Identifying and Overcoming the Factors Limiting Growth and Substrate Utilization in *Caldicellulosiruptor bescii*

A Dissertation Presented for the Doctor of Philosophy Degree

The University of Tennessee, Knoxville

Punita Manga

August 2018

Copyright © 2018 by Punita Manga

All rights reserved.

Dedication

Dedicated to my beloved grandfather
Late Mr. Kundan Lal Manga
and my parents, the pillars of my being

Dr. Vinod K. Manga and **Mrs. Sudershan Manga**

Acknowledgements

I would like to express my sincere gratitude to Genome Science and technology graduate program at University of Tennessee, Knoxville for providing me with this opportunity to fulfill my dream to do my Ph.D. I would also like to thank all the staff members of GST and ORNL for their time and cooperation. My Ph.D. journey has been nothing but enlightening, full of fun and the adventure of science. This ride has been one through which I have learned and grown a lot and it was much smoother because of my advisors Dr. Steven D. Brown and Dr. Brian Davison, who gave me the inspiration, motivation and guidance throughout the course of my Ph.D. and without whose support none of this would be possible.

I would like to express my heartfelt thanks and sincere gratitude to my committee members Dr. Robert L. Hettich, Dr. Margaret E. Staton and Dr. James G. Elkins for their constructive criticism, suggestions, critical inputs, patience and constant support through inspiration for the successful completion of this project. A special note of thanks to Dr. Elizabeth Howell and Dr. Janet Westpheling for being very supportive and generously allowing me to use the required lab equipments and loaning *C. bescii* strains respectively for some of my experiments. Their valuable advice, suggestions and comments throughout have also contributed to my intellectual development.

My Ph.D. journey would not be as much fun nor would it be complete without contributions from my lab members Miguel Jr. Rodriguez, Kyle Sander, Dawn Klingeman and Meredith Yeary. I would also like to thank all my wonderful friends who have supported me in so

many ways in my time away from home. Especially, Paromita Nath, Arkadipta Bakshi, Ujjal Jaggi and Srijata Chakravorti who have been like family to me standing beside me through thick and thin and have always encouraged me.

Finally, I would like to thank the pillars of my life, my father Dr. Vinod Kumar Manga my mother Mrs. Sudershan Manga, my brother Ankur Manga, sister-in-law Gaganpreet Manga and niece Kiara Manga for their unconditional love, support and contributions which helped me in every step of my Ph.D. journey. Everything that I am today is all because of my family. There is are no words to express my gratitude towards them, for all the sacrifices my parents have made to provide me with all that I have ever needed and to support me in my endeavors as well as affection and support I have received from both my brother and sister-in-law.

I feel extremely lucky to have been blessed with such amazing people that are my family, friends and mentors in life. Whenever I count my blessings I will always count them twice!

Abstract

Caldicellulosiruptor bescii is an anaerobic extreme thermophile being studied for production of lignocellulosic biofuels due to its potential for plant biomass deconstruction. It can grow on a wide range of substrates and co-metabolize C5 and C6 sugars. However, incomplete biomass utilization and low cell growth, among other bottlenecks, majorly limit its bioconversion potential. The work in this dissertation aimed at identifying and overcoming the factors that hinder growth and substrate utilization in *C. bescii* and focused on low pH and high osmolarity as the investigated conditions that may serve as inhibitors. An RNA-seq data analysis pipeline was developed using a *Bacillus thuringiensis* data-set that determined essential parameters such as required number of reads and replicates for achieving results with high statistical confidence. This was further used for examining the physiological and systems level responses of *C. bescii* to acidic pH using integrated omics. In this study, lowering pH from 7.2 to 6.0 in mid-log and post stationary growth phases demonstrated lowered membrane potential/proton motive force (PMF) as a cause of these limitations. Dramatic increase in growth, improved substrate utilization and higher product generation was observed upon alleviating the PMF limitations post-acid addition. Patterns of elevated membrane potential and higher ATP pools further supported the hypothesis. In a follow-up study using liquid and crystalline cellulose it was demonstrated that *C. bescii* also benefits from the lowered pH on solid substrates indicating PMF limitation exists irrespective of the substrate and alleviation of the limitation under lower pH improves growth. Moreover, this study revealed osmolarity as the next immediate factor limiting the bioconversion potential of *C. bescii* once PMF

limitation is alleviated. The ability of *C. bescii* to maintain growth at pH 5.5 (0.1 hr-1 dilution rate) in chemostat on Avicel was also displayed here which has not been previously reported, extending its growth pH range (5.5-7.3). Finally, an attempt to expand the genetic tools available for *C. bescii* was made, exploring RNA interference (RNAi) technology as a basis for developing a genome-wide screening tool in the future, which would aid to identify genetic elements that could confer robustness under various stress conditions.

Table of Contents

Chapter 1.0 Introduction **1**
1.1 Background and Motivation 1
1.1.1 Bio-Fueling Our Future 1
1.1.2 Biofuels: Categories and Trends 2
1.1.3 Lignocellulosic Biofuels 4
1.1.4 Why Consolidated Bioprocessing? 6
1.1.5 Microorganisms as Biocatalysts for Biofuels 8
1.1.*6 Caldicellulosiruptor species*: The Overview 12
1.1.7 *C. bescii* and the Three C's: Capabilities, Current limitations and Challenges 16
1.1.8 Recent Advancements: Taking Advantage of Advanced Tools and Techniques 22
1.1.9 New Tools to Meet New Challenges 23
1.2 Overview of the thesis 24

Chapter 2.0 Replicates, Read Numbers, and Other Important Experimental Design Considerations for Microbial RNA-Seq. Identified Using *Bacillus thuringiensis* Datasets **26**
2.1 Introduction 27
2.2 Materials and Methods 30
2.2.1 Organism Growth and Sampling 30
2.2.2 RNA Extraction and cDNA Library Preparation 31
2.2.3 Data Analysis 34
2.2.4 Differential Gene Expression Analysis: DESeq2 35
2.2.5 RT-qPCR Validation of RNA-seq Results 36
2.2.6 Determination of Iron Content in Media and Water 36
2.2.7 Alteration of Sequence Read and Biological Replicate Numbers 36
2.3 Results 38
2.3.1 RNA-seq Experiments 38
2.3.2 Differential Gene Expression Analysis using DESeq2: Medium Lot and Date Effect 42
2.3.3 Real Time-Quantitative PCR Validation (RT-qPCR) 46
2.3.4 Effect of Reduced Number of Replicates and Reads on Differential Gene Expression Detection 52
2.4 Discussion 59
2.4.1 Recommended Number of Replicates and Reads 60
2.4.2 Differentially Expressed Genes and Experimental Design Considerations 62
2.4.3 Significance of the Available Data Set 63

Chapter 3.0 Improved *Caldicellulosiruptor bescii* growth and substrate utilization by alleviating membrane potential limitations **64**
3.1 Introduction 64

3.2 Materials and Methods 67
3.2.1 Strains, media and growth conditions 67
3.2.2 Growth and product profiling 67
3.2.3 Integrated omics sampling, measurements and analysis 68
3.2.4 Membrane potential estimation and ATP measurements 73
3.3 Results 74
3.3.1 Characterization of *C. bescii* growth patterns at varying pH 74
3.3.2 Batch fermentations to capture physiological and systems level response to medium acidification 77
3.3.3 Hypothesis for the observed growth improvement upon lowering pH 89
3.3.4 Replicated batch fermentations to test proposed hypothesis 91
3.3.5 Membrane potential and ATP pool trends, pre and post-acid addition 95
3.3.6 Effect of post-stationary phase acid addition on *C. bescii* strain with inactive [NiFe] hydrogenase 99
3.3 Discussion 102

Chapter 4.0 High Osmolarity inhibits the fermentation potential of *Caldicellulosiruptor bescii* on both liquid and solid substrates after PMF stress is relieved 107
4.1 Introduction 107
4.2 Materials and Methods 110
4.2.1 Strains, Media and Growth Conditions 110
4.2.2 Growth assessments 110
4.2.3 Quantification of Substrates and Products 111
4.2.4 Osmolarity Measurements 112
4.3 Results 112
4.3.1 Bottle growth assessments with sodium chloride 112
4.3.2 High liquid substrate loadings and osmolarity at pH 6.0 113
4.3.3 PMF limitation alleviation, Solid substrate conversion and Osmolarity 118
4.4 Discussion 123

Chapter 5.0 Investigating small RNA interference and its prospects in genome-wide screening in *C. bescii* 127
5.1 Introduction 127
5.2 Materials and Methods 133
5.2.1 Strains, Media and Culture conditions 133
5.2.2 Investigative Design and Methodology 133
5.2.3 Construction of Vectors bearing RNAi Fragments 139
5.2.4 Transformation, Screening, Plasmid Purification, and Sequence Verification 141
5.2.5 Growth and Product Analysis of Mutants for Phenotypic Effects 142
5.2.6 RNA Extraction and Expression Quantification 142
5.3 Results 143

5.3.1 Growth and Product Profiles for Targeted RNAi 143
5.3.2 Target Gene Expression Comparisons and Expression of Insert Fragments 145
5.3.3 Attributes of RNAi based Genome-wide Screening Tool Development 149
5.4 Discussion 154

Chapter 6.0 Conclusions, Significance and Future Directions 160

References 167

Appendix 184

Vita 186

List of Tables

Table 1: Sample collection and culture details for all samples from strains CT43 and ATCC10792. 33

Table 2: List of RT-qPCR primers used for validation of RNA-Seq results for medium lots and culture dates. 37

Table 3a: Reads and mapping details for strain ATCC10792. 39

Table 3b: Reads and mapping details for strain CT43. 40

Table 3c: Example genome coverage calculation shown sample 1A from strain ATCC10792 as performed for all samples in the study. 41

Table 4: Genes related to iron acquisition and metabolism differentially expressed in strain ATCC10792 and CT43 grown in medium lot #1091744 over #7220443. 47

Table 5: Elemental analysis of the two media lots and water sources. 50

Table 6: Effect of decreasing number of replicates on significantly differentially expressed genes while maintaining 100 and 25% of the reads. 53

Table 7: Effect of decreasing number of reads on significantly differentially expressed genes while maintaining all four replicates. 57

Table 8: Cell counts, and product profiles monitored to determine the wash out pH. 78

Table 9: Substrate and product profiles (HPLC) from treated and control fermenter samples validate observed growth differences. 80

Table 10: Proteins significantly differentially detected in treated batch fermenters at 60 min post acid addition and correlating gene expression changes from RNA-seq. 85

Table 11: Metabolites differentially detected in treated batch fermenters at 90 min post acid addition 87

Table 12: Substrate and product profiles (HPLC) from post-stationary phase growth recovery batch fermenters, depicting complete substrate utilization of ~15 g/L maltose. 93

Table 13: HPLC product profiles of cells grown in batch fermenters starting at pH 6.0, show similar product accumulation and sugar utilization patterns as for fermenters starting at pH 7.2 and the lowered to 6.0. 95

Table 14: End-point HPLC product profiles for fermentations with inactive [NiFe] hydrogenase strain, JWCB038 and its parent strain, JWCB036 at pH 6.0 show similar patterns post-acidification. 101

Table 15: HPLC product profiles and corresponding osmolarity values for maltose fermentations as depicted in figure 22. 117

Table 16: HPLC product profiles depict products accumulated, substrate additions along fermentation and utilization patterns. 122

Table 17: List of primers used in the study (*gibson overlap regions are in written in red). 140

List of Figures

Figure 1: Baseline schemes depicting the steps involved in the bioprocessing of lignocellulose to ethanol, comparing currently used methodology with consolidated bioprocessing (CBP). 7

Figure 2: Simplified depiction of plant biomass deconstruction facilitated by the large and diverse inventory of glycoside hydrolases with cellulolytic and hemicellulolytic abilities as well as muti-funtional modular enzymes such as CelA, to produce simple sugars (both C6 and C5) which can be further utilized in fermentation to produce biofuels and other desired products by *Caldicellulosiruptor spp.* 14

Figure 3: Potential growth inhibitory compounds that are generated by breakdown and utilization of pretreated or un-pretreated biomass, or as sugar degradation products of metabolism. 20

Figure 4: Sampling and potential sources of variation in RNA seq data. 32

Figure 5: Variation analysis of raw read count data for strain ATCC10792 and strain CT43. 43

Figure 6a: Multivariate correlation analysis summarizing variation among biological replicates of strain ATCC10792. 44

Figure 6b: Multivariate correlation analysis summarizing variation among biological replicates of strain CT43 45

Figure 7: RNA seq data validation: Correlation between RNA seq and RT qPCR results for differential gene expression in *Bt* strain ATCC10792 grown on different medium lots (#744–443) and dates (2/23/12–3/6/12). 51

Figure 8: Venn analysis of DE genes detected with varying replicate numbers. 55

Figure 9: Venn diagram depicting effect of reducing number of reads on DE gene numbers. 58

Figure 10: Similar growth patterns observed for *C. bescii* DSM6725 (wild type) and JWCB049 (ΔpyrFA Δldh CIS1:: PS-layerTeth39_0206) in LOD and LODU media (50 mL) respectively with 5g/L maltose at pH 7.2-6.5. 75

Figure 11: Chemostat setup: 1.5L applikon fermenter with working volume of 750 mL. 76

Figure 12: Increased cell growth observed in treated batch fermenters (pH lowered to 6.0 at mid-log phase): T1, T2, T3 as compared to controls: C1, C2, C3 maintained at pH 7.2. 79

Figure 13: Sulfate addition showed minimal effect on cell growth over 96 hrs. in LOD + 5g/L maltose (controls– C1, C2, C3) compared to LOD + 5g/L maltose + 5mM Na2SO4 (treated - T1, T2, T3). 82

Figure 14: Heat map depicting hierarchical clustering of genes differentially expressed post-acid addition across time-points sampled (5, 10, 30, 60. 90 min). 83

Figure 15: Hypothesis Model (*for depicting conditions in hypothesis model: OD data points are actual; regions are hypothesized). 90

Figure 16: Growth curve depicting post-stationary phase growth recovery upon lowering pH to 6.0 and again after addition of more sugar at pH 6. 92

Figure 17: Cells grown in batch fermenters starting at pH 6.0 show similar growth pattern to cells after pH is lowered from 7.2 to 6.0. 95

Figure 18: Membrane potential changes estimated across the growth curve in figure 16 at different stages (overlaid blue line graph). 96

Figure 19: ATP pool changes across the growth curve in figure 16 at different stages (overlaid blue line graph). 98

Figure 20: Strains JWCB036 (parent) and JWCB038 (mutant with inactive [Ni-Fe]) grown in batch fermenters show similar growth patterns after pH is lowered from 7.2 to 6.0. 100

Figure 21: Effects of high osmolarities (200-350mOsm) on growth of *C. bescii* cultures (triplicates) due to added NaCl in LOD medium. 114

Figure 22: *C. bescii* cultures grown starting at pH 6.0 (controlled), with subsequent additions of substrate (maltose, 10 g/L) and base salts equivalent of per liter show growth limitation effects at high osmolarities. 115

Figure 23: *C. bescii* depicts the ability to maintain cell viability at high osmolarities for long durations by slowly growing back to an OD of ~ 0.6 post-medium dilution. 119

Figure 24: Alkali addition curve from *C. bescii* fermentation representing growth of cultures starting at pH 7.2 on 10g/L Avicel grown over 1000 hours with subsequent addition of substrate and base salts (equivalent to per liter media). 120

Figure 25: Overview of the antisense fragment designs depicting the in-built parameters tested in the two broad approaches (EGS, Approach I & random genomic RNAi library, Approach II) used in this study and the possible RNA interference causing mechanisms that may follow. 132

Figure 26 a: Design and methodology detailing the RNaseP directed transcript degradation (Approach I**) used in the study to screen *C. besci*i for RNAi effect, possible mechanism(s) and the random genome fragment based RNAi library approach (Approach II) attempting development of RNAi based genome wide screening tool. 134

Figure 26b: Continuation of the design and methodology from figure 26a, depicting steps following transformation of the RNAi fragment carrying plasmids into *C.bescii*. 135

Figure 27: RNAi vector construct designs. 137

Figure 28: Effects of RNAi on target genes of interest *acetate kinase (ack)* and *phosphate trasacetylase (pta)* anticipated to result in reduced acetate and increased ethanol levels as depicted using hypothetical values (pathway modified from [239]). 144

Figure 29: No significant growth (a) or product (b) phenotype was observed in the RNAi plasmid transformed single colony isolates (P1, P2, P3, A1, A2) in comparison to the negative control (UI) and parent (JWCB032). 146

Figure 30: All replicates of the single colony transformants from the growth curves with RNAi fragments for targeted genes *pta* (P1, P2, P3) and *ack* (A1, A2) show presence of stable RNAi fragments upon PCR screen using primer pair BHSB006/ BHSB007 and NEB 100 bp ladder (L). 147

Figure 31: qRT-PCR estimated mRNA expression levels of target genes pta (a) ak (b) showed no significant expression changes for the gene *pta* but transformant A2 for *ack* depicted significant mRNA reduction (~20%) in comparison to both parent and negative control. 148

Figure 32: Unique identifier sequence (87 bp) amplified using cDNA (primer pair UI_01 &UI_02) from three biological replicates of the single colony isolate and run on agarose gel against 100 bp ladder (L). 150

Figure 33: Extracted gDNA (a) on 0.8% and fragments generated via bioruptor sonication method (b) on 1 % agarose gel. 150

Figure 34: Fifty transformed *C. bescii* colonies screened for inserted genomic fragment between unique identifier sequence overhangs in the plasmid construct. 151

Figure 35: PCR bands for four single colony isolates #11, 16, 21 and 35 (primer pair UI01 and UI_02), Sanger sequenced. 153

Figure 36: Fragment expression and insert sequence length differences validated using RT-qPCR amplification and melt curves respectively. 155

List of Equations

Equation 1.0 89

Equation 2.0 116

Equation 3.0 116

Equation 4.0 152

List of Abbreviations

CBP	Consolidated Bio-Processing
DE	Differentially Expressed
PMF	Proton Motive Force
MFI	Mean Fluorescence Intensity
BDL	Below Detection Limit
QS	Quantitative Saccharification
GOI	Gene Of Interest
OD	Optical Density
TE	Transformation efficiency
CFU	Colony Forming Units
RNAi	RNA interference
EGS	External Guide Sequence
asRNA	Antisense RNA

Chapter 1.0

Introduction

1.1 Background and Motivation

1.1.1 Bio-Fueling Our Future

Over the past century, an increasing standard of living accompanied by population explosion has led to the excessive and fast use of fossil fuels [1, 2]. This extent of fuel burning has not only affected the finite fuel reserves on the planet but has also lead to a startling 40% increase in atmospheric carbon dioxide (CO_2), which is a recognized greenhouse gas [2, 3]. It is estimated that transportation alone accounts for 30-40% of emissions of this greenhouse gas into the atmosphere [International Energy Outlook 2017, U.S. Energy Information Administration (EIA)]. The Intergovernmental Panel on Climate Change (IPCC) stated that around 2010 the world was already 0.6 ^{o}C warmer owing to greenhouse gas emissions [4]. As a result, we step into the future with depleting fossil fuel reserves and growing concern for global warming [4-6]. The need for cleaner, renewable and sustainable fuels is thus, ever increasing with the realization that at the current pace of advancement our energy needs would soon, no longer be met by fossil fuels. Moreover, there is also the issue of the rapidly and drastically changing climate due to increasing atmospheric CO_2 that requires attention to ensure the survival of life for years to come, regardless of the availability of fuels. Therefore, the fuels of the future

must not only be renewable but also recyclable in terms of the carbon released and aimed at approaching carbon neutrality [7].

The current global energy supply is still dominated by fossil fuels (81% in 2017, IEA report), but alternative energy sources such as nuclear power, hydropower, wind, solar, geothermal and bioenergy are now although smaller but active contributors [8]. Bioenergy has become one of the most important renewable energy sources by contributing about 10% of the total energy supplies [4]. An important form of harnessing bioenergy are biofuels. These fuels are mainly derived from biomass (hydrocarbon-rich living organisms such as plants or microalgae) or bio waste [9, 10] by thermal, chemical or biochemical conversion processes [4]. Biofuels can be combusted to generate energy and used for many purposes just as conventional fuels and thus are a promising alternative to fossil fuels [11]. The use of biofuels in industry and transport sectors which are two major energy consumers that have increased 1.6% oil demand and 3% natural gas consumption [IEA world outlook, 2017] and are estimated to generate more than two-thirds of the projected increase in future oil demand, could make a huge difference.

1.1.2 Biofuels: Categories and Trends

Generally, the realization of biofuel bioprocessing concepts is classified into three generations. The conventional or 1st generation biofuels utilize food crops or a feedstock that can be consumed as food such as starches from wheat, corn, sugar cane, molasses, potatoes, other fruits or vegetable oil. On the other hand, the 2nd generation biofuels also referred to as “advanced biofuels” are produced from sustainable, non-food crop

feedstocks. The 3rd generation biofuels are the ones generated using algae as feedstock. Algae are considered a low-cost, high energy and entirely renewable feedstock that can be grown using land and water unsuitable for food production. Moreover, algae are also predicted to have higher energy potential per acre in comparison to conventional crops [12, 13]. Recently, another category referred to as the 4^{th} generation biofuels has been introduced. The production of biofuels in this category is aimed at a sustainable energy production method that also allows for carbon capture and storage such as in the form of CO_2. Therefore, unlike the second and third generation biofuels which are potentially carbon neutral, the carbon capture in this process makes it a carbon negative method. This system thus, does not only provide biofuels replacing fossil fuels but also reduces CO_2 emissions atmosphere [14]. Biofuels produced from any of these methods can come in several forms such as liquid (bioalcohols, biodiesel, bioethers), gaseous (biogas, biohydrogen, syngas) or solid biofuels (wood, dried plants, bagasse) and can meet diversified energy needs [1, 15]. Most of the currently available biofuels belong to the first or second generation for example, U.S. and Brazil commercially produce bioethanol to be added to gasoline using corn and sugar cane respectively, which accounts for 87% of the world's fuel ethanol [16].

In America and Europe, it is a common practice to blend most gasoline and diesel fuels with biofuels. Biodiesel is popular in Europe whereas, bioethanol prevails in the Americas. Ethanol is one of the most widely used biofuels today with more than 22 billion gallons of fuel ethanol produced each year. It currently constitutes 94% of the biofuel produced in the world [Biofuels issues and trends (2012), US EIA (1, appendix)]. Ethanol blended with gasoline can increase octane and improve the emissions quality, a 10% ethanol and 90%

gasoline (E10) blend is used in the United States [2, 4]. It can be used in even higher concentrations, such as 85% (E85) or in its pure form 100% (E100) replacing petroleum completely with required engine modifications [2]. Since the commercially available bioethanol today is produced using food crops like corn, sugarcane etc., it entails additional carbon and energy costs as well as competes with the food supply [4, 17]. Besides the cost of production that in turn depends on the land availability and competition when food crops are involved, the food versus fuel feud over the utilization of the produce to be directed towards each, is the major challenge with industrial production of the 1st generation biofuels [11].

1.1.3 Lignocellulosic Biofuels

Biotechnology offers the potential for reducing the cost and dramatically increasing the production of bioethanol as well as rerouting it to non-food plants via cellulosic biofuels [18, 19]. These biofuels are produced using the cellulose and hemicellulose-based carbohydrate components of plants unlike the starch or sugar rich components. The potential of lignocellulosic biofuels is reflected in the estimated light energy stored in plants annually. Based on the global net primary production of land plants (~65 Gt carbon/year) if this energy were to be converted into chemical energy it would exceed the world's energy demand by a factor of three to four [2]. Crops like switchgrass and poplar which are rich in cellulosic biomass, form the potential feedstock for such fuels [2, 5, 20]. These substrates offer a potentially renewable source of sugars that can be fermented to the desired biofuels. Yet, it is not easy to harness this energy trapped in the plant fibers.

To do so in a sustainable and industrially cost-effective manner methods of effective deconstruction and utilization are needed [18]. The plant structure is complex, it has evolved to present a strong barrier to chemical and enzymatic attacks as may come from microbes or animals [21].This constitutes one of the most difficult challenges in producing industrially relevant cellulosic biofuel and is termed “lignocellulosic recalcitrance” [21-23]. Several factors can contribute to this barrier including structural aspects of the plant cell wall such as density of the vascular bundles, sclerenchymatous or thick wall tissue, degree of lignification, structural heterogeneity [24, 25] and presence and/or release of chemical fermentation inhibitors [26].

The phenomenon of lignocellulosic recalcitrance typically requires thermochemical pretreatment and addition of enzymes before biological processing of the feedstock [27, 28]. This mainly adds to the cost of production of these fuels and to the list of challenges to be overcome [19]. The steps involved in the current biochemical conversion of such biomass to liquid biofuels such as ethanol include: physiochemical pretreatments to increase accessibility of cellulose to enzymes and solubilization of hemicellulose sugars, enzymatic hydrolysis of sugars and fermentation to ethanol [29]. In these processes either an entire step is solely dedicated to enzyme production or commercially obtained fungal enzyme cocktails are added for hydrolysis. The cellulase cocktails used generally consist of cellobiohydrolases, endoglucanases and β-glucosidases [5]. It is estimated that the elimination of the pretreatment and enzymes addition to aid hydrolysis steps from the procedure will reduce the cost of production by 40% [22]. This would require the use of both high-yielding plants with reduced recalcitrance and employing microbes capable of saccharification as well as fermentation of the biomass to provide high ethanol titers. So

far, of the several proposed ways of processing, consolidated bioprocessing (CBP) technology is an important one that comes the closest in concept to providing what is required to make cellulosic biofuels an economic reality at the industrial level.

