to DOE Joint Genome Institute (JGI, Walnut Creek, CA). JGI prepared an indexed, directional, amplified Illumina RNA_seq library for each sample using the RiboZero kit for rRNA removal [84]. Three RNA samples were pooled into a single channel and sequenced using 2x100 bp chemistry. Sequence reads were mapped to the reference genome. A report from JGI was returned with a correlation analysis between replicate samples and contained counts (raw and normalized to the gene length and total number of counts for the sample) for known gene models. Gene

expression analysis was conducted using JMP Genomics Version 6.0 (SAS, Cary, NC). Raw count data was log-2 transformed and normalized by the Upper Quartile Scaling method [89]. Quality control analysis revealed the number of reads for two samples was too low and these data sets were removed from subsequent analyses. The statistical significance were determined via t-test for each independent variable and the three independent variables together and the false discovery rate method of nominal α, $p < 0.05$. Genes considered significantly differently expressed at a log-2 difference greater than 1 (representing a 2-fold change in expression) were further analyzed using pathway analysis and Hierarchical Clustering. Further analysis was done by placing genes into 20 different categories based on Cluster-of-Orthgolous groups (COGs) and gene function [49]. An odds ratio of the number of up- or down-regulated genes for either the mid-log or late-log phase in a category versus the total number of up- or down- regulated genes for the mid-log or late-log phase with a normally distributed 95% confidence interval (α=0.05) was conducted to determine the significance of each category. Further odds ratios of certain subsets of genes were conducted to further determine significance[116].

3.4 Results and Discussion

3.4.1 Fermentative growth

Batch fermentations were conducted for the *Populus* mutant (PM) and wild type (WT) strains of *C. thermocellum* as previously reported in Linville et al. [105]. The WT grown in the presence of 17.5% v/v *Populus* hydrolysate exhibited severely inhibited growth in an unpredictable manner (data not shown). Samples were taken at regular intervals from each fermentation unit based on their growth rate. Samples were analyzed for optical density (OD_{600}) and metabolite concentration by HPLC. The dry

cell weight (DCW) of the samples was determined by calibration curve (data not shown). Figure 16 show the average DCW (Figure 16A) ethanol yield (Figure 16B) and acetic acid yield (Figure 16C) for each of the test conditions. End product yields are calculated as a function of the amount of glucan utilized from the fermentation broth since the hydrolysate contained additional cellobiose and glucose. In brief, the PM had an increased growth rate when compared to the WT in the same condition [115]. The PM also produced more ethanol and the same amount of acetic acid than the WT under the same test conditions. The samples for RNA analysis were harvested from the fermentors during the mid-log and late-log phase based on the maximum growth (horizontal bars in (Figure 16A).

3.4.2 RNA-seq analysis

An analysis of variance (ANOVA) was conducted on each of the independent variables separately: strain, *Populus* hydrolysate concentration, and time. Differentially expressed genes were determined by a 2-fold statistical significance change in expression with a false discovery rate of less than 5% ($p<0.05$). Of the 3,236 genes in *C. thermocellum*, roughly 18% (n = 574) of the total predicted genes showed a change in expression for the comparison between strains. Furthermore, approximately 16% (n = 505) of the genes showed a change in expression between the three concentration comparisons and 0% (n = 0) of the genes showed a change in expression between the two time points. Since, there were no statistically significant changes in expression between the time points, the mid-log and late-log data were grouped together in further analyses.

The ANOVA of the three independent variables in combination revealed approximately 55% (n = 1795) of the genes were differentially expressed in at least one of the simple combinations (Supplemental File S6). Simple combinations only look at the difference between one of the three

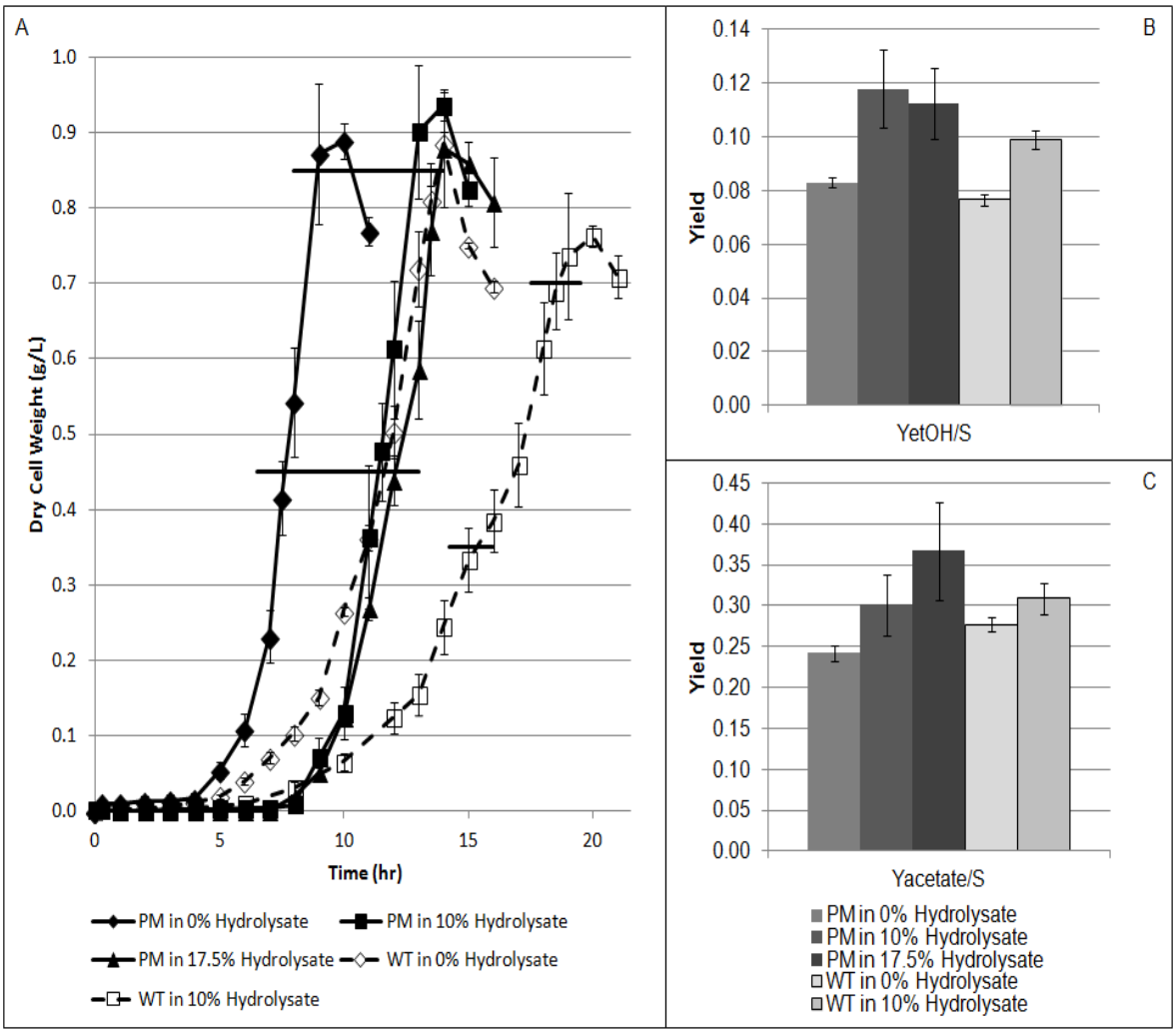


Figure 16: Growth Comparison for WT and PM in various concentrations of Populus hydrolysate

(A) Dry cell weight versus time. The horizontal bars indicate sample points for RNA-seq analysis. (B) Yield of ethanol produced (g/L)/glucan utilized (g/L). (C) Yield of acetic acid produced (g/L)/ glucan utilized (g/L).

variables at a time: strain, *Populus* hydrolysate concentration or time. Two comparisons are the focus of this paper. The first analysis will focus on the strain comparison: the WT versus PM in 0% v/v *Populus* hydrolysate. A positive differential expression represents a higher expression level in the PM strain and a negative differential expression represents a lower expression level in the PM strain when compared to the WT strain. The second analysis will focus on the three *Populus* hydrolysate comparisons: the PM in 0% versus 10%v/v *Populus* hydrolysate and 0% versus 17.5% v/v *Populus* hydrolysate, and the WT in 0% versus 10% v/v *Populus* hydrolysate. For these comparisons a positive differential expression represents an increase in expression level and a negative differential expression represents a decrease in expression level in the *Populus* hydrolysate conditions compared to standard conditions. Table 10 shows the number of number of genes that were increased and decreased in each comparison. Of the 1795 differentially expressed genes, 1506 are represented by these four comparisons.

The 2,312 genes with known function in *C. thermocellum* were manually placed into one of 20 categories based on function. Of the 1506 genes represented by the four comparisons, 492 encoding for proteins classified as hypothetical, uncharacterized and unknown function which were removed from further analysis (Supplemental File 7). The remaining 1014 differentially expressed genes were then divided into the 20 categories (Supplemental File 8). The odds ratio of the number of up- or down-regulated genes in each category versus the total number of up- or down- regulated genes was conducted to determine the significance of the genes differentially expressed for each of the categories. Table 11 shows the total number of genes in each category and indicates the direction of significant differential expression in the categories for each comparison.

Table 10: Number of positive and negative differentially expressed genes by at least 2-fold.

	Number of Genes	
	Positive	Negative
WT versus PM in 0% v/v *Populus* hydrolysate	186	393
PM in 0% versus 10% v/v *Populus* hydrolysate	39	53
PM in 0% versus 17.5% v/v *Populus* hydrolysate	229	260
WT in 0% versus 10% v/v *Populus* hydrolysate	593	447

Includes both strain and concentration simple comparisons involving the WT and PM strain and various concentrations of *Populus* hydrolysate.

Table 11: Total number of gene, and the number of positive and negative differentially expressed genes by odds ratio for each comparison.

	Category	Number of Genes	PM vs. WT in 0% hydrolysate	PM 0 vs. 10% hydrolysate	PM 0 vs. 17.5% hydrolysate	WT 0 vs. 10% hydrolysate
Gene Expression and Protein Production						
Transcription	1	161		Down	Down	
Translation, Ribosomal Structure and Biogenesis	2	185				
Posttranslational Modification, Protein Turnover and Chaperones	3	70				
Cellular Growth Processes						
Cell Division and Sporulation	4	116	Down			
DNA Replication, Recombination and Repair	5	255				
Cell Defense Mechanisms	6	94	Down	Down		Up
Cell Wall						
Cell Envelope Biogenesis, Outer membrane	7	179	Down			Down
Signal Transduction Mechanisms	8	79				
Cell motility	9	106	Down			Up
Cellulosome	10	99	Down		Up	Up

For the PM vs. WT in 0% hydrolysate comparison 'up' indicates the category was upregulated and 'down' indicated the category was downregulated in the PM compared to the WT. For the standard media (0%) versus *Populus* hydrolysate media (10 or 17.5%) comparisons, 'up' indicates the category was upregulated and 'down' indicates that the category was down regulated in the hydrolysate media compared to standard media.

Table 11: Total number of gene, and the number of positive and negative differentially expressed genes by odds ratio for each comparison.

	Category	Number of Genes	PM vs. WT in 0% hydrolysate	PM 0 vs. 10% hydrolysate	PM 0 vs. 17.5% hydrolysate	WT 0 vs. 10% hydrolysate
Metabolism and Transport						
Energy Production and Conversion	11	152	Up	Up		
Carbohydrate Transport and Metabolism	12	83				
Amino Acid Transport and Metabolism	13	163	Up			Down
Inorganic Ion Transport and Metabolism	14	60	Down	Up	Up and Down	Down
Coenzyme Metabolism	15	88				
Nucleotide Transport and Metabolism	16	63				
Lipid Metabolism	17	45				Down
Secondary Metabolites biosynthesis Transport and Catabolism	18	20				
General Transport and Secretion	19	143		Up		
Miscellaneous						
Miscellaneous	20	151	Down			
Total		2312				

For the PM vs. WT in 0% hydrolysate comparison 'up' indicates the category was upregulated and 'down' indicated the category was downregulated in the PM compared to the WT. For the standard media (0%) versus *Populus* hydrolysate media (10 or 17.5%) comparisons, 'up' indicates the category was upregulated and 'down' indicates that the category was down regulated in the hydrolysate media compared to standard media.

3.5 Discussion

3.5.1 Strain Comparison

The strain comparison analyzes the difference in expressed genes between the WT and PM in standard media (0% v/v *Populus* hydrolysate) to elucidate the effect of the mutations on the PM strain. The smaller number of genes with increased versus decreased expression for the PM compared to the WT in standard media (186 versus 393) supports the hypothesis that the gene expression profile of the PM results in a more efficient cellular metabolism which leads to increased robustness. The *Populus* hydrolysate tolerant phenotype of the PM is the result of several simultaneous mechanisms of action, including increases in cellular repair, and altered energy metabolism (Table 9) [105]. These and related mechanisms are apparently independent from each other and can involve genes or gene clusters widely dispersed on the chromosome [1]. These mutation alter the growth rate in standard media suggesting a more global change in gene expression which will be evaluated by comparing the change in expression between the PM and WT strain in standard media (0% v/v *Populus* hydrolysate).

The PM has a non-synonymous SNP in a strongly conserved amino acid of the single copy of the *rpoB* gene which encodes for a DNA directed RNA polymerase, beta subunit (Table 9) [105]. The beta subunit of the RNA polymerase interacts directly with both the DNA and has weak binding sites for the sigma factor [117]. This mutation potentially changes the specificity, activity and/or stability of the RNA polymerase [105]. Since, the RNA polymerase contacts every promoter in the genome, a single amino acid substitutions in the critical portions of the enzyme may lead to global changes in gene expression and have significant impact of cellular functions [92]. As the set of genes controlled by a single sigma

factor can number in the hundreds, sigma factors provide effective mechanisms for simultaneously regulating large numbers of genes [54]. The PM differentially expresses multiple sigma factors when compared to the WT in standard media which can be directly linked to the overall change in expression for certain categories of genes. The sigma factors can be seen in Table 12 and will be discussed with the genes under their control.

3.5.1.1 Categories of gene with increased expression in the PM

The PM increases the gene expression in only two categories compared to the WT in standard media. The increased expression in the energy production and conversion may allow for the increased growth phenotype observed in the PM strain. Increases in glycolysis would lead to increases in reducing power (in the form of NADH) being available for downstream electron transport and ethanol production. The increase in ethanol production and increase in electron flux may generate sufficient NAD+ to ensure increased cellular metabolism [111].

The assemblage of genes encoding proteins involved in pyruvate metabolism and end-product synthesis dictate, in part, how carbon and electrons flux is distributed between the catabolic, anabolic, and energy producing pathways of the cell [97]. The flow of carbon and electrons from phosphoenolpyruvate (PEP) towards end-products may be separated into branch-points or nodes which include (i) the PEP/oxaloacetate/pyruvate node, (ii) the pyruvate/lacate/acetyl-CoA node, (iii) the acetyl-CoA/acetate/ethanol/node, and the (iv) ferredoxin/NAD(P)H/H_2 node [97]. Several different enzymes

Table 12: Arithmetic mean of differentially expressed sigma factors.

Gene Name	Product	PM vs. WT 0		PM 0 vs. 10		PM 0 vs. 17.5		WT 0 vs. 10	
		ML	LL	ML	LL	ML	LL	ML	LL
Cthe_1272	sigma-70 region 2 domain protein	2.34	1.24	-2.20	-1.64	-1.38	1.94	**6.00**	**2.72**
Cthe_0195	Sigma-70 region 4 type 2	**2.80**	1.61	-2.06	-1.23	-1.44	1.49	**3.37**	**1.86**
Cthe_1438	RNA polymerase sigma factor, sigma-70 family	**2.68**	**2.06**	**-2.26**	-1.76	**-2.95**	**-2.42**	-1.43	1.61
Cthe_1809	RNA polymerase sigma factor, sigma-70 family	**18.26**	**16.44**	-1.69	**-2.11**	**-4.55**	**-4.06**	**-2.25**	-1.68
Cthe_0446	sigma-E processing peptidase SpoIIGA	-1.86	**-2.21**	-1.10	1.45	-1.03	1.51	-1.78	-1.92
Cthe_0447	RNA polymerase sigma-E factor	1.90	**2.58**	-1.56	-1.19	-1.30	**-2.65**	-1.77	1.14
Cthe_0120	RNA polymerase sigma-F factor	1.71	**2.01**	1.01	1.15	-1.03	-1.22	-1.43	1.18
Cthe_0448	RNA polymerase sigma-G factor	-1.79	-2.55	-2.10	-1.23	-1.56	-1.06	**-4.11**	**-2.73**
Cthe_1012	RNA polymerase sigma-K factor	**-3.94**	**-4.74**	1.13	1.20	1.07	**3.57**	-1.21	-1.33
Cthe_2059	RNA polymerase sigma-H factor	1.45	1.65	-1.30	-1.52	-1.41	**-2.13**	-1.66	1.05
Cthe_0315	RNA polymerase sigma-I factor	-1.40	-2.19	-1.13	-1.17	-1.00	1.92	**2.52**	1.36
Cthe_2975	RNA polymerase sigma-I factor	1.24	1.47	-2.09	-1.94	1.15	1.45	**4.32**	1.96
Cthe_0495	RNA polymerase, sigma 28 subunit	**-3.04**	**-3.47**	1.18	1.43	1.37	1.53	**3.87**	1.83
Cthe_2100	transcriptional regulator, AbrB family	2.21	2.48	-2.67	-1.16	**-5.28**	**-13.66**	**-10.68**	1.66

Bold values indicate significantly differential expression by ANOVA. For the PM vs. WT in 0% v/v *Populus* hydrolysate a positive differential expression represents a higher expression in the PM and a negative differential expression represents a lower expression in the PM compared to the WT. For the standard media (0%) versus *Populus* hydrolysate media (10 or 17.5%) a positive value represents a higher expression and a negative value represents a lower expression in the hydrolysate media compared to standard media.

may be involved within these nodes. The oxidation of electrons carriers (NADH and/or ferredoxins) is required for maintaining glycolytic flux and leads to the ultimate production of reduced products (ethanol, lactate, and hydrogen) [97]. *C. thermocellum* catabolizes glucose via the Embden-Meyerhof pathway [118,119]. The conversion of PEP to pyruvate occurs via a "malate shunt" consisting of reaction of PEP to oxaloacetate (OAA), OAA to malate, and malate to pyruvate mediated [119,120].