1.1.4 Why Consolidated Bioprocessing?

Consolidated bioprocessing (CBP) methodology offers to combine saccharolytic enzyme production and secretion, polysaccharide hydrolysis and fermentation of available sugars to achieve the conversion of lignocellulosic feedstock into desired products in a single-unit operation (Figure 1) [30, 31]. There are two generations of CBP biofuel production envisioned, the first generation uses the pretreated biomass whereas the advanced CBP biofuels or the second generation aims at eliminating this additional step and utilizing unpretreated biomass directly for conversion into products to further reduce the cost-factor at the industrial level. Thus overall, CBP strategy bestows two major advantages over the currently used methods (Figure 1), notably of improving cellulose conversion efficiency and decreasing lignocellulosic biomass processing costs and capital costs for producing biofuels and other value-added products [32].

The success of cellulosic biofuels produced via CBP relies as much on the biomass as on the microbe for conversion, utilization and fermentation to the desired biofuel [30]. Therefore, the research in this field is directed both towards the feedstock and the microorganism of choice. Since no single organism, plant or microbe yet known displays all CBP desired characteristics, it is a challenge to be able to choose the right players.

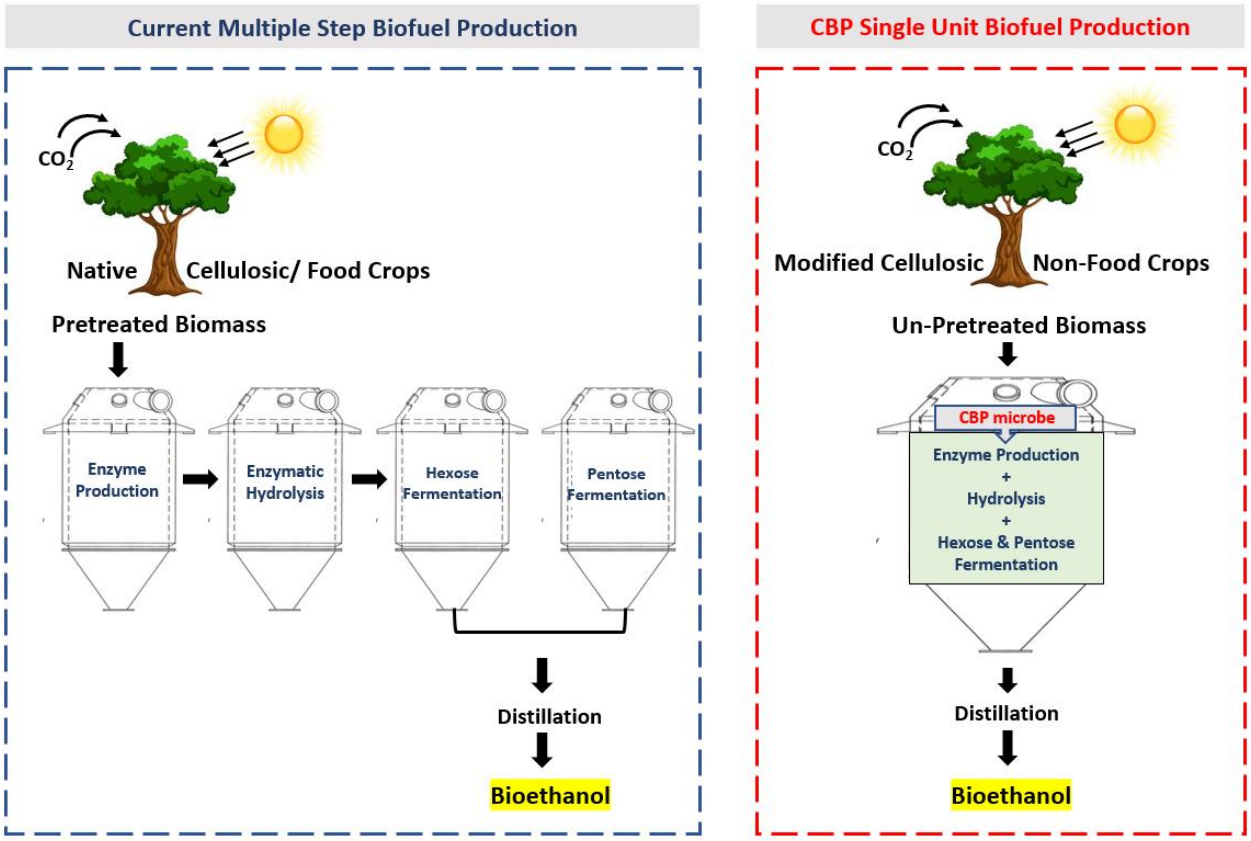


Figure 1: Baseline schemes depicting the steps involved in the bioprocessing of lignocellulose to ethanol, comparing currently used methodology with consolidated bioprocessing (CBP). Current methodologies require physiochemically pretreated biomass and involve separate units for production of saccharolytic enzymes, biomass hydrolysis to release mixed sugars followed by separation and individual fermentation of hexose and pentose sugars in a multi-step process to produce ethanol. CBP on the other hand, offers a much more economical alternative, single unit conversion of solid biomass to liquid biofuel. The key to CBP is engineering of a microorganism that can efficiently produce enzymes required for deconstruction of unpretreated biomass to hydrolyze lignocellulosic polysaccharides into fermentable sugars and ferment this mixed-sugar hydrolysate to produce high yields and titers of ethanol under industrial conditions.

One can either continue screening organisms in nature, which could go on indefinitely or engineer organisms with some existing favorable characteristics. To do so having the right set of genetic engineering tools and the ability to use those tools to efficiently get high product yields, titers and robustness at industrial level is yet another challenge to be met [30]. In research aimed towards reduced recalcitrance in plants several high sugar release and low lignin content natural variants or transgenic plant lines, for crops such as aspen [33], alfa-alfa [34], switchgrass [35] and *Populus* [36, 37] have been developed and/or studied. For example, in switchgrass transgenic lines such as GAUT4 [38], COMT [39] with knock-down of an enzyme in the pectin and lignin biosynthesis pathway respectively and the miRNA156 [40], MYB4 [41] (lignin biosynthetic pathway repressor) overexpression lines have been developed. Similarly, in *Populus*, the sugar release of some natural variants have been compared and studied with respect to their lignin content to explore better suited options for CBP [42]. More of such crop lines continue to emerge with time and as research progresses. The overall goal is to produce a set of “top lines” for plants that offer high carbon content as well as sugar release capabilities and have reduced lignin content so that they are best suited as feedstocks for economical CBP based biofuel production [43].

1.1.5 Microorganisms as Biocatalysts for Biofuels

It is estimated that up to 48% of the minimum ethanol selling price is contributed by the enzyme costs [44, 45]. The use of whole microorganisms as biocatalysts in the biofuel production process instead of using the purified enzymes produced by them, would thus

understandably enable cost reductions. Moreover, the use of microbes is also an environmentally friendly approach. CBP proposes to cash in on these advantages by utilizing microbes that ideally would possess the ability to solubilize lignocellulosic biomass efficiently to produce high yield and titers of the desired product under industrial conditions [30]. In nature certain microbes have evolved with capabilities to deconstruct various complex plant cell wall polysaccharides and have native enzyme systems dedicated towards this cause [29, 46, 47]. Majority of these microorganisms that possess the ability to degrade cellulose are fungi and bacteria that survive in diverse environments ranging from mesophilic to thermophilic ecosystems, with the availability of abundant plant-derived substrates or waste [47]. The diversity of enzymes revealed to be produced by such biomass degraders [48-55], inspired by the complexity and heterogeneity of plants in these varied ecosystems, holds promise for an approach like CBP where biofuels or other products can be directly obtained using microbes with little to no pretreatment of the biomass.

Different microorganisms with inherent biomass degrading (cellulolytic and hemicellulolytic) and/or fermentative capabilities have been under review for years in pursuit of discovering or being able to engineer an ideal microbe for industrial level, CBP-derived ethanol production [56]. Engineering microorganisms with either of these traits has its benefits and challenges [29]. Where natively fermentative microbes need metabolic engineering (recombinant cellulolytic strategy) to introduce the required genes for the upstream processes such as, enzymes for biomass degradation and sugar transport, the cellulolytic microbes (native cellulolytic strategy) need engineering for the downstream processes mainly involving re-direction of the carbon to produce the biofuel

or desired products [57, 58]. Another strategy proposed towards the same aim would involve starting with a ‘minimal’ microorganism [29] with all the necessary genetic elements for survival and then engineering it from scratch into a synthetic ‘designer CBP-microbe’ with all the upstream and downstream processing pathways for producing the desired biofuel [59]. This alternative ‘synthetic strategy’ could enable surpassing some challenges that may arise from the resistance encountered while re-directing native or introducing new metabolic pathways in existing organisms. It may however, also be as challenging or more, as the other strategies.

Both mesophiles and thermophiles with one of the above inherent abilities have been explored under the recombinant and/or native engineering strategies [30]. Among mesophilic, non-saccharolytic microbes, *E. coli* [60]*, B. subtilis* [61] and *S. cerevisiae* are some notable examples where recombinant strategy has been employed, with the most advancements made with the latter [30]. Some engineered strains of yeast have been reported to be able to produce as high as 98.7% theoretical yield of ethanol in sugarcane molasses medium [62] proving its abilities in bioethanol production. *S. cerevisiae* is thus a candidate for being industrially exploited to produce lignocellulosic biofuels and has been genetically modified for various abilities towards this cause [57, 63-65] [66]. Although, it still faces the challenge of producing sufficient enzymes for industrially relevant polysaccharide degradation [6] some research has also focused on engineering CBP yeast capable of releasing as well as utilizing sugars from plant biomass [67]. On the other hand, among cellulolytic microorganisms that are a target for the native strategy the candidates belong to the following categories, namely, fungi, free-enzyme bacteria and cellulosome-forming bacteria [30, 47]. Since cellulolytic microorganisms isolated from

diverse ecosystems employ varied enzyme systems for biomass conversion the above-mentioned categorizations are based on the major enzyme system paradigms identified [47]. Among organisms from these categories, there has been a recent increase in interest towards thermophilic bacteria, mainly, extremely thermophilic anaerobes (optimum growth temperatures ≥ 70^{o}C) [2]. Thermophilic anaerobes with biomass degrading enzyme systems form attractive and promising candidates due to the various possible advantages of carrying out biofuel production at elevated temperatures under anoxic conditions [29, 56]. Minimized biological contamination, higher reaction rates due to the reduced substrate and product viscosity with increased bioavailability [68], the possibility of co-distillation of ethanol via auto-evaporation [69-71] , reduced cooling effort as required with mesophiles and aided temperature maintenance from biological energy input [29] are some of the noted ones. Moreover, not only are the thermophilic enzymes more stable and often more tolerant to various other harsh conditions like high pressure and denaturing solvents that they may encounter during the process [72], but the closeness of their optimum temperatures for growth to the saccharolytic enzymes increases the efficiency of the entire production process [73, 74]. Furthermore, carrying out the biofuel production process under anaerobic conditions allows for efficient product yields due to better conservation of the reducing equivalents in the metabolic reactions involved and avoiding substrate gained energy loss to aerobic respiration, as well as saves expense of higher agitation requirements for aeration [30, 56].

Thermophilic microbes with cellulolytic abilities currently being studied are mostly from the genera *Clostridium, Thermoanaerobacter, Thermoanaerobacterium* and *Caldicellulosiruptor*, which belong to the phylogenetic group of Firmicutes [2]. The

optimum growth temperatures for the organisms from these genera under study for biofuel production range from a T_{opt} of 60°C to 78°C. *Dictyoglomi, Spirochaeta thermophila* (DSM 6192 and 6578) and *Bacteriodetes* are some other phlya with thermophilic cellulose degrading members which are currently relatively less characterized [75, 76]. Most of the studies in the field have focused on *Clostridium thermocellum*, which grows optimally at 60^{o}C (moderately thermophilic) and has the capability of hydrolyzing cellulose to produce ethanol as a fermentation product using a multienzyme complex called the “cellulosome” [56, 77, 78]. While this does make it a promising CBP organism, *C. thermocellum* lacks the ability to utilize hemicellulosic sugars [79, 80] and has yet to report industrially relevant yields of ethanol from engineered strains or co-cultures from lignocellulosic substrates [46, 77, 81]. An interest in another species, *Caldicellulosiruptor* has risen in the past years for its potential towards CBP biofuels [82]. Microorganisms from this genus not only have been recognized for their potential in plant biomass deconstruction but are also known to grow at the highest temperature optima among other cellulolytic microbes. [29, 30, 83].

1.1.6 *Caldicellulosiruptor* species: The Overview

The genus *Caldicellulosiruptor* from the phylum Firmicutes, is placed under the class Clostridia, order Thermoanaerobacterales and Thermoanaerobacterales Family III. Bacteria in this genus are gram-positive, asporogenous, anaerobic and extremely thermophilic with optimum growth temperatures between 65-78^{o}C [2]. Globally distributed across North America, Russia, New Zealand, Japan and Iceland [84] a total of ten species

have been isolated to date. The first member of the species was isolated from a hot spring in New Zealand in 1987 and originally named *"Caldocellum saccharolyticum*" which is now known as "*Caldicellulosiruptor saccharolyticus"* [85].. The other isolated species include, *C. lactoaceticus* [86], *C. owensensis* [87], *C. kristjanssonii* [88], *C. acetigenus* [89], *C. kronotskyensis, C. hydrothermalis* [90], *C. bescii* [91] and *C. obsidiansis* [92] and *C. changbaiensis* [93] in their chronological order of discovery. The genomes of eight of the ten species excluding *C. acetigenus* and the recent *C. changbaiensis* have been sequenced [50]. The genome size between species ranged from 2.4-2.97 Mb with a 35-36% G+C content and inter-species 16s rRNA gene sequence similarities of 94.8-99.4% [50]. The core genome with only the genes commonly shared within species totals to 1548 genes. The compiled pangenome (comprising of all genes from sequenced species) on the other hand is open or incomplete with a total of 4009 genes and displays further scope for the discovery of new genes from new species [50].

Genomic analysis also revealed the encoded vast diversity of carbohydrate metabolizing enzymes also known as carbohydrate-active enzymes (CAZymes) [29, 84]. These comprise of glycoside hydrolases (GH) (endo and exoglucanases) as well as carbohydrate easterases and polysaccharide lyases. The diversity and specificity of these enzymes confer the lignocellulolytic abilities to the organism enabling it to effectively deconstruct plant biomass. *Caldicelluolosiruptor* species produce GHs with both cellulase and hemicellulase abilities enabling hydrolysis of β-1,4-glycoside and β-1,4-xyloside linkages in the complex carbohydrates, cellulose and hemicellulose [84] resulting in a variety of hexose as well as pentose monosaccharides [81, 94, 95] (Figure 2). The released C6 and C5 sugars are then also co-metabolized by these species without

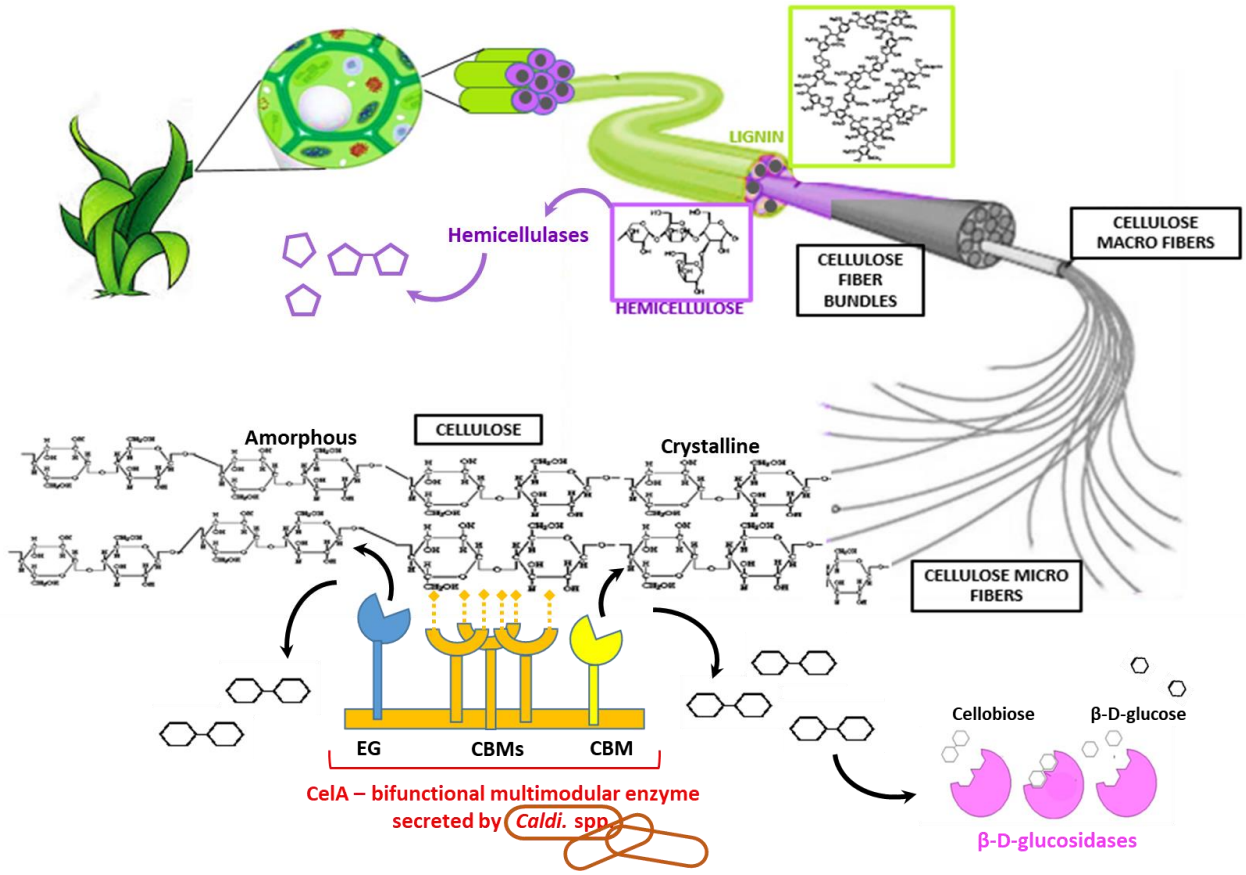


Figure 2: Simplified depiction of plant biomass deconstruction facilitated by the large and diverse inventory of glycoside hydrolases with cellulolytic and hemicellulolytic abilities as well as muti-funtional modular enzymes such as CelA, to produce simple sugars (both C6 and C5) which can be further utilized in fermentation to produce biofuels and other desired products by *Caldicellulosiruptor* spp. CelA is composed of three carbohydrate-binding modules (CBMs) that promote proximity to substrate and two catalytic domains, one with endoglucanase (EG) and another with exoglucanase /cellobiohydrolase activity (CBH) to help beakdown crystalline cellulosic structure by acting on the amorphous region and the reducing /non-reducing ends of cellulose molecule respectively.

experiencing carbon catabolite repression and fermented to acetate, lactate, CO_2, hydrogen and trace amounts of ethanol [96]. Another interesting but more so, unique and distinguishing feature of this genus is the presence of large multimodular, multifunctional CAZymes [29, 50, 51]. These enzymes consist of various GHs associated with Carbohydrate Binding Modules, CBMs (mostly CBM3 domains) [97] in different combinations. Some commonly found combinations among the species are: GH9-CBM3-CBM3-CBM3-GH48, GH74-CBM3-CBM3-GH48 and GH9-CBM3-CBM3-CBM3-GH5 [2]. This system of secreted, large modular, multi-domain, complex enzymes involved in cellulose hydrolysis at thermophilic temperatures presents a paradigm separate from the other known system with membrane bound multi-enzyme complexes called cellulosomes [47, 78, 98]. The first of these modular enzymes to be discovered and studied was CelB [99], followed by other enzymes such as manA and Cel A [100, 101].

CelA is the largest multi-modular, bifunctional enzyme that is also one of the most-thermostable cellulose degrading GH enzymes [2] known today. The multi-domain architecture of this enzyme reveals three CBM domains and two catalytic domains (GH9-CBM3-CBM3-CBM3-GH48) [2, 5, 97]. One of the catalytic domains has endoglucanase activity (EG) that cleaves bonds in the amorphous regions of cellulose, generating ends further acted upon by cellobiohydrolases and the other domain with exoglucanase or cellobiohydrolase (CBH) activity [102] that cleaves disaccharide units from the reducing and/or non-reducing ends of cellulose (Figure 2). Recent studies exploring the mechanism of action of this cellulase report its ability to create cavities in the middle of cellulose microfibrils as opposed to the end-degradation mechanism seen with previously

characterized cellulases [102]. Such a mechanism increases the overall efficiency of cellulose deconstruction by promoting the formation of exposed cellulose fiber chains open to digestion by other cellulases in the system on the side (Figure 2).

The presence of multiple copies of GH48 domains in combination with other GHs is yet another feature that is not often seen in many other organisms and bestows the species in *Caldicellulosiruptor* genus with diverse functionalities as well as higher cellulolytic abilities [103]. Glycoside hydrolases belonging to families 5, 9, 10, 43, 44, 48 and 74 are found in several but not all species in the genus [97]. The presence, absence, combinations as well as the genomic loci encoding these enzymes can thus, differentiate the strongly cellulolytic species from the weaker ones [84]. Examples of varying cellulolytic capability can be witnessed within species in the *Caldicellulosiruptor* genus where *C. lactoaceticus, C. kronotskyensis, C. obsidiansis, C. saccharolyticus and C. bescii* display a higher cellulolytic ability than other species [50]. *C. bescii* among these has been established to be the most thermophilic species in the *caldicellulosiruptor* genus with the highest growth temperatures of 90^0C and Topt of 78^0C [91]. Gaining attention due to the several characteristics that make it favorable for biofuel production as discussed below, it has become one of the more extensively studied species in the genus.

1.1.7 *C. bescii* and the Three C's: Capabilities, Current limitations and Challenges

The hyperthermophilic *Caldicellulosiruptor bescii* is not only one of the most cellulolytic species within its genus but also upholds the reputation among other organisms. In a study comparing the cell-free extracellular cellulase systems of *T. reesei* with *C. bescii,* it

was observed that *C. bescii* degraded twice the amount of cellulose than *T. reesei* (a fungus used to produce the commercially used cellulose degrading enzymes) [104-106]. *C. bescii* can grow and utilize a variety of substrates including cellulose, hemicellulose, xylan and sugars like maltose, cellobiose, xylose, glucose [49]. It can internalize and co-metabolize both C5 and C6 sugars without experiencing carbon catabolite repression giving it an edge over other organisms from the genera *Clostridium* and *Thermoanaerobacter* that suffer from such limitations [50]. *C. bescii* is capable of degrading crystalline cellulose and hemicellulose in plant biomass by the virtue of its diverse inventory of GHs including modular, multifunctional enzymes such as Cel A that contain CBMs to assist in binding to the substrate for efficient degradation [50, 51]. Other than the CBMs, the S-layer binding domains, two putative adhesins and a type 4 pilus are present and associated with biomass adherence [2, 50, 51]. It further uses the Meyerhof-Parnas pathway to convert glucose released from plant biomass to pyruvate producing lactate, acetate, CO_2 and hydrogen as the major fermentation end products [83].

Alongside the above-mentioned features what gives *C. bescii* an enormous advantage is its ability to grow on and degrade un-pretreated plant biomass i.e., with no physiochemical or thermal pretreatment which has not yet been demonstrated in any of the other microbes under study [83, 107] for similar purposes. This demonstrates the potential for eliminating the need for harsh and expensive physiochemical treatments in the future [6] which would bestow huge economic benefits. Studies have shown that *C. bescii* can tolerate and grow on high loads of un-pretreated biomass, including industrially relevant loads of 200 gL^{-1} (switchgrass) and digest ~85% in the process [107]. Other similar studies under pH-controlled and optimized media environments have also demonstrated *C. bescii* to

completely degrade ~30 gL^{-1} crystalline cellulose and ~10 gL^{-1} switchgrass from 50gL^{-1} loads [77]. This ability to readily grow on more than 100 g/L unpretreated biomass using it as the sole carbon source and not being inhibited by this level of substrate loading is a promising quality displayed by *C. bescii* towards becoming a biofuel microorganism [2]. Moreover, *C. bescii* is very effective in channeling reductants towards molecular hydrogen (H_2) during sugar fermentation and both ferredoxin and NAD(P)H produced by pyruvate ferredoxin oxidoreductase and in glycolysis respectively can be re-oxidized by hydrogen [96, 108]. Its genome encodes for a bifurcating hydrogenase that can potentially account for most of the H_2 produced [109] by the system. A close relative *C. saccharolyticus* has been shown to produce high hydrogen yields approaching the Thauer limit (4 moles H_2/mole glucose) [110]. This expands the horizons for *C. bescii* to be used as a biofuel producing microbe as it can not only be optimized for liquid biofuels but also holds potential for production of commercial biohydrogen. These characteristics make *C. bescii* an attractive, potential CBP microorganism and merits further exploration and development of the species.