In general, the PM had a higher expression of genes belonging to energy production and conversion than the WT did in standard media with 23 genes being significantly increased. The PM upregulated 8 genes specific to the central metabolism and mix-acid fermentation compared to the WT in standard media (Figure 17 and Table 13). A study of the effect of cellulose fermentation found that these genes are typically upregulated during cellulose fermentation compared to cellobiose fermentation [49,97]. The native upregulation of these genes in the PM may allow for the phenotypic faster growth rate and possibly further increased growth on cellulose.

C. thermocellum uses the hydrogenase-mediated pathway for production of molecular hydrogen to dispose the excess reducing equivalents generated during carbohydrate catabolism [49,120]. In the process, the complexes pump H^+/Na^+ ions across the cell membrane and create proton gradients for powering ATP synthesis by ATP synthase (ATPase) [49]. *C. thermocellum* has two ATPase clusters: a F_1-F_0-ATP synthase (Cthe_2602-Cthe_2609) and a V_1-V_0-ATP synthase (Cthe_2261-Cthe_2269) [49,113]. Only 44 genomes contain both types of ATPase [113]. The PM has a mutation in the non-coding region upstream of the F-type ATPase (Table 9) which was believed to cause an increase in the expression of this gene cluster in the PM compared to the WT in standard media (Table 14) [105]. The PM also increases the expression of four Ech-type hydrogenases compared to the WT in standard media

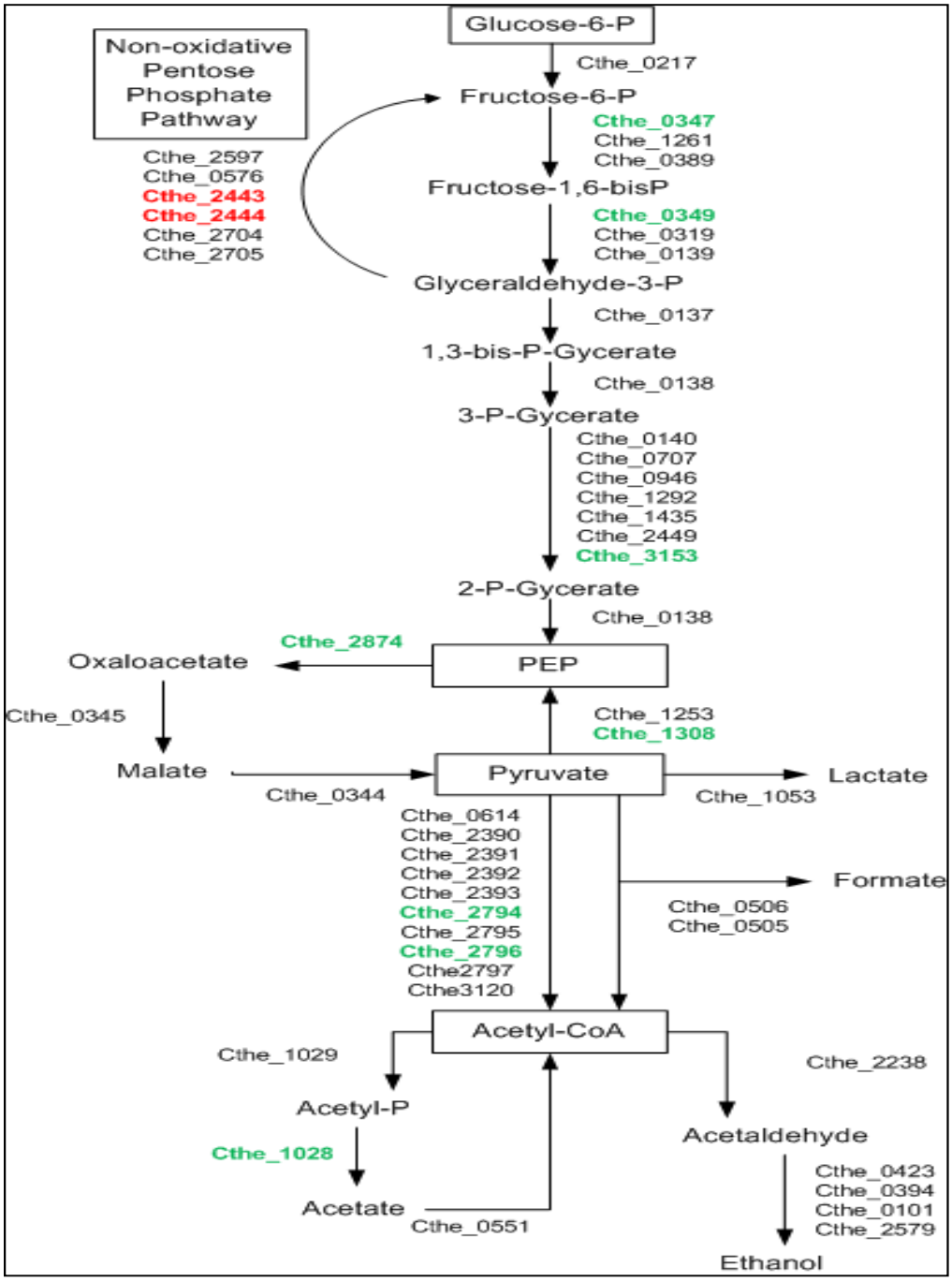


Figure 17: The PM has increased expression of genes in the central metabolism pathway compared to the WT in standard media

Genes colored green indicate upregulation and genes colored red indicate downregulation for the Pm compared to the WT. The extent of gene expression change and expression levels in other comparisons can be seen in Table 13. This figure was adapted from [49,109].

Table 13: Arithmetic mean of differentially expressed genes along the central metabolism and mixed acid fermentation pathways

Gene	Product	PM vs. WT 0		PM 0 vs. 10		PM 0 vs. 17.5		WT 0 vs. 10	
		ML	LL	ML	LL	ML	LL	ML	LL
glucose-6-phosphate to PEP									
Cthe_0347	phosphofructokinase	1.77	**2.59**	-1.35	-1.31	-1.14	-1.49	**-2.27**	-1.07
Cthe_0349	fructose-1,6-bisphosphate aldolase, class II	1.60	**2.49**	-1.52	-1.41	-1.49	**-2.03**	**-3.14**	-1.42
Cthe_2449	Phosphoglycerate mutase	-2.46	-1.85	-1.48	-1.90	-2.01	-2.88	**-5.18**	**-2.03**
Cthe_3153	alpha-ribazole phosphatase	**2.11**	**2.33**	1.21	1.23	1.42	-1.14	1.82	**2.04**
Cthe_0143	enolase	-1.23	-1.04	-1.11	-1.13	-1.05	**-2.96**	**-2.22**	-1.16
Non-oxidative Pentose Phosphate pathway									
Cthe_2443	Transketolase domain-containing protein	**-3.24**	**-4.70**	1.14	-1.75	-1.90	-1.52	-2.88	-2.74
Cthe_2444	Transketolase domain-containing protein	**-3.47**	**-4.63**	1.26	-1.72	-1.63	-1.57	-2.41	-2.36
Cthe_2705	Transketolase central region	-1.44	-1.33	1.24	1.21	1.17	-1.81	**-2.60**	-1.32
PEP to Pyruvate									
Cthe_2874	Phosphoenolpyruvate carboxykinase [GTP]	1.39	**2.46**	-1.05	-1.04	1.30	1.38	-1.07	1.14
Cthe_0344	malic protein NAD-binding	-1.68	1.06	-1.01	-1.14	1.20	-1.27	**-2.13**	1.02
Cthe_1308	pyruvate, phosphate dikinase	1.64	**2.29**	-1.05	1.07	1.30	1.10	**2.03**	1.49
Pyruvate to Lactate/Formate/Acetyl-CoA									
Cthe_1053	L-lactate dehydrogenase	-1.78	-1.25	-1.41	-1.27	-1.33	-1.16	**-3.30**	-1.55
Cthe_2794	pyruvate/ketoisovalerate oxidoreductase, gamma subunit	**4.30**	1.48	1.92	2.56	2.45	2.78	1.61	-1.05
Cthe_2796	pyruvate flavodoxin/ferredoxin oxidoreductase domain protein	**3.13**	1.47	1.98	2.05	1.88	1.94	1.49	1.02
Cthe_0505	formate acetyltransferase	-1.95	-1.91	1.24	1.04	1.11	-1.81	**-2.31**	-1.76

Bold values indicate significantly differential expression by ANOVA. For the PM vs. WT in 0% v/v *Populus* hydrolysate a positive differential expression represents a higher expression in the PM and a negative differential expression represents a lower expression in the PM compared to the WT. For the standard media (0%) versus *Populus* hydrolysate media (10 or 17.5%) a positive value represents a higher expression and a negative value represents a lower expression in the hydrolysate media compared to standard media.

Table 13: Arithmetic mean of differentially expressed genes along the central metabolism and mixed acid fermentation pathways

Gene	Product	PM vs. WT 0		PM 0 vs. 10		PM 0 vs. 17.5		WT 0 vs. 10	
		ML	LL	ML	LL	ML	LL	ML	LL
Acetyl-CoA to Ethanol/Acetate									
Cthe_1028	Acetate kinase	1.67	**2.57**	**2.12**	1.26	**2.76**	1.50	-1.02	1.06
Cthe_1029	phosphate acetyltransferase	1.54	1.79	**2.42**	1.33	**2.73**	1.63	-1.08	-1.61
Cthe_2238	Aldehyde Dehydrogenase	1.06	1.04	1.20	1.36	1.36	1.02	**2.30**	1.83
Cthe_0101	iron-containing alcohol dehydrogenase	-1.35	-1.19	1.22	-1.04	-1.18	-1.82	**-2.43**	-1.48
Cthe_0423	iron-containing alcohol dehydrogenase	1.12	1.07	1.26	1.06	1.28	1.45	**-3.36**	**-4.42**

Bold values indicate significantly differential expression by ANOVA. For the PM vs. WT in 0% v/v *Populus* hydrolysate a positive differential expression represents a higher expression in the PM and a negative differential expression represents a lower expression in the PM compared to the WT. For the standard media (0%) versus *Populus* hydrolysate media (10 or 17.5%) a positive value represents a higher expression and a negative value represents a lower expression in the hydrolysate media compared to standard media.

Table 14: Arithmetic mean of differentially expressed genes involving redox.

		PM vs. WT 0		PM 0 vs. 10		PM 0 vs. 17.5		WT 0 vs. 10	
		ML	LL	ML	LL	ML	LL	ML	LL
Redox transcriptional repressor									
Cthe_0422	Redox-sensing transcriptional repressor rex	1.13	-1.08	1.04	-1.02	-1.04	-1.11	**-5.96**	**-6.08**
Ech-type hydrogenases									
Cthe_3019	4Fe-4S ferredoxin iron-sulfur binding domain-containing protein	**2.89**	**3.12**	1.14	1.03	1.54	**2.02**	1.87	1.08
Cthe_3020	NADH-ubiquinone oxidoreductase chain 49kDa	**2.96**	**3.83**	1.02	-1.03	1.55	**2.07**	1.64	1.18
Cthe_3021	ech hydrogenase, subunit EchD, putative	**4.29**	**4.79**	-1.12	-1.09	1.16	1.71	1.79	1.46
Cthe_3024	NADH/Ubiquinone/plastoquinone (complex I)	**2.04**	**2.26**	-1.12	-1.10	-1.17	1.05	-1.54	-1.02
ATP synthase									
Cthe_2602	ATP synthase subunit a	**2.77**	**3.55**	-1.21	1.10	-1.35	-1.72	**-2.44**	1.01
Cthe_2603	ATP synthase subunit c	**2.27**	**2.68**	1.11	1.04	-1.04	-1.22	-1.06	-1.66
Cthe_2604	ATP synthase subunit b	**2.48**	**2.18**	-1.01	1.10	-1.23	-1.32	-1.64	-1.80
Cthe_2605	ATP synthase F1, delta subunit	**3.55**	**2.04**	-1.06	-1.05	-1.77	**-2.01**	-1.15	-1.53
Cthe_2606	ATP synthase F1, alpha subunit	**2.40**	**2.00**	1.60	1.31	1.20	1.25	1.40	-1.24
Cthe_2607	ATP synthase F1, gamma subunit	**2.63**	**2.09**	1.17	1.20	1.09	1.49	1.50	-1.18
Cthe_2608	ATP synthase F1, beta subunit	**2.67**	**2.65**	1.40	1.37	1.04	1.05	-1.00	-1.20
Cthe_2609	ATP synthase epsilon chain	**2.94**	**2.87**	1.12	1.33	-1.21	-1.11	-1.24	-1.26

Bold values indicate significantly differential expression by ANOVA. For the PM vs. WT in 0% v/v *Populus* hydrolysate a positive differential expression represents a higher expression in the PM and a negative differential expression represents a lower expression in the PM compared to the WT. For the standard media (0%) versus *Populus* hydrolysate media (10 or 17.5%) a positive value represents a higher expression and a negative value represents a lower expression in the hydrolysate media compared to standard media.

(Table 14). It is possible that the increased expression of the F-type ATPase and Ech-type hydrogenases helps to maintain the electron flux in the cell allowing for more efficient growth [105].

Furthermore, sigma factor σ^A (σ^{70}) is the principle sigma factor present in vegetatively growing *B. subtilis* and other Gram-positive bacteria [55] and it directs transcription of genes important to metabolism [54]. There are 10 genes that encode for σ^A subunits in *C. thermocellum*. Three of the genes that for encode σ^A (Cthe_0195, Cthe_1438 and Cthe_1809) are upregulated in the PM compared to the WT in standard conditions (Table 12). Cthe_1809 had 16-fold greater expression level in the PM strain than the WT strain. The change in expression of these three sigma factors were considered significant based on the subset odds ratio of the total number of σ^A. The higher expression level may be the reason for the higher observed growth phenotype in the PM strain under standard conditions (Figure 16). This might also be the reason for the increased energy production and conversion for the PM versus the WT in standard media.

Of the 163 genes that encode for various parts of the amino acid transport and metabolism, the PM upregulated a significant number of genes (23 genes) compared to the WT in standard media. The PM increased the expression of 10 of the 15 genes along the histidine metabolism compared to the WT in standard media (**Table 15**). Cthe_2880-Cthe_2889 is a single operon and is among the most highly differentially expressed genes in the PM versus WT comparison with an average 18-fold increase in expression. The PM decreases the expression of one gene in this pathway, Cthe_3028 which converts histidine to histamine (Figure 18). De novo biosynthesis of histidine during fermentation may be constrained by the high $NADH/NAD^+$ ratio during anaerobic growth and the requirement for further reduction of NAD^+ in the two terminal steps of biosynthesis [109]. Histidine may be limited by the

Table 15: Arithmetic mean of differentially expressed genes involved in histidine metabolism.

Gene	Product	PM vs. WT 0		PM 0 vs. 10		PM 0 vs. 17.5		WT 0 vs. 10	
		ML	LL	ML	LL	ML	LL	ML	LL
Cthe_2880	ATP phosphoribosyltransferase regulatory subunit	**98.42**	**29.12**	1.25	1.01	1.12	-1.06	1.25	**-2.73**
Cthe_2881	ATP phosphoribosyltransferase	**78.48**	**23.79**	1.64	1.24	1.35	-1.01	1.52	**-3.40**
Cthe_2882	histidinol dehydrogenase	**35.86**	**18.44**	1.49	1.33	1.37	1.40	1.88	-1.83
Cthe_2883	histidinol-phosphate aminotransferase	**38.12**	**19.61**	1.15	1.22	1.19	1.42	1.89	-1.69
Cthe_2884	Imidazoleglycerol-phosphate dehydratase	**7.45**	**7.71**	1.23	1.25	1.18	1.27	-1.89	-1.77
Cthe_2886	Imidazole glycerol phosphate synthase subunit hisH	**11.99**	**12.29**	1.19	1.12	1.09	-1.01	-1.04	-1.16
Cthe_2887	1-(5-phosphoribosyl)-5-[(5-phosphoribosylamino)methylideneamino] imidazole-4-carboxamide isomerase	**13.46**	**11.01**	1.44	1.20	1.29	1.13	1.93	-1.10
Cthe_2888	Imidazole glycerol phosphate synthase subunit hisF	**12.46**	**14.23**	1.61	1.30	1.54	1.24	1.99	1.02
Cthe_2889	Histidine biosynthesis bifunctional protein hisIE	**12.71**	**10.80**	1.48	1.29	1.51	1.28	**3.08**	1.11
Cthe_3028	Pyridoxal-dependent decarboxylase	**-11.35**	**-13.46**	-2.37	-1.04	**-3.78**	**-2.89**	**-3.79**	**-2.02**
Cthe_3149	aminoacyl-histidine dipeptidase	**3.34**	**4.23**	1.15	1.05	1.39	1.37	**4.09**	**2.72**
Cthe_1332	Histidyl-tRNA synthetase	-1.58	-1.89	1.10	-1.03	-1.15	-1.62	**-2.38**	-1.64

Bold values indicate significantly differential expression by ANOVA. For the PM vs. WT in 0% v/v *Populus* hydrolysate a positive differential expression represents a higher expression in the PM and a negative differential expression represents a lower expression in the PM compared to the WT. For the standard media (0%) versus *Populus* hydrolysate media (10 or 17.5%) a positive value represents a higher expression and a negative value represents a lower expression in the hydrolysate media compared to standard media.

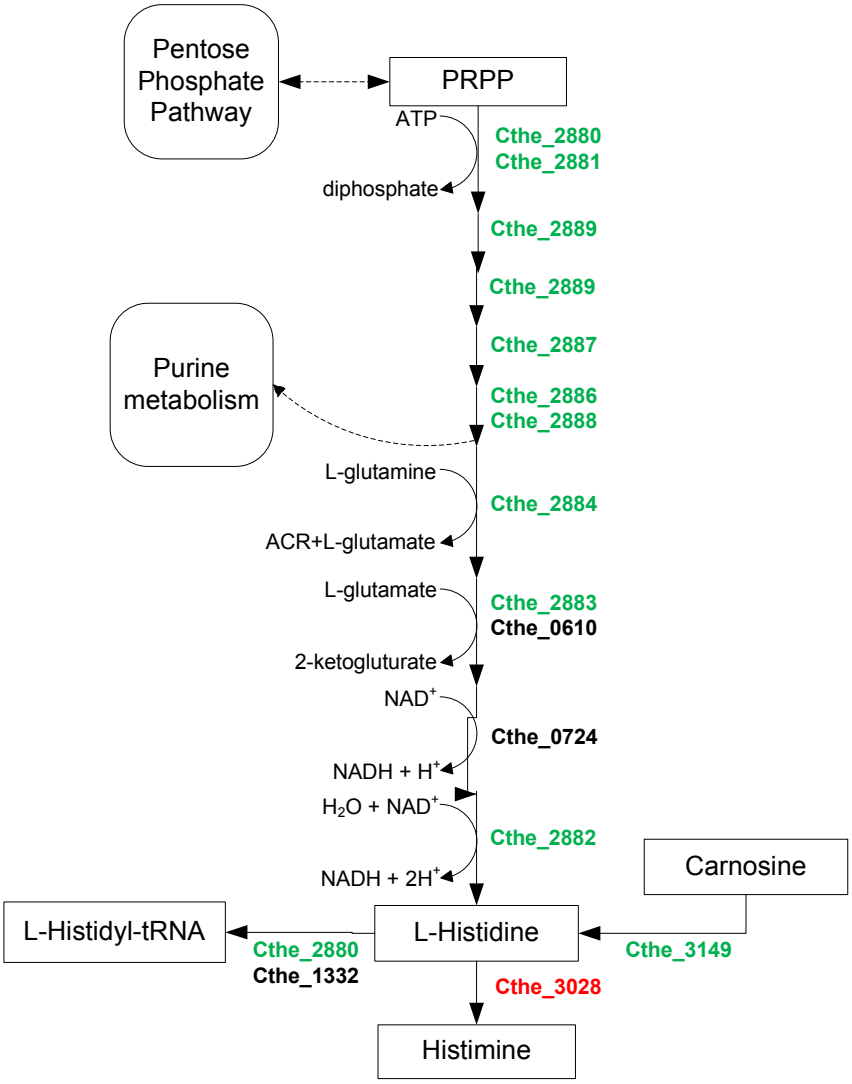


Figure 18: The PM has increased expression of genes in the hisidine biosynthesis pathway compared to the WT in standard media

Genes colored green indicate upregulation and genes colored red indicate downregulation. The extent of gene expression change and expression levels in other comparisons can be seen in Table 15. PRPP, 5-phosphoribosyl 1-pyrophosphate. ACR, aminoimidazole carboxamide ribonucleotide. This figure was adapted from [109].

addition of furfural [109]. The PM has two mutations involved with glutamate catabolism; a possible gain in function in *argD* (Cthe_1866) and a possible loss in function in (Cthe_1766) (**Table 9**) [105]. These two mutations suggest a shift from proline production to glutamate and arginine production in PM which seems to be a beneficial mutation that commonly occurs[105]. The shift in amino acid production may also assist in the increased expression in the histidine pathway since glutamate is utilized in the pathway.

3.5.1.2 Categories of gene with decreased expression in the PM

There are a number of categories with decreased expression level for the PM when compared to the WT in standard media. The decreased expression of these genes may be a result of trying to conserver cellular resources and redirected them in such a way as to increase the growth rate for the PM. The categories will be discussed briefly below.

Cell Division and Sporulation Genes

The downregulation of the cell division and sporulation genes by the PM compared to the WT in standard media may seem counterintuitive with the faster growth rate of the PM strain. However, the genes in this category can be broken down into cell division genes and sporulation genes. Independent odds ratios on the subsets of genes showed that only the subset of sporulation genes were significantly downregulated by the PM in standard media (data not shown). There are two possible reasons that the PM downregulates the sporulation genes.

The first possible reason for the reduction in sporulation genes is that the PM has a stop codon placed early in the coding region of Cthe_3087 which should disrupt the gene function (Table 9) [105]. Cthe_3087 is a distantly related homologue of the *spo0A* gene [105]. There is a second copy of *spo0A* in *C. thermocellum,* Cthe_0812. The gene expression of Cthe_3087 does not change significantly. However, Cthe_0812 is significantly downregulated by an unknown mechanism in standard conditions compared to the WT. The spo0A protein is activated when phosphorylated and has been shown to regulate sporulation in a number of clostridia [121]. The signal transduction mechanism activates sporulation in *B. subtilis* through a chain of four proteins in which the phosphoryl group is transferred before activating the *Spo0A* transcription factor [99,100,122]. However, the clostridia sporulation uses a different signaling which is not fully understood [121]. In *B. subtilis Spo0A* has been shown to repress transcription of the transcriptional pleiotropic regulator of transition state genes (*abrB)* and to activate operons encoding stage II and strange III sporulation genes [44,99]. *Spo0A* contains a DNA binding site that stimulates *spo0F*, *spo0B*, *spoIIA*, *spoIIE*, and *spoIIG* promoter activity [99,100]. Although, it is rare for *C. thermocellum* to go into sporulation, it has been shown that sporulation will occur under vitamin limitation, oxygen stress and switching between soluble and insoluble substrates [123]. *Spo0A* mutants which are unable to go into sporulation are thought to be locked in exponential growth since they continue to grow under nutritional conditions that would normally induce sporulation. Essentially they appear to maintain growth until the nutrients are exhausted, whereupon cell lysis occurs [100]. The PM growth kinetics is consistent with other *spo0A* defective mutants [41,43,44]. The cell density of the PM strain does not show a normal reduction of growth into stationary phase and then the decay phase. Instead the PM maintains its growth rate until it utilizes all of the sugar available [115].

The second reason for the reduction in sporulation genes may be that the PM differentially expresses the sigma factors that control sporulation. The five-know sporulation sigma factors in *B.*

subtilis are σ^E, σ^F, σ^G, σ^H and σ^K [55,121]. In *B. subtilis*, σ^H is the earliest sporulation sigma factor [121]. σ^E is the mother cell-specific sigma factor and is also involved in the synthesis of σ^K, the late acting mother cell sigma factor [55]. Furthermore, σ^F – dependent transcription appears to be limited to the early expression of forespore-specific genes and σ^G appears to encode products that are synthesized within the forespore compartment during the later stages of sporulation to enhance spore survival and facilitate germination [55].There are six genes that encode the various sporulation sigma factors in *C. thermocellum*. The PM has increased expression in σ^E (Cthe_0447) and σ^F (Cthe_0120) for the late-log time point, and decreased expression in σ^E (Cthe_0446) for the late-log time point and σ^K (Cthe_1012) for both time points (Table 12). The greater reduction in expression of sigma factor σ^K may result in the lower expression of the sporulation genes for the PM. The mechanism that down regulates the sigma factors may also be the mechanism for the downregulation of Cthe_0812 for the PM in standard media compared to the WT.

Going into sporulation is an energy intensive function which requires the transcription of a large number of genes. By reducing the expression of the sporulation genes, the PM has more cellular resources to dedicate to a faster growth rate. Furthermore, it has been shown that *C. thermocellum* forms L-forms in the depletion of substrate during growth [123]. It is possible the PM prefers to form L-forms since it may require less energy than sporulation and has a faster recovery time in rich media than spores [123]. Once the genes that control the switch into L-forms have been discovered, this hypothesis can be tested.

Cellular Defense Mechanisms

During their lifetime, microorganisms are faced with the constant threat of invading foreign DNA, mainly genetic elements such as phages, plasmids, transposons and genomic islands [124]. Microbes have devised various strategies that allow them to survive exposure to foreign genetic elements in natural environments [125]. However, in industrial environments and in the controlled laboratory condition used during directed evolution of this strain, these defense mechanisms may play a less important role in survival. Of the genes which encode for various cell defense mechanisms, the PM downregulated the expression of 29 genes compared to the WT in standard media. There are three main groups of genes that represent the majority of the cellular defense genes: CRISPR associated proteins, Hedgehog/intein hint domain proteins and phage related proteins. Together these three subgroups make up 65 of the 94 cellular defense genes. Odds ratios were conducted on each of the three subsets of genes separately to determine their independent significance.

One adaptive microbial immune system termed Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and CRISPR-associated (*cas*) genes has been identified [125,126]. CRISPR loci typically consist of several noncontiguous direct repeats separated by stretches of variable sequences call spacers [126]. The spacers mostly correspond to segments of captured viral and plasmid DNA and thus function as a memory bank that will recognize the same genetic material upon repeated infection [124,126]. The CRISPR spacers interfere with both plasmid conjugation and transformation [125]. It was found in *C. thermocellum* that individual spacer sequences had strong effects on transcription and processing patterns of a CRISPR cluster [126]. It has also been found that internal spacers can be deleted, perhaps allowing the host to limit the expansion of the CRISPR locus so that the relative size of the locus does not increase to a detrimental level [125]. The PM has an almost 6-fold

lower average expression level for 15 of the 19 CRISPR-associated genes compared to the WT in standard media.

The Hint (Hog/intein) superfamily are another type of immune system and includes several types of auto-catalytic protein domains [127]. The domains process the proteins in which they are present by protein-splicing, self-cleavage and ligation activities [127]. Inteins are embedded in-frame within highly conserved regions of essential proteins and help stabilize them [127]. The intein domain is translated with its protein host, and excises itself by protein splicing [128,129]. Inteins are currently known in more than 50 types of proteins with diverse functions including metabolic enzymes, DNA and RNA polymerases, proteases, ribonucleotide reductases, and vaculor-ATPases [129]. Hog Hint domains have a role in the normal maturation process of their protein hosts [127]. *C. thermocellum* has 12 genes that encode Hedgehog/intein hint domain proteins, of which the PM has almost a 6-fold lower expression for 10 genes compared to the WT in standard media.

Although, defense mechanisms have their advantages, the PM may reduce the expression of the CRISPR-associated genes and Hedgehog/ intein hint domain protein in an effort to conserve cellular resources. Since the PM did not delete the CRISPR-associated regions, it still has the ability to recognize the foreign DNA. However, the reduced expression of these two groups of genes may come at the expense of increased expression of phage associated genes. *C. thermocellum* has 34 genes which encode for various phage-associated proteins. These genes are not typically considered part of the cell defense mechanisms. The PM has an average 2-fold increased expression of 6 of the phage associated genes compared to the WT in standard media which was deemed significant by the odds ratio.

Cellulosome

C. thermocellum has one of the fastest growth rates on crystalline cellulose, the major component in biomass due to a membrane bound complex, termed the cellulosome, included cellulases and other polysaccharide degrading enzymes assembled together in large protein complexes [48,49]. The cellulosome complex has a primary scaffolding protein where various carbohydrate degrading enzymes bind to and then attaches to the cell surface via the cell wall anchor protein [49]. During growth on insoluble substrates, the cells are tightly attached to the substrate via the carbohydrate binding module (CBM) [49]. However, the composition of the cellulosome differ as a function of the growth conditions [46]. Of the 99 genes that encode for the cellulosome components, the PM has a decreased expression of 19 genes (Supplemental File 8). The decrease expression in cellulosome genes by the PM may be an attempt to conserve energy since the cells were mutated in media containing cellobiose and soluble glucans present in the hydrolysate. It has been hypothesized that the downregulation of the cellulosome on soluble substrate such as cellobiose occurs via catabolite repression [48]. The PM has a synonymous SNP in Cthe_2119 (*RsgI6*) which is an anti-σ^I factor and is involved in regulating the expression of cellulosomal genes in the presence of xylans and cellulose (Table 9) [105]. It is possible that this mutation changes the specificity of the anti-σ^I factor and reduces the expression of the cellulosomal genes over and above he reduction that would be achieved by catabolite repression alone. This additional repression would not be active in the WT.

Other categories with decreased expression

The PM has a lower expression of 31 genes that encode for cell envelope and outer membrane in standard media than the WT does (Supplemental File 8). The average change in expression for theses

31 genes is 3-fold lower than the WT in standard media. The PM downregulated 21 genes that encode for cell motility in standard media compared to the WT. It has been proposed that the σ^D in *B. subtilis* controls flagellin production and possibly has a role in the expression of the methyl-accepting chemotaxis proteins [55]. Sigma factor σ^D (Cthe_0495) is downregulated in the PM compared to the WT in standard media by 3-fold (Table 12). This may be the reason for the decrease in cell motility genes for the PM versus the WT in standard media. The PM also downregulated 12 genes that encode for various inorganic ion transport and metabolism compared to the WT in standard media. However, the downregulated genes do not belong to any specific pathway. The PM also downregulated 26 genes in the miscellaneous category compared to the WT in standard media. The benefits of reducing the expression level of these genes are not clear besides reducing the consumption of cellular resources.

3.5.2 Hydrolysate Comparison

The *Populus* hydrolysate concentration comparison analyzes the difference in gene expression for various hydrolysate concentrations within a given strain. Inhibitory compounds from the *Populus* hydrolysate may affect the cell by damaging and denaturing biological molecules, resulting in adverse outcomes, including the improper unfolding of proteins, DNA damage, improper RNA unfolding and degradation, and the impairment of biophysical changes to cell membranes necessary for energy generation and the proper functioning of molecular pumps [1,32]. The small number of differentially expressed genes (total of 92 genes) for the PM in 10% v/v *Populus* hydrolysate compared to standard media indicates that the PM strain requires relatively few changes in gene expression to adapt to the hydrolysate media (Table 10).This is not entirely surprising since the PM was adapted to the during the directed evolution process. Even when the PM strain is placed in 17.5% v/v *Populus* hydrolysate it still changes the expression of less than half the genes of the WT in 10% v/v *Populus* hydrolysate (489 genes

versus 1040 genes) (Table 10). All of the differentially expressed genes can be found in Supplemental File 8. The lower change in expression suggests that the PM has a more targeted response to the *Populus* hydrolysate than the WT does further supporting the cellular efficiency hypothesis. Furthermore, the symmetry between induced and repressed genes in the standard versus hydrolysate conditions (Table 10) suggests that a global conservation principle, possibly competition for cellular resources, is involved in the dynamics of the genetic regulatory system [40]. Analysis of the categories with a significant number of differentially expressed genes may provide insight to the differences in these two strains. The PM s to increases the expression level of very few genes in response to the *Populus* hydrolysate. The PM upregulates genes related to growth and downregulates genes related to adaptation or survival where as the WT upregulates genes related to survival and downregulates growth genes. In summary, the hydrolysate initiates a stress-link response in the WT, but not in the PM. Only one category of genes is similarly regulated between the two strains.

3.5.2.1 Upregulated genes in the PM in hydrolysate media

The genes that are significantly upregulated by the PM in hydrolysate conditions belong to energy production and conversion, amino acid transport and metabolism, inorganic ion transport and metabolism, and general transport and secretion.

Energy production and conversion

The PM increased the expression of five energy production and conversion genes in 10% v/v *Populus* hydrolysate which represents a significant increase in this category as determined by the odds ratio. Although, not significant by the odds ratio, the PM increased the expression of 12 genes in this

category in 17.5% v/v *Populus* hydrolysate. Specific differentially expressed genes related to the central metabolism can be seen in Table 13. An NADPH-dependant alcohol dehydrogenase (ADH6p) was identified as one of the enzymes responsible for HMF and furfural reduction in *S. cerevisiae.* Furthermore, gene deletion in the pentose phosphate pathway (PPP) genes exhibited growth deficiency in the presence of furfural indicating that *S. cerevisiae* tolerance to furfural was associated with the activity of PPP. The increased PPP level in the PM strain in hydrolysate might assist in protecting against and repairing furfural induced damage [29]. There are four possible reasons for the increased growth rate related to the expression of the energy production and conversion genes.

The first reason is that the PM has a mutation in Cthe_0158 (a deletion near the 3' end) which encodes for the RNase E/G family of endoribonucleases and has a central role in RNA degradation and processing (Table 9) [95]. The RNase E/G has been shown to be required for the normal decay of several transcripts including the functional form of enolase (*eno*) and acetaldehyde dehydrogenase (*adhE*) mRNA which are involved in glycolysis [95]. Both of these genes have a constant expression level where as the WT downregulates them in 10% v/v *Populus* hydrolysate (Table 13). The second possible reason is a mutation in the non-coding region upstream of Cthe_0422-Cthe_0423 operon which encodes the *Rex* (redox) repressor and *adhE* (Table 9) [105]. This operon is not differentially expressed in any of the PM hydrolysate conditions; however, it is severely downregulated in the WT in 10% v/v Populus hydrolysate compared to standard media (Table 13 contains the expression for Cthe_0423 and Table 14 contains the expression of Cthe_0422). These two mutations may allow for the increased expression of these genes and could result in greater tolerance to the hydrolysate [105].

The third possible reason is that in 10 and 17.5% v/v *Populus* hydrolysate, the PM decreases the expression of two sigma factors σ^A genes (Cthe_1438 and Cthe_1809) (Table 12). The PM reduces the

expression of Cthe_1438 by 2-fold in both concentration which would bring the expression back to the level of the WT in standard media. The PM also reduces the expression of Cthe_1809 by 2-fold in 10% v/v *Populus* hydrolysate and by 4-fold in 17.5% v/v *Populus* hydrolysate; however, since there was greater than a 16-fold increase in expression for the PM compared to the WT strain in standard media, the expression is still higher than the WT in standard media. The slight downregulation of the sigma factor σ^A for the PM in hydrolysate media may be the cause of the observed slightly slower growth phenotype in hydrolysate conditions (Figure 16). However, since the PM still has a faster growth rate in hydrolysate than the WT does in standard media, as expected the expression level for the sigma factor σ^A are still higher than in the WT in standard media.