As it is known, that there is no ideal CBP organism known or engineered yet, all microbes under study experience limitations and harbor challenges of their own that need to be overcome before one or more of them could be ready for industrial application. Currently alongside other challenges pertaining to CBP derived biofuels like biomass recalcitrance and its microbial deconstruction, incomplete substrate utilization and unexplained growth inhibition continue to be the major limitations with *C. bescii* [77]. An ideal CBP organism must not only possess the ability to efficiently deconstruct biomass making it available for bioconversion, but also be able to completely utilize the available substrate for high

product yields. Despite the ability of *C. bescii* to utilize a wide range of substrates, fermentation inhibition can occur mid-way due to the presence of growth inhibitory factors. Growth inhibitors may be present in biomass, generated through pretreatment or as a product of metabolism (Figure 3). Presence of such inhibitors can generate unfavorable conditions for growth and thus limit bioconversion to products and result in incomplete substrate utilization. Some potential inhibitory chemicals generated during the fermentation process can be sugar degradation products such as furfural and hydroxymethylfurfural (HMF); weak acids such as acetic, formic, and levulinic acids; lignin degradation products such as the substituted phenolics, vanillin and lignin monomers (Figure 3). These compounds can affect the organism both directly and indirectly due to their cellular toxicity or by altering the overall growth environment. Compounds like furans, phenols, carboxylic acids and inorganic salts released during fermentation can have several negative effects on the cell membrane function, growth and glycolysis of the microorganisms [111]. Although the cause for a concentration-independent nature of the growth inhibition seen under the above mentioned conditions in *C. bescii* remains unclear and challenging to pinpoint [2].

Underlying these inhibitory effects are changes in osmolarity and pH which are further two conditions most prone to change in effect of these compounds and are critical for the growth of the organism, such as accumulation of organic acids and/or its ionic derivatives can affect both these parameters. Previous studies in *C. bescii* have shown and at other instances indicated the potential for growth inhibition due to pH and osmolarity. The optimal growth pH for *C. bescii* is reported to be a narrow range between pH 6.8-7.3 [112]. The accumulation of acetic acid, a major fermentation product of *C. bescii* has been seen

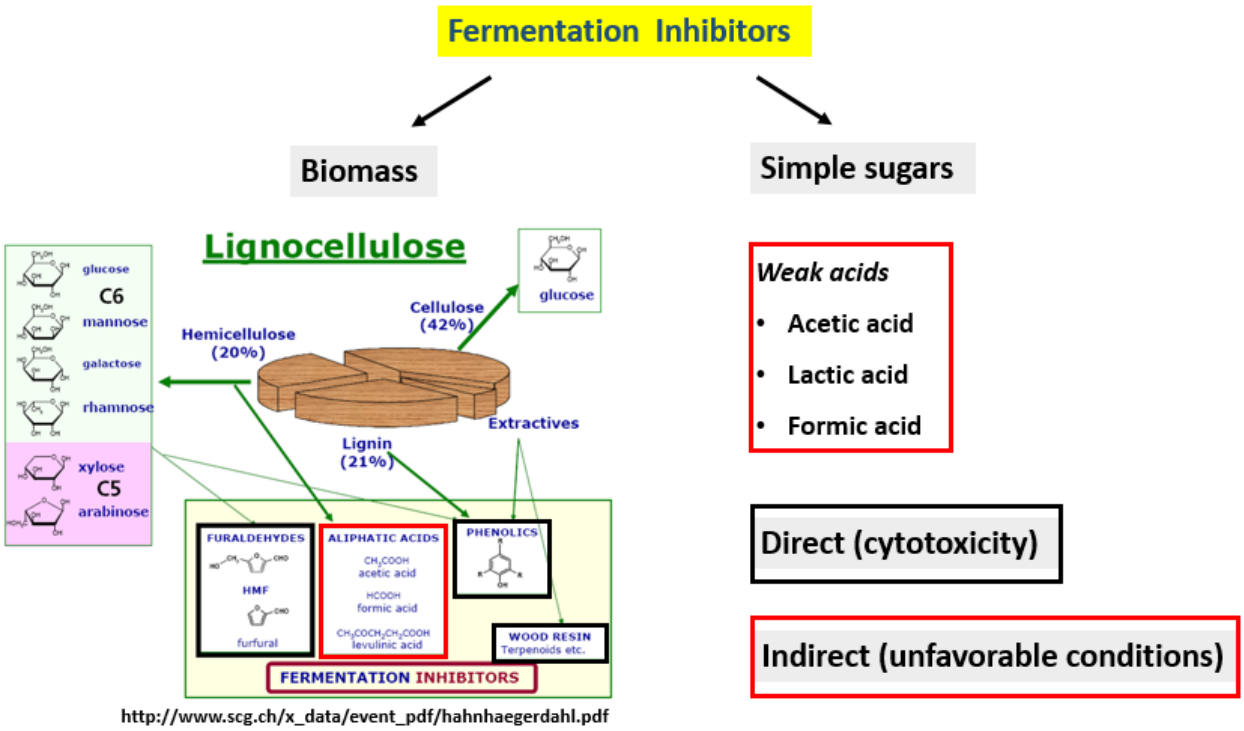


Figure 3: Potential growth inhibitory compounds that are generated by breakdown and utilization of pretreated or un-pretreated biomass, or as sugar degradation products of metabolism. Presence of such inhibitors can limit growth and bioconversion to products as they can affect the organism both directly and indirectly due to their cellular toxicity or by altering the overall growth environment resulting in unfavorable growth conditions.

along the process with increasing acetate levels towards declining growth and the final medium pH reported at ~5.0 [6, 112]. Moreover, the addition of a buffer (40 mM, MOPS) to the medium was shown to significantly improve *C. bescii* growth (~twofold) [112] suggesting the significance of pH in relation to cell yield. On the other hand, development of the minimal growth media for *C. bescii*, LOD (low osmolarity defined) medium reported enhanced growth in this medium and attributed it to reduced solute load leading to lower osmolarity in comparison to the other rich media. The same study showed various levels of growth inhibition with increasing amounts of NaCl in the medium and complete inhibition of growth was reported at 50mM NaCl (~250 mOsm) [113]. Long lag phases indicative of stress has also been reported for *C. bescii* at high sugar loadings of 90 gL-1 glucose (>50h) and 150mM acetate [6].

Other challenges with *C. bescii* are the complete degradation of high loads of unpretreated biomass [77, 107] for production of industrially relevant titers of ethanol [114, 115]. Since the inhibition of growth, of microbes like *C. bescii* due to stress factors such as high osmolarity and low pH along the fermentation process can limit their ability to reach higher cell densities and in turn diminish product yields, which can be a major issue in reaching the goal for industrial biofuel production and needs increased attention. Thus, identification of major inhibitors or inhibitory factors resulting in stressed growth conditions form potential immediate targets for research.

1.1.8 Recent Advancements: Taking Advantage of Advanced Tools and Techniques

In addition to the inherent properties of *C. bescii* making it an organism of interest for CBP based biofuel production, several recent advancements in the system have helped push it up the capabilities ladder. Advanced tools such as next-generation sequencing and omics including transcriptomics, proteomics and metabolomics are being used to explore *C. bescii* [50, 97, 116]. The value of reliable genetic methods is similarly paramount for increasing our knowledge of an organism. The recent development of a genetic system [113, 117], that now enables addition and deletions of genes to and from the organism [118] is thus one of the most important advancements in *C. bescii.* This has opened the gates to potentially expressing foreign genes and pathways to engineer the future "CBP-*Cb"* (CBP optimized *C. bescii*) by manipulating it to produce the desired biofuel related products while rerouting flux from unwanted products. This development has already given way to a number of engineered strains of *C. bescii*, for instance strains capable of producing 12.8 mM ethanol directly from 2% (wt/vol) switchgrass, as opposed to the trace amounts to none produced natively [6, 112]. Other strains have also been reported such as, JWCB005, which is a uracil auxotroph (*ΔpyrFA, ura*$^{-}$ */5-FOA*^{R}) also JWCB018 (*ΔpyrFA ldh::ISCbe4 Δcbe1*) [119] which contains a deletion of the lactate dehydrogenase (ldh) wild type gene in the chromosome and lacks the insertion element, ISCbe4 [120] to facilitate improved transformations and JWCB038 [121] where the NiFe hydrogenase maturation gene cluster has been deleted and can be used in studies related to redox and metabolism of the organism.

The low osmolarity defined (LOD) minimal and complex media (LOC) [113] formulations alongside have also assisted understanding and further research of the growth and

physiological responses of the organism to various environmental stimuli. It has also led to improvement in culture techniques, overcome issues with previous media and allows for selection of nutritional markers, assessment of substrate utilization and more efficient DNA transformation by electroporation [113, 122-124]. Moreover, now there are at least two stable selection markers available for *C. bescii* including the nutritional selection marker for uracil auxotrophy and an antibiotic selection marker for kanamycin resistance [125]. Such tools inspire and support various studies towards increasing our knowledge of this hyperthermophile at physiological, genetic as well as systems biology levels and would keep doing so in the future where newer and more robust strains will be engineered.

1.1.9 New Tools to Meet New Challenges

Another approach towards meeting new and existing challenges with understanding any organism is to develop new tools that can help explore further, faster and deeper. Several challenges with *C. bescii* such as screening for various genetic elements that confer robustness under high osmolarity, low pH or in the presence of biomass released inhibitors call for high-throughput techniques for improved and expedited results. Genome wide screening tools are one such approach that can be successfully used to explore the newer, lesser characterized organisms. *C. bescii* and other similar hyperthermophiles constantly fall short of the tools tailored for use to understand their overall physiology and metabolism. Moreover due to the challenges associated with their growth, their characterization and screening under various environmental stresses is ever challenging

[126, 127]. Transposon based random mutagenesis and RNA interference based tools if developed and optimized for use with lesser characterized (hyper)thermophilic organisms like *C. bescii* would enable fast growth in our knowledge providing newer genes and pathways to target for robustness under various stress conditions [128].

1.2 Overview of the thesis

On the basis of literature review and the existing knowledge/research gaps the work in this dissertation addresses the following questions:

- What factor(s) limit the growth and substrate utilization ability of *Caldicellulosiruptor bescii*?
- How can the limitations be overcome or surpassed to improve the overall performance of the organism?
- Can RNA interference be used to modify gene expression in *C. bescii* and form the basis for the development of a genome-wide screening tool to screen for genetic elements that might confer robustness against the limiting or inhibitory conditions in a fast and high throughput manner?

The central aim of the work is attempting to enhance the overall productivity of *C. bescii* and developing a genome wide screening tool to be used for improving our understanding of the organism as a metabolic model.

Chapter 2 focuses on the objective of developing a sound RNA-seq based transcriptomics pipeline that can be further used for achieving statistically relevant results from a microbial

transcriptomics dataset. A RNA-seq dataset from *Bacillus thuringenesis* was used to addresses the present questions regarding this omics tool such as number of reads or biological replicates needed for achieving universally comparable results with statistical confidence as well as various sources of variation that can go unnoticed and lead to noise in the data, while analyzing large data sets. The study also outlines some experimental design parameters for an efficient and cost effective microbial transcriptomics study. These published findings and developed pipeline from this chapter were further utilized in analyzing RNA-seq data for chapter three.

Chapters 3 and 4 of the thesis address the existing growth and substrate utilization limitations in *C. bescii*. In these chapters, pH and osmolarity are the two major factors investigated for their contribution in limiting the fermentation potential of *C. bescii.* One of the major growth and substrate utilization limiting factors was revealed along with the way of alleviating it and improving the overall productivity, followed by the secondary factor prone to becoming the next immediate cause limiting fermentation potentials.

Chapter 5 offers a potential method for fast and high-throughput discovery of target genetic elements that can confer robustness under the limiting or stressful conditions. Here the attempts at validating and developing an RNAi based approach to genome wide screening in *C. bescii* is described and detailed.

Chapter 2.0

Replicates, Read Numbers, and Other Important Experimental Design Considerations for Microbial RNA-Seq. Identified Using *Bacillus thuringiensis* Datasets

The research presented in this chapter was published in Frontiers in Microbiology under the title “Replicates, Read Numbers, and Other Important Experimental Design Considerations for Microbial RNA-seq Identified Using *Bacillus thuringiensis* Datasets” (Manga, P., et al., Replicates, Read Numbers, and Other Important Experimental Design Considerations for Microbial RNA-seq Identified Using Bacillus thuringiensis Datasets. Frontiers in Microbiology, 2016. 7(794). In this publication, Punita Manga contributed to the following aspects of the work: generated and analyzed RT-qPCR data; processed (trimmed, filtered and mapped to reference genome) raw RNA-Seq reads, wrote and executed DESeq2 RNA-Seq analysis pipeline using R to generate read count data and generated all figures in the manuscript; along with Loren J. Hauser, Charlotte M. Wilson and Steven D. Brown analyzed the generated data for biological inferences and wrote the manuscript.

2.1 Introduction

Ever decreasing next-generation sequencing (NGS) costs, continued technical and analytical advances, along with diverse applications have made RNA-sequencing (RNA-seq) an ever-increasing choice for transcriptome studies [129-131]. RNA-seq applications include differential gene expression studies, the detection of strand specific expression or transcript fusions, determination of alternative splicing isoforms, identification of specific SNP's and their locations, long and small RNAs, genome guided, and de novo transcript assemblies and start sites analyses [132-134]. It also enables detection of weakly expressed genes and does not have to be limited by previously sequenced genome knowledge [130].

Various NGS platforms, assembly and statistical tools can be used to generate RNA-seq datasets, but the overall methodology across platforms is similar [131]. While direct RNA sequencing is possible [135], for the majority of expression studies RNA is isolated from cells and usually undergoes rRNA depletion or poly(A) enrichment. The transcript enriched material is then used as template material to generate complementary DNA (cDNA) libraries via a reverse transcription enzymatic reaction, which represents the transcripts within each sample. Library creation may include the addition of barcodes/adaptors so samples from multiple conditions can be pooled, run together, and then data attributed appropriately. In the case of indirect RNA-seq methods an amplification step is required. Sequence data in the form of raw reads are quality filtered/trimmed, most often aligned to a reference genome, then the number of reads mapped to individual genes in the reference genome are counted and then further used

to estimate differential gene expression using a range of statistical methods [130, 136, 137].

While RNA-Seq has a number of advantages over DNA microarrays, it is still a developing technology that faces a number of challenges [134, 137-139]. Variation, errors, and biases may be introduced in any of the multiple steps used to generate and analyze the datasets [140]. Technical and biological factors that contribute to variation, errors, and biases include experimental design, RNA extraction procedures, sample handling, differences in amount of starting RNA, library preparation steps such as PCR amplification, sample storage, GC content, and read number differences [138, 141]. A number of different of normalization methods have been developed for NGS data to remove unwanted variance [142, 143]. Normalization methods include examples such Total Count (TC), Upper Quartile (UQ), Reads Per Million base pairs (RPM), Reads Per Kilobase per Million base pairs (RPKM), Trimmed Mean of M-values (TMM), Kernel Density Mean of M-component (KDMM), and analysis packages like DESeq and edgeR have inbuilt normalization algorithms [142-146]. There is no clear consensus on which normalization is the best suited for RNA-seq data. Although, studies that have compared some of these methods to one another show that UQ, TMM, and DESeq normalization result in similar qualitative characteristics of the normalized dataset and differential expression analysis [142, 147]. Recent studies have shown that RNA-seq data often fits well to a negative binomial distribution [148-151]; and this method is being more widely adopted. While a well-designed experiment and normalization are important, they may be insufficient if there is large unknown variance [138]. A recent study analyzed RNA-seq data from 48 samples obtained from seven Illumina HiSeq 2000 lanes and concluded that

“bad” replicates risk skewing data interpretation and that increasing biological replicates beyond the typical two or three is beneficial [151].

Other important considerations for an RNA-seq study include the choice of quality trimming/filtering tools, mapping algorithm, statistical test, required number of reads, or genome coverage, number of biological replicates and cost [138, 140, 152]. NGS technologies generate large datasets that may be computationally challenging for smaller laboratories to store, retrieve, and analyze. Thus, there is a demand for bioinformatics tools that are proficient in data handling i.e., are fast and have reduced error rates, have a broad consensus and are easy to use [136, 139, 153, 154]. Cost is an essential factor for most laboratories, which is directly related to the number of reads generated per sample and the number of replicates used. Thus, it is important to establish an acceptable trade-off between number of reads and replicates for an efficient, powerful, yet cost effective experiment [152].

The aim of the present study was to better understand the required number of reads and replicate numbers for statistically confident results in the context of a typical experimental laboratory. Transcriptomic profiles were generated for two closely related *Bacillus thuringiensis* strains (serovar berliner ATCC10792 and CT43) under similar experimental conditions and since the outcomes for each strain were similar to one another we mainly present ATCC10792 analyses but the analyses for strain CT43 are also shown. *B. thuringiensis* is a Gram-positive, spore and Cry toxin producing bacterium [155] that has been applied for biocontrol of different insects [156-158] and a number of genome sequences are available for study [159, 160]. The ATCC 10792 genome (NCBI accession NZ_CM000753) is 6,260,142 bp, was recently reannotated with 6330 genes and 13

copies for the 5S, 16S, and 23S rRNA genes predicted. The data from this well-replicated study with 32 samples, each from one Illumina HiSeq 2000 lane, generated a large number of reads per sample, and significantly differentially expressed genes were detected using DESeq2 [146] .Differentially expressed (DE) genes were validated by Real Time quantitative RT-qPCR. This study provides insights into sample and read numbers required to derive biologically meaningful results and will be useful to others looking to develop or assess different bioinformatics and/or statistical approaches for RNA-seq studies.

2.2 Materials and Methods

2.2.1 Organism Growth and Sampling

Bacillus thuringiensis serovar berliner strain ATCC 10792 and *Bacillus thuringiensis serovar chinensis* strain CT-43 were obtained from the Bacillus Genetic Stock Center (www.bgsc.org) and have Average Nucleotide Identity (ANI)-values of ≥99.63% in reciprocal genome analyses based on BLAST (ANib). Each strain was plated on Luria Bertani (LB) medium and cultured at 30°C. Single colonies were used to inoculate 5 mL LB starter cultures, which were grown at 30°C with shaking at 200 rpm (New Brunswick Scientific, Innova 4430) overnight. For RNA-seq experiments, 1 mL aliquots of overnight cultures were used to inoculate 500 mL baffled flasks containing 200 mL of LB medium. Cultures were grown for 3 h at 30°C with shaking at 200 rpm and harvested at approximately mid-log phase (OD600, ~0.42). To harvest cells for RNA extraction, 40 mL culture aliquots were collected by rapid centrifugation (Sorvall, Evolution RC) at 7649 × g

at 4°C for 5 min. Cell pellets were frozen in liquid nitrogen for 10 min and then stored at −80°C. All cultures were grown and harvested under similar conditions. A total of 16 samples were collected per strain, with four biological replicates for each strain, collected on four different dates and using media from two different LB broth lots (lot #1091744 and 7220443) using water from two different buildings to generate 32 samples for RNA-seq analysis (Figure 4 and Table 1) Difco Lennox LB medium was used in this study (Becton, Dickinson and Company, Franklin Lakes, NJ, USA).

2.2.2 RNA Extraction and cDNA Library Preparation

High quality RNA (Bioanalyzer RNA integrity numbers (RINs) >8.5) was isolated from strain CT43 and strain ATCC 10792 using the TRIzol reagent (Invitrogen, Carlsbad, CA, USA) combined with a bead beating step, essentially as described previously [161]. Briefly, cell pellets from each sample were resuspended in TRIzol reagent, then TRIzol/cell mixtures were added to tubes containing 800 mg of 0.1 mm glass beads (Biospec Products Inc., Bartlesville, OK, USA) and cells were lysed by bead beating on a Precellys 24 high-throughput tissue homogenizer (Bertin Technologies, Montigny-le-Bretonneux, France) with the following settings; 3 × 20 s at 6500 rpm. Chloroform was added post-lysis, mixed by vortexing, and the mixture was centrifuged at 20,817 × g (Centrifuge 5417R, Eppendorf) for 15 min at 4°C. The aqueous phase was collected and

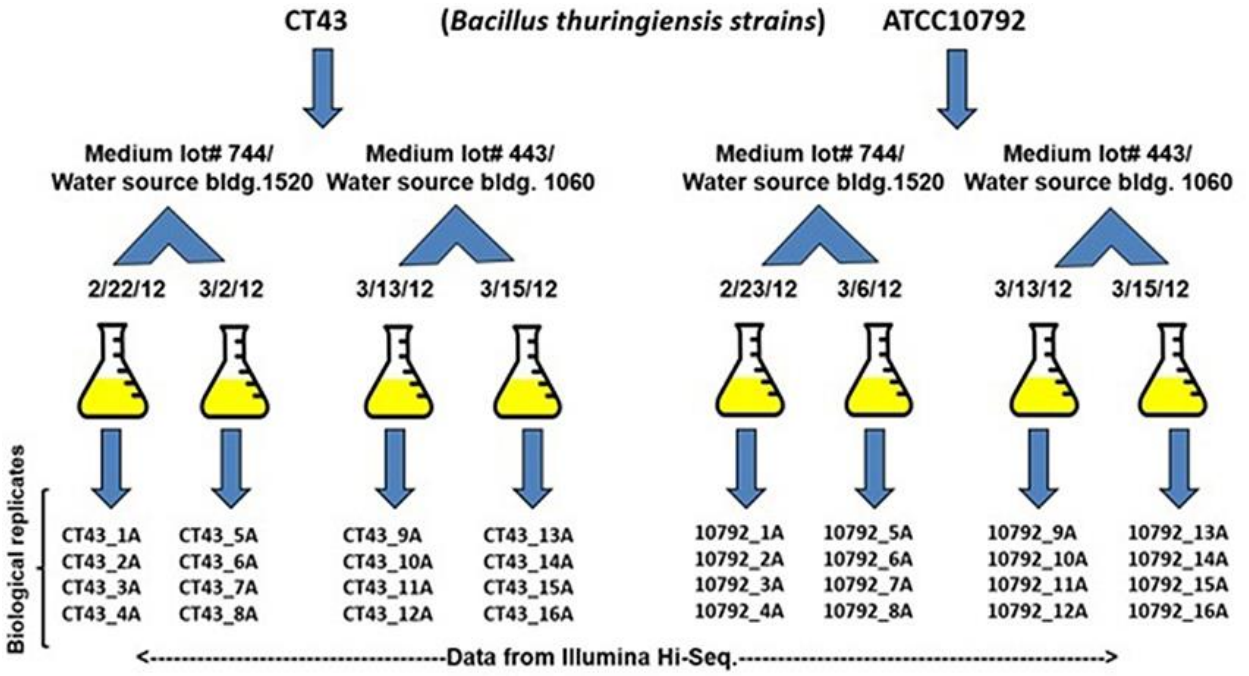


Figure 4: Sampling and potential sources of variation in RNA seq data. Strains CT43 and ATCC10792 grown in two medium lots # 1091744 and 7220443 in water taken from two different sources. Bacteria were cultured on four different dates and four biological replicates were grown for each date, followed by sequencing using Illumina HiSeq 2000.

Table 1: Sample collection and culture details for all samples from strains CT43 and ATCC10792.

Sample	Medium lot #	Water source	Culture date	Sample OD (600 nm)
10792_1A	1091744	1520	2/23/2012	0.394
10792_2A	1091744	1520	2/23/2012	0.398
10792_3A	1091744	1520	2/23/2012	0.384
10792_4A	1091744	1520	2/23/2012	0.404
10792_5A	1091744	1520	3/6/2012	0.398
10792_6A	1091744	1520	3/6/2012	0.42
10792_7A	1091744	1520	3/6/2012	0.406
10792_8A	1091744	1520	3/6/2012	0.452
10792_9A	7220443	1060	3/13/2012	0.384
10792_10A	7220443	1060	3/13/2012	0.386
10792_11A	7220443	1060	3/13/2012	0.398
10792_12A	7220443	1060	3/13/2012	0.392
10792_13A	7220443	1060	3/15/2012	0.482
10792_14A	7220443	1060	3/15/2012	0.476
10792_15A	7220443	1060	3/15/2012	0.504
10792_16A	7220443	1060	3/15/2012	0.468
CT43_1A	1091744	1520	2/22/2012	0.464
CT43_2A	1091744	1520	2/22/2012	0.349
CT43_3A	1091744	1520	2/22/2012	0.352
CT43_4A	1091744	1520	2/22/2012	0.54
CT43_5A	1091744	1520	3/2/2012	0.388
CT43_6A	1091744	1520	3/2/2012	0.398
CT43_7A	1091744	1520	3/2/2012	0.412
CT43_8A	1091744	1520	3/2/2012	0.396
CT43_9A	7220443	1060	3/13/2012	0.408
CT43_10A	7220443	1060	3/13/2012	0.392
CT43_11A	7220443	1060	3/13/2012	0.38
CT43_12A	7220443	1060	3/13/2012	0.396
CT43_13A	7220443	1060	3/15/2012	0.426
CT43_14A	7220443	1060	3/15/2012	0.438
CT43_15A	7220443	1060	3/15/2012	0.422
CT43_16A	7220443	1060	3/15/2012	0.476

mixed with ethanol and purified using the RNeasy Mini kit (Qiagen, Waltham, MA, USA) in accordance with the manufacturer's instructions and using the optional on column DNaseI treatment. The quantity and quality of RNA was assessed using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE) and Agilent Bioanalyzer (Agilent, Santa Clara, CA, USA). Ribosomal RNA was depleted from the samples using a Ribo-Zero rRNA removal kit for Gram-Positive bacteria (Epicentre, Madison, WI, USA) and cDNA libraries were prepared and barcoded using a ScriptSeq v2 RNA-seq Library preparation kit. The final libraries were quantified with a Qubit double stranded broad range assay kit and fluorometer (Invitrogen) and quality assessed using a Bioanalyzer (Agilent). Samples were diluted according to manufacturer's recommendations using the Illumina dilution calculator, and sequence data was generated via two runs using an Illumina HiSeq 2000 instrument with SR50 sequencing kits (50 bp single end reads) and using phiX control DNA (Illumina, Inc., San Diego, CA, USA), as previously described (Wilson et al., 2013).

2.2.3 Data Analysis

Mapping, Clustering, and Quality Control

Raw reads were imported into the CLC Genomics Workbench 7.0.4 (CLC Bio, a Qiagen company) and were filtered and trimmed based on quality assessments. Sequence reads < 20 nucleotides were discarded. A modified-Mott trimming algorithm, which incorporates quality scores on a Phred scale, was used for quality trimming in CLC Bio with the quality trimming parameter set to 0.02. Trimmed and filtered reads were then aligned to their

respective reference genomes, using the default “prokaryote genomes” and unique reads settings. Raw data variance was observed by Principle Component Analysis (PCA) and by cluster analysis using JMP Genomics 6.0 (SAS Institute, Cary, NC, USA).

Data Access

The genomes used in this study have been described [159, 160] and are available from the NCBI GenBank database under accession numbers for NZ_CM000753 and NC_017208 for strains ATCC10792 and CT43, respectively. RNA-seq data have been deposited in NCBI Gene Expression Omnibus (GEO) database under accession number GSE71189 and raw sequence data deposited at the NCBI Sequence Read Archive (SRA) under accession number SRP041628.

2.2.4 Differential Gene Expression Analysis: DESeq2

Uniquely mapped reads were incorporated into a tabular format [Data Sheet 3 (2, Appendix)] and analyzed using the DESeq2 differential expression analysis pipeline [146]. Differentially expressed (DE) genes were identified based on comparisons between medium lots and culture dates for each strain using a 5% False Discovery Rate (FDR) and a two-fold expression difference to detect significantly DE genes (data available in Geo repository).

2.2.5 RT-qPCR Validation of RNA-seq Results

RNA-seq data for the differentially expressed genes was validated using real-time quantitative PCR (RT-qPCR) as previously described [161]. Six *B. thuringiensis* strain ATCC10792 genes that represented a range of differential expression values from RNA-seq data were chosen for validation. Primers used to validate medium and the date effects are listed in Table 2.

2.2.6 Determination of Iron Content in Media and Water

Iron content for the LB medium and for the water sources from the two different building sources were quantified by elemental analysis using a Perkin Elmer ELAN 6100 Inductively Coupled Plasma-Mass Spectrometer (ICP-MS), as previously described [162]. Water source and media lot pH were measured using colorpHast pH-indicator strips (EMD Millipore, Billerica, MA, USA).

2.2.7 Alteration of Sequence Read and Biological Replicate Numbers

To observe the effect of using fewer biological replicates and lower sequence read numbers on differential gene expression detection, data available from all biological replicates per strain within each condition (i.e., same medium lot and date of culture) were grouped in sets with replicates for another condition (Figure 4). The number of biological replicates varied from two to four and the number of differentially expressed genes were

Table 2: List of RT-qPCR primers used for validation of RNA-Seq results for medium lots and culture dates.

Sequence Description	**Primer Sequence**
Medium effect validation genes	
bthur0008_13460_FP	TCGTTGAGGGAGGAAAAGAA
bthur0008_13460_RP	CGCACCTGAAGATCGGAATA
bthur0008_6770_FP	AGGCGCTGTTAATACGATGC
bthur0008_6770_RP	CTCCAACTCCGCCATACACT
bthur0008_28360_FP	ACAGCTTGGTTTTTCGGAGA
bthur0008_28360_RP	CCCATACTCCGACATGCATAC
bthur0008_34530_FP	AGGGATTGCAATTGGATCAG
bthur0008_34530_RP	TATGCTTCCAGACACCCACA
bthur0008_41720_FP	GTGCTGATGTAGCGAAGCAA
bthur0008_41720_RP	GTGGACCTCTTCCGAACTCA
bthur0008_34520_FP	TACAACTGGTAGCCCGGAAG
bthur0008_34520_RP	GCCCCGTTATACGGATCAAT
bthur0008_38600_FP	TGCGGCATTAACAGAGGTGT
bthur0008_38600_RP	ACTTTCCCAAGCACAAGAGC
Date effect validation genes	
bthur0008_61310_FP	TCATGCACAGGTGAACGTCG
bthur0008_61310_RP	ACGGGTAATTTCTCCAACCCC
bthur0008_38200_FP	CCAACGTGTTGCGTTAGGTA
bthur0008_38200_RP	GCATTGCAACACGGAGTTTA
bthur0008_3370_FP	ATAATCCGAAGCGTGATTGG
bthur0008_3370_RP	ATCATATTCCCATGCCGAAC
bthur0008_4210_FP	ACAAGTTGGCCCTGAGTACG
bthur0008_4210_RP	TAGACCCGCTAACCATTCCA
bthur0008_52220_FP	GTGGCTATGGATGGCTTGTT
bthur0008_52220_RP	AAGGGGATTTTCCCTTCTTG
bthur0008_17630_FP	TGCGATTGCAAACTTTATGG
bthur0008_17360_RP	CCCAAGTAACGCAAATGGAT

determined using DESeq2 [Data Sheet 8, (3, Appendix)]. For example, a set of two replicates 1A/2A were grouped with 9A/10A for differential expression estimation. When analyzing the effect that the total number of reads per sample had on the number of differentially expressed genes, the original number of reads obtained for each sample with genome coverage ranges from 87 to 465x was considered as 100% reads (Table 3a, b and c). Subsets with randomly reduced reads of 75, 50, 25, 10, and 5% of the original number of reads were generated using the “sample reads” option in the genome finishing module of CLC Genomics Workbench 7.0.4 (CLCBio). Each subset was remapped with the respective reference genome prior to performing differential gene expression analysis via DESeq2 [Data Sheet 9, (4, Appendix)].

2.3 Results

2.3.1 RNA-seq Experiments

Samples were harvested for all cultures during mid-exponential growth. The average culture turbidity, as measured by optical density at OD600nm, was 0.422 ± 0.04 (range 0.384–0.504) for strain ATCC10792 and 0.415 ± 0.05 for CT43 at the time of sample collection (Table 1). For each sample 15–30 M raw reads were generated, and the resulting genome coverages were between 87 and 465X post-quality filtering and trimming. Post-trimming and mapping results for strain ATCC10792 and CT43 are provided in Tables 3a, b and c.

Table 3a: Reads and mapping details for strain ATCC10792

ATCC10792 (Ref. genome size= 6260142)	Total No. of Reads (trimmed)	*Genome Coverage	Total Mapped Reads	Unique Gene Reads
10792_1A	26,986,606	202x	19,762,858	19,708,701
10792_2A	11,787,714	87x	7,579,068	7,534,508
10792_3A	27,315,600	203x	19,174,331	19,098,676
10792_4A	53,643,496	400x	37,447,561	37,326,511
10792_5A	51,109,286	380x	37,229,691	37,112,202
10792_6A	57,636,652	430x	41,566,889	41,466,848
10792_7A	49,915,906	370x	34,855,294	34,747,534
10792_8A	52,689,519	392x	37,104,818	36,999,659
10792_9A	20,291,318	160x	9,640,962	9,590,311
10792_10A	13,356,476	105x	6,185,482	6,154,691
10792_11A	22,487,034	177x	9,762,915	9,698,237
10792_12A	25,052,676	197x	10,408,386	10,361,349
10792_13A	21,857,603	172x	9,475,191	9,444,912
10792_14A	27,286,565	215x	10,871,957	10,828,524
10792_15A	26,104,043	206x	9,558,505	9,484,938
10792_16A	30,818,722	243x	21,037,052	20,987,988

Table 3b: Reads and mapping details for strain CT43.

CT-43 (Ref. genome size = 6260142)	Total No. of Reads (trimmed)	*Genome Coverage	Total Mapped Reads	Unique Gene Reads
CT43_1A	47,788,640	407x	33,410,382	33,277,242
CT43_2A	54,445,876	465x	37,563,633	37,442,782
CT43_3A	48,368,479	411x	31,998,788	31,871,485
CT43_4A	51,206,012	437x	33,759,084	33,631,091
CT43_5A	26,873,919	228x	18,821,043	18,737,093
CT43_6A	16,109,350	137x	10,823,017	10,777,024
CT43_7A	25,742,903	217x	17,529,810	17,431,969
CT43_8A	15,094,710	128x	10,551,146	10,515,619
CT43_9A	35,539,162	210x	12,471,721	12,405,463
CT43_10A	24,180,859	217x	12,195,076	12,140,964
CT43_11A	23,290,427	209x	10,692,596	10,628,423
CT43_12A	21,950,705	197x	9,343,385	9,296,130
CT43_13A	12,875,156	116x	5,350,598	5,318,644
CT43_14A	24,636,162	221x	10,717,848	10,662,619
CT43_15A	23,706,467	213x	23,706,467	23,706,467
CT43_16A	22,706,616	204x	10,263,296	10,216,842

*Genome coverage= Σ(Avg. trimmed read length x No. of reads after trimming)/Reference genome

Table 3c: Example genome coverage calculation shown sample 1A from strain ATCC10792 as performed for all samples in the study.

Name	Number of reads	Avg.length	Number of reads after trim	Percentage trimmed	Avg.length after trim		
10792	28769	50	28769	100	46	Avg.length after trim x Number of reads after trim ➔	1323374
10792	5113107	50	5113015	~100%	46.8		239289102
10792	5088912	50	5088833	~100%	46.7		237648501
10792	1828539	50	1828511	~100%	47		85940017
10792	2783	50	2783	100	47.3		131635.9
10792	5014708	50	5014650	~100%	46.9		235187085
10792	5036856	50	5036773	~100%	47		236728331
10792	4873350	50	4873272	~100%	47		229043784
			26986606			Total Reads	1265291830
		Genome size	6260142		**Coverage**	Total Reads/ Genome Size➔	202.12

The ribosomal RNA depletion strategy worked well and similarly for both strains as indicated by an analysis showing that on average for both strains only 0.07% of trimmed, mapped reads aligned to the 5S, 16S, and 23S rRNA genes (S.D. ± 0.05 and 0.06 for ATCC 10792 and CT43, respectively). For both strains, medium lot and culture date were identified as important variance sources during Principal Component Analysis (PCA) and cluster analysis of raw data (Figure 5). Variation across biological replicates was low with the linear (Pearson) correlation values within like replicates for both ATCC10792 and CT43 ranging from between 0.95 and 0.99 (Figure 6 a and b).

2.3.2 Differential Gene Expression Analysis using DESeq2: Medium Lot and Date Effect

As variation based on medium lots and culture dates was detected, differential expression analysis was conducted to examine the effect of different media, and culture dates on transcriptomic profiles. When all replicates per strain and 100% reads were applied to the analysis, 735 and 1086 genes (5% False Discovery Rate (FDR) and two-fold change) were observed to be differentially expressed between medium lots (#1091744 vs. #7220443) for strains ATCC10792 and CT43, respectively. Files with a complete list of altered gene expression based on medium lot difference could be found in Geo repository for both strains ATCC10792 and CT43. In response to the different medium lots, genes related to iron acquisition and metabolism were consistently differentially expressed for both strains. A summary of iron related genes that passed both 5% FDR and two-fold expression difference significance thresholds is shown (Table 4). Based on the differential

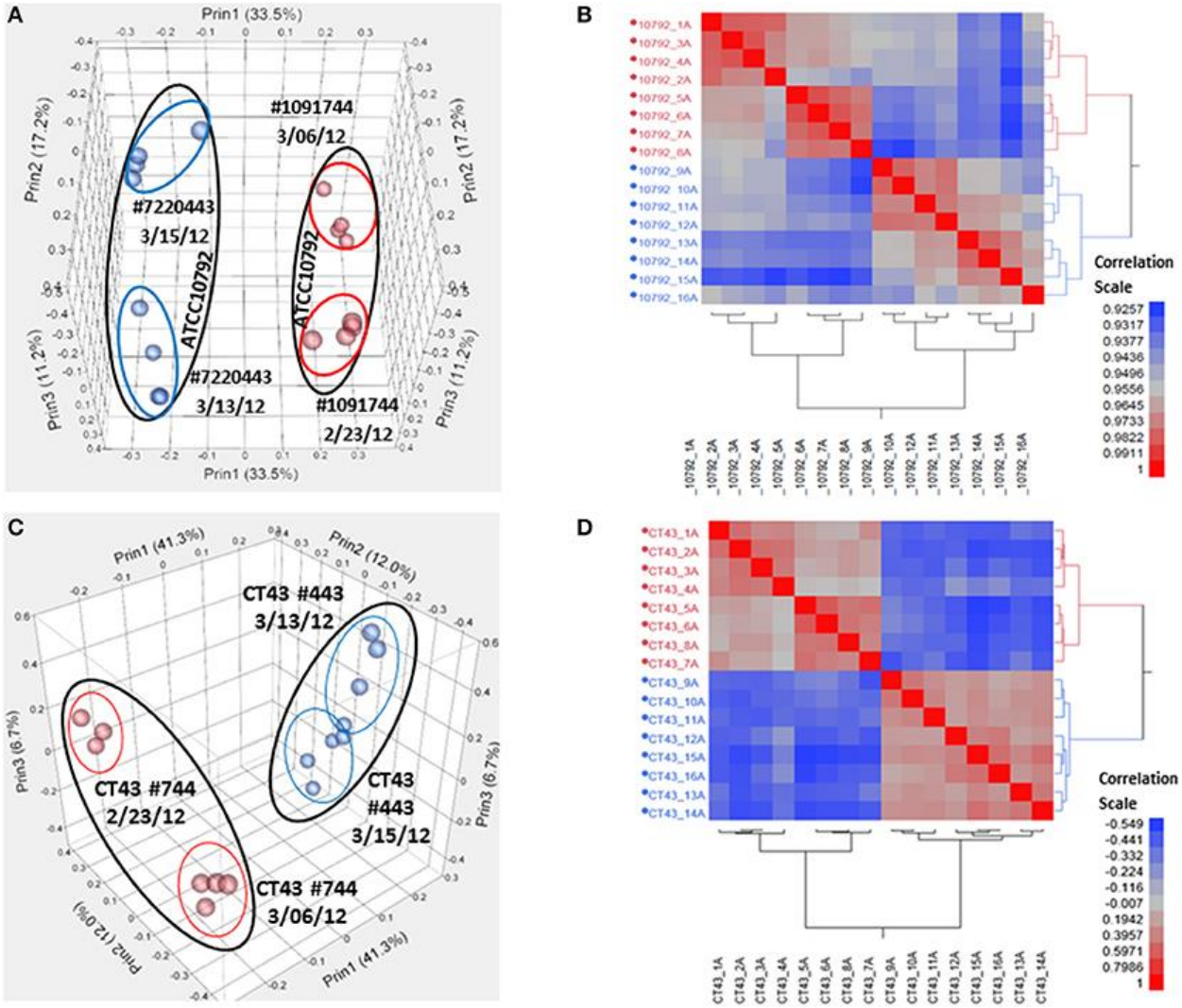


Figure 5: Variation analysis of raw read count data for strain ATCC10792 and strain CT43. (A) Principal Component Analysis (PCA) for ATCC10792 using a Pearson correlation coefficient and colored by media, (B) Hierarchical cluster analysis of the same data for strain ATCC10792, (C) PCA for CT43 and (D) CT43 cluster analysis.

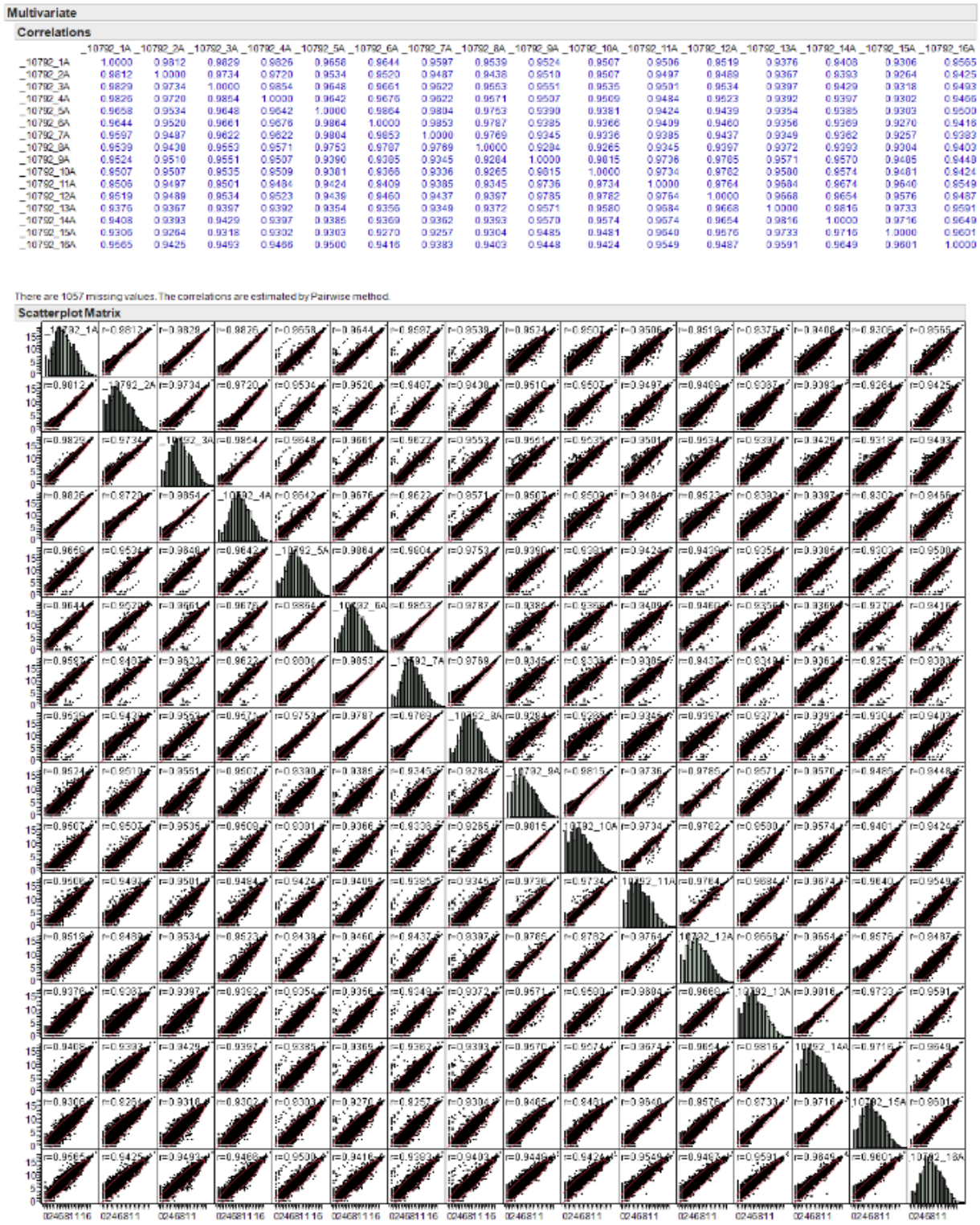


Figure 6a: Multivariate correlation analysis summarizing variation among biological replicates of strain ATCC10792.

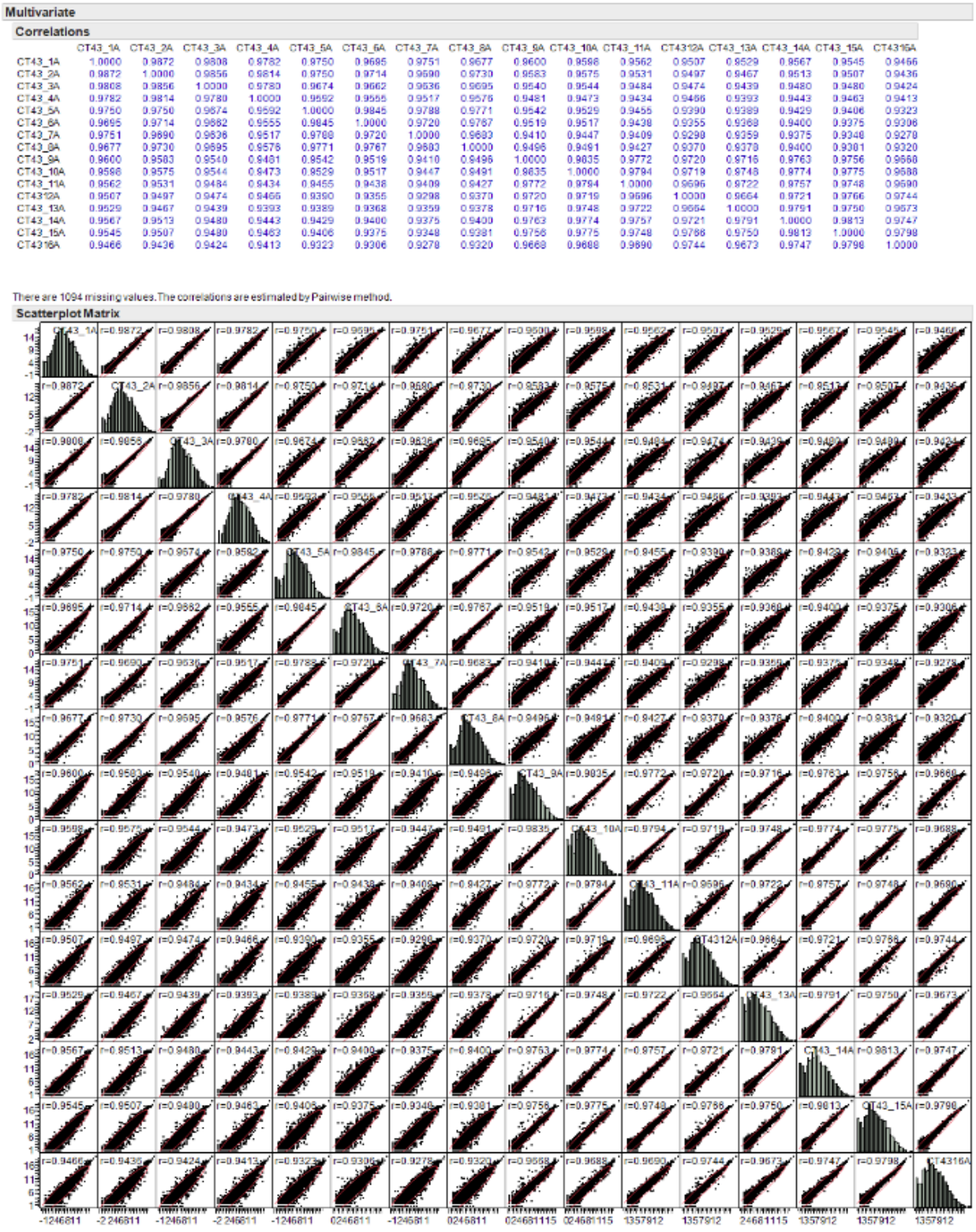


Figure 6b: Multivariate correlation analysis summarizing variation among biological replicates of strain CT43

expression results, it was hypothesized that iron had become limiting for cultures grown in medium from lot 1091744 compared to cells grown in medium prepared from the other lot. An elemental analysis of sterile media prepared from different lots revealed that higher amounts of total iron were indeed present in medium lot #7220443 compared to medium lot #1091744 (Table 5). Both media were prepared at pH ~ 7.0.

When analyzing the data based on different culture dates within a medium lot (the date effect) for ATCC10792, 403 genes were identified as differentially expressed for cultures in medium lot #1091744 when culture from the date 2/23/12 was compared with 3/6/12. Similarly, for cultures grown in medium lot #7220443 when comparing cultures from dates 3/13/12 and 3/15/12, 458 genes were identified as differentially expressed in ATCC10792. Similar results were obtained for strain CT43 when differential gene expression analysis was conducted for the culturing date effect within a medium lot.

2.3.3 Real Time-Quantitative PCR Validation (RT-qPCR)

Six genes exhibiting a broad range of strain ATCC10792 expression differences for both medium lot and date effect comparisons were selected for confirmation by RT-qPCR. Comparison of DESeq2 estimated expression differences with measurements determined by RT-qPCR showed that the two different data sets had correlation coefficient (R2)-values of 0.90 and 0.92 for genes chosen for medium lots and culture dates, respectively (Figure 7).

Table 4: Genes related to iron acquisition and metabolism differentially expressed in strain ATCC10792 and CT43 grown in medium lot #1091744 over #7220443.