Finally the differential expression of the transcriptional regulator *abrB* gene (Cthe_2100) is worth noting. Since there is a disruption in expression of both *spo0A* genes in the PM [105], the expected significant increase in expression of the *abrB* is observed under hydrolysate conditions. The PM increases the expression of *abrB* by 5- to 13-fold in 17.5% v/v *Populus* hydrolysate (Table 14). The WT in 10% v/v *Populus* hydrolysate also increases the expression of *abrB* at the mid-log time point but reduces the expression by the late-log time point. The competitive advantage that the *abrB* product provides the cell is to minimize the expression of extraneous gene products that would interfere with the maximal rate of growth of the cell under nutrient-excess conditions [100].

<u>Amino acid and inorganic ion transport and metabolism</u>

The expression levels of the amino acid transport and metabolism genes do not change expression levels for the PM in the hydrolysate conditions. Since the PM upregulated these genes in standard media compared to the WT, this means that the amino acid transport and metabolism genes

remain elevated in the hydrolysate conditions. The significantly upregulated histidine metabolism remains elevated in the hydrolysate condition with the exception of one gene Cthe_3028 which is down regulated. Histidine may be limited under furfural conditions so the further reduction of Cthe_3028 stops the conversion of histidine into histamine. The two terminal steps in histidine biosynthesis involve the reduction of NAD^+ to NADH, a reaction that may be slowed by the high NADH/NAD^+ ratio associated with fermentation [109].

The PM in 10% v/v *Populus* hydrolysate increased the expression of 3 genes, and in 17.5% v/v *Populus* hydrolysate increased the expression of 6 genes of the 27 genes along the cysteine and methionine metabolism pathway (Table 16). There are two copies (Cthe_1569 and Cthe_1842) of the *metY* gene in *C. thermocellum*, which produces L-cystathionie, both of which have mutations in the PM that seem to inactiviate the genes [105]. Cthe_1569 has a stop sequence inserted early in the gene and Cthe_1842 has a non-synonymous SNP in a highly conserved region. These two mutations essentially cut off the conversion of homoserine and hydrogen sulfide to homocyctiene. There is an increased expression in both of these genes in hydrolysate media in possibly an attempt to overcome these mutations. Cthe_1569 and Cthe_1842 also convert serine to cysteine. There are also a second set of genes (Cthe_1560 and Cthe_1840) which function along the same pathway. Cthe_1560 and Cthe_1840 are upregulated in the hydrolysate media as well. Cysteine contains a sulfhydryl group (-SH) which can link one chain of amino acids to another by disulfide linkage (R-S-S-R) through two cysteine molecules, one from each chain [130]. It has been shown the *metY* is not essential for *E. coli* cell growth [98]; the mutation may be an attempt to reduce acetate and H^+ formation.

The PM upregulated 3 genes belonging to inorganic ion transport and metabolism in 10% v/v *Populus* hydrolysate compared to standard media and both up- and down-regulated 9 genes in this

Table 16: Arithmetic mean of differentially expressed genes involved in the cysteine and methionine metabolism.

		PM vs. WT 0		PM 0 vs. 10		PM 0 vs. 17.5		WT 0 vs. 10	
		ML	LL	ML	LL	ML	LL	ML	LL
Cthe_0580	aminotransferase class I and II	1.22	1.48	1.03	-1.00	1.44	**2.03**	1.64	1.26
Cthe_0715	S-adenosylmethionine decarboxylase proenzyme	1.21	1.33	-1.51	-1.64	-1.87	**-2.76**	**-3.67**	-1.10
Cthe_0755	aminotransferase class I and II	-2.42	-1.28	-1.75	-1.37	-1.60	-2.06	**-6.77**	-1.25
Cthe_0961	aspartate-semialdehyde dehydrogenase	**-2.51**	**-2.11**	1.18	1.15	1.47	**2.34**	-1.01	-1.34
Cthe_1053	L-lactate dehydrogenase	-1.78	-1.25	-1.41	-1.27	-1.33	-1.16	**-3.30**	-1.55
Cthe_1200	Adenosylhomocysteinase	-1.26	1.07	1.39	1.18	1.01	-1.62	**-2.02**	-1.39
Cthe_1559	Cys/Met metabolism pyridoxal-phosphate-dependent protein	**-9.22**	**-5.72**	**4.66**	**3.12**	**16.05**	**8.31**	2.39	2.17
Cthe_1560	Pyridoxal-5'-phosphate-dependent protein beta subunit	**-6.16**	**-2.97**	**6.25**	**3.41**	**15.55**	**6.43**	**3.77**	**3.05**
Cthe_1569	Cys/Met metabolism pyridoxal-phosphate-dependent protein	1.02	1.09	**3.94**	**2.46**	**5.21**	**4.42**	**8.24**	**4.90**
Cthe_1728	DNA-cytosine methyltransferase	**2.09**	**2.38**	1.03	-1.01	1.59	1.80	**2.60**	1.04
Cthe_1749	DNA-cytosine methyltransferase	1.08	-1.12	-1.08	1.20	1.13	1.46	**5.95**	**2.58**
Cthe_1840	cysteine synthase A	-1.52	-1.21	1.37	1.27	1.83	-1.27	**-3.48**	**-2.07**
Cthe_1842	O-acetylhomoserine/O-acetylserine sulfhydrylase	-1.68	-1.54	1.51	1.16	**2.46**	1.75	-1.02	-2.01

Bold values indicate significantly differential expression by ANOVA. For the PM vs. WT in 0% v/v *Populus* hydrolysate a positive differential expression represents a higher expression in the PM and a negative differential expression represents a lower expression in the PM compared to the WT. For the standard media (0%) versus *Populus* hydrolysate media (10 or 17.5%) a positive value represents a higher expression and a negative value represents a lower expression in the hydrolysate media compared to standard media.

category in 17.5% v/v *Populus* hydrolysate compared to standard media all of which are considered significant via the odds ratio. For the PM in 17.5% v/v *Populus* hydrolysate, four of the upregulated genes belonged to the sulfate ABC transporter and four of the downregulated genes belonged to the phosphate ABC transporter. This suggests an increase in sulfate metabolism within the PM cell. The sulfur metabolism is linked to the cysteine and methionine metabolism pathway through the genes Cthe_1559, Cthe_1560 and Cthe_1840. These results suggest that the addition of furfural results in intercellular deficit in sulfur-containing amino acids (cysteine and methionine) which may be associated with the high NADPH requirements in this pathway [109]. Furfural could inhibit sulfur amino acids biosynthesis by limiting the availability of reduced sulfur (H_2S) from sulfate [109].

The increased transcript levels for histidine biosynthesis and cysteine and methionine metabolism may indicate an attempt to overcome a shortage of these amino acids in the cell [109]. The WT does not attempt to increase the transcription of these pathways. Depletion of cysteine and methionine by furfural would be expected to initiate a cascade of cellular events including stalled ribosomes that trigger a stringent response [109].

The genes that belong to the general transport category are basic ABC transporter and glycosyl transferase groups which are labeled with multiple COG designations. ABC transporters utilize ATP energy to transport inorganic ions, amino acids, hydrocarbons, polypeptides or hydrophobic compounds [1]. In some Gram-positive organisms, the ATP-binding subunit of an ABC system is not part of a specific transporter complex; instead, it is shared by multiple transporters [76] increasing the efficiency of the cell. The PM in 17.5% v/v *Populus* hydrolysate upregulated 16 of the 143 genes that encode for various transport genes compared to standard media. This may allow for faster transport of compounds into the cell, allowing the faster growth phenotype (Supplemental File 8).

3.5.2.2 Downregulated genes in the PM in hydrolysate media

A change in the environment causes a response of the genetic network which in turn allows efficient plastic adaptation of cellular metabolism to a broad range of unforeseen challenges [40]. Increased transcriptional flexibility allows the cells to address challenges on physiological timescales (not through new mutations) [40]. The PM in both the 10% and 17.5% v/v Populus hydrolysate downregulated the expression of transcription genes compared to the expression in standard media. The PM in 10% v/ v *Populus* hydrolysate decreases the expression of 8 transcription genes, and in 17.5% v/v *Populus* hydrolysate it decreases the expression of 22 genes (Supplemental File 8). The decrease in transcriptional genes may increase the amount of time needed to respond to changes in the environment for the PM which may not be beneficial in a natural environment. However, the PM decreased expression in transcriptional genes may be an attempt to conserve energy as a result of being well adapted to the particular media. The PM in 10% v/v *Populus* hydrolysate decreases the expression of four genes in the cell defense mechanism category which was determined significant by the odds ratio because of the small total number of genes being differentially expressed.

3.5.2.3 Upregulated genes in the WT in hydrolysate

The WT in hydrolysate media significantly upregulates two categories of genes that relates to survival mechanisms: cell defense mechanisms and cell motility genes. The WT already had a higher expression of the cell defense mechanism genes compared to the PM in standard media which is increases further in hydrolysate media. In 10% v/v *Populus* hydrolysate the WT increased the expression 38 genes in the cellular defense mechanism category compared to standard conditions (Supplemental File 8). The WT has an average 2-fold higher expression of 8 genes that encode Hedgehog/intein hint

domain proteins compared to standard media. Perhaps more significant is the fact that the WT increases the expression of 18 phage-associated proteins in 10% v/v *Populus* hydrolysate compared to standard media. This is possibly due to a programmed cell response to the general deterioration of the cell health in hydrolysate conditions. Although, increasing the expression of cellular defense mechanism genes in a stress environment may help the cell to survive in a natural environment, it takes resources away from the central metabolism and ethanol production. The WT in 10% v/v *Populus* hydrolysate increases the expression of 44 cell motility genes. The WT also upregulated the expression of sigma factor σ^D by 3-fold in 10% v/v *Populus* hydrolysate media (Table 12). The increase in motility of the WT in response to hydrolysate may be an attempt for the cell to swim away from unfavorable environments (Supplemental File 8). The PM may not see the hydrolysate conditions as an unfavorable environment and further conserves energy by reducing the expression of the cell motility genes. However, a transcriptional analysis of *Clostridium beijerinckii* found that the genes were downregulated during the switch from acidogenesis to solventogenesis during fermentation [114].

3.5.2.4 Downregulated genes in the WT in hydrolysate

The WT in 10% v/v *Populus* hydrolysate media downregulates the expression of the sigma factors σ^A gene Cthe_1809 by 2-fold compared to standard media. The further downregulation of Cthe_1809 by the WT in 10% v/v *Populus* hydrolysate may be the cause of the observed slower growth phenotype. Since the change in expression for Cthe_1809 is the most closely related to the observed growth rates in both the WT and PM, it may be one of the more important genes that encode for sigma factor σ^A in *C. thermocellum*. The WT in 10% v/v Populus hydrolysate does upregulated a sigma 70 region 2 domain protein; however, the protein is approximately half the length of the genes encoding for the RNA polymerase sigma factors; therefore, its exact function is unknown.

The WT in 10% v/v *Populus* hydrolysate decreases the expression of 37 genes in the cell envelope and outer membrane category compared to standard media (Supplemental File 8). The PM in standard media also decreases the genes in this category which supports the hypothesis that reducing the expression of cell envelope and outer membrane genes might be a way to conserve energy within the cell. The cytoplasmic membranes of all cells contains lipids [130] and lipids are important source of metabolic energy in all organisms [131]. Therefore, lipid degradation and biosynthesis pathways must be switched on and off according to the availability of fatty acids to maintain membrane lipid homeostasis [131]. The WT in 10% v/v *Populus* hydrolysate downregulated 11 of the 45 genes belonging to lipid metabolism compared to standard media (Supplemental File 8). The reduction of lipid degradation and biosynthesis pathways suggests that the WT does not have the energy required to exert the elaborate and highly sophisticated regulation of these pathways in 10% v/v Populus hydrolysate[131]. This may be the reason for the downregulation of genes belonging to the cell envelope and outer membrane in 10% v/v *Populus* hydrolysate compared to standard media.

The WT also downregulated a significant number of amino acid transport and metabolism genes (20 genes) in 10% v/v *Populus* hydrolysate compared to the standard media (Supplemental File 8). However, the change in gene expression did not belong to a specific pathway. The WT downregulates 19 genes in 10% v/v *Populus* hydrolysate compared to standard media including all 8 ABC transporter genes that were increased in the PM hydrolysate comparisons (Supplemental File 8). This possibly reduces the amount of sulfur and phosphate available in the cell. It has been shown the sulfur metabolism to cysteine increases tolerance to furfural [109].

3.5.2.5 Similarly expressed category

The PM in 17.5 % v/v *Populus* hydrolysate increases the expression level of 14 genes that encode of parts of the cellulosome. The WT in 10% v/v *Populus* hydrolysate also increases the expression level of 30 genes which encode for parts of the cellulosome. The majority of the genes that have an increased expression belong to various glycoside hydrolase (GH) families. The various GH families encode for endo- and exoglucanases used to degrade the cellulose components [48,49]. The PM in 17.5% v/v *Populus* hydrolysate increases the expression of 8 GH family proteins, and the WT in 10% v/v *Populus* hydrolysate increases the expression of 18 GH family proteins. This may be due to the fact that the concentrated *Populus* hydrolysate contains significant amounts of other sugars from the original pretreated biomass such as 22.7 g/L glucose, 42.7 g/L xylose, 1.84 g/L arabinose and 6.34 g/L mannose [105] and these molecules may be used as signaling mechanisms in the regulation of cellulosomal gene activity.

3.6 Conclusion

The greater number of genes with decreased expression versus increased expression for the PM compared to the WT in standard media suggests increased cellular efficiency in the PM strain of *C. thermocellum*. The PM increases expression in energy production and conversion and in the histidine biosynthesis pathway. The PM has a decreased expression in a number of categories allowing for greater efficiency. Also, the small number of differentially expressed genes for the comparisons of the PM strain in hydrolysate media to standard media compared with the WT in hydrolysate media versus standard media suggests that there is a more targeted response to the *Populus* hydrolysate by the PM than the WT strain. The PM upregulates genes related to growth processes and downregulates genes related to

survival mechanism in the hydrolysate conditions. The WT had the opposite response when placed in the hydrolysate media. These expression level changes may be detrimental in natural environments but may allow for the better growth in the industrial environments to which the strain was evolved.

Conclusion

In this work, a 17.5% v/v *Populus* hydrolysate tolerant mutant strain of *Clostridium thermocellum* was developed through directed evolution from the wild type strain ATCC 27405 on acid pretreated *Populus* hydrolysate. The *Populus* hydrolysate contained a wide range of inhibitory compounds from the pretreatment process. The mutant strain's phenotype included increased growth rate in standard as well as hydrolysate media when compared to the wild type stain. The increased growth in hydrolysate media indicated an increase tolerance to the inhibitory compound from the pretreatment. The mutant also had an increased ethanol to acetate ratio and increased ethanol yield. The strain showed improved industrial robustness through the changes in its phenotype. Two hypotheses were considered for the tolerant phenotype: mutations to specific genes allowing the mutant strain to detoxify specific compounds in the hydrolysate media or mutations that modify the overall metabolism thereby increasing the growth rate of the mutant strain.

In understanding the mechanism of tolerant phenotypes, comparisons of genomic and transcriptomic characterizations of the wild type and mutant strains can be useful for inferring molecular scale differences. Although, not every mutation or differentially expressed gene would be expected to provide tolerance upon engineered expression, it may give insight to the possible mechanism of tolerance. The genome of the wild type strain, intermediate population samples and single colony isolate samples were sequenced in order to conduct a longitudinal genomic and pan-genomic analysis. Also, a transcriptomic analysis was conducted on one of the single colony isolates (isolate #6) and the wild type strain in various concentrations of *Populus* hydrolysate (0, 10 and 17.5% v/v *Populus* hydrolysate) with samples taken at the mid-log and late-log phase. Statistical analysis of the difference in gene expression was used to determine the significantly differentially expressed genes between the wild type and Populus mutant in standard condition, as well as the difference in expression for each strain in standard and hydrolysate media. The differentially expressed genes were then placed

into functional categories and the significance determined by odds ratio. Kinetic modeling was use to quantitatively describe the changes in phenotype during those fermentations. A Monod based model which included ethanol product inhibition and cellular decay was used to model the growth in standard media. A general inhibition term was added to the growth rate and the degradation rates of two inhibitors were added to the model for growth in standard media.

The kinetic models showed the extent of the change in phenotype between the wild type and *Populus* mutant strains. The *Populus* mutant has an almost 10% higher maximum growth rate than the wild type strain. The *Populus* mutant also has a higher yield of ethanol per gram sugar and lower yield of acetate per gram sugar than the wild type. The changes in yield make the ethanol-to- acetate ratio more desirable for the *Populus* mutant strain in an industrial setting Furthermore, the general inhibition term is lower for the wild type stain than the *Populus* mutant strain showing the wild type is more inhibited by the *Populus* hydrolysate than the mutant strain. The degradation of readily measured inhibitors (furfural and 5-HMF) is based on the cell mass and current inhibitor concentration for both strains. Since the *Populus* mutant has a faster growth rate, it is able to degrade the inhibitors more quickly which is shown by the decreased degradation constant. In 10% v/v *Populus* hydrolysate, the degradation terms for the Populus mutant are roughly a quarter of the wild type's degradation rate. The *Populus* mutant also had better growth compared to the wild type on individual inhibitors.