Differentially expressed Iron genes [medium lot #744 vs #443]			
Locus tag	**Product**	**log2Fold Change**	**Padj (FDR=5%)**
BTHUR0008_RS01670	iron ABC transporter permease	2.67	<0.001
BTHUR0008_RS01675	ferrichrome ABC transporter permease	2.70	<0.001
BTHUR0008_RS01680	ABC transporter substrate-binding protein	2.94	<0.001
BTHUR0008_RS01685	ferredoxin--NADP reductase	2.56	<0.001
BTHUR0008_RS02820	iron-enterobactin transporter ATP-binding protein	1.29	<0.001
BTHUR0008_RS02825	iron ABC transporter permease	1.45	<0.001
BTHUR0008_RS02835	iron siderophore-binding protein	1.23	<0.001
BTHUR0008_RS03465	iron transporter FeoA	-1.23	<0.001
BTHUR0008_RS06975	ferredoxin	-1.04	<0.001
BTHUR0008_RS10095	Fe-S oxidoreductase	-1.43	<0.001
BTHUR0008_RS10345	iron(III) dicitrate-binding protein	1.97	<0.001
BTHUR0008_RS15775	ferrichrome ABC transporter permease	2.48	<0.001
BTHUR0008_RS15780	iron ABC transporter permease	2.34	<0.001
BTHUR0008_RS15785	iron-hydroxamate ABC transporter substrate-binding protein	2.51	<0.001
BTHUR0008_RS17445	iron-uptake system-binding protein	3.46	<0.001
BTHUR0008_RS17450	ferrichrome ABC transporter permease	2.88	<0.001
BTHUR0008_RS17455	iron ABC transporter permease	3.72	<0.001
BTHUR0008_RS17460	ABC transporter ATP-binding protein	3.41	<0.001
BTHUR0008_RS17465	IroE protein	2.48	<0.001
BTHUR0008_RS20850	iron ABC transporter ATP-binding protein	3.69	<0.001
BTHUR0008_RS20855	iron ABC transporter permease	3.25	<0.001
BTHUR0008_RS20860	iron-hydroxamate ABC transporter substrate-binding protein	4.02	<0.001

Table 4: Continued.

Differentially expressed Iron genes [medium lot #744 vs #443]			
Locus tag	**Product**	**log2Fold Change**	**Padj (FDR=5%)**
BTHUR0008_RS21120	ferrichrome ABC transporter substrate-binding protein	2.69	<0.001
BTHUR0008_RS21675	ferrichrome ABC transporter substrate-binding protein	1.57	<0.001
BTHUR0008_RS21745	heme-degrading monooxygenase IsdG	2.42	<0.001
BTHUR0008_RS21760	ABC transporter permease	1.37	<0.001
BTHUR0008_RS21765	heme ABC transporter substrate-binding protein	2.32	<0.001
BTHUR0008_RS23110	iron transporter FeoA	1.13	<0.001
BTHUR0008_RS23575	ferritin	-1.58	<0.001
BTHUR0008_RS25020	iron ABC transporter substrate-binding protein	2.17	<0.001
BTHUR0008_RS25025	iron ABC transporter permease	1.71	<0.001
BTHUR0008_RS25030	iron ABC transporter permease	1.31	<0.001
BTHUR0008_RS25035	iron ABC transporter ATP-binding protein	1.11	<0.001
BTHUR0008_RS25920	ferrichrome ABC transporter permease	1.75	<0.001
BTHUR0008_RS25930	iron-dicitrate ABC transporter ATP-binding protein	1.31	<0.001
BTHUR0008_RS25935	ferrichrome ABC transporter substrate-binding protein	2.73	<0.001
CT43_RS01110	ferrichrome ABC transporter permease	1.39	<0.001
CT43_RS01900	iron ABC transporter permease	2.96	<0.001
CT43_RS01905	ferrichrome ABC transporter permease	2.84	<0.001
CT43_RS01915	ferredoxin--NADP reductase	2.62	<0.001
CT43_RS03040	iron siderophore-binding protein	1.10	<0.001
CT43_RS03050	iron ABC transporter permease	1.33	<0.001
CT43_RS03055	iron-enterobactin transporter ATP-binding protein	1.28	<0.001

Table 4: Continued.

Differentially expressed Iron genes [medium lot #744 vs #443]			
Locus tag	**Product**	**log2Fold Change**	**Padj (FDR=5%)**
CT43_RS03470	iron ABC transporter permease	1.30	<0.001
CT43_RS03815	iron transporter FeoB	-1.07	<0.001
CT43_RS03825	iron transporter FeoA	-1.95	<0.001
CT43_RS07445	ferredoxin	-1.02	<0.001
CT43_RS08155	iron-sulfur cluster assembly protein HesB	1.58	<0.001
CT43_RS11170	iron(III) dicitrate-binding protein	1.88	<0.001
CT43_RS17225	iron-hydroxamate ABC transporter substrate-binding protein	3.15	<0.001
CT43_RS17230	iron ABC transporter permease	2.91	<0.001
CT43_RS17235	ferrichrome ABC transporter permease	3.18	<0.001
CT43_RS18585	iron ABC transporter permease	4.25	<0.001
CT43_RS18590	ferrichrome ABC transporter permease	3.74	<0.001
CT43_RS18595	iron-uptake system-binding protein	4.46	<0.001
CT43_RS18875	heme ABC transporter ATP-binding protein	-1.15	<0.001
CT43_RS22070	iron ABC transporter ATP-binding protein	3.71	<0.001
CT43_RS22075	iron ABC transporter permease	3.11	<0.001
CT43_RS22080	iron-hydroxamate ABC transporter substrate-binding protein	3.71	<0.001
CT43_RS22340	ferrichrome ABC transporter substrate-binding protein	2.18	<0.001
CT43_RS22890	ferrichrome ABC transporter substrate-binding protein	2.30	<0.001
CT43_RS22960	heme-degrading monooxygenase IsdG	2.88	<0.001

Table 4: Continued.

Differentially expressed Iron genes [medium lot #744 vs #443]			
Locus tag	**Product**	**log2Fold Change**	**Padj (FDR=5%)**
CT43_RS22980	heme ABC transporter substrate-binding protein	2.59	<0.001
CT43_RS25985	iron ABC transporter ATP-binding protein	1.28	<0.001
CT43_RS25990	iron ABC transporter permease	1.95	<0.001
CT43_RS25995	iron ABC transporter permease	1.97	<0.001
CT43_RS26000	iron ABC transporter substrate-binding protein	2.67	<0.001
CT43_RS27330	ferrichrome ABC transporter substrate-binding protein	3.35	<0.001
CT43_RS27335	iron-dicitrate ABC transporter ATP-binding protein	1.45	<0.001
CT43_RS27340	iron ABC transporter permease	1.01	<0.001
CT43_RS27345	ferrichrome ABC transporter permease	1.78	<0.001

Table 5: Elemental analysis of the two media lots and water sources.

Media sample	**Total Fe (ppm)**	**Fe2+ (ppm)**
Lot # 1091744 1520	0.15 ± 0.01	0.02 ± 0.02
Lot # 7220443 1060	0.30 ± 0.01	0.07 ± 0.02
Water Source (Building)		
1520	0.01 ± 0.01	
1060	0.01 ± 0.00	

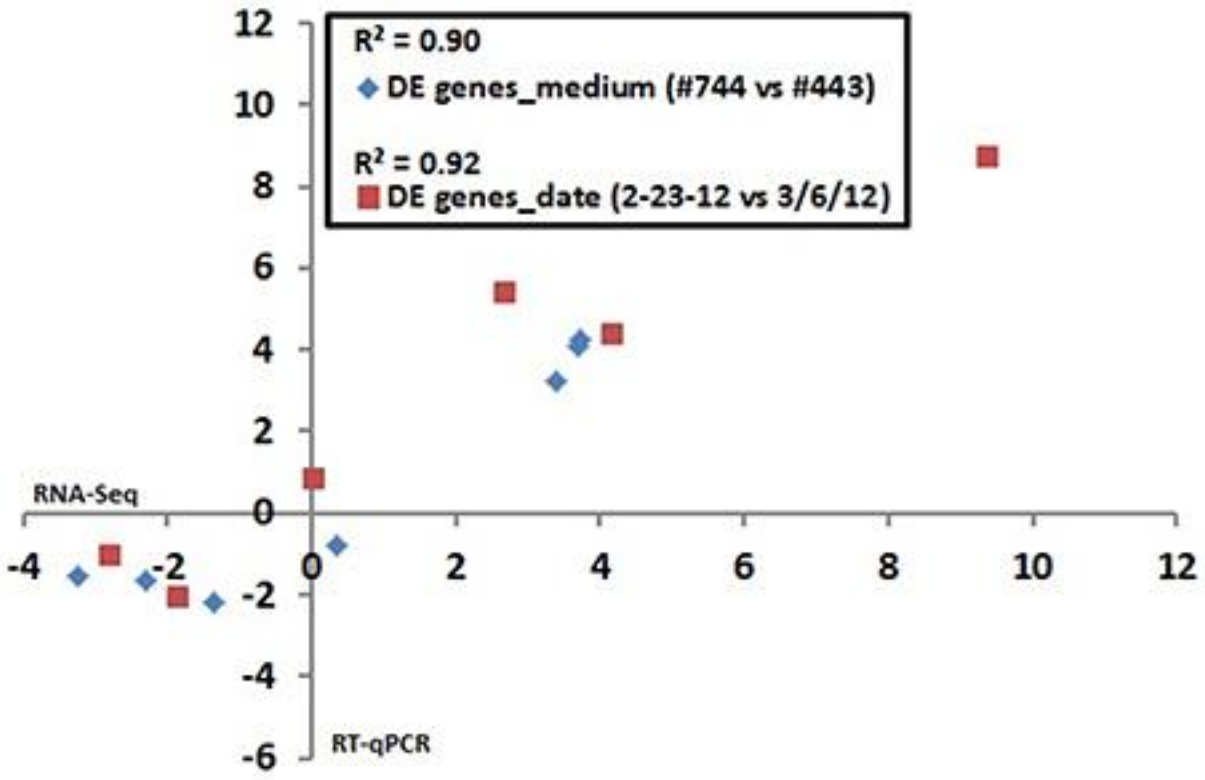


Figure 7: RNA seq data validation: Correlation between RNA seq and RT qPCR results for differential gene expression in Bt strain ATCC10792 grown on different medium lots (#744–443) and dates (2/23/12–3/6/12). The log2 transformed expression ratio values from RNA seq (x-axis) and RT qPCR (y-axis) were plotted against each other and correlation coefficient (R2)-values were calculated. Seven genes plotted for medium effect: BTHUR0008_RS06920, BTHUR0008_RS03645, BTHUR0008_RS15085, BTHUR0008_RS17455, BTHUR0008_RS20850, BTHUR0008_RS17460 and BTHUR0008_RS19345. Six genes plotted for date effect in samples from medium Lot #744 (2/23/12 vs. 3/6/12): BTHUR0008_RS30620, BTHUR0008_RS19140, BTHUR0008_RS01820, BTHUR0008_RS21040, BTHUR0008_RS26070 and BTHUR0008_RS08955.

2.3.4 Effect of Reduced Number of Replicates and Reads on Differential Gene Expression Detection

To investigate the number of reads and replicates required to detect differentially expressed genes with confidence, the number of reads as well as replicates were varied, and the subsequent outcome on differential gene detection was determined. Based on knowledge from the literature as well as a realistic range for number of replicates in any biological study considering time, money and sample availability, here we focused on two to four replicates.

A set of four replicates from each medium lot and any one of the two culture dates with 100% of the trimmed reads (~12–58 M) was analyzed (Table 6). A total of 887 genes were detected as significantly differentially expressed upon analyzing the differential gene expression between medium lots for ATCC10792 at thresholds of two-fold differential gene expression and FDR ≤ 0.05. Analyses that included sets of three and two replicates led to the detection of 885 and 720 differentially expressed genes, respectively (Table 6). The significantly differentially expressed genes with three replicates had 743 and two replicates had 607 genes common with the results from the four replicates analysis (Figure 8A). To examine how fewer biological replicate numbers affected differential gene expression results for an experiment containing a modest number of reads, the 25% read dataset was selected for further analysis. The 25% read dataset was created by randomly removing 75% of the total reads that had been filtered and trimmed for quality (100%; ~12–58 M reads). The sets of replicates and their reduced (25%) read coverages were

Table 6: Effect of decreasing number of replicates on significantly differentially expressed genes while maintaining 100 and 25% of the reads.

Number of replicates (ATCC10792)*	**Differentially expressed genes (FDR 5%, two-fold)**		**Genes commonly detected with all four reps**	
	100%Reads	**25%Reads**	**100%Reads**	**25%Reads**
4 (1A,2A,3A,4A)/ (9A,10A,11A,12A)	887	696	100% (887)	100% (696)
3 (1A, 3A, 4A)/ (9A, 11A, 12A)	885	689	83.5% (741)	84.9% (591)
2 (1A, 4A)/ (9A,12A)	720	501	68.4% (607)	59.3% (413)

*All combinations of available replicates were tested. Results for replicates with most similar read numbers are shown

1A, 2A, 3A, 4A (~3–13 M) vs. 9A, 10A, 11A, 12A (~3–6 M) for analysis with all four replicates; 1A, 3A, 4A (~7–13 M) vs. 9A, 11A, 12A (~5–6 M) for three replicates and 1A, 4A (~7–13 M) vs. 9A, 12A (~5–6 M) two replicates, which gave 696, 689, and 501 significantly differentially expressed genes, respectively (Table 6). There were 591 genes detected with three replicates and 413 genes detected with two replicates that were in common with the set of four replicates when the 25% subset of reads was analyzed for all (Figure 8B). Four out of the seven RT-qPCR validated genes (BTHUR0008_RS03645, BTHUR0008_RS17455, BTHUR0008_RS20850, and BTHUR0008_RS17460) were among the genes considered significant for all conditions in 25% read dataset analysis. Moreover, the same four genes were also considered significant for analyses that included four, three, and two replicates with 100% of the available reads; as well as for analyses that contained 5–100% reads and compared four replicates (Figures 7, 8). Selection of an appropriate sequencing depth or genome coverage is a concern in the field, which impacts sensitivity, detection of weakly expressed genes as well as considerations such as cost and replicate numbers [131, 141, 152]. The outcome of reducing read numbers on the detection of differentially expressed genes was examined. in this study. The initial quality trimmed and filtered reads for strain ATCC10792 (~12–58 M/sample, Table 3a) are referred to as 100% of the reads, which were randomly subsampled to generate input files with 75, 50, 25, 10, and 5% (Data sheet 8, Appendix) of the total available reads for a four replicate differential gene expression analysis of cells grown in different media lots. A trend of fewer genes being considered as differentially expressed was observed as fewer reads were incorporated into the analysis (Table 7) and there was a core of 328 DE genes regardless of differing read levels (Figure

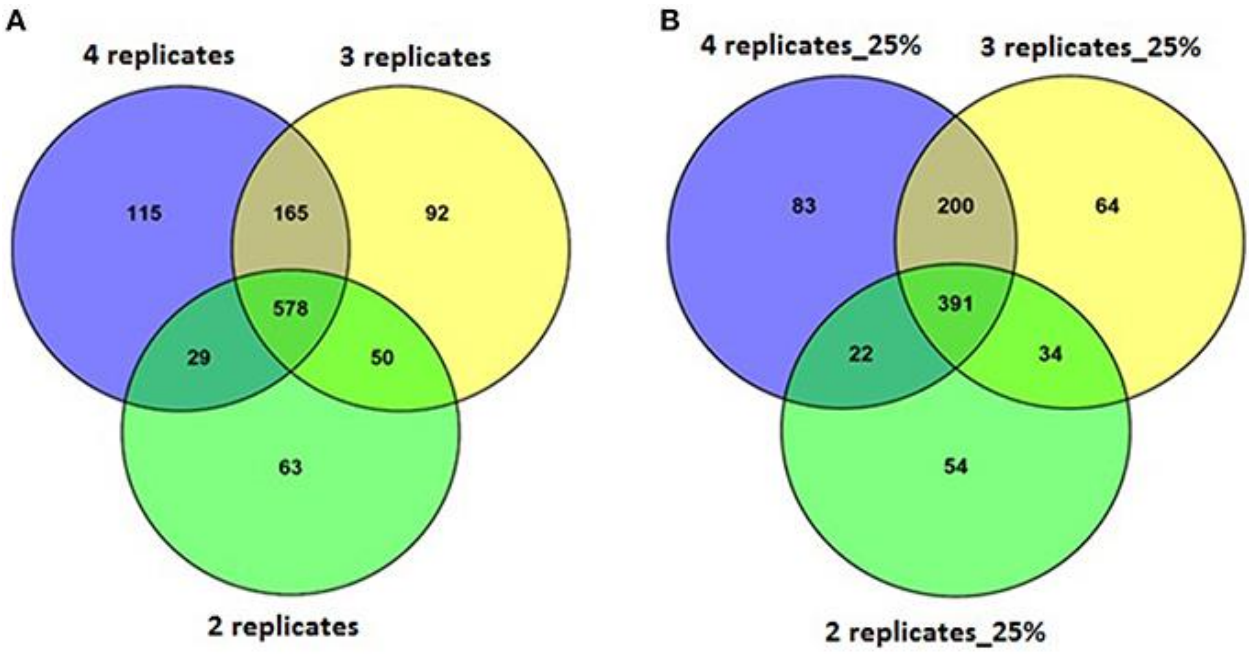


Figure 8: Venn analysis of DE genes detected with varying replicate numbers. (A) Venn diagram depicting the effect of 2–4 replicates while maintaining 100% of the reads, on significantly differentially expressed genes and the genes commonly detected within sets of varying replicates for strain ATCC10792. (B) Venn diagram for differentially expressed genes detected with 25% reads (~5–10 M) and 2–4 replicates for strain ATCC10792.

9). Along with the genes that were commonly detected as significantly differentially expressed between datasets of varying number of biological replicates and/or numbers of reads, the genes that were exclusively detected within each of these datasets was also important. Between datasets of four and three replicates using 100% of the reads, 144 genes were exclusively detected as differentially expressed within the four replicates dataset (Figure 8A). When these 144 genes were examined in the dataset consisting of three replicates, it was found that ~60% (87 of 144 genes) were not detected in the differentially expressed gene list as they fell below the two-fold expression threshold. Approximately 24% (34 of 144 genes) dropped below the FDR threshold due to minor deviations from the set limits. For example, one of the genes “BTHUR0008_RS07395” had a difference of 0.04 between the log 2-fold values from four and three replicates and thus fell below the fold-change cut-off with a difference of 0.02; yet another gene “BTHUR0008_RS29930” had an adjusted p-value difference of –0.04 between values from four and three replicates and dropped out at the FDR set threshold with a difference of 0.01. The remaining 16% (23 of 144 genes) did not show up in the differentially expressed genes as their p-adjusted values were set to “NA” by automatic independent filtering based on low mean normalized counts in DESeq2. Similarly, for datasets of four and two replicates (Figure 8A), out of the 280 genes that were exclusively detected as differentially expressed with four replicates, almost half (139/280) of the genes fell below the two-fold threshold, ~11% (32/280) dropped below FDR cut-off and ~40% (109/280) were left out due to independent filtering in DESeq2. Those genes that were exclusive to sets of two and three replicates but were not detected as differentially expressed with four replicates (Figure 8A), were also seen to be left out because of the same reasons as

Table 7: Effect of decreasing number of reads on significantly differentially expressed genes while maintaining all four replicates.

Number of reads (ATCC10792) (%)	**Differentially expressed genes (FDR 5%, two-fold)**	**Genes commonly detected with 100% reads**
100	887	100% (887)
75	843	95% (803)
50	793	89.4% (722)
25	696	78.5% (611)
10	574	64.7% (487)
5	449	50.6% (376)

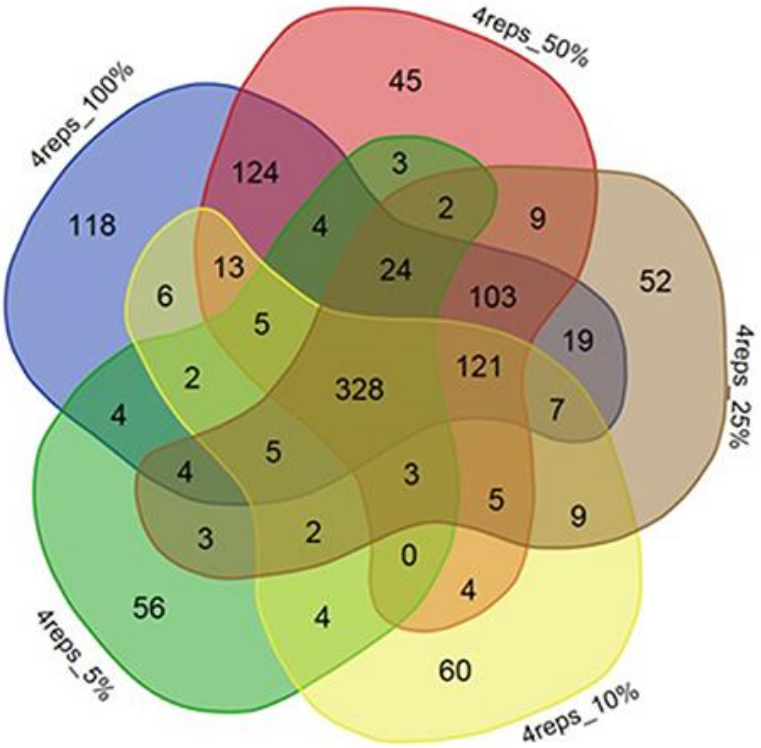


Figure 9: Venn diagram depicting effect of reducing number of reads on DE gene numbers. The effect of decreasing of read numbers on significantly differentially expressed genes and the number of genes commonly detected within sets of 100, 50, 25, 10, 5 % reads for strain ATCC10792.

mentioned above, which led to their not meeting the significant differential expression criteria. A similar trend was also observed for the genes exclusively detected to datasets of four, three and two replicates with 25% of the reads (Figure 8B) as well as for such genes between datasets maintaining all four replicates but reads varying from 5 to 100% (Figure 8).

2.4 Discussion

In this study, we analyzed large transcriptomic datasets from *B. thuringiensis* ATCC10792 and CT43 using negative binomial distribution in DESeq2 for the assessment of differential gene expression. Analysis was performed in different combinations of the data sets to better understand the major challenges of experimental design, variation, required number of replicates and adequate sequencing depth associated with RNA-seq data analysis. Based on our results, as well as previously reported comparisons, we outline important considerations and provide design recommendations for cost-effective RNA-seq results with sufficient statistical power. Improving NGS technologies and instrumentation has led to reproducible results with little technical variation [163] and the preference of the field has shifted toward biological replicates over technical replicates just as array-based transcriptomics evolved earlier [141]. Biological replication is important as without estimating the variability within a group it is not possible to estimate true differences between the groups under observation and conclusions from such results cannot be generalized [136, 164]. In transcriptomic studies one aims for an adequate

trade-off between the number of replicates as well as reads such that it is cost-effective and provides sufficient statistical confidence for interpretation [141, 152, 165].

2.4.1 Recommended Number of Replicates and Reads

Higher numbers of biological replicates provide a better representation of biological variance across samples for transcriptomic analysis [141, 166, 167]. It also reduces chances of any “bad replicates” skewing results or adding unwanted variation [151]. However, it is not practical nor is it always possible to have very large number of replicates for each condition in biological studies due to time, sample availability and cost constraints and the addition of replicates beyond a certain level has diminishing returns and will depend on the nature of the study's goal [152]. Our results are similar to previous studies that show a greater proportional increase in the number of differentially expressed genes when moving from two to three replicates compared to from three to four. Similar trends of decreasing DE genes were observed when replicates and reads were removed from analysis. Having read numbers in the same range across samples is a consideration. Similarly, higher genome coverages with large number of reads have also been shown to result in higher DE gene detections but within limits [168]. It has been reported that with reads below 5 M a considerable drop in DE gene numbers is observed, but increasing reads beyond 10 M results in only slight increments in detection and adding replicates at this point has a much higher effect instead [152, 168]. Our analyses with sets of randomly sampled 5–75% of the total available reads while maintaining four replicates depicted a similar trend in consensus with previous findings. In addition, here we performed a combined analysis with reduced reads (25% subset) and reducing replicate numbers

[130-132]. The results depicted that the set of three replicates and 25% reads (~7.5–13 M) detected a similar number of DE genes as four replicates and 25% reads with most of the genes commonly detected between the two sets. Moreover, the number of DE genes detected with three and four replicates with 100% reads was not much greater than when 25% reads with the most similar numbers were used. Thus, based on all the above it is recommended that designing an experiment with as low as three good replicates (high correlation coefficients, indicating reproducibility across samples) and at least 5–10 million reads per sample would be the most efficient and cost-effective design for a microbial transcriptomics study. It is not just the “number” of DE genes detected with increasing number of replicates and reads, but also whether the genes detected with a set of fewer replicates and/or reads forms a subset of the genes detected with the set of higher number of replicates and reads or if they are “newly” identified genes. We performed an overlap assessment for the DE genes detected from all combinations of analysis to look at core as well as exclusive subsets of genes and used RT-qPCR for validating a representative few. The major reasons for exclusively detected genes were related to FDR and p-value cut-offs and filtering out of the lowly expressed genes via the independent filtering in DESeq2. Genes with expression and p-values very close to the cut-off range moved above and below the limits with changes in the numbers of reads and replicates. Due to these reasons such genes varied between data subsets and were therefore not found among the commonly detected core genes. Thus, by applying well considered cut-offs tailored to the aim of the study as well as the expected gene expression profile one can potentially avoid missing out on genes relevant to the analyses. This signifies the importance of choosing appropriate threshold limits set for

each study and examining the broader context of expression changes by considering pathway and operons as two examples. Deeper sequencing coverage (more replicates and reads) may still be required for studies that are more discovery-based than hypothesis-driven [161, 169]. A higher coverage may also be desirable to deal with sequencing errors and polymorphisms larger than single base differences [131, 139] or to detect a rare transcript or variant or lowly expressed genes.