There were multiple mutations identified with the possibility of causing the observed phenotype, as expected. There were 24 core mutations that belonged to all seven single colony isolates and 8 mutational hotspots within the genome. There were mutations that showed the observed phenotype was due to increased cellular efficiency instead of increased rate of inhibitory compound degradation. The increased growth phenotype is likely due to a mutation in the distantly related

sporulation transcriptional activator *Spo0A* homologue. There was a stop codon placed early in the coding region of Cthe_3087 which likely disrupted the functionality of this gene. Literature shows that a mutation in *Spo0A* gene locks the strain into the exponential growth phases allowing for increased growth rate under nutrient limited conditions. The *Populus* mutant strain also had multiple mutations dealing with the redox balance in the cell allowing it to have more reducing power and increased central metabolism. One mutation in the non-coding region upstream of the operon Cthe_0422-Cthe_0423 may also be responsible for the increased ethanol-to-acetate ratio. Mutations that deal with DNA repair, and RNA transcription and translation may allow for increased cellular efficiency. Other mutations included mutations along amino acid pathways which allow the cell to regulate which amino acids are produced. Lastly, mutations involving solute transport allow for greater uptake of required nutrients into the cell.

The transcriptomic analysis further supports the hypothesis that the tolerant phenotype is the results of increased cellular efficiency. The wild type has an increased expression as twice the number of genes as the *Populus* mutant does in standard media. The categories of genes with increased expression for the *Populus* mutant also points to increased cellular efficiency with increased expression in transcription, energy production and conversion and amino acid transport compared to the wild type in standard media. One of the most significantly differentially expressed sets of genes belongs to histidine metabolism. The wild type has increased expression belonging to cell division and sporulation, cell defense mechanisms, cell envelope and outer membrane, cell motility, cellulosome, and inorganic ion transport and metabolism compared to the *Populus* mutant in standard media. These genes can also be thought of as having decreased expression for the *Populus* mutant when compared to the wild type. Although, these genes are required for survival in natural environments, decreasing the expression of these genes in a stable predictable environment such as the one used to drive the mutations, may allow for the fast growth rate that is required for an industrially robust strain.

The *Populus* mutant further demonstrates its increased cellular efficiency by the lack of change in gene expression when placed in hydrolysate media. The *Populus* mutant strain differentially expressed one tenth of the genes that the wild type does in 10% v/v *Populus* hydrolysate compared to each strain in standard media. Even when the *Populus* mutant strain is placed in 17.5% v/v *Populus* hydrolysate, the *Populus* mutant still differentially expresses less than half the genes that the wild type does in 10% v/v *Populus* hydrolysate. In general the *Populus* mutant increases the expression of genes encoding for of energy production and conversion, inorganic ion transport and metabolism, cellulosome and general transport of solutes, and decreases the expression of genes belonging to cell defense mechanisms and transcription when places in hydrolysate media. Conversely, the wild type strain increases the expression of genes encoding for cell defense mechanisms, cell motility, and the cellulosome, and decreases the expression of genes encoding for cell envelope and outer membrane, amino acid transport and metabolism, inorganic ion transport and metabolism, and lipid metabolism. This show a targeted response by the Populus mutant strain compared to the wild type strain in hydrolysate media due to increased cellular efficiency.

Evidence to disprove the hypothesis that the tolerant phenotype is a result of mutations to specific genes capable of detoxifying the *Populus* hydrolysate comes from the detoxification experiment. Both strains have the ability to degrade two readily measured inhibitors, furfural and 5-HMF. When both strains are regrown on media that has had furfural and 5-HMF degraded by either strain there was no difference in growth profiles for the two strains. The wild type and *Populus* mutant grew equally as well on the media detoxified by the wild type strain as the *Populus* mutant strain.

Although, the exact mechanism of tolerance is still unknown, the analysis conducted in this works supports the hypothesis that the *Populus* mutant strain tolerant phenotype is due to increased

cellular efficiency and increased growth rate. One benefit that the *Populus* mutant has over other tolerant mutant strains is that the *Populus* mutant strain does not suffer from poor growth in standard media like typical tolerant mutant strains. This study is the first investigation of the mechanism of tolerance for *C. thermocellum* to inhibition by a complex hydrolysate generated during pretreatment, in this case using *Populus* hydrolysate and lays the ground work for producing an industrially robust microorganism for the construction of an industrially robust organism for economical ethanol production.

Future work

Although this work has done much of the ground work, there are many points left undiscovered. Future work may include strain engineering to verify and further elucidate the genetic mutations responsible for the change in phenotype. The proteins with mutations of unknown functions such as mutation not inside of a domain may be isolated and the change in specificity, activity or conformation determined. Other mutations not considered in this work such as mutations belonging to a single isolate or shared by a subset of isolates may also give insight to the tolerant phenotype. The kinetic modeling and transcriptomic analysis may also be repeated with other isolates to further determine the most important set of differentially expressed genes. Growth experiments in other conditions based on the change in expression data may also provide insightful such as determining the optimal temperature range for the *Populus* mutant since the extreme heat sigma factor had a change in expression (data not shown). The *Populus* mutant may also be grown in other pretreated biomass hydrolysates to determine the specificity of the tolerance to the *Populus* hydrolysate and in real biomass to determine as of yet unseen limitations to the mutant strain.

List of References

1. Nicolaou SA, Gaida SM, Papoutsakis ET (2010) A comparative view of metabolite and substrate stress and tolerance in microbial bioprocessing: From biofuels and chemicals, to biocatalysis and bioremediation. Metabolic Engineering 12: 307-331.
2. DOE/SC-0095 (2006) Breaking the biological barriers to cellulosuc ethanol: a joint research agenda. U.S. Department of Energy.
3. Varga E, Klinke H, Reczey K, Thomsen AB (2004) High Solid Simultaneous Saccharification and Fermentation of Wet Oxidized Corn Stover to Ethanol. Biotechnology and Bioengineering. pp. 567-574.
4. Lynd LR, van Zyl WH, McBride JE, Laser M (2005) Consolidated bioprocessing of cellulosic biomass: an update. Current Opinion in Biotechnology 16: 577-583.
5. Miller R, Keller M (2009) The DOE BioEnergy Science Center-a US Department of Energy Bioenergy Research Center. In Vitro Cellular & Developmental Biology-Plant 45: 193-198.
6. la Grange DC, den Haan R, van Zyl WH (2010) Engineering cellulolytic ability into bioprocessing organisms. Applied Microbiology and Biotechnology 87: 1195-1208.
7. Pienkos PT, Zhang M (2009) Role of pretreatment and conditioning processes on toxicity of lignocellulosic biomass hydrolysates. Cellulose 16: 743-762.
8. Elkins JG, Raman B, Keller M (2010) Engineered microbial systems for enhanced conversion of lignocellulosic biomass. Current Opinion in Biotechnology 21: 657-662.
9. Hendriks A, Zeeman G (2009) Pretreatments to enhance the digestibility of lignocellulosic biomass. Bioresource Technology 100: 10-18.
10. Geddes CC, Mullinnix MT, Nieves IU, Peterson JJ, Hoffman RW, et al. (2010) Simplified process for ethanol production from sugarcane bagasse using hydrolysate-resistant *Escherichia coli* strain MM160.
11. Du BW, Sharma LN, Becker C, Chen SF, Mowery RA, et al. (2010) Effect of Varying Feedstock-Pretreatment Chemistry Combinations on the Formation and Accumulation of Potentially Inhibitory Degradation Products in Biomass Hydrolysates. Biotechnology and Bioengineering 107: 430-440.
12. Palmqvist E, Hahn-Hagerdal B (2000) Fermentation of lignocellulosic hydrolysates. II: inhibitors and mechanisms of inhibition. Bioresource Technology 74: 25-33.
13. Klinke HB, Thomsen AB, Ahring BK (2004) Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass. Applied Microbiology and Biotechnology 66: 10-26.
14. Chen SF, Mowery RA, Castleberry VA, van Walsum GP, Chambliss CK (2006) High-performance liquid chromatography method for simultaneous determination of aliphatic acid, aromatic acid and neutral degradation products in biomass pretreatment hydrolysates. Journal of Chromatography A 1104: 54-61.
15. Joh LD, Rezaei F, Barabote RD, Parales JV, Parales RE, et al. (2010) Effects of Phenolic Monomers on Growth of *Acidothermus cellulolyticus*. American Institute of Chemical Engineers. pp. 1-8.
16. Palmqvist E, Hahn-Hagerdal B (2000) Fermentation of lignocellulosic hydrolysates. I: inhibition and detoxification. Bioresource Technology 74: 17-24.
17. Franden MA, Pienkos PT, Zhang M (2009) Development of a high-throughput method to evaluate the impact of inhibitory compounds from lignocellulosic hydrolysates on the growth of Zymomonas mobilis. Journal of Biotechnology 144: 259-267.
18. Geng AL, Wang ZK, Lai KS, Tan MWY (2010) Bioenergy II: Comparison of Laboratory and Industrial Saccharomyces cerevisiae Strains for their Stress Tolerance. International Journal of Chemical Reactor Engineering 8.
19. Barber MS, McConnell VS, DeCaux BS (2000) Antimicrobial intermediates of the general phenylpropanoid and lignin specific pathways. Phytochemistry 54: 53-56.

20. Zaldivar J, Martinez A, Ingram LO (2000) Effect of alcohol compounds found in hemicellulose hydrolysate on the growth and fermentation of ethanologenic Escherichia coli. Biotechnology and Bioengineering 68: 524-530.
21. VanderGheynst JS, Rezaei F, Dooley TM, Berry AM (2010) Switchgrass Leaching Requirements for Solid-State Fermentation by Acidothermus cellulolyticus. Biotechnology Progress 26: 622-626.
22. Ezeji T, Qureshi N, Blaschek HP (2007) Butanol production from agricultural residues: Impact of degradation products on Clostridium beijerinckii growth and butanol fermentation. Biotechnology and Bioengineering 97: 1460-1469.
23. Rani KS, Seenayya G (1999) High ethanol tolerance of new isolates of Clostridium thermocellum strains SS21 and SS22. World Journal of Microbiology & Biotechnology 15: 173-178.
24. Dar RD, Karig DK, Cooke JF, Cox CD, Simpson ML (2010) Distribution and regulation of stochasticity and plasticity in Saccharomyces cerevisiae. Chaos 20.
25. Conrad TM, Lewis NE, Palsson BO (2011) Microbial laboratory evolution in the era of genome-scale science. Molecular Systems Biology 7: 11.
26. Zhu LC, Wang Q, Tang P, Araki H, Tian DC (2009) Genomewide Association between Insertions/Deletions and the Nucleotide Diversity in Bacteria. Molecular Biology and Evolution 26: 2353-2361.
27. Alsaker KV, Paredes C, Papoutsakis ET (2010) Metabolite Stress and Tolerance in the Production of Biofuels and Chemicals: Gene-Expression-Based Systems Analysis of Butanol, Butyrate, and Acetate Stresses in the Anaerobe Clostridium acetobutylicum. Biotechnology and Bioengineering 105: 1131-1147.
28. Medini D, Serruto D, Parkhill J, Relman DA, Donati C, et al. (2008) Microbiology in the post-genomic era. Nature Reviews Microbiology 6: 419-430.
29. Almeida JRM, Modig T, Petersson A, Hahn-Hagerdal B, Liden G, et al. (2007) Increased tolerance and conversion of inhibitors in lignocellulosic hydrolysates by Saccharomyces cerevisiae. Journal of Chemical Technology and Biotechnology 82: 340-349.
30. Miller EN, Turner PC, Jarboe LR, Ingram LO (2010) Genetic changes that increase 5-hydroxymethyl furfural resistance in ethanol-producing Escherichia coli LY180. Biotechnology Letters 32: 661-667.
31. Yang SH, Pelletier DA, Lu TYS, Brown SD (2010) The Zymomonas mobilis regulator hfq contributes to tolerance against multiple lignocellulosic pretreatment inhibitors. Bmc Microbiology 10.
32. Zhang Y, Han B, Chukwuemeka Ezeji T (2011) Biotransformation of furfural and 5-hydroxymethylfurfural (HMF) by *Clostridium acetobutylicium* ATCC 824 during butanol fermentation. New Biotechnology.
33. Brown SD, Guss AM, Karpinets TV, Parks JM, Smolin N, et al. (2011) Mutant alcohol dehydrogenase leads to improved ethanol tolerance in Clostridium thermocellum. Proceedings of the National Academy of Sciences of the United States of America 108: 13752-13757.
34. Williams TI, Combs JC, Lynn BC, Strobel HJ (2007) Proteomic profile changes in membranes of ethanol-tolerant Clostridium thermocellum. Applied Microbiology and Biotechnology 74: 422-432.
35. Brown SD, Raman B, McKeown CK, Kale SP, He Z, et al. (2007) Construction and Evaluation of a *Clostridium thermocellum* ATCC 27405 Whole-Genome Oligonucleotide Microarray. Humana Press Inc. pp. 663-674.
36. Dharmadi Y, Gonzalez R (2004) DNA microarrays: Experimental issues, data analysis, and application to bacterial systems. Biotechnology Progress 20: 1309-1324.
37. Causton HC, Ren B, Koh SS, Harbison CT, Kanin E, et al. (2001) Remodeling of yeast genome expression in response to environmental changes. Molecular Biology of the Cell 12: 323-337.

38. Hirasawa T, Furusawa C, Shimizu H (2010) Saccharomyces cerevisiae and DNA microarray analyses: what did we learn from it for a better understanding and exploitation of yeast biotechnology? Applied Microbiology and Biotechnology 87: 391-400.
39. Gasch AP, Spellman PT, Kao CM, Carmel-Harel O, Eisen MB, et al. (2000) Genomic expression programs in the response of yeast cells to environmental changes. Molecular Biology of the Cell 11: 4241-4257.
40. Stern S, Dror T, Stolovicki E, Brenner N, Braun E (2007) Genome-wide transcriptional plasticity underlies cellular adaptation to novel challenge. Molecular Systems Biology 3.
41. Tomas CA, Alsaker KV, Bonarius HPJ, Hendriksen WT, Yang H, et al. (2003) DNA array-based transcriptional analysis of asporogenous, nonsolventogenic Clostridium acetobutylicum strains SKO1 and M5. Journal of Bacteriology 185: 4539-4547.
42. Helmann JD, Wu MFW, Kobel PA, Gamo F-J, Wilson M, et al. (2001) Global Transcriptional Response of Bacillus subtilis to Heat Shock. pp. 7318-7328.
43. Shi Z, Blaschek HP (2008) Transcriptional Analysis of Clostridium beijerinckii NCIMB 8052 and the Hyper-Butanol-Producing Mutant BA101 during the Shift from Acidogenesis to Solventogenesis. Applied and Environmental Microbiology 74: 7709-7714.
44. Fawcett P, Eichenberger P, Losick R, Youngman P (2000) The transcriptional profile of early to middle sporulation in Bacillus subtilis. Proceedings of the National Academy of Sciences of the United States of America 97: 8063-8068.
45. Tomas CA, Beamish J, Papoutsakis ET (2004) Transcriptional analysis of butanol stress and tolerance in Clostridium acetobutylicum. Journal of Bacteriology 186: 2006-2018.
46. Dror TW, Morag E, Rolider A, Bayer EA, Lamed R, et al. (2003) Regulation of the cellulosomal celS (cel48A) gene of Clostridium thermocellum is growth rate dependent. Journal of Bacteriology 185: 3042-3048.
47. Zhang YHP, Lynd LR (2005) Regulation of cellulase synthesis in batch and continuous cultures of Clostridium thermocellum. Journal of Bacteriology 187: 99-106.
48. Raman B, Pan C, Hurst GB, Rodriguez M, McKeown CK, et al. (2009) Impact of Pretreated Switchgrass and Biomass Carbohydrates on Clostridium thermocellum ATCC 27405 Cellulosome Composition: A Quantitative Proteomic Analysis. Plos One 4.
49. Raman B, McKeown CK, Rodriguez M, Brown SD, Mielenz JR (2011) Transcriptomic analysis of Clostridium thermocellum ATCC 27405 cellulose fermentation. Bmc Microbiology 11: 15.
50. Riederer A, Takasuka TE, Makino S, Stevenson DM, Bukhman YV, et al. (2011) Global Gene Expression Patterns in Clostridium thermocellum as Determined by Microarray Analysis of Chemostat Cultures on Cellulose or Cellobiose. Applied and Environmental Microbiology 77: 1243-1253.
51. Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y (2008) RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. Genome Research 18: 1509-1517.
52. Bergemann TL, Wilson J (2011) Proportion statistics to detect differentially expressed genes: a comparison with log-ratio statistics. Bmc Bioinformatics 12: 10.
53. Oshlack A, Robinson MD, Young MD (2010) From RNA-seq reads to differential expression results. Genome Biology 11: 10.
54. Boor KJ (2006) Bacterial stress responses: What doesn't kill them can make them stronger. Plos Biology 4: 18-20.
55. Haldenwang WG (1995) THE SIGMA-FACTORS OF BACILLUS-SUBTILIS. Microbiological Reviews 59: 1-30.
56. van Schaik W, van der Voort M, Molenaar D, Moezelaar R, de Vos WM, et al. (2007) Identification of the sigma(B) regulon of Bacillus cereus and conservation of sigma(B)-regulated genes in low-GC-content gram-positive bacteria. Journal of Bacteriology 189: 4384-4390.