2.4.2 Differentially Expressed Genes and Experimental Design Considerations

Most typical laboratory experiments would control for much of the variation identified in this study by using media prepared from all the same stock, by using defined media, and/or by randomizing treatment/control samples if they had to be cultured on different dates. However, there may be circumstances such as long term continuous growth studies where rich media cannot be prepared from the same lot or batch, in which case being able to build such variables into statistical, or other, tests is an important consideration. RNA-seq analysis identified significantly differentially expressed genes based on medium lot and culture date differences, two major variation sources in the dataset. The highly differentially expressed iron acquisition and metabolism genes, were a likely consequence of differing amounts of iron in the two media lots. Indeed, in this study RNA-seq was a tool for predictive biology since we hypothesized and confirmed the two LB medium lots had different iron contents. The large (~two-fold) difference in measured iron contents was surprising and may be of broader interest to the research communities that use this media. Our results are in agreement with earlier studies

showing the importance of culture medium choices in transcriptomics [170]. In this study, other experimental factors beyond the type of media used were shown to be important.

2.4.3 Significance of the Available Data Set

This dataset has relevance to researchers interested in Bacillus biology. The genus Bacillus contains representatives such *B. thuringiensis* (e.g., BMB171, Bt407), *Bacillus subtilis* (e.g., BSn5), *Bacillus anthracis* (e.g., Ames), and others [171-173] that occupy diverse ecological niches and have important biotechnological roles. Moreover, this is a rich dataset with four biological replicates and high genome coverage (85–465X), which may interest researchers in testing, developing, and evaluating bioinformatics software for RNA-seq analyses in future for example in testing/developing normalization algorithms and mapping tools etc. [138, 174, 175]. This data set could also be utilized toward generating and testing new globally acceptable RNA-seq analysis pipelines such as the recently developed “PANDORA” [176], which could then permit further comparisons and developments and improving existing RNA-seq analysis pipelines.

In conclusion this study outlines the significance of a well-controlled experimental design, choice of threshold parameters, adequate number of reads and replicates toward an efficient and cost-effective transcriptomics study. Moreover, the depth and complexity of this RNA-seq data will be useful to others for a range of studies such as for insights into Bacillus physiology and for further developments in the field of bioinformatics for microbial transcriptomics.

Chapter 3.0

Improved *Caldicellulosiruptor bescii* growth and substrate utilization by alleviating membrane potential limitations.

3.1 Introduction

The need for renewable and sustainable transportation biofuels is increasing and biotechnology offers potential solutions to meet the future demand [6]. *Caldicellulosiruptor* species are thermophilic anaerobes being studied for production of biofuels and are recognized for their potential in plant biomass deconstruction [29]. *Caldicellulosiruptor bescii* can utilize a wide range of substrates including cellulose, hemicellulose and lignocellulosic plant biomass as well as co-metabolize both C_5 and C_6 sugars [6]. Major fermentation end products include lactate, acetate, CO_2 and hydrogen [124], and although *Caldicellulosiruptor* species natively only produces trace amounts of ethanol some recently engineered strains can produce increased yields of ethanol [6, 29, 77]. An ideal biocatalyst must be able to utilize the vast majority of the available substrate to achieve high product yields and titers. Currently alongside other challenges, incomplete biomass utilization and low *C. bescii* cell growth are important areas for improved understanding and for strain development [2, 177].

Toxic compounds released from biomass pretreatment, microbial biomass deconstruction or released as products of metabolism have the potential to inhibit productivity. Potential inhibitory compounds include a wide range of compounds such as sugar degradation products like furfural and hydroxymethylfurfural (HMF), lignin degradation products such

as substituted phenolics, vanillin and lignin monomers and weak acids such as acetic, lactic, formic, and levulinic acids [178] [179]. Osmolarity and pH are two further conditions prone to change during the fermentation and are critical for continued metabolic activity. High solute concentrations including sugars, salts, ionic moieties, derivatives of inorganic salts or organic acids can affect the osmolarity of the growth medium. Long lag phases and inhibited growth are indicative of stress and have been reported for *C. bescii* at ~160mM total organic acids and ~500 mOsm of osmolarity [6, 77]. Similar reports are also available for other thermophilic organisms, like severe growth inhibition of C. saccharolyticus at 200mM acetate concentrations [11, 12].

However, members of the *Caldicellulosiruptor* family including *C. bescii* have been observed to grow poorly reaching low cell densities (OD 680nm ~ 0.1-0.3 in unbuffered medium) previously [112, 177, 180], in the absence of most of the above-mentioned inhibitory conditions that are now under consideration with respect to biofuel production. Improving the cell yields to increase their productivity has been one the many nuances under their research[181]. *C. bescii's* growth cessation occurs with 1-2g/L of soluble sugar still available and with no biomass derived inhibitors present while growing on soluble sugars (e.g. 5g/L of maltose) in a low osmolarity medium (LOD), or while growing under controlled conditions, in gas sparged fermenters without hydrogen accumulation. Moreover, in these cases neither total organic acids nor the osmolarity reaches inhibitory concentrations of ~160mM and 500 mOsm respectively, yet growth ceases.

To discover the underlying factor(s) contributing to the inhibition at this point, it bodes well to start at the basic growth determining attributes that could be easily overlooked while focusing on higher level inhibitory/limiting factors. To this end, there are indications from

the literature that the limitation experienced at this preliminary stage is related to pH. The growth pH currently reported for *C. bescii* is a narrow range between pH 6.8-7.3 [13] and with the accumulation of acidic byproducts along fermentation the medium pH declines and so does the growth rate [14]. An important observation though, is that the final medium pH when measured/reported at fermentation end-point is always ~5.0, irrespective of the substrate [1, 13]. Another result strongly supporting the above-mentioned rationale comes from a study where the addition of a buffer (40 mM, MOPS) to the medium resulted in almost twofold increase in *C. bescii's* growth [13] indicating the significance of pH in relation to cell yield.

Accumulating organic acids and derivatives during fermentation not only affect the medium/ intracellular pH but also the membrane potential of the cell, ultimately influencing the operating proton motive force (PMF). Maintaining an optimal PMF is an important factor for an organism's growth and productivity [15] [16] but has generally remained neglected in studies involving anaerobes. The recent interest in anaerobic fermentative bacteria for industrial biofuel production, now calls for better understanding to find ways for increasing product yields and titers. Here, we studied pH and PMF in detail for the first time in a member of *Caldicellulosiruptor* spp. to demonstrate that membrane potential limitation hampers growth, product yields as well as substrate utilization ability in *C. bescii* and alleviation dramatically improves its overall productivity.

3.2 Materials and Methods

3.2.1 Strains, media and growth conditions

Experiments were performed using *Caldicellulosiruptor bescii* strains DSM6725 (wild-type), JWCB049 (*ΔpyrFA Δldh CIS1*:: $P_{S\text{-}layer}$Teth39_0206), JWCB038 (Δ*pyrFA* Δ*ldh CIS1*::$P_{S\text{-}layer}$ *Cthe-adhE ΔhypADFCDE*/(*ura*−/5-FOA^R)) and JWCB036 (Δ*pyrFA* Δ*ldh CIS1*::$P_{S\text{-}layer}$ *Cthe-adhE*//(*ura*−/5-FOA^R)) parent strain to JWCB038 [13][182]. Cells were grown anaerobically at 75 °C in liquid LOD medium [14] with 5 g/L maltose (catalog no. M5895, Sigma) as the carbon source, unless otherwise noted. The pH varied as per experiment and ranged from 5.5 - 7.2 as individually specified. Cultures in anaerobic 50 mL serum bottles were grown using a 0.5% inoculum and incubated at 75 °C, with shaking at 150 rpm. In buffered bottle cultures, LOD was supplemented with 100 mM BIS-TRIS (catalog no. B9754, Sigma). Applikon bioreactors (Applikon Biotechnology, Inc.) with total volumes of 1.5 and 5.0 L were used with 1.0-3.0 L working volume respectively. Fermenters were seeded with a 5% inoculum, maintained at 75 °C, 300 rpm (impeller speed) and constantly sparged with nitrogen filtered using 0.2 µm filters (Millipore) attached to the gas inlet port, at low flow rates between 0.5-1 L/min for both reactor volumes.

3.2.2 Growth and product profiling

Growth was estimated by monitoring optical cell density using a spectrophotometer (Thermo Scientific™ GENESYS™ 20), measuring absorbance at 680 nm. For samples from avicel grown cultures, cell enumeration by direct counting using microscopy was performed as previously described [183]. Fermentation products maltose, glucose,

acetate, lactate and ethanol, were monitored and analyzed using high-performance liquid chromatography (HPLC) quantification with an Aminex™ HPX-87H column (Bio-Rad Laboratories Inc.) at 0.5 mL/min flow rate of 5 mm H_2SO_4 mobile phase and column temperature of 60 °C. The signal was quantified on a Shimadzu Prominence LC-20A Series system (Shimadzu Scientific Instruments, Columbia, MD) equipped with a refractive index detector (model RID-20A) at 35 °C. Integrated total peak areas were compared to peak areas and retention times of known standards for each compound of interest.

3.2.3 Integrated omics sampling, measurements and analysis

Systems biology samples were collected from fermenters using 60mL syringes attached to the sampling port, centrifuged at 9000 rpm for 5min in a benchtop centrifuge (Thermo Scientific™ Sorvall™ Legend™ XTR) and the pellets were flash frozen in liquid nitrogen before placing them in -80 °C until further use. A time course sampling at 5, 10, 30, 60 and 90 min post-acid addition was done as described above from the batch pH-controlled fermenters, where the pH was lowered at mid-log from 7.2 to 6.0 to study response medium acidification using integrated omics analysis.

Transcriptomics: Samples from time points, 5-90 min post pH decrease from the batch fermenters were used for RNA-seq. High quality RNA was isolated from the wild-type strain DSM6725 growing in the control and treated fermenters sampled for omics, described as follows. Briefly, cell pellets from each sample were incubated in 250 µL of 20 mg/mL lysozyme (Sigma Aldrich part number L-7651, St. Louis, MO) resuspended in SET buffer (50 mM Tris-HCl pH 8.0 50 mM EDTA, 20% w/v Sucrose) followed by

incubation at 37ºC for eight minutes, vertexing briefly every two minutes. After incubation, 900 µL of buffer RLT from a Qiagen RNeasy Kit (Qiagen, Hilden, Germany) was added to the lysozyme cell mixture and vortexed. Then 650 µL of 100% ethanol was added to each aliquot and mixed by pipetting. Each sample was then applied to a RNeasy spin column (Qiagen, Hilden, Germany) and processed following the manufacturer's protocol included the optional on column Dnase I treatment. Samples were eluted in 35 µL RNAse free water (Qiagen, Hilden, Germany). The quantity and quality of RNA was assessed post-elution using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE) and Agilent Bioanalyzer (Agilent, Santa Clara, CA, USA). Ribosomal RNA was depleted from the samples using a RiboZero rRNA Removal Kit for bacteria (Illumina Inc. San Diego, CA) and cDNA libraries were prepared and barcoded using a ScriptSeq v2 RNA-Seq library preparation kit (Illumina) following the manufacturer's protocols. The final libraries were quantified with a Qubit double stranded broad range assay kit and fluorometer (Invitrogen) and quality assessed using a Bioanalyzer (Agilent). Samples were diluted and pooled according to manufacturer's recommendations. The library was then sequenced using a SR50 sequencing protocol on an Illumina HiSeq 2500 platform (HudsonAlpha Genomic Services Laboratory; Huntsville, AL) using V4 chemistry.

Analysis of the RNA-Seq data obtained was performed as previously described [184]. Reads were mapped to the reference genome, CP001393.1 (NCBI GenBank accession number) [185], using CLC Genomics Workbench (CLC bio, Aarhus, Denmark; version 8.0) as previously described [184] . Mapped reads were analyzed for differential gene expression using DESeq2 [146] and filtering of genes was performed for an expression

fold change ≥ log2 with a 5% FDR. The raw RNA-Seq data was submitted to NCBI Sequence Read Archive (SRA), available under accession number: SRP131057 and gene expression data to NCBI Gene Expression Omnibus (GEO), available under accession number: GSE109442.

Proteomics: Samples from the 60 min time-point were used for mass spectrometry-based proteomics analysis. Frozen samples were thawed and resuspend in SDC (sodium deoxycholate) lysis buffer [4% SDC, 10mM DTT in 100 mM ABC (ammonium bicarbonate) buffer pH 7.8]. An aliquot of 500 µL of the resuspended pellet was then boiled at 95ºC for 5min and sonicated. The samples were boiled again at 95ºC for 5min following this sonication step. Iodoacetamide (IAA) was added to the cell lysates to a final concentration of 30 mM using a 500 mM stock and kept in dark for 20 min at room temperature to alkylate cysteines. Samples were then centrifuged at max speed 14000g for 20 min to remove cell debris and sample protein concentrations were measured via bicinchoninic acid assay (Pierce™ BCA Protein Assay Kit, Thermo Scientific™). The supernatant was then added to the 10 kDa separation filter assembly, MWCO spin filter (Vivaspin 500, Sartorius) and centrifuged at max speed 14000g for 20min. The filter was washed with 1mL ABC buffer and the flow through was discarded. The salt concentration in samples was diluted to ~2% SDC by adding 500 µL of ABC buffer. Samples were digested using sequencing-grade trypsin as previously described [186]. Tryptic peptides were then collected by centrifugation (14,000 rpm, 15 min) SDC was precipitated with 1% formic acid, and the precipitates were removed from the peptide solution using water-saturated ethyl acetate. The samples were then concentrated using a Heto DNA Mini

Centrifugal Evaporator (Cole-Parmer) and protein concentrations were measured using BCA assay.

Two-dimensional LC-MS/MS analysis: Peptide samples were analyzed by automated 2D LC-MS/MS analysis using a Vanquish UHPLC with autosampler plumbed directly in-line with a Q Exactive Plus mass spectrometer (Thermo Scientific) outfitted with a triphasic back column (RP-SCX-RP; reversed-phase [5 µm Kinetex C18] and strong-cation exchange [5 µm Luna SCX] chromatographic resins; Phenomenex) coupled to an in-house pulled nanospray emitter packed with 30 cm Kinetex C18 resin. For each sample, 6 µg of peptides were autoloaded, desalted, separated and analyzed across three successive salt cuts of ammonium acetate (35, 50, and 500 mM), each followed by 105 min organic gradient. Eluting peptides were measured and sequenced by data-dependent acquisition on the Q Exactive as previously described [186].

Data analysis: MS/MS spectra were searched with MyriMatch v.2.2 [187] against the *C. bescii* DSM6725 proteome concatenated with common protein contaminants and reversed entries to estimate false-discovery rates (FDR). Peptide spectrum matches (PSM) were required to be fully tryptic with any number of missed cleavages; a static modification of 57.0214 Da on cysteine and a dynamic modification of 15.9949 Da on methionine residues. PSMs were filtered using IDPicker v.3.0 [188] with an experiment-wide FDR controlled at < 1% at the peptide-level. Peptide intensities were assessed by chromatographic area-under-the-curve (label-free quantification option in IDPicker). To remove cases of extreme sequence redundancy, the *C. bescii* proteome was clustered at 90% sequence identity (UCLUST [189]) and peptides intensities summed to their respective protein groups/seeds to estimate overall protein abundance. Protein

abundance distributions were then normalized across samples and missing values imputed to simulate the mass spectrometer's limit of detection. Differentially detected proteins between control and treatment for each time-point were attained from the list of identified proteins and abundances using a cut off threshold of ≥ log2 fold change and 5% FDR.

Metabolomics: Samples from the 90 min time-point were used for mass spectrometry-based metabolomics analysis. Frozen cell pellets were weighed into centrifuge tubes containing 5 mL of 80% ethanol, and 75 µL sorbitol (1 mg/mL) added as an internal standard. Samples were sonicated (Q500 Sonicator, QSonica sonicators) for 3 min (30s on, 30s off with an amplitude of 30%) while being chilled with liquid nitrogen. Samples were then centrifuged at 4500 rpm for 20 min, and the supernatant was decanted into scintillation vials and stored at -20 °C. A 1 mL aliquot was dried under a nitrogen stream, dissolved in 0.5 mL acetonitrile, and silylated to generate trimethylsilyl derivatives, as described elsewhere [190]. After 2 days, 1 µL aliquots were injected into an Agilent 5975C inert XL gas chromatograph-mass spectrometer (GC-MS). The standard quadrupole GC-MS was operated in electron impact (70 eV) ionization mode, targeting 2.5 full-spectrum (50-650 Da) scans per second, as previously described [190]. Metabolite peaks were extracted using a key selected ion, characteristic m/z fragment to minimize integration of co-eluting metabolites. The extracted peaks of known metabolites were scaled back to the total ion current (TIC) using scaling factors previously calculated. Peaks were quantified by area integration and normalized to the quantity of internal standard recovered, amount of sample extracted, derivatized, and injected. A large user-created database and the Wiley Registry 10th Edition/NIST 2014 Mass Spectral Library was used

to identify the metabolites of interest to be quantified. Unidentified metabolites were represented by their retention time and key m/z ratios.

3.2.4 Membrane potential estimation and ATP measurements

Flow cytometry was used to estimate membrane potential in cells sampled from different growth phases and pH using diethyloxacarbocyanine (DiOC2) as previously described [191]. Samples from pH-controlled fed-batch fermenters were extracted and incubated in a Coy anaerobic glove bag (COY, Grass Lake, Mich., USA). Triplicate 1mL samples were incubated with 30 µM DiOC2 for 45 min at 75 °C the dye for each fermenter. Negative controls were incubated at the same temperature with 100 µM CCCP for 30 min followed by incubation with 30 µM $DiOC_2$ 45 min both at 75 °C. All measurements were made using a Cytopeia flow cytometer instrument (BD, Franklin Lakes, NJ, USA). Each set of triplicate technical replicates was measured, and the values were averaged to get the value of the red and green mean fluorescence intensity (MFI) ratios (R/G). Fluorescence values were normalized against the number of cells gated per measurement.

Samples for ATP measurements were collected from three biological replicates of batch, pH-controlled fermenters in the mid-log and stationary growth phases at pH 7.2 and 6.0. ATP was extracted from cell pellets immediately using boiling water (molecular grade) as previously described [192]. The concentration of ATP in the extracts from each sample were measured following the ENLITEN® ATP assay kit protocol from Promega. BioTek Synergy2 luminescence plate reader (BioTek® Instruments, Inc.) was used under the luminescence detection method on the instrument. All samples and standards were

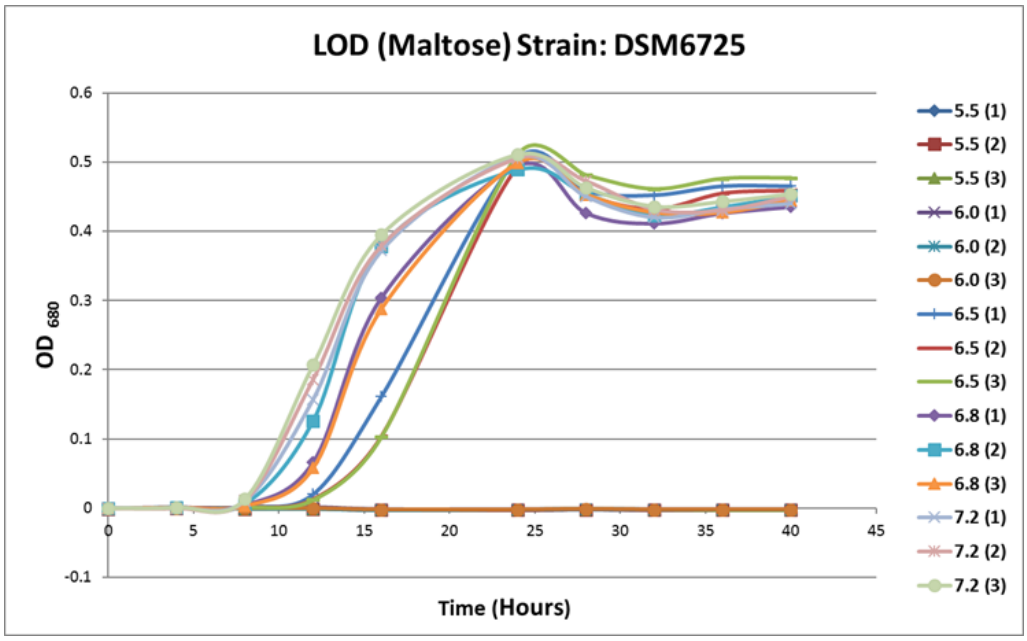
measured in triplicate. The ATP values were normalized to optical density measured or cell dry weight as per representation in results.

3.3 Results

3.3.1 Characterization of *C. bescii* growth patterns at varying pH

C. bescii strains, DSM6725 (wildtype) and JWCB049 (*ΔpyrFA Δldh* CIS1:: PS-layerTeth39_0206) an engineered ethanol producing strain [112] were inoculated from seed culture bottles (starting pH 7.2) and grown in bottles with buffered medium and 5g/L maltose across a pH range of 7.2-5.5 with three biological replicates per pH. Similar growth patterns and final cell densities (OD 680 nm) were observed with both strains, across replicates in the pH range 7.2-6.5. No growth past the starting OD range of ~ 0.0-0.003 (0.2% inoculum), was observed in bottles at pH 6.0 or below for ≥40 hours with both strains and all replicates (Figure 10).

Following this, to further characterize *C. bescii's* growth maintenance ability with declining pH along fermentation, we set up a single replicate solid substrate chemostat. Here the strain JWCB049 was grown on 5 g/L avicel in 1.5 L Applikon fermenters (Figure 11) with a 700mL working volume and the dilution rate was set at 0.1 hr^{-1}. This experiment was designed with the aim of monitoring cell wash-out upon gradually lowering the medium pH at steady-state (cell count over at least 3-4 vessel volume changes/ residence time). The pH was lowered after steady state at each pH level until cells could no longer support or maintain growth for the set dilution rate at a given pH, which triggered a wash-out over time indicated by reducing cell numbers in the system. *C. bescii* displayed the ability to



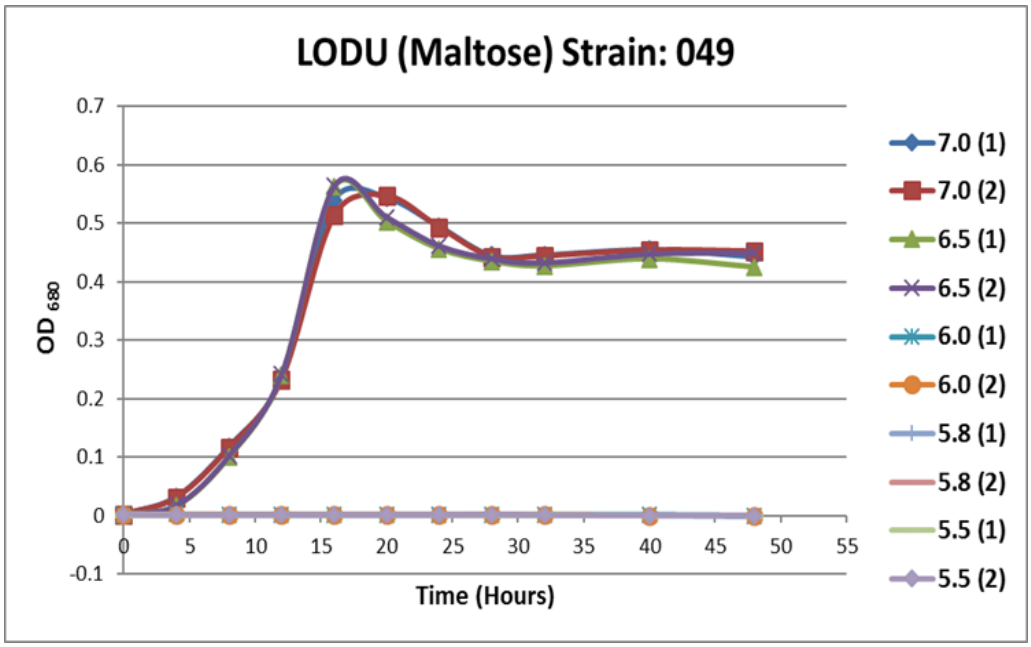


Figure 10: Similar growth patterns observed for *C. bescii* DSM6725 (wild type) and JWCB049 (ΔpyrFA Δldh CIS1:: PS-layerTeth39_0206) in LOD and LODU media (50 mL) respectively with 5g/L maltose at pH 7.2-6.5. No growth seen at pH 6 and below over 40 hours post-inoculation.



Figure 11: Chemostat setup: 1.5L applikon fermenter with working volume of 750 mL.

Media: LODU + Avicel 5 g/L; Strain: JWCB049

maintain growth with a dilution rate of 0.1 hr^{-1} at a pH as low as 5.5, which has not been shown before. The growth and product profile from the chemostat also showed increased growth and product formation at pH 6.5 and 6.0 (Table 8). Cells showed no visible (i.e., morphological) signs of stress at ≥ pH 6.0 whereas elongated cells were seen at pH 5.5 when observed under the microscope.

3.3.2 Batch fermentations to capture physiological and systems level response to medium acidification

To determine the effect of reduced pH on *C. bescii* at a physiological and molecular level, the wild type strain was grown under control and treated conditions in fermenters with 2L (working volume in 5L applikons) LOD medium in triplicates, starting at pH 7.2. In control experiments the pH was maintained at 7.2 whereas in treated experiments, pH was lowered to 6.0 at mid-log phase using 1M sulfuric acid. Following this the fermenters were sampled for growth, products and integrated omics. An improved cell growth phenotype was observed in treated fermenters. The optical densities from the treated fermenters were three times those of the controls and the growth rates were almost double (Figure 12). Moreover, HPLC product profiles of the treated fermenters showed higher products, supporting the observed higher growth. Acetate produced was almost 2.5 times, and lactate was 100 times higher than the controls (Table 9). Also, the treated fermenters almost achieved complete substrate utilization whereas the control fermenters still had ~1g/L of sugars unutilized (Table 9). The possibility of improved growth due to addition of sulfate [193] as a result of using sulfuric acid was not supported as bottle cultivations

Table 8: Cell counts, and product profiles monitored to determine the wash out pH.

pH	Cell count (cells/ mL)	Acetate (g/L)	Ethanol (g/L)
7.00	4.49E+06	0.74	0.04
7.00	4.75E+06	0.72	0.05
6.50	5.00E+06	0.83	0.10
6.50	5.60E+06	0.82	0.11
6.00	3.11E+06	0.85	0.06
6.00	3.65E+06	0.90	0.05
5.80	3.94E+06	0.58	0.06
5.80	4.03E+06	0.64	0.03
5.50	3.22E+06	0.43	0.01
5.50	3.12E+06	0.43	0.02
5.20	5.05E+04	BDL	BDL
5.20	4.21E+04		

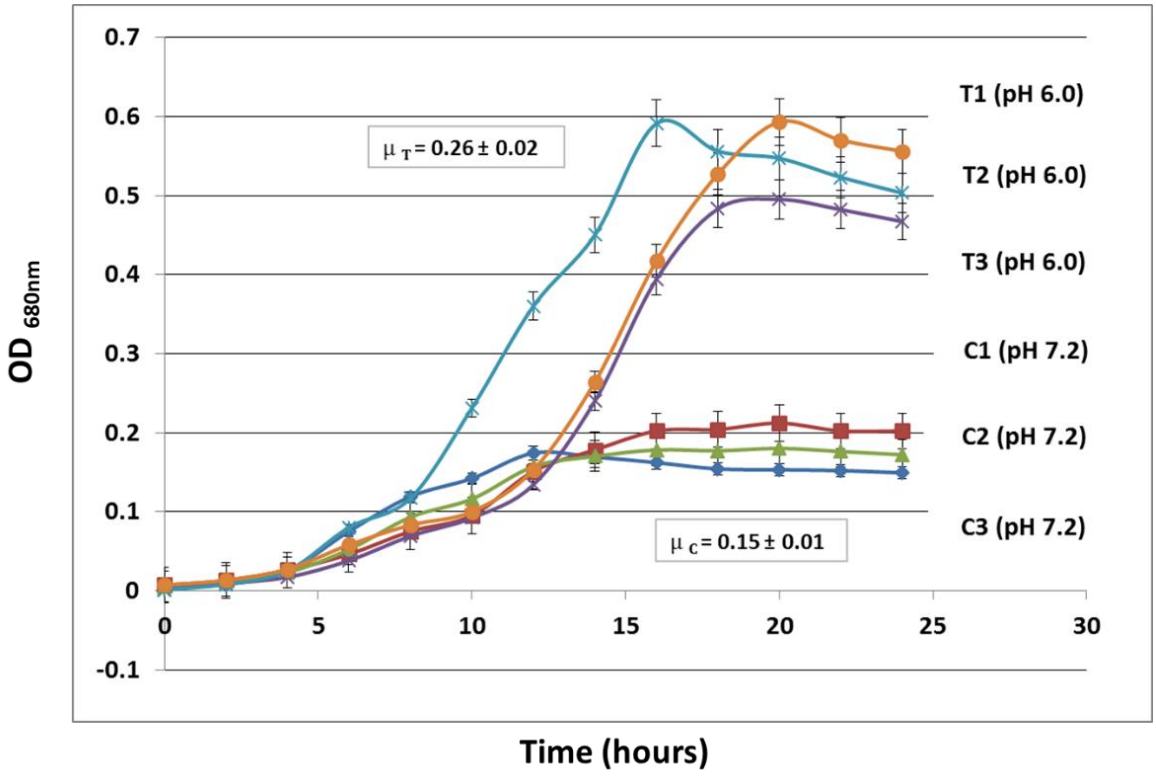


Figure 12: Increased cell growth observed in treated batch fermenters (pH lowered to 6.0 at mid-log phase): T1, T2, T3 as compared to controls: C1, C2, C3 maintained at pH 7.2.

Table 9: Substrate and product profiles (HPLC) from treated and control fermenter samples validate observed growth differences.

Samples	Maltose (g/L)	Glucose (g/L)	Acetate (g/L)	Lactate (g/L)
C1 (pH 7.2)	0.89	0.75	0.01	0.52
C2 (pH 7.2)	0.88	0.76	0.01	0.65
C3 (pH 7.2)	0.98	0.71	0.01	0.49
T1 (pH 6.0)	0.22	0.71	0.59	1.22
T2 (pH 6.0)	0.10	0.77	0.81	1.29
T3 (pH 6.0)	0.05	0.57	0.59	1.29

with and without added 5mM Na_2SO_4 (the equivalent to the amount that went into the treated fermenters) showed similar growth (Figure 13)

Omics Summary: The time-course omics samples collected from these fermenters were used for integrated omics profiling that included transcriptomics, metabolomics and proteomics for all time-points, 60 min and 90 min samples respectively. The transcriptomic profiles were clustered to analyze co-expression patterns, pre or post-acid addition (Figure 14). In the transcriptomics data a total of 115, 332, 372, 404 and 361 genes were detected to be differentially expressed (threshold: $p \leq 0.05$; log2 fold change ≥ 1) at 5, 10, 30, 60 and 90 min respectively. Two top most overexpressed clusters post-acid treatment included a sugar ABC transporter operon (*ATHE_RS0515-0525*: annotated as xylose ABC transporter) and a cluster of hypothetical genes (*Athe_RS09865-09885*) with log2 fold expression values > 5. The hypothetical genes in the cluster were found to be co-located in the genome and seemed to be a part of a single transcriptional unit/operon. A preliminary BLAST analysis revealed homology and conserved domains in three of five genes from this hypothetical genes cluster to the PQQ (Pyrroloquinoline quinone) redox cofactor biosynthesis pathway. This cluster also depicted structural similarity to the PQQ biosynthesis pathway operon. Moreover, upon deeper metabolomic analysis, of peaks below the threshold, a peak at the exact m/z for the PQQ standard peak was detected when run with the PQQ standard. The presence of this cofactor has not yet been reported in the genus *Caldicellulosiruptor*. The transcriptomics data also overall indicated several genetic elements indicating various possible mechanisms that may be involved and employed in pH homeostasis by *C. bescii*. For example: differentially expressed (DE) genes such as ATPases and ABC transporters

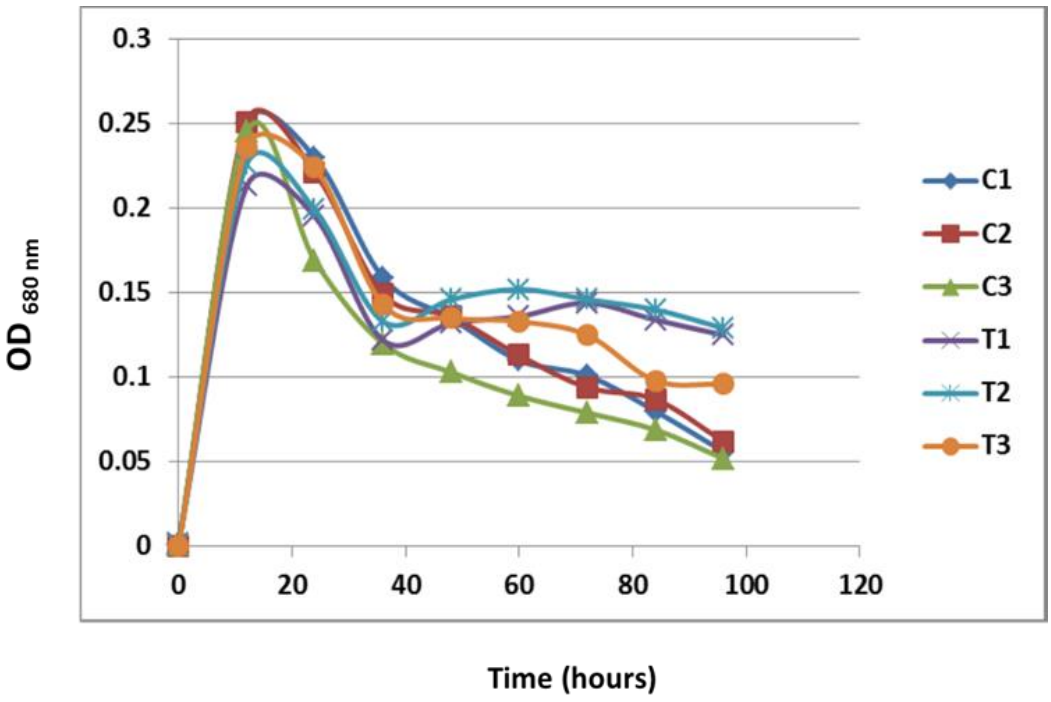


Figure 13: Sulfate addition showed minimal effect on cell growth over 96 hrs. in LOD + 5g/L maltose (controls– C1, C2, C3) compared to LOD + 5g/L maltose + 5mM Na_2SO_4 (treated - T1, T2, T3).

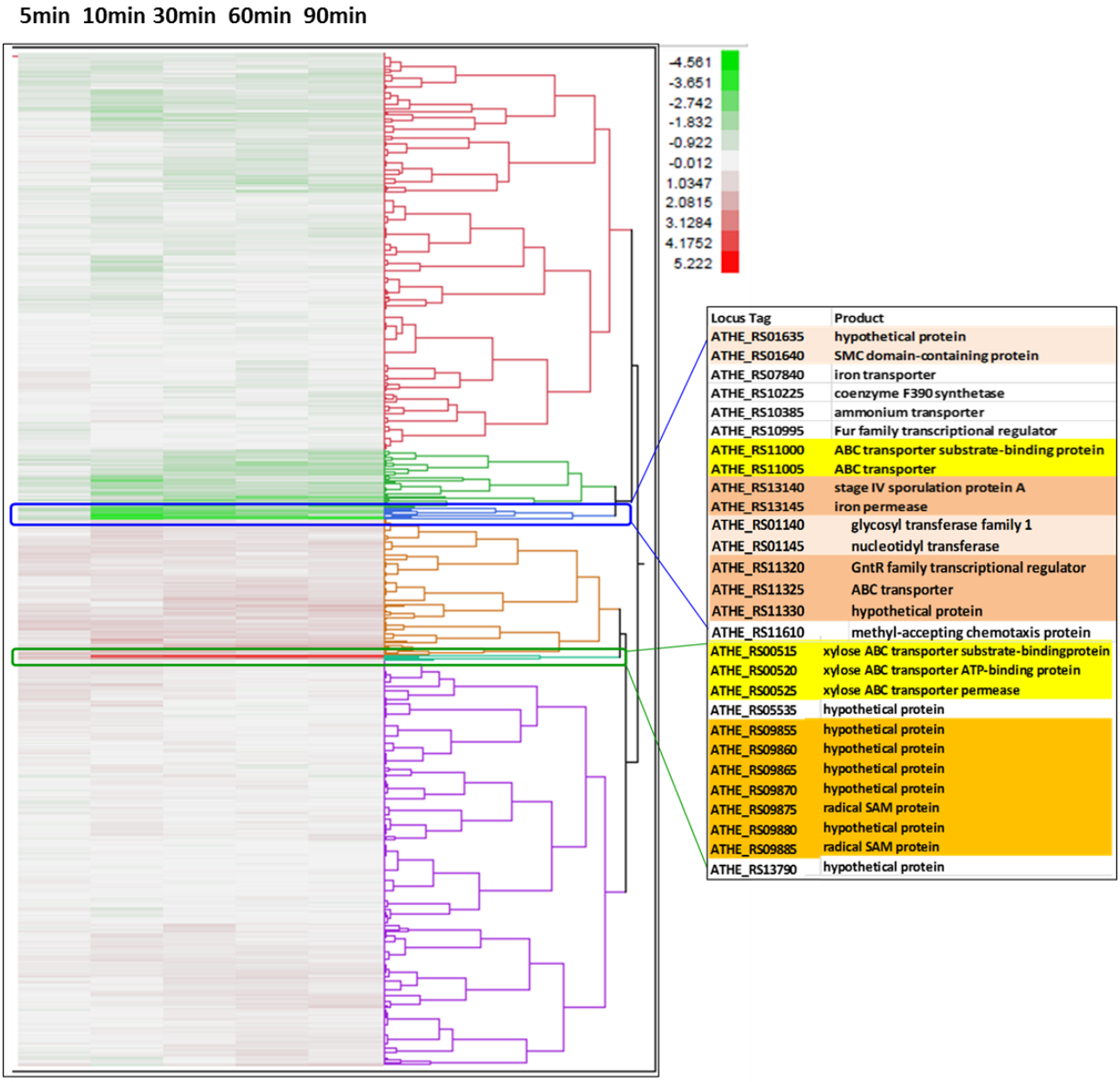


Figure 14: Heat map depicting hierarchical clustering of genes differentially expressed post-acid addition across time-points sampled (5, 10, 30, 60. 90 min). Clusters marked 1 and 2 have interesting expression patterns

indicate active proton transport; decarboxylases, and dehydrogenases suggest re-routing of metabolic fluxes; regulated multidrug transporters, lipid metabolism genes and chaperons were observed which may be involved in passive pH homeostasis mechanisms and transcriptional regulators of the *LacI/ GntR/ DeoR* family may control expression of genes that participate in pH sensing or signaling elements. Another interesting aspect revealed from this data, was the downregulated stress related genes (e.g. iron metabolism genes), suggesting cells were not eliciting the general stress responses and were thus probably not under stress at this pH.

Using mass spectrometry-based proteomics more than a total of 1650 proteins were detected. Upon data analysis using $p \leq 0.05$ and log2 fold change ≥ 1 as the threshold, 36 proteins were differentially detected (Table 10) between treated and controls. Proteins belonging to the highly DE genes observed in the transcriptomics data such as sugar ABC transporter operon, were also significantly differentially detected in the proteomics data.

The metabolomics data analysis revealed 62 differential metabolites (Table 11). Two aspects from this data that were majorly of interest were, the detection of the PQQ m/z peak and the significantly differentially detected high inorganic phosphates in the samples from the treated fermenters. The presence of inorganic phosphates suggested a response involving the phosphate species in the system which indicated the possibility of changes in energy metabolism.

Table 10: Proteins significantly differentially detected in treated batch fermenters at 60 min post acid addition and correlating gene expression changes from RNA-seq.

Proteins	Locus Tag	p value	Log2Fold (Proteomics)	Log2Fold (RNAseq 60min)
Radical SAM domain protein	ATHE_RS09885	0.00107	4.22	3.14
D-xylose ABC transporter, periplasmic substrate-binding protein	ATHE_RS00515	0.00020	2.63	5.22
Integrase family protein	ATHE_RS10175	0.00642	2.57	1.57
Hypothetical protein	ATHE_RS09880	0.00022	2.46	2.78
Hypothetical protein	ATHE_RS12415	0.00346	2.35	0.56
Hydrogenase assembly chaperone hypC/hupF	ATHE_RS05460	0.01738	2.18	-0.09
Extracellular solute-binding protein family 1	ATHE_RS04125	0.00012	2.13	0.26
Hypothetical protein	ATHE_RS05535	0.00134	2.10	3.06
Glucosylceramidase	ATHE_RS00475	0.00220	1.95	0.76
Phosphoenolpyruvate carboxykinase (GTP)	ATHE_RS01910	0.00216	1.89	2.22
D12 class N6 adenine-specific DNA methyltransferase	ATHE_RS12280	0.02314	1.84	1.03
ABC transporter related	ATHE_RS04135	0.01044	1.77	0.61
Hypothetical protein	ATHE_RS09870	0.01567	1.71	3.53
Iron-containing alcohol dehydrogenase	ATHE_RS11300	0.01044	1.68	0.96
Radical SAM domain protein	ATHE_RS09875	0.03568	1.66	3.05
Molybdenum cofactor biosynthesis protein C	ATHE_RS04155	0.00527	1.60	1.67
Hypothetical protein	ATHE_RS00790	0.00505	1.56	0.75
Hemerythrin-like metal-binding protein	ATHE_RS12945	0.01322	1.54	0.06
Rubredoxin-type Fe(Cys)4 protein	ATHE_RS07195	0.00472	1.32	-0.05
ABC transporter related	ATHE_RS00520	0.00546	1.31	4.49

Table 10: Continued

Proteins	**Locus Tag**	**p value**	**Log2Fold (Proteomics)**	**Log2Fold (RNAseq 60min)**
Hypothetical protein	ATHE_RS01090	0.01817	1.29	0.93
Molybdenum cofactor synthesis domain protein	ATHE_RS04145	0.00445	1.29	1.46
S-layer domain protein	ATHE_RS00380	0.02009	1.26	-0.14
Hypothetical protein	ATHE_RS04575	0.00004	1.08	0.47
Transcriptional regulator, ArsR family	ATHE_RS01535	0.00157	1.07	-1.28
Prephenate dehydrogenase	ATHE_RS04225	0.01596	1.07	1.28
aldo/keto reductase	ATHE_RS02295	0.04898	1.06	-0.41
Radical SAM domain protein	ATHE_RS10975	0.00718	1.06	1.95
Two component transcriptional regulator, AraC family	ATHE_RS04270	0.01469	-1.08	0.51
Hypothetical protein	ATHE_RS01085	0.03198	-1.12	-2.06
Transcriptional regulator, GntR family	ATHE_RS11320	0.02146	-1.14	-4.09
Protein of unknown function DUF464	ATHE_RS05600	0.02262	-1.16	0.52
FeS cluster assembly scaffold protein NifU	ATHE_RS08605	0.00681	-1.28	-0.95
Hypothetical protein	ATHE_RS14020	0.02166	-1.51	-0.52
Hypothetical protein	ATHE_RS08120	0.01066	-1.57	-0.45
Flagellar FlbD family protein	ATHE_RS10865	0.04119	-1.57	-0.20

Table 11: Metabolites differentially detected in treated batch fermenters at 90 min post acid addition.

Metabolite	Acid Treated vs Control	p-value
[Retention time (min); Key m/z]	Fold Change	
8.96 102 75 85 158 204 219	3.34	0.000
uracil	2.99	0.098
8.82 102 75 85 158 204 219	2.86	0.004
phosphate	2.50	0.009
heptadecanoic acid	2.41	0.050
9.77 174 276 N-metabolite M+291	2.09	0.140
glycerol-1/3-phosphate	1.71	0.076
nonanoic acid	1.46	0.202
glucose	1.45	0.134
thymine	1.44	0.423
5.95 155 170	1.33	0.289
5.81 155 170	1.33	0.269
13.99 325 dipeptide	1.29	0.309
8.02 259 189	1.29	0.398
azelaic acid	1.27	0.339
tridecane	1.27	0.345
8.60 184 300 285	1.25	0.421
12.12 404 375 321 307	1.25	0.452
12.98 320 307 361	1.24	0.350
11.39 177 248	1.24	0.421
lactic acid	1.22	0.240
12.64 320 307 361	1.21	0.465
3-phosphoglyceric acid	1.21	0.568
palmitic acid	1.19	0.478
mimosine	1.14	0.481
fumaric acid	1.12	0.674
9.83 276 174	1.10	0.899
stearic acid	1.04	0.873
benzoic acid	1.02	0.932
9.72 155 170 243 258	1.02	0.961
3,4-dihydroxyphenylacetic acid	0.99	0.985
galactose	0.98	0.903
lysine	0.96	0.830

Table 11: Continued.

Metabolite [Retention time (min); Key m/z]	Acid Treated vs Control Fold Change	p-value
citramalic acid	0.96	0.922
glycine	0.95	0.709
formamide	0.91	0.798
cysteine	0.91	0.901
malic acid	0.91	0.552
glycerol	0.88	0.640
catechol	0.86	0.675
3,4-dihydroxybenzoic acid	0.86	0.620
11.61 368 353	0.86	0.627
11.81 368 353	0.84	0.558
11.79 376 273 170	0.83	0.357
p-hydroxybenzoic acid	0.78	0.514
valine	0.78	0.376
9.56 155 170 243 258	0.77	0.382
11.72 376 273 170	0.76	0.278
alanine	0.74	0.395
5-oxo-proline	0.67	0.122
14.76 318 246 glycoside	0.66	0.327
10.64 347 257 359	0.62	0.120
citric acid	0.59	0.275
cystine	0.59	0.092
N-carbamoyl-aspartic acid	0.52	0.034
N-acetylglutamic acid	0.47	0.020
11.24 218 235 252	0.33	0.010
14.25 611 521 glycoside	0.42	0.092
phenylalanine	0.22	0.007
10.44 223 238 297 282	0.21	0.007
glutamic acid	0.21	0.027
13.16 424 410 245 308	0.00	0.001

3.3.3 Hypothesis for the observed growth improvement upon lowering pH

We hypothesized that, *C. bescii* becomes bioenergetically limited under the conditions it is currently grown, due to insufficient/ sub-optimal proton motive force (PMF) generation. Growth and productivity would improve with higher membrane potentials driven by PMF and increased ATP generation. Upon alleviation of the membrane potential/PMF limitation, given continuous substrate availability, growth would continue until a secondary limiting factor takes over such as high total osmolarity or inhibitory acetate concentration. Thus, *C. bescii's* growth rate could then be defined as a discontinuous function of PMF (Δp) and the secondary limiting factor e.g. inhibitory acetate concentration [Ai].

$$\mu = \begin{cases} f\ (\sim \Delta \mathrm{p}), & [\mathrm{A}] < [\mathrm{Ai}] \\ f\ (\sim \mathrm{A}), & [\mathrm{A}] > [\mathrm{Ai}] \end{cases} \qquad \text{[Equation 1.0]}$$

Where: μ = Specific growth rate; [A] = Acetate concentration; [Ai] = Inhibitory acetate concentration; Δp = Proton Motive Force (PMF).

The hypothesis model in Figure 15, addresses the phenotype observed in the batch reactors upon lowering pH from 7.2 to 6.0 based on the proposed hypothesis. *Condition I:* When cells are grown in a batch system without pH control or in a controlled system but at a non-optimal pH. The cells generate and set up a PMF as they grow and use various mechanisms to maintain it within optimal growth range utilizing the charged species in the system. These mechanisms are energy intensive and use up ATP generated by the cells, most of which is derived via substrate level phosphorylation (SLP-derived) in fermentative organisms. Under these conditions the cells slowly lose their potential to maintain PMF (Δp) and cellular functions are affected once the cells are outside the growth supporting Δp range.