57. Price CW, Fawcett P, Ceremonie H, Su N, Murphy CK, et al. (2001) Genome-wide analysis of the general stress response in Bacillus subtilis. Molecular Microbiology 41: 757-774.
58. Petersohn A, Brigulla M, Haas S, Hoheisel JD, Volker U, et al. (2001) Global analysis of the general stress response of Bacillus subtilis. Journal of Bacteriology 183: 5617-5631.
59. Kahel-Raifer H, Jindou S, Bahari L, Nataf Y, Shoham Y, et al. (2010) The unique set of putative membrane-associated anti-Sigma factors in Clostridium thermocellum suggests a novel extracellular carbohydrate-sensing mechanism involved in gene regulation. Fems Microbiology Letters 308: 84-93.
60. Bahari L, Gilad Y, Borovok I, Kahel-Raifer H, Dassa B, et al. (2011) Glycoside hydrolases as components of putative carbohydrate biosensor proteins in Clostridium thermocellum. Journal of Industrial Microbiology & Biotechnology 38: 825-832.
61. Huang WH, Wang FS (2010) Kinetic modeling of batch fermentation for mixed-sugar to ethanol production. Journal of the Taiwan Institute of Chemical Engineers 41: 434-439.
62. Bouguettoucha A, Balannec B, Amrane A (2011) Unstructured Models for Lactic Acid Fermentation - A Review. Food Technology and Biotechnology 49: 3-12.
63. Nasrollahzadeh HS, Najafpour GD, Pazouki M, Younesi H, Zinatizadeh AA, et al. (2010) BIODEGRADATION OF PHENANTHRENE IN AN ANAEROBIC BATCH REACTOR: GROWTH KINETICS. Chemical Industry & Chemical Engineering Quarterly 16: 157-165.
64. Becker P, Markl H (2000) Modeling of olive oil degradation and oleic acid inhibition during chemostat and batch cultivation of Bacillus thermoleovorans IHI-91. Biotechnology and Bioengineering 70: 630-637.
65. Boyer LJ, Vega JL, Klasson KT, Clausen EC, Gaddy JL (1992) THE EFFECTS OF FURFURAL ON ETHANOL-PRODUCTION BY SACCHAROMYCES-CEREVISIAE IN BATCH CULTURE. Biomass & Bioenergy 3: 41-48.
66. Hidaka T, Horie T, Akao S, Tsuno H (2010) Kinetic model of thermophilic L-lactate fermentation by Bacillus coagulans combined with real-time PCR quantification. Water Research 44: 2554-2562.
67. Shuler ML, Kargi F (2002) Bioprocess Engineering: Basic Concepts; Amundson NR, editor. Upper Saddle River, NJ: Prentice Hall PTR.
68. Ljunggren M, Willquist K, Zacchi G, van Niel EWJ (2011) A kinetic model for quantitative evaluation of the effect of hydrogen and osmolarity on hydrogen production by Caldicellulosiruptor saccharolyticus. Biotechnology for Biofuels 4: 15.
69. Gnanapragasam G, Senthilkumara M, Arutchelvan V, Velayutham T, Nagarajan S (2011) Bio-kinetic analysis on treatment of textile dye wastewater using anaerobic batch reactor. Bioresource Technology 102: 627-632.
70. Rial D, Vazquez JA, Murado MA (2011) Effects of three heavy metals on the bacteria growth kinetics: a bivariate model for toxicological assessment. Applied Microbiology and Biotechnology 90: 1095-1109.
71. Desvaux M, Guedon E, Petitdemange H (2001) Carbon flux distribution and kinetics of cellulose fermentation in steady-state continuous cultures of Clostridium cellulolyticum on a chemically defined medium. Journal of Bacteriology 183: 119-130.
72. Palmqvist E, Almeida JS, Hahn-Hagerdal B (1999) Influence of furfural on anaerobic glycolytic kinetics of Saccharomyces cerevisiae in batch culture. Biotechnology and Bioengineering 62: 447-454.
73. Modig T, Liden G, Taherzadeh MJ (2002) Inhibition effects of furfural on alcohol dehydrogenase, aldehyde dehydrogenase and pyruvate dehydrogenase. Biochemical Journal 363: 769-776.
74. Shao XJ, Raman B, Zhu MJ, Mielenz JR, Brown SD, et al. (2011) Mutant selection and phenotypic and genetic characterization of ethanol-tolerant strains of Clostridium thermocellum. Applied Microbiology and Biotechnology 92: 641-652.

75. Roberts SB, Gowen CM, Brooks JP, Fong SS (2010) Genome-scale metabolic analysis of Clostridium thermocellum for bioethanol production. Bmc Systems Biology 4.
76. Yu TT, Xu XP, Peng YF, Luo YM, Yang KQ (2012) Cell wall proteome of Clostridium thermocellum and detection of glycoproteins. Microbiological Research 167: 364-371.
77. Nataf Y, Bahari L, Kahel-Raifer H, Borovok I, Lamed R, et al. (2010) Clostridium thermocellum cellulosomal genes are regulated by extracytoplasmic polysaccharides via alternative sigma factors. Proceedings of the National Academy of Sciences of the United States of America 107: 18646-18651.
78. Stevenson DM, Weimer PJ (2005) Expression of 17 genes in Clostridium thermocellum ATCC 27405 during fermentation of cellulose or cellobiose in continuous culture. Applied and Environmental Microbiology 71: 4672-4678.
79. Zhang YH, Lynd LR (2003) Quantification of cell and cellulase mass concentrations during anaerobic cellulose fermentation: Development of an enzyme-linked immunosorbent assay-based method with application to Clostridium thermocellum batch cultures. Analytical Chemistry 75: 219-227.
80. Ozkan M, Desai SG, Zhang Y, Stevenson DM, Beane J, et al. (2001) Characterization of 13 newly isolated strains of anaerobic, cellulolytic, thermophilic bacteria. Journal of Industrial Microbiology & Biotechnology 27: 275-280.
81. Schell DJ, Farmer J, Newman M, McMillan JD (2003) Dilute-sulfuric acid pretreatment of corn stover in pilot-scale reactor - Investigation of yields, kinetics, and enzymatic digestibilities of solids. Applied Biochemistry and Biotechnology 105: 69-85.
82. Jung HW, Tschaplinski TJ, Wang L, Glazebrook J, Greenberg JT (2009) Priming in Systemic Plant Immunity. Science 324: 89-91.
83. Miller EN, Jarboe LR, Yomano LP, York SW, Shanmugam KT, et al. (2009) Silencing of NADPH-Dependent Oxidoreductase Genes (yqhD and dkgA) in Furfural-Resistant Ethanologenic Escherichia coli. Applied and Environmental Microbiology 75: 4315-4323.
84. Grigoriev IV, Nordberg H, Shabalov I, Aerts A, Cantor M, et al. (2012) The Genome Portal of the Department of Energy Joint Genome Institute. Nucleic Acids Research 40: D26-D32.
85. Li H, Ruan J, Durbin R (2008) Mapping short DNA sequencing reads and calling variants using mapping quality scores. Genome Research 18.
86. Syed MH, Karpinets TV, Parang M, Leuze MR, Park BH, et al. (2012) BESC knowledgebase public portal. Bioinformatics 28: 750-751.
87. Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, et al. (2006) Pfam: clans, web tools and services. Nucleic Acids Research 34: D247-D251.
88. Bateman A, Birney E, Cerruti L, Durbin R, Etwiller L, et al. (2002) The Pfam Protein Families Database. Nucleic Acids Research 30: 276-280.
89. Bullard JH, Purdom E, Hansen KD, Dudoit S (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. Bmc Bioinformatics 11: 13.
90. Cotter PD, Hill C (2003) Surviving the acid test: Responses of gram-positive bacteria to low pH. Microbiology and Molecular Biology Reviews 67: 429-+.
91. Malik S, Willby M, Sikes D, Tsodikov OV, Posey JE (2012) New Insights into Fluoroquinolone Resistance in Mycobacterium tuberculosis: Functional Genetic Analysis of gyrA and gyrB Mutations. Plos One 7: 10.
92. Moeller R, Vlasic I, Reitz G, Nicholson WL (2012) Role of altered rpoB alleles in Bacillus subtilis sporulation and spore resistance to heat, hydrogen peroxide, formaldehyde, and glutaraldehyde. Archives of Microbiology 194: 759-767.
93. Hosokawa K, Park NH, Inaoka T, Itoh Y, Ochi K (2002) Streptomycin-resistant (rpsL) or rifampicin-resistant (rpoB) mutation in Pseudomonas putida KH146-2 confers enhanced tolerance to organic chemicals. Environmental Microbiology 4: 703-712.

94. Lee ASG, Ong DCT, Wong JCL, Siu GKH, Yam WC (2012) High-Resolution Melting Analysis for the Rapid Detection of Fluoroquinolone and Streptomycin Resistance in Mycobacterium tuberculosis. Plos One 7: 8.
95. Jourdan SS, McDowall KJ (2008) Sensing of 5 ' monophosphate by Escherichia coli RNase G can significantly enhance association with RNA and stimulate the decay of functional mRNA transcripts in vivo. Molecular Microbiology 67: 102-115.
96. Wietzke M, Bahl H (2012) The redox-sensing protein Rex, a transcriptional regulator of solventogenesis in Clostridium acetobutylicum. Applied Microbiology and Biotechnology 96: 749-761.
97. Carere CR, Rydzak T, Verbeke TJ, Cicek N, Levin DB, et al. (2012) Linking genome content to biofuel production yields: a meta-analysis of major catabolic pathways among select H-2 and ethanol-producing bacteria. Bmc Microbiology 12: 21.
98. Krin E, Laurent-Winter C, Bertin PN, Danchin A, Kolb A (2003) Transcription regulation coupling of the divergent argG and metY promoters in Escherichia coli K-12. Journal of Bacteriology 185: 3139-3146.
99. Grimsley JK, Tjalkens RB, Strauch MA, Bird TH, Spiegelman GB, et al. (1994) SUBUNIT COMPOSITION AND DOMAIN-STRUCTURE OF THE SPO0A SPORULATION TRANSCRIPTION FACTOR OF BACILLUS-SUBTILIS. Journal of Biological Chemistry 269: 16977-16982.
100. Hoch JA (1993) REGULATION OF THE PHOSPHORELAY AND THE INITIATION OF SPORULATION IN BACILLUS-SUBTILIS. Annual Review of Microbiology 47: 441-465.
101. Zeller T, Klug G (2006) Thioredoxins in bacteria: functions in oxidative stress response and regulation of thioredoxin genes. Naturwissenschaften 93: 259-266.
102. Ichinose H, Diertavitian S, Fujimoto Z, Kuno A, Lo Leggio L, et al. (2012) Structure-based engineering of glucose specificity in a family 10 xylanase from Streptomyces olivaceoviridis E-86. Process Biochemistry 47: 358-365.
103. Jeng WY, Wang NC, Lin MH, Lin CT, Liaw YC, et al. (2011) Structural and functional analysis of three beta-glucosidases from bacterium Clostridium cellulovorans, fungus Trichoderma reesei and termite Neotermes koshunensis. Journal of Structural Biology 173: 46-56.
104. Pao SS, Paulsen IT, Saier MH (1998) Major facilitator superfamily. Microbiology and Molecular Biology Reviews 62: 1-+.
105. Linville JL, Rodriguez Jr. M, Land M, Syed MH, Engle NL, et al. (Submitted) Industrial Robustness: Understanding the mechanism of tolerance for the *Populus* hydrolysate-tolerant mutant strain of *Clostridium thermocellum*. Plos One.
106. Baskaran S, Ahn HJ, Lynd LR (1995) INVESTIGATION OF THE ETHANOL TOLERANCE OF CLOSTRIDIUM-THERMOSACCHAROLYTICUM IN CONTINUOUS-CULTURE. Biotechnology Progress 11: 276-281.
107. Holwerda EK, Lynd LR (2013) Testing Alternative Kinetic Models for Utilization of Crystalline Cellulose (Avicel) by Batch Cultures of *Clostridium thermocellum*. Biotechnology and Bioengineering.
108. Marquardt DW (1963) AN ALGORITHM FOR LEAST-SQUARES ESTIMATION OF NONLINEAR PARAMETERS. Journal of the Society for Industrial and Applied Mathematics 11: 431-441.
109. Miller EN, Jarboe LR, Turner PC, Pharkya P, Yomano LP, et al. (2009) Furfural Inhibits Growth by Limiting Sulfur Assimilation in Ethanologenic Escherichia coli Strain LY180. Applied and Environmental Microbiology 75: 6132-6141.
110. He MX, Wu B, Shui ZX, Hu QC, Wang WG, et al. (2012) Transcriptome profiling of Zymomonas mobilis under furfural stress. Applied Microbiology and Biotechnology 95: 189-199.

111. Yang SH, Giannone RJ, Dice L, Yang ZMK, Engle NL, et al. (2012) Clostridium thermocellum ATCC27405 transcriptomic, metabolomic and proteomic profiles after ethanol stress. Bmc Genomics 13.
112. Yang SH, Land ML, Klingeman DM, Pelletier DA, Lu TYS, et al. (2010) Paradigm for industrial strain improvement identifies sodium acetate tolerance loci in Zymomonas mobilis and Saccharomyces cerevisiae. Proceedings of the National Academy of Sciences of the United States of America 107: 10395-10400.
113. Peng YF, Luo YM, Yu TT, Xu XP, Fan KQ, et al. (2011) A Blue Native-PAGE analysis of membrane protein complexes in Clostridium thermocellum. Bmc Microbiology 11: 10.
114. Wang Y, Li XZ, Mao YJ, Blaschek HP (2012) Genome-wide dynamic transcriptional profiling in clostridium beijerinckii NCIMB 8052 using single-nucleotide resolution RNA-Seq. Bmc Genomics 13.
115. Linville JL, Rodriguez Jr. M, Mielenz JR, Cox CD (Submitted) Kinetic modeling of batch fermentation for Populus hydrolysate tolerant mutant and wild type strains of Clostridium thermocellum. Bioresource Technology.
116. Morris JA, Gardner MJ (1988) CALCULATING CONFIDENCE-INTERVALS FOR RELATIVE RISKS (ODDS RATIOS) AND STANDARDIZED RATIOS AND RATES. British Medical Journal 296: 1313-1316.
117. Burgess RR, Anthony L (2001) How sigma docks to RNA polymerase and what sigma does. Current Opinion in Microbiology 4: 126-131.
118. Lamed R, Zeikus JG (1980) ETHANOL-PRODUCTION BY THERMOPHILIC BACTERIA - RELATIONSHIP BETWEEN FERMENTATION PRODUCT YIELDS OF AND CATABOLIC ENZYME-ACTIVITIES IN CLOSTRIDIUM-THERMOCELLUM AND THERMOANAEROBIUM-BROCKII. Journal of Bacteriology 144: 569-578.
119. Deng Y, Olson DG, Zhou JL, Herring CD, Shaw AJ, et al. (2013) Redirecting carbon flux through exogenous pyruvate kinase to achieve high ethanol yields in Clostridium thermocellum. Metabolic Engineering 15: 151-158.
120. Rydzak T, Levin DB, Cicek N, Sparling R (2011) End-product induced metabolic shifts in Clostridium thermocellum ATCC 27405. Applied Microbiology and Biotechnology 92: 199-209.
121. Paredes CJ, Alsaker KV, Papoutsakis ET (2005) A comparative genomic view of clostridial sporulation and physiology. Nature Reviews Microbiology 3: 969-978.
122. Appleby JL, Parkinson JS, Bourret RB (1996) Signal transduction via the multi-step phosphorelay: Not necessarily a road less traveled. Cell 86: 845-848.
123. Mearls EB, Izquierdo JA, Lynd LR (2012) Formation and characterization of non-growth states in Clostridium thermocellum: spores and L-forms. Bmc Microbiology 12.
124. Fineran PC, Charpentier E (2012) Memory of viral infections by CRISPR-Cas adaptive immune systems: Acquisition of new information. Virology 434: 202-209.
125. Horvath P, Barrangou R (2010) CRISPR/Cas, the Immune System of Bacteria and Archaea. Science 327: 167-170.
126. Richter H, Zoephel J, Schermuly J, Maticzka D, Backofen R, et al. (2012) Characterization of CRISPR RNA processing in Clostridium thermocellum and Methanococcus maripaludis. Nucleic Acids Research 40: 9887-9896.
127. Burglin TR (2008) The Hedgehog protein family. Genome Biology 9: 9.
128. Mann RK, Beachy PA (2004) Novel lipid modifications of secreted protein signals. Annual Review of Biochemistry 73: 891-923.
129. Hall TMT, Porter JA, Young KE, Koonin EV, Beachy PA, et al. (1997) Crystal structure of a hedgehog autoprocessing domain: Homology between hedgehog and self-splicing proteins. Cell 91: 85-97.
130. Madigan MT, Martinko JM, Dunlap PP, Clark DP (2009) Brock Biology of Microorganisms: Pearson Benjamin Cummings.

131. Fujita Y, Matsuoka H, Hirooka K (2007) Regulation of fatty acid metabolism in bacteria. Molecular Microbiology 66: 829-839.

Appendix

A1: Amount inhibited for concentration ranges of inhibitory compounds found in pretreated biomass for various microorganisms.

	S. cerevisiae				S. pombe				S. Roseus				B. subtilis				E. coli			
	Low		High		Low		High		Low		High		Low		High		Low		High	
Inhibitory compound[a]	C[a]	I	C	I	C	I	C	I	C	I	C	I	C	I	C	I	C	I	C	I
glucuronic acid																				
acetic acid	1.1	0[b]	4.5	48[b]																
Acetate																	10.0 (10.0)	36[f] (15)	20.0 (20.0)	100[f] (100[f])
formic acid	0.9	0[b]	3.5	48[b]																
Formate																	4.0 (4.0)	46[f] (51[f])	12.0 (12.0)	100[f] (100[f])
3,4-dihydroxybenzoic acid																				
vanillic acid																				
caffeic acid	1.4	8[c]			1.4	25[c]					1.4	95[c]			0.7	95[c]			1.4	95[c]
syringic acid																				
ferulic acid			0.8	95[c]	1.6	8[c]					0.4	95[c]			0.4	95[c]			0.4	95[c]
hydrocinnamic acid																				
coumaric acid																				
trans - p - Courmaic acid			1.3	93[c]			1.3	95[c]			1.3	95[c]			0.3	95[c]			0.3	95[c]
sinapic acid*	1.8	51[c]			1.8	0[c]					1.8	15[c]			0.4	95[c]			0.4	95[c]
5-hydroxymethylfurfural (5-HMF)	0.9	0[b]	3.8	48[b]													2.0 (2.0)	70[f] (5[f])	3.0 (4.0)	100[f] (100[f])
Furfural	0.7	0[b]	2.9	48[b]													0.5 (0.5)	12.5[f] (>0[f])	1.5 (2.0)	100[f] (100[f])
3,4-dihydroxybenzaldehyde																				
Vanillin																				
Syringaldehyde																				

Table 17: Concentrations at which compounds are inhibitory for selected organisms.

A)C is concentration in g/L and I is inhibition in percent, b) Six strains of yeast. All inhibitors present in a single cocktail. Averaged percent inhibited taken as [1 - (OD_{600} with inhibitors/OD_{600} w/o inhibitors)]*100 for all strains [18], c) Each compound was tested at six concentrations (8.0, 4.0, 2.0, 1.0, 0.5 and 0.25 mM). The minimum inhibitory concentration (MIC) that gave 95% inhibition was reported for most cases but actual % inhibition was given when the MIC was greater than 8.0 mM [19], d) The concentration inhibiting 25%, 50%, and 100% of growth were reported [20], e) The concentration inhibiting 25%, 50%, and 100% of growth were reported [17], f) Percent inhibited taken as [1 - (OD_{550} with inhibitors/OD_{550} without inhibitors)]*100 for each compound. The values given in parenthesis are for a sugar bagasse hydrolysate tolerant mutant [10], g) Percent inhibited taken as [1 - (OD_{550} with inhibitors/highest reported OD_{590}adj)]*100 for each compound[15], h) Percent inhibited taken as [1 - (cell conc. with inhibitors/cell conc. without inhibitors)]*100 for each compound [22].

	S. cerevisiae				S. pombe				S. Roseus				B. subtilis				E. coli			
	Low		High		Low		High		Low		High		Low		High		Low		High	
Inhibitory compound[a]	C[a]	I	C	I	C	I	C	I	C	I	C	I	C	I	C	I	C	I	C	I
p-Coumaraldehyde*	0.3	95[c]	0.6	100[c]	0.3	95[c]	0.6	100[c]	0.3	95[c]	0.6	100[c]			0.3	95[c]			0.3	95[c]
coniferaldehyde*	0.1	95[c]	0.4	100[c]	0.1	95[c]	0.4	100[c]	0.2	95[c]	0.4	100[c]			0.4	95[c]			0.7	95[c]
sinapaldehyde*	0.8	95[c]	1.7	100[c]	0.8	95[c]	1.7	100[c]	0.8	95[c]	0.8	100[c]			0.8	95[c]			1.7	95[c]
p-Coumaraldehyde*	0.3	95[c]	0.6	100[c]	0.3	95[c]	0.6	100[c]	0.3	95[c]	0.6	100[c]			0.3	95[c]			0.3	95[c]
Guaiacol																	0.35	25[d]	3.0	100[d]
Hydroquinone*																	0.50	25[d]	3.0	100[d]
methylcatechol*																	0.20	25[d]	1.5	100[d]
vanillyl alcohol*																	1.00	25[d]	9.0	100[d]
Catechol*																	0.35	25[d]	3.0	100[d]
furfuryl alcohol*																	2.00	25[d]	20.0	100[d]
p-Coumaryl alcohol*	1.2	45[c]			1.2	0[c]			1.2	0[c]					1.2	95[c]	1.20	80[c]		80[c]
coniferyl alcohol*	1.4	33[c]					1.4	90[c]	1.4	50[c]					1.4	84[c]	1.44	0[c]	3.0	100[d]
sinapyl alcohol*	1.7	32[c]					1.7	51[c]			1.7	80[c]			1.7	80[c]	1.68			81[c]

Table 17: Concentrations at which compounds are inhibitory for selected organisms.

A)C is concentration in g/L and I is inhibition in percent, b) Six strains of yeast. All inhibitors present in a single cocktail. Averaged percent inhibited taken as [1 - (OD_{600} with inhibitors/OD_{600} w/o inhibitors)]*100 for all strains [18], c) Each compound was tested at six concentrations (8.0, 4.0, 2.0, 1.0, 0.5 and 0.25 mM). The minimum inhibitory concentration (MIC) that gave 95% inhibition was reported for most cases but actual % inhibition was given when the MIC was greater than 8.0 mM [19], d) The concentration inhibiting 25%, 50%, and 100% of growth were reported [20], e) The concentration inhibiting 25%, 50%, and 100% of growth were reported [17], f) Percent inhibited taken as [1 - (OD_{550} with inhibitors/OD_{550} without inhibitors)]*100 for each compound. The values given in parenthesis are for a sugar bagasse hydrolysate tolerant mutant [10], g) Percent inhibited taken as [1 - (OD_{550} with inhibitors/highest reported OD_{590}adj)]*100 for each compound[15], h) Percent inhibited taken as [1 - (cell conc. with inhibitors/cell conc. without inhibitors)]*100 for each compound [22].

	Z. mobilis				P. syringae				A. cellulolyticus				C. beijerinckii			
	Low		High		Low		High		Low		High		Low		High	
Inhibitory compound[a]	C	I	C	I	C	I	C	I	C	I	C	I	C	I	C	I
glucuronic acid													1.00	7.0[h]	3.00	11.0[h]
acetic acid																
acetate	11	25[e]	25	100[e]												
formic acid																
formate																
3,4-dihydroxybenzoic acid									0.0015	2[g]	3.08	98.9[g]				
vanillic acid									0.0017	7.7[g]	3.36	98.8[g]				
caffeic acid							0.7	95[c]								
syringic acid									0.0020	7.5[g]	3.96	99.2[g]				
ferulic acid							0.4	95[c]	0.0194	34.8[g]	3.88	99[g]	0.25	60.0[h]	2.00	100[h]
hydrocinnamic acid									0.0015	1.1[g]	3.02	98.3[g]				
coumaric acid													0.25	7[h]	2.00	100[h]
trans - p - Courmaic acid							0.3	95[c]	0.0016	8.2[g]	3.28	98.9[g]				
sinapic acid*							0.9	95[c]								
5-hydroxymethylfurfural (5-HMF)	1.2	25[e]	8	100[e]									0.25	< 0[h]	3.00	< 0[h]
furfural	0.7	25[e]	5	100[e]									0.25	< 0[h]	3.00	< 0[h]
3,4-dihydroxybenzaldehyde									0.0014	7.6[g]	2.76	99[g]				
Vanillin									0.0015	12.7[g]	3.04	99[g]				
syringaldehyde													0.50	2.2[h]	3.00	33.3[h]

Table 17: Concentrations at which compounds are inhibitory for selected organisms.

A) C is concentration in g/L and I is inhibition in percent, b) Six strains of yeast. All inhibitors present in a single cocktail. Averaged percent inhibited taken as [1 - (OD600 with inhibitors/OD600 without inhibitors)]*100 for all strains [18], c) Each compound was tested at six concentrations (8.0, 4.0, 2.0, 1.0, 0.5 and 0.25 mM). The minimum inhibitory concentration (MIC) that gave 95% inhibition was reported for most cases but actual % inhibition was given when the MIC was greater than 8.0 mM [19], d) The concentration inhibiting 25%, 50%, and 100% of growth were reported [20], e) The concentration inhibiting 25%, 50%, and 100% of growth were reported [17], f) Percent inhibited taken as [1 - (OD550 with inhibitors/OD550 without inhibitors)]*100 for each compound. The values given in parenthesis are for a sugar bagasse hydrolysate tolerant mutant [10], g) Percent inhibited taken as [1 - (OD550 with inhibitors/highest reported OD590adj)]*100 for each compound[15], h) Percent inhibited taken as [1 - (cell conc. with inhibitors/cell conc. without inhibitors)]*100 for each compound [22].

	Z. mobilis				P. syringae				A. cellulolyticus				C. beijerinckii			
	Low		High		Low		High				Low		High		Low	
Inhibitory compound[a]	C	I	C	I	C	I	C	I	C	I	C	I	C	I	C	I
p-Coumaraldehyde*							0.3	95[c]								
coniferaldehyde*							0.4	95[c]								
sinapaldehyde*							0.8	95[c]								
phenol													0.25	38[h]	3.00	100[h]
guaiacol																
Hydroquinone*																
methylcatechol*																
vanillyl alcohol*																
Catechol*																
furfuryl alcohol*																
p-Coumaryl alcohol*							1.2	72[c]								
coniferyl alcohol*							1.4	55[c]								
sinapyl alcohol*							1.7	37[c]								

Table 17: Concentrations at which compounds are inhibitory for selected organisms.

A) C is concentration in g/L and I is inhibition in percent, b) Six strains of yeast. All inhibitors present in a single cocktail. Averaged percent inhibited taken as [1 - (OD600 with inhibitors/OD600 without inhibitors)]*100 for all strains [18], c) Each compound was tested at six concentrations (8.0, 4.0, 2.0, 1.0, 0.5 and 0.25 mM). The minimum inhibitory concentration (MIC) that gave 95% inhibition was reported for most cases but actual % inhibition was given when the MIC was greater than 8.0 mM [19], d) The concentration inhibiting 25%, 50%, and 100% of growth were reported [20], e) The concentration inhibiting 25%, 50%, and 100% of growth were reported [17], f) Percent inhibited taken as [1 - (OD550 with inhibitors/OD550 without inhibitors)]*100 for each compound. The values given in parenthesis are for a sugar bagasse hydrolysate tolerant mutant [10], g) Percent inhibited taken as [1 - (OD550 with inhibitors/highest reported OD590adj)]*100 for each compound[15], h) Percent inhibited taken as [1 - (cell conc. with inhibitors/cell conc. without inhibitors)]*100 for each compound [22].

A2: Matlab Files used for Kinetic Modeling

A2.1 Optimization function

File Name: myObj.m

```
function obj = myObjective(x, Yobs1, Yobs2, Yobs3 ,tobs)
%Current function line is for use with Growth_short.m
%Alternative function line: function obj = myObjective(x, q, Yobs1, Yobs2,
Yobs3, tobs)

%My Objective finds the optimal parameter values for either the standard or
%hydrolysate conditions based on the observed data from the triplicate
%fermentations by using a function minimization solver to minimize the
%difference between the observed value and calculated value based on the
%minimization of SSE/SST. SSE/SST was used to reduce the bias of the
%different magnitudes in concentration.

%Yobs is a 2D matrix for each of the triplicate fermentation runs of the
observed values at each sample time point
%tobs is a 1D matrix of the times that the sample time points
% x is a 1D matrix of the parameters to be optimized
% x = [Ks, Kp, Yx/s, Ye/s, Ya/s, mu_max, Kd]
% These are found in the specific .m file for each of the conditions and
% can be copied into the work space.

%Calculates the average value from the triplicate fermentations
Yavg = (Yobs1 +Yobs2 +Yobs3)/3;

%Calculating the unbiased standard deviation. The fermentations were run in
%triplicate so n-1 = 2

sig1 = (Yavg-Yobs1).^2;
sig2 = (Yavg-Yobs2).^2;
sig3 = (Yavg-Yobs3).^2;
SD = ((sig1+sig2+sig3)/2).^(1/2);

%Y0 is a 1D matrix of the avergae initial concentrations;
Y0 = Yavg(1,:);

%y are the modeled components
%Y0 is a 1D matrix of the guess of initial concentrations

[T Y] = ode45(@(t,y) growth_short(t, y, x),[tobs],[Y0]);
%Alternate for hydrolysate conditions:
%[T Y] = ode45(@(t,y) growth(t,y,x,q),[tobs],[Y0])

%Calculates weighted sum squared error
WSE = ((Y - Yavg)./SD).^2;
WSE2 = sum(WSE);
WSSE = WSE2(1)+WSE2(2)+WSE2(3)+WSE2(4)
%WSSE = WSE2(1)+WSE2(2)+WSE2(3)+WSE2(4)+WSE2(5)+WSE2(6)
obj2=WSSE;
end
```

```
%end of the objective function file.

%On the command line on the main matlab page you will populate the
%matricies for Yobs, tobs, Y0. Then type:

%For standard conditions:
%ObjectiveFunction = @(x)myObj(x, Yobs1, Yobs2, Yobs3, tobs);

%For hydrolysate conditions:
%ObjectiveFunction = @(q)myObj(x, q, Yobs1, Yobs2, Yobs3, tobs);

% this uses an anonomous function as described here:
%http://www.mathworks.com/help/optim/ug/passing-extra-parameters.html
          %
% X0 = [0.9, 3, 0.2, 0.6...  ] %these are your initial guesses of x, Type
        %them all instead of ...

% lb = [0,0,0,0,0,0];
% ub = [10,10,1,1,10,1];
% [x,fval,exitFlag,output] = simulannealbnd(ObjectiveFunction,X0,lb,ub);

% you can also specify options if needed:

% options = saoptimset('simulannealbnd');
% options = saoptimset('MaxFunEvals',2000000);
% [x,fval,exitFlag,output] =
simulannealbnd(ObjectiveFunction,X0,lb,ub,options);

%for hydrolysate conidtions:
% X0=q;
%[q,fval,exitFlag,output] = simulannealbnd(ObjectiveFunction,X0,lb,ub);
```

A2.2 Growth calculations for standard conditions

File Name: Growth_short.m

```
function dy = growth_short( t, y, x)
%Modeling of an anaerobic bacteria in batch fermentations
%The short version of the growth files will be used for the standard media
%conditions where the variables are listed below will be calculated for the
%sample points of the fermentation unit.

% Variables
% y(1) is biomass (g/L)
% y(2) is cellbiose (g/L)
% y(3) is ethanol (g/L)
% y(4) is acetic acid (g/L)

% Initial Guess for parameters -
%mu_max = 0.5
%Ks =0.9
%Kp = 3
%Yxs = 0.23
%Yex = 0.4
%Yax = 1.2

%[t] = [time points]
%[y] = [variables above]
%[x] = [umax, Ks, Yx/s, Ye/x, Ya/x]

%[T Y] = ode45(@(t,y) growth(t,y,x),[0 15],[0.005, 6, 0, 1.3, 0.5, 1.5])

mu = x(1) *(y(2)/(x(2) + y(2)));
dy = zeros(4,1);
dy(1) = (mu-0.12) * y(1);
dy(2) = - mu *y(1)/x(3);
dy(3) = (x(4)/x(3))*mu*y(1);
dy(4) = (x(5)/x(3))*mu*y(1);


end
```

A2.3 Growth calculations for hydrolysate conditions

File name: Growth.m

```
function dy = growth( t, y, x, q)
%Modeling of an anerobic bacteria in batch fermentations
%The full version of the growth files will be used for the hydrolysate media
%conditions where the variables are listed below will be calculated for the
%sample points of the fermentation unit.

% Variables
% y(1) is biomass (g/L)
% y(2) is cellbiose (g/L)
% y(3) is ethanol (g/L)
% y(4) is acetic acid (g/L)
% y(5) is HMF (g/L)
% y(6) is Furfural (g/L)

%[t] = [time points]
%[y] = [variables above]
%[x] = [umax, Ks, Yx/s]
% x variables are being held constant for these calculations
%[q] = [KiHMF, KiF, Kigen, Ye/s, Ya/s]
% q variables are being optimized.

mu = x(1)*(y(2)/(x(2) + y(2)))*q(3);

dy = zeros(6,1);
dy(1) = (mu-0.12) * y(1);
dy(2) = - mu *y(1)/x(3);
dy(3) = (q(4)/x(3))*mu*y(1);
dy(4) = (q(5)/x(3))*mu*y(1);
dy(5) = -q(1)*y(5)*y(1);
dy(6) = -q(2)*y(6)*y(1);
end
```

A2.4 Calculations of R^2, confident intervals and parameter correlations in standard media

File Name: Jacob.m

```
function jacob = myJacobian(x,tobs,Yobs1, Yobs2, Yobs3, num)

%calculates the average and standard deviation for experimental data
Yavg = (Yobs1 +Yobs2 +Yobs3)/3;

sig1 = (Yavg-Yobs1).^2;
sig2 = (Yavg-Yobs2).^2;
sig3 = (Yavg-Yobs3).^2;
SD = ((sig1+sig2+sig3)/2).^(1/2);

%calculates the initial values
Y0 = Yavg(1,:);

%calculates the Jacobian matrix for the optimized parameters
    J= zeros(num*4,5);

%Calculates the optimum value of Y
[T Y] = ode45(@(t,y) growth_short(t, y, x),[tobs],[Y0]);
 Y
 Yopt=(Y-Yavg)./SD;

%Calculates R2 for each variable
%Calculates SSE
se1 = (Y-Yobs1).^2;
sse1 = sum(se1);
se2 = (Y-Yobs2).^2;
sse2 = sum(se2);
se3 = (Y-Yobs3).^2;
sse3 = sum(se3);

ssem = sse1 + sse2 + sse3; %ssem = Sum squared Error Matrix

%Calculates SST
Y1avg = (sum(Yobs1))/num;
Y2avg = (sum(Yobs2))/num;
Y3avg = (sum(Yobs3))/num;
Yavgs = (Y1avg + Y2avg + Y3avg)/3;

for a = 1:num %Matrix have to be the same size to subtract
    Yavg(a,:)= Yavgs;
end

st1= (Yobs1 - Yavg).^2;
sst1 = sum(st1);
st2= (Yobs2 - Yavg).^2;
sst2 = sum(st2);
st3= (Yobs3 - Yavg).^2;
sst3 = sum(st3);
```

```
sstm = sst1 + sst2 + sst3;
sse_t = ssem./sstm;
R2 = 1-sse_t
R2tot = 1 -(sum(ssem) /sum(sstm))

for a = 1:5 %changes the ath parameter by 0.1% and evlauates the 4 variables
for the time points
    z=x;
    z(a)= x(a)*1.001;

    [T Y] = ode45(@(t,y) growth_short(t, y, z),[tobs],[Y0]);
     Ydiff = (Yavg-Y)./SD;

    diff = (Yopt-Ydiff)/(x(a)-z(a));
    % Fills in the Jacobian matrix
    J(1:num,a) = diff(:,1);
    J(num+1:2*num,a) = diff(:,2);
    J(2*num+1:3*num,a)= diff(:,3);
    J(3*num+1:4*num,a)= diff(:,4);

end
    jacob=((J.')*J)^-1;
    output = zeros(3,5);

    %Calculates the 95% confidence intervale
for b= 1:5
    output(1,b) = x(b);
    output(2,b) = x(b)+1.96*sqrt(jacob(b,b));
    output(3,b) = x(b)-1.96*sqrt(jacob(b,b));

end

    output

    %Calcualtes the covariance of the parameters
    Cor = zeros (5,5);
    for a = 1:5
        for b = 1:5
            Cor(a,b) = jacob(a,b)/sqrt(jacob(a,a)*jacob(b,b));
        end
    end

    Cor
end
```

A2.5 Calculations of R^2, confident intervals and parameter correlations in hydrolysate media

File Name: Jacob_hydrol.m

```
function jacob_hydrol = myJacobian(q, x, tobs, Yobs1, Yobs2, Yobs3, num)

%calculates the average and standard deviation for experimental data
Yavg = (Yobs1 +Yobs2 +Yobs3)/3;

sig1 = (Yavg-Yobs1).^2;
sig2 = (Yavg-Yobs2).^2;
sig3 = (Yavg-Yobs3).^2;
SD = ((sig1+sig2+sig3)/2).^(1/2);

%calculates the initial values
Y0 = Yavg(1,:);

%calculates the Jacobian matrix for the optimized parameters
J= zeros(num*6,5);

%Calculates the optimum value of Y
[T Y] = ode45(@(t,y) growth(t, y, x, q),[tobs],[Y0]);
 Y
Yopt=(Y-Yavg)./SD;

%Calculates R2 for each variable
%Calculates SSE
se1 = (Y-Yobs1).^2;
sse1 = sum(se1);
se2 = (Y-Yobs2).^2;
sse2 = sum(se2);
se3 = (Y-Yobs3).^2;
sse3 = sum(se3);

ssem = sse1 + sse2 + sse3; %ssem = Sum squared Error Matrix

%Calculates SST
Y1avg = (sum(Yobs1))/num;
Y2avg = (sum(Yobs2))/num;
Y3avg = (sum(Yobs3))/num;
Yavgs = (Y1avg + Y2avg + Y3avg)/3;

for a = 1:num %Matrix have to be the same size to subtract
    Yavg(a,:)= Yavgs;
end

st1= (Yobs1 - Yavg).^2;
sst1 = sum(st1);
st2= (Yobs2 - Yavg).^2;
sst2 = sum(st2);
st3= (Yobs3 - Yavg).^2;
sst3 = sum(st3);
```

```
sstm = sst1 + sst2 + sst3;
sse_t = ssem./sstm;
R2 = 1-sse_t
R2tot = 1 -(sum(ssem) /sum(sstm))

for a = 1:5 %changes the ath parameter by 10% and evlauates the 4 variables
for the time points
    z=q;
    z(a)= q(a)*1.001;

    [T Y] = ode45(@(t,y) growth(t, y, x, z),[tobs],[Y0]);
     Ydiff = (Yavg-Y)./SD;

    diff = (Yopt-Ydiff)/(q(a)-z(a));

    % Fills in the Jacobian matrix
    J(1:num,a) = diff(:,1);
    J(num+1:2*num,a) = diff(:,2);
    J(2*num+1:3*num,a)= diff(:,3);
    J(3*num+1:4*num,a)= diff(:,4);
    J(4*num+1:5*num,a)=diff(:,5);
    J(5*num+1:6*num,a)=diff(:,6);
end
    jacob_hydrol=((J.')*J)^-1;
    output = zeros(3,5);

    for b= 1:5
        output(1,b) = q(b);
        output(2,b) = q(b)+1.96*sqrt(jacob_hydrol(b,b));
        output(3,b) = q(b)-1.96*sqrt(jacob_hydrol(b,b));

    end

    output

    %Calcualtes the covariance of the parameters
    Cor = zeros (5,5);
    for a = 1:5
        for b = 1:5
            Cor(a,b) =
jacob_hydrol(a,b)/sqrt(jacob_hydrol(a,a)*jacob_hydrol(b,b));
        end
    end

    Cor
    end
```

A2.6 Wild type in 0% v/v *Populus* hydrolysate observed values

File: YobsWT.0

```
%Contains the time that the observation data points were taken and the data
%from the triplicate fermentations for the wild type strain of C.
%thermocellum in 0% v/v Populus Hydrolysate.

% Observational data given as [biomass (g/L), cellbiose (g/L), ethanol (g/L),
% acetic acid (g/L)]

%Adjusted time points when the samples were taken
tobs = [0; 1; 2; 3; 4; 5; 6; 7; 8; 9; 9.5; 10; 11; 12];

%Number of data points
num=14;

Yobs1 = [0.032  4.921   0.04    0.049
0.041   4.815   0.042   0.049
0.060   4.741   0.041   0.105
0.094   4.581   0.054   0.102
0.126   4.36    0.065   0.157
0.175   4.196   0.067   0.305
0.282   3.683   0.105   0.382
0.387   2.886   0.143   0.508
0.544   2.287   0.195   0.734
0.772   1.238   0.228   1.151
0.858   0.498   0.334   1.262
0.863   0.293   0.379   1.288
0.745   0.18    0.383   1.359
0.685   0.128   0.393   1.405
];

Yobs2 = [0.033  4.991   0.05    0.039
0.042   4.865   0.058   0.052
0.063   4.752   0.06    0.08
0.092   4.589   0.066   0.11
0.124   4.406   0.076   0.16
0.169   4.074   0.083   0.243
0.277   3.677   0.114   0.348
0.370   3.021   0.157   0.485
0.494   2.498   0.198   0.659
0.687   1.38    0.287   0.943
0.786   0.896   0.305   1.213
0.894   0.33    0.394   1.277
0.743   0.195   0.409   1.364
0.697   0.115   0.41    1.431
];

Yobs3 = [0.030  4.929   0.036   0.034
0.036   4.881   0.046   0.069
0.054   4.858   0.027   0.094
0.081   4.588   0.053   0.098
0.110   4.361   0.063   0.138
```

```
0.155   4.114   0.072   0.223
0.271   3.716   0.097   0.316
0.356   3.213   0.125   0.519
0.480   2.409   0.172   0.628
0.689   1.436   0.256   0.904
0.766   0.972   0.247   1.233
0.869   0.37    0.352   1.323
0.749   0.215   0.363   1.364
0.697   0.148   0.354   1.448
];
```

A2. 7 Wild type in 10% v/v *Populus* hydrolysate observed values

File name: YobsWT10.m

```
%Contains the time that the observation data points were taken and the data
%from the triplicate fermentations for the Wild type strain of C.
%thermocellum in 10% v/v Populus Hydrolysate.

% Observational data given as [biomass, cellbiose (g/L), ethanol (g/L),
% acetic acid (g/L), HMF (g/L) and Furfural (g/L)]

%Adjusted time points when the samples were taken
tobs = [0; 2; 4; 6; 7; 8; 9; 10; 11; 12; 12.5; 13; 14; 15];

%Number of data points
num = 14;


Yobs1 = [0.032  5.692   0.023   1.444   0.047591    0.082339
0.052   5.499   0.043   1.483   0.038593    0.042913
0.088   5.222   0.06    1.547   0.019654    0.017408
0.149   4.916   0.087   1.77    0.008467    0.008202
0.181   4.538   0.095   1.736   0.00675 0.007495
0.269   4.253   0.124   1.941   0.004462    0.004999
0.361   3.82    0.175   2.105   0.003524    0.00003953
0.398   3.753   0.174   2.414   0.00311 0.00003188
0.494   3.148   0.196   2.588   0.002708    0.00003199
0.645   2.347   0.271   2.77    0.001984    0.00002683
0.699   2.188   0.303   3.23    0.002084    0.00002157
0.779   1.51    0.443   2.751   0.001912    0.00003016
0.772   0.311   0.517   2.985   0.001554    0.00002635
0.699   0.354   0.555   3.112   0.001393    0.000008
];

Yobs2 = [0.034  5.706   0.033   1.434   0.047078    0.0773
0.057   5.383   0.051   1.49    0.03717 0.040308
0.092   5.158   0.07    1.566   0.017742    0.016844
0.155   4.933   0.093   1.843   0.007958    0.007588
0.190   4.461   0.113   2.202   0.007712    0.007515
0.285   4.462   0.149   2.161   0.00454 0.0005451
0.371   3.714   0.2 2.012   0.00397 0.00004728
0.430   3.199   0.258   2.173   0.003093    0.00004339
0.497   2.716   0.297   2.47    0.002372    0.00003576
0.652   2.02    0.354   2.698   0.00194 0.00003143
0.731   1.526   0.44    2.705   0.001926    0.0000345
0.779   1.197   0.497   2.793   0.001923    0.00003265
0.758   0.352   0.579   3.018   0.001584    0.00002877
0.683   0.37    0.597   3.338   0.001495    0.00002608
];

Yobs3 = [0.025  5.631   0.026   1.45    0.048219    0.091794
0.040   5.543   0.042   1.478   0.042318    0.056599
0.071   5.362   0.054   1.586   0.024422    0.020609
```

```
0.117   5.106   0.074   1.713   0.010294    0.009493
0.140   4.837   0.083   1.686   0.008478    0.008639
0.218   4.526   0.105   1.756   0.00614 0.000664
0.297   4.198   0.148   1.871   0.004577    0.00005119
0.350   3.826   0.186   2.027   0.003391    0.00004299
0.403   3.295   0.229   2.13    0.002987    0.00004485
0.547   2.785   0.285   2.33    0.002344    0.00003742
0.634   2.435   0.328   2.449   0.002165    0.00003632
0.638   2.077   0.364   2.524   0.002172    0.00003412
0.745   1.352   0.453   2.774   0.001721    0.00003019
0.736   0.332   0.521   2.974   0.001495    0.00002687
];
```

A2.8 *Populus* mutant in 0% v/v *Populus* hydrolysate observed values

File Name: YobsPM0.m

```
%Contains the time that the observation data points were taken and the data
%from the triplicate fermentations for the Populus Mutant strain of C.
%thermocellum in 0% v/v Populus Hydrolysate.

% Observational data given as [biomass (g/L), cellbiose (g/L), ethanol (g/L),
% acetic acid (g/L)]

%Adjusted time points when the samples were taken
tobs = [0; 1; 2; 3; 3.5; 4; 5; 6; 7];

%Number of data points
num = 9;

Yobs1 = [0.041  4.594   0.046   0.043
0.078   4.254   0.075   0.133
0.137   3.863   0.085   0.218
0.264   3.017   0.122   0.311
0.438   2.39    0.16    0.442
0.585   1.851   0.181   0.598
0.905   0.392   0.371   1.022
0.858   0.245   0.402   1.177
0.780   0.144   0.408   1.255
];

Yobs2 = [0.040  4.588   0.049   0.037
0.079   4.282   0.054   0.081
0.138   4.04    0.065   0.19
0.263   3.07    0.102   0.306
0.458   2.747   0.142   0.497
0.587   1.934   0.173   0.585
0.924   0.419   0.382   1.012
0.877   0.241   0.424   1.136
0.767   0.18    0.424   1.132
];

Yobs3 = [0.029  4.653   0.043   0.029
0.059   4.384   0.05    0.059
0.101   3.993   0.058   0.126
0.205   3.292   0.084   0.248
0.369   2.769   0.117   0.368
0.464   2.358   0.133   0.466
0.761   0.899   0.294   0.852
0.904   0.315   0.38    1.147
0.745   0.233   0.395   1.184
];
```

A2.9 *Populus* mutant in 10% v/v *Populus* hydrolysate observed values

File Name YobsPM10.m

```
%Contains the time that the observation data points were taken and the data
%from the triplicate fermentations for the Populus Mutant strain of C.
%thermocellum in 10% v/v Populus Hydrolysate.

% Observational data given as [biomass, cellbiose (g/L), ethanol (g/L),
% acetic acid (g/L), HMF (g/L) and Furfural (g/L)]

%Adjusted time points when the samples were taken
tobs = [0; 1; 2; 3; 4; 5; 5.5; 6; 7; 8; 9];
%Number of data points
num = 11;

Yobs1 = [0.010  5.743   0.024   1.383   0.053962    0.12297
0.022   5.613   0.021   1.397   0.053809    0.1015
0.043   5.44    0.022   1.412   0.053442    0.088817
0.117   5.232   0.074   1.538   0.040848    0.030387
0.182   4.745   0.119   1.662   0.03716 0.021291
0.452   3.949   0.221   1.926   0.012936    0.014611
0.547   4.041   0.256   2.359   0.010698    0.013499
0.704   3.095   0.34    2.534   0.009007    0.011364
0.988   1.347   0.572   2.76    0.006446    0.00008
0.910   0.288   0.654   3.024   0.005091    0.00007
0.835   0.291   0.72    3.184   0.004057    0.00005
];

Yobs2 = [0.005  5.818   0.028   1.385   0.055631    0.12885
0.010   5.827   0.028   1.409   0.054797    0.111093
0.022   5.674   0.031   1.446   0.052814    0.078906
0.083   5.447   0.062   1.506   0.046038    0.040414
0.141   5.08    0.096   1.617   0.034898    0.02444
0.392   4.387   0.178   1.83    0.016723    0.016353
0.478   3.96    0.207   1.963   0.011779    0.014149
0.603   3.447   0.275   2.125   0.009561    0.012161
0.855   2.032   0.476   2.575   0.007149    0.00007
0.945   0.297   0.616   2.975   0.005704    0.00009
0.822   0.314   0.666   3.15    0.004687    0.00008
];

Yobs3 = [0.005  5.637   0.04    1.376   0.055516    0.128674
0.010   5.564   0.019   1.39    0.054713    0.114524
0.022   5.558   0.03    1.474   0.051134    0.063563
0.066   5.319   0.046   1.469   0.047229    0.045347
0.116   4.981   0.07    1.578   0.038174    0.026448
0.271   4.493   0.122   1.771   0.020828    0.017778
0.420   4.021   0.145   1.918   0.014024    0.015142
0.529   3.585   0.2 2.059   0.011591    0.013103
0.831   2.544   0.366   2.927   0.008061    0.00008
0.918   0.299   0.534   2.926   0.00633 0.00007
0.794   0.29    0.592   3.119   0.005065    0.00004
];
```

A2.10 *Populus* mutant in 17.5% v/v *Populus* hydrolysate observed values

File Name: YobsPM175.m

```
%Contains the time that the observation data points were taken and the data
%from the triplicate fermentations for the Populus Mutant strain of C.
%thermocellum in 17.5% v/v Populus Hydrolysate.

% Observational data given as [biomass, cellbiose (g/L), ethanol (g/L),
% acetic acid (g/L), HMF (g/L) and Furfural (g/L)]

%Adjusted time points when the samples were taken
tobs = [0; 1; 2; 3; 4; 5; 6; 7; 7.5; 8; 9; 10];

%Number of data points
num = 12;

Yobs1 = [0.007  5.97    0.026   2.431   0.087229    0.179548
0.019   5.893   0.012   2.472   0.086064    0.153018
0.043   5.75    0.016   2.485   0.085887    0.108673
0.076   5.672   0.038   2.568   0.076769    0.055287
0.153   5.348   0.071   2.628   0.0654  0.025972
0.297   4.91    0.124   2.785   0.037051    0.018314
0.481   4.236   0.204   3.065   0.014098    0.001292
0.657   3.277   0.318   3.373   0.009366    0.00007
0.820   2.39    0.435   3.727   0.006327    0.00007
0.951   0.507   0.501   3.913   0.005786    0.00007
0.823   0.497   0.567   4.155   0.005489    0.00007
0.736   0.511   0.638   4.387   0.004182    0.00003
];

Yobs2 = [0.006  5.821   0.015   2.388   0.09499 0.197942
0.013   5.747   0.016   2.446   0.091952    0.168681
0.037   5.702   0.019   2.494   0.090735    0.120602
0.070   5.569   0.039   2.534   0.083777    0.063633
0.133   5.374   0.061   2.634   0.071482    0.028556
0.273   4.898   0.118   2.763   0.043062    0.018436
0.423   4.289   0.201   2.991   0.015336    0.00126
0.534   3.698   0.302   3.154   0.008857    0.00008
0.706   3.032   0.414   3.405   0.006778    0.00008
0.803   2.603   0.465   3.539   0.006012    0.00007
0.881   0.511   0.528   4.001   0.005568    0.00009
0.841   0.504   0.654   4.129   0.004505    0.00005
];

Yobs3 = [0.004  5.814   0.039   2.374   0.093708    0.19432
0.010   5.793   0.028   2.415   0.091386    0.161992
0.035   5.739   0.03    2.465   0.089317    0.116335
0.061   5.566   0.047   2.51    0.082968    0.061797
0.133   5.278   0.068   2.609   0.070183    0.026753
0.270   4.971   0.121   2.792   0.042548    0.018043
0.428   4.517   0.176   3.175   0.014986    0.001120
0.569   3.587   0.304   3.269   0.009157    0.00009
0.771   2.811   0.422   3.521   0.006808    0.00009
```

```
0.857   2.426   0.429   3.865   0.00594 0.00008
0.845   0.515   0.49    4.222   0.005436    0.00008
0.827   0.567   0.493   4.712   0.005022    0.00004
];
```

Vita

Jessica L. Linville was born in St. Petersburg, FL, to the parents of Lorna and Gregory Linville. She has an older sister, Cindy, and a twin sister, Melissa. She attended Pinellas Park High school where she was part of the Criminal Justice Academy. She also belonged to the Environmental Concerns club, the drama club and several honor societies. After graduation, she attended the University of South Florida St. Petersburg Campus where is joined the Honors College. She transferred to the University of South Florida Tampa Campus to take her engineering classes although she remained in the Honors College at the St. Petersburg Campus. In her junior year she joined Tau Beta Pi, the engineering honor society. She was also an undergraduate research assistant. She obtained her Bachelor's of Science degree from the University of South Florida in August of 2008 in Chemical Engineering. She accepted a graduate research assistantship at the University of Tennessee, Knoxville, in the Civil and Environmental Engineering Department. Jessica graduated with a Doctorate of Philosophy in Civil Engineering in May 2013.