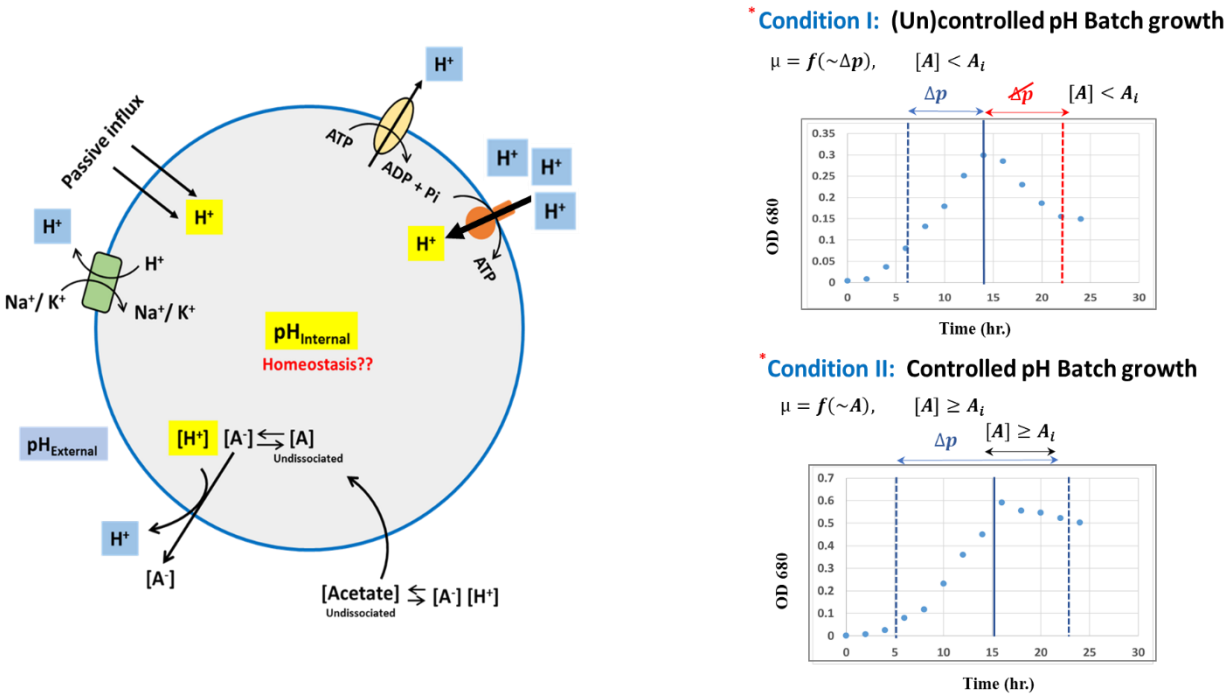
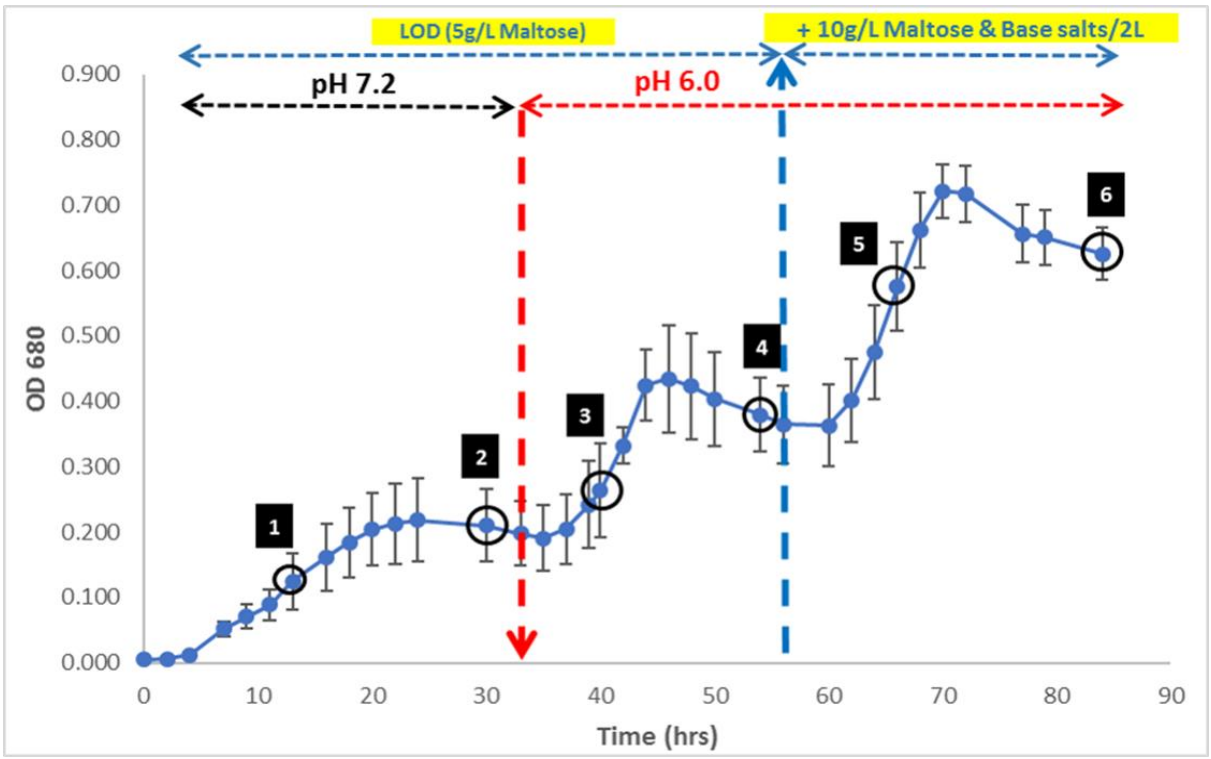


Figure 15: Hypothesis Model (*for depicting conditions in hypothesis model: OD data points are actual; regions are hypothesized).

Condition II: When cells are grown in a controlled batch system at an optimal pH (Hypothesis for growth effects seen in fermenters at pH 6.0). When grown in a controlled pH system at a lower pH, a higher potential for PMF maintenance is provided by the additional protons available in the system due to lowering pH with acid. This leads to cells staying in the Δp range for extended periods, with lower energy costs and/or increased ATP generation via PMF driven ATP synthases in addition to the SLP-derived energy. This would result in improved growth and productivity.

3.3.4 Replicated batch fermentations to test proposed hypothesis

If growth cessation was due to PMF limitation alone and the improved growth upon lowering the pH with acid (introducing additional protons to the system) was due to alleviation of this limitation; then growth which ceased in the stationary phase at pH 7.2 with residual sugars, should be revivable/recoverable if the pH was dropped post-growth cessation in stationary phase. Therefore, to test our hypothesis that growth cessation was in fact due to PMF, controlled batch reactors were set up in triplicate with 5g/L Maltose and allowed to grow at pH 7.2 until stationary phase (OD_{max} ~ 0.2; acetate ~8mM; lactate ~0.1mM) and then the pH was lowered to 6.0. It was observed that the cells grew further reaching higher ODs (~0.8), accumulated more products (acetate ~30mM; lactate ~8mM) and completely utilized the 5g/L maltose within ≤ 60 hours (Figure 16, Table 12). Complete substrate utilization even at 5g/L sugars within this time frame has not been previously reported with *C. bescii.* To further test if the cessation of growth at pH 6.0 observed at this point now was due to sugar limitation, maltose was added to reach an additional 10g/L along with the equivalent base salts. Recovery of cell growth was



1. Mid-Log: pH 7.2
2. Stationary: pH 7.2
3. Mid-Log: pH 6.0
4. Stationary: pH 6.0
5. Mid-Log +10g/L Maltose: pH 6.0
6. Stationary +10g/L Maltose: pH 6.0

Figure 16: Growth curve depicting post-stationary phase growth recovery upon lowering pH to 6.0 and again after addition of more sugar at pH 6.

Samples were collected at different stages along the growth curve to measure products, membrane potential and ATP pools. Stages 1, 3 and 5: depict cells in mid-log phase at pH 7.2, after lowering pH to 6.0 and post additional sugar supply respectively. Similarly, stages 2, 4 and 6 depict when the cells in stationary phase at pH 7.2, after lowering pH to 6.0 and post additional sugar supply respectively.

Table 12: Substrate and product profiles (HPLC) from post-stationary phase growth recovery batch fermenters, depicting complete substrate utilization of ~15 g/L maltose.

Sampling Stage		Maltose (g/L)	Glucose (g/L)	Acetate (g/L)	Lactate (g/L)
1	ML 7.2	3.25	0.31	0.18	0.01
2	ST 7.2	1.50	0.97	0.75	0.02
3	ML 6.0	0.89	1.38	0.93	0.03
4	ST 6.0	BDL	0.58	1.76	0.04
5	MLM 6.0	5.98	2.67	2.04	0.07
6	STM 6.0	0.04	4.70	3.62	0.25

observed again and cells reached even higher ODs, made more products and completely utilized the additional 10g/L maltose as well (Figure 16, Table 12). This was a novel observation, since now the system depicted the ability to completely utilize ≥ 10g/L sugar and produce ~ 100 mM total organic acids.

In a separate experiment, growth starting at pH 6 was also tested. It was observed that cells grown in batch, pH-controlled fermenters starting at pH 6.0 showed similarly improved growth and product profiles as cells in the system where pH is lowered from 7.2 to 6.0 at stationary/mid-log phase (Figure 17, Table 13).

3.3.5 Membrane potential and ATP pool trends, pre and post-acid addition

The fermenters were sampled for membrane potential estimation at the mid-log and stationary phases at pH 7.2 and after lowering the pH to 6.0. At pH 7.2 the membrane potential was observed to be higher in the mid-log phase with an overall R/G ratio of 0.96, as the cells rapidly grew than the stationary phase with lowered overall R/G ratio of 0.93, suggesting a drop-in membrane potential (Figure 18: 1 and 2). Upon acid addition (pH now 6.0), alongside the recovery of cell growth the R/G ratio was also observed to have increased to 1. This indicated increased membrane potential values at pH 6 which were also in fact slightly higher than what they were at pH 7.2 at the mid-log phase. Moreover, the most important observation was that the R/G ratio in the stationary phase at pH 6 was

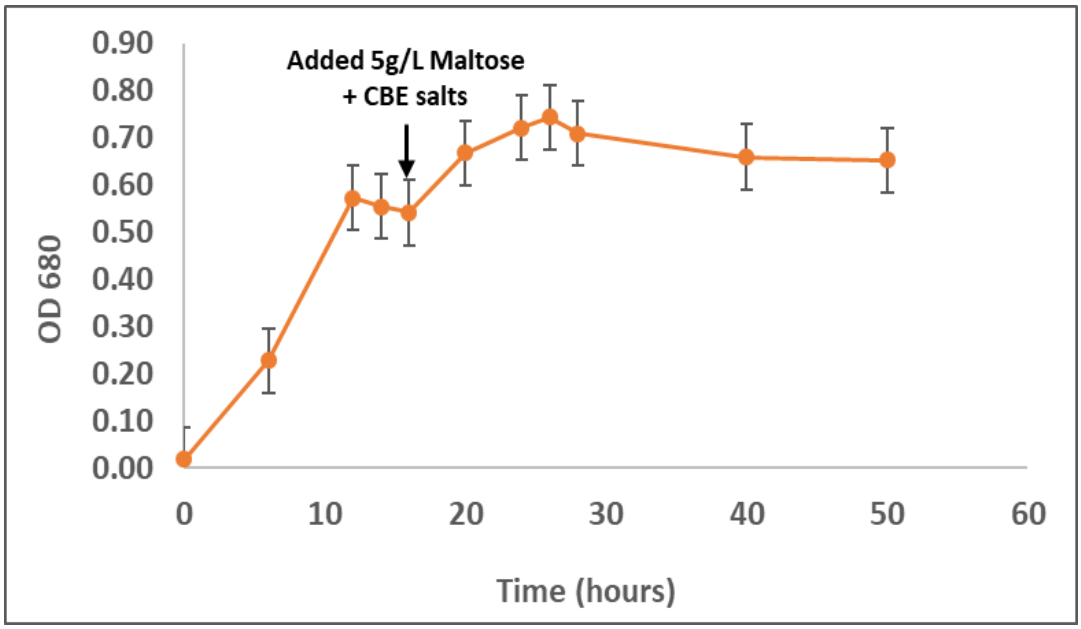


Figure 17: Cells grown in batch fermenters starting at pH 6.0 show similar growth pattern to cells after pH is lowered from 7.2 to 6.0.

Table 13: HPLC product profiles of cells grown in batch fermenters starting at pH 6.0, show similar product accumulation and sugar utilization patterns as for fermenters starting at pH 7.2 and the lowered to 6.0.

Hours	Maltose (g/L)	Glucose (g/L)	Acetate (g/L)	Lactate (g/L)
16 (pre-sugar addition)	0.43	0.79	1.33	0.50
16 (post-sugar addition)	5.07	0.81	1.28	0.48
50	BDL	1.73	2.79	1.26

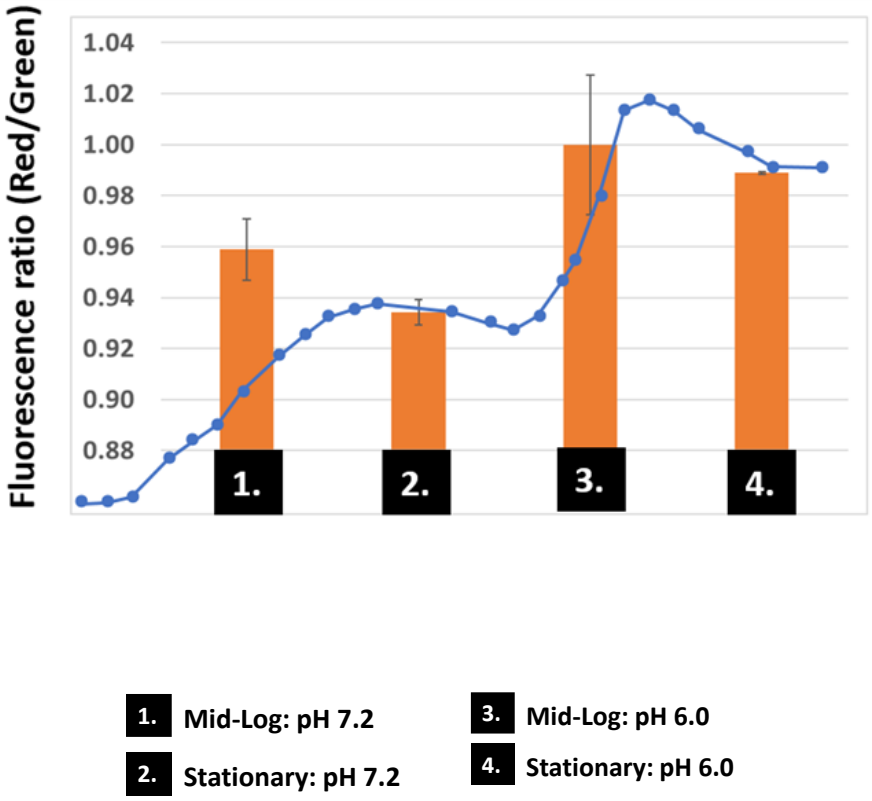


Figure 18: Membrane potential changes estimated across the growth curve in figure 16 at different stages (overlaid blue line graph). The ratios were calculated using red and green MFI from each sample normalized using CCCP treated samples at each stage.

0.99, which is similar to the mid-log phase ratio indicating that the membrane potential did not drop in the stationary phase here as opposed to at pH 7.2 (Figure 18: 3 and 4). Although, the difference the in absolute ratio values between stages were not huge, these differences were consistent across sample readings and replicates. Three independent measurements were made per sample from each fermenter and averaged to get the final values for the sample. Sample values for a pH phase were then averaged across the three replicate fermenters to get the ratio values reported. Moreover, it has been previously reported that membrane potentials detected with DiOC2, can vary in magnitude with species, for many organisms the ratios have been shown to vary with the intensity of the proton gradient and for some they do not appear to be proportional to proton gradient intensity. Thus, we derive our conclusions following along the consistently observed ratiometric shifts in potential measured conservatively, alongside the observed phenotype post-acid addition.

The measured ATP trends during growth at pH 7.2 indicated a lower ATP pool at the mid-log phase, ~17 nM/g cdw (cell dry mass) and an almost 4 times bigger pool ~74 nM/g cdw in the stationary phase (Figure 19: 1 and 2). On the other hand, post-acid addition as the growth recovered, it was observed that the ATP pool in mid-log at pH 6.0 was almost 5 times bigger, ~30 nM g cdw than stationary phase which was ~ 6nM/ g cdw at this pH, but also twice as much in comparison to the measured ATP pool in the mid-log phase at pH 7.2 (Figure 19: 3 and 4).

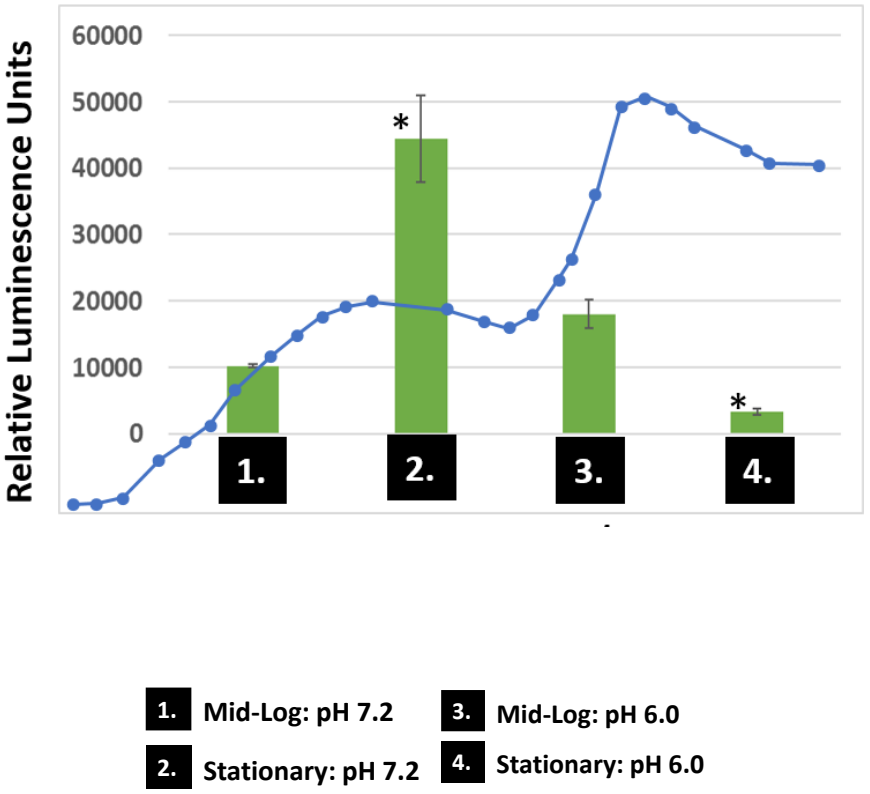


Figure 19: ATP pool changes across the growth curve in figure 16 at different stages (overlaid blue line graph).

3.3.6 Effect of post-stationary phase acid addition on *C. bescii* strain with inactive [Ni-Fe] hydrogenase

This experiment was planned to explore the involvement of the [Ni-Fe] hydrogenase in the observed improved growth phenotype by looking for a loss of this response when the strain with an inactive [Ni-Fe] hydrogenase was subjected to similar conditions. *C. bescii* strains with the deletion of a cluster of genes, *hypABFCDE*, required for maturation of the [Ni-Fe] hydrogenase, JWCB038 (Δ*pyrFA* Δ*ldh CIS1*::$P_{S\text{-}layer}$ *Cthe-adhE ΔhypADFCDE*/(*ura*−/5-FOA^R)) and its parent strain JWCB036 (Δ*pyrFA* Δ*ldh CIS1*::$P_{S\text{-}layer}$ *Cthe-adhE*//(*ura*−/5-FOA^R)) [182] were grown in batch, pH-controlled fermenters starting at pH 7.2, 75^{o}C in triplicate. The pH was lowered to 6.0 once cells stopped growing at pH 7.2, as done in the previous batch fermentations that showed growth recovery after lowering pH in stationary phase. All three fermenter replicates with the deletion mutant (JWCB038) showed a similar response as the parent (JWCB038) and as was seen with the wild-type previously upon lowering the pH from 7.2 to 6.0 in the stationary phase. The cells started to regrow within 2-4 hours of lowering the pH and reached higher ODs than at pH 7.2, completely utilizing the remaining sugar out of the total 5g/L supplied in the beginning, as noted from the HPLC measurements (Figure 20, Table 14). Thus, the loss of improved growth phenotype upon acid addition was not observed in JWCB038. This indicates that, [NiFe] hydrogenase is not critically involved in utilizing the added protons under the conditions tested in this study to bring about the growth phenotype that was observed. The strain identities in each fermenter for this experiment were confirmed via PCR using the primers MC017 and MC020, as previously used [182].

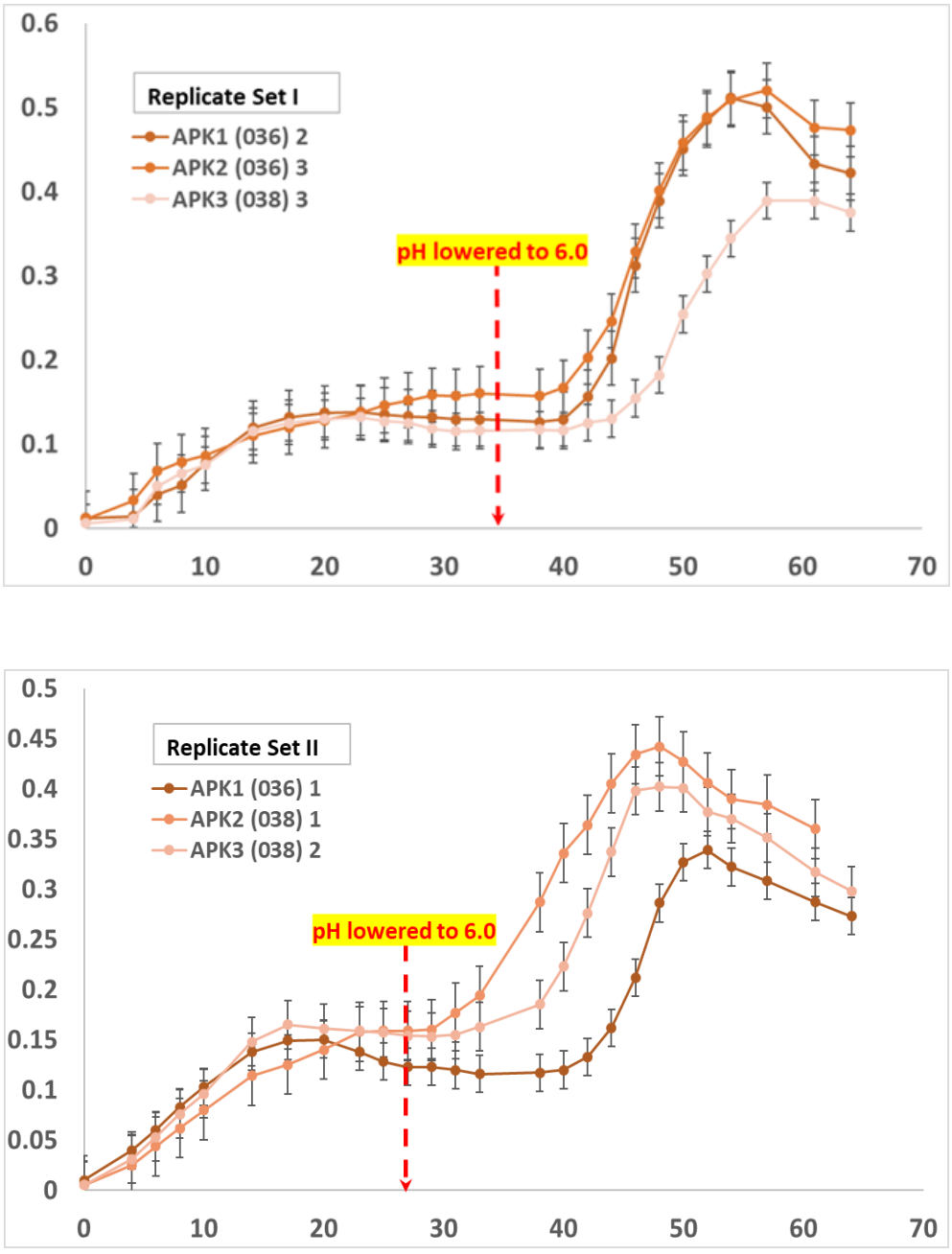


Figure 20: Strains JWCB036 (parent) and JWCB038 (mutant with inactive [Ni-Fe]) grown in batch fermenters show similar growth patterns after pH is lowered from 7.2 to 6.0.

Table 14: End-point HPLC product profiles for fermentations with inactive [NiFe] hydrogenase strain, JWCB038 and its parent strain, JWCB036 at pH 6.0 show similar patterns post-acidification.

Strain	**Phase/pH**	**Maltose (g/L)**	**Glucose (g/L)**	**Acetate (g/L)**
JWCB036 (parent)	Stationary/ 7.2	2.08	0.99	0.61
	Stationary/ 6.0	0.02	0.62	2.00
JWCB038 (NiFe mutant)	Stationary/ 7.2	1.57	1.40	0.67
	Stationary/ 6.0	0.01	0.55	2.03

3.4 Discussion

This work explores *C. bescii*'s growth and molecular level responses to medium acidification and more importantly provides insights into what seems to be a major factor limiting its growth and productivity. It was observed that when grown in batch fermenters cells grew better after pH was lowered to 6.0 from 7.2 in the mid-log phase. Sulfate addition due to use of sulfuric acid was experimentally ruled out as a cause for the response. The time course omics from these fermenters post-acid addition also did not reveal any specific genetic elements differentially expressed that could directly explain the observed phenotype. This anti-correlated response led us to think that it could be a physiochemical response, tied to the availability of additional protons in the system due to acid addition. It is well known that the exchange of protons across cellular membrane affects the membrane potential of the cell and the proton gradient thus formed generates a proton motive force (PMF). The established PMF is a critical component in driving various cellular functions like ATP generation, substrate/ion transport and various metabolic reactions [194]. Our hypothesis visualized PMF as the common factor for enhanced ATP production, substrate transport and other energy linked functions affecting cell growth, product yield and substrate utilization in *C. bescii* post-acid addition. As described in the results we tested our hypothesis in batch fermenters by adding acid to the medium (pH lowered to 6.0) once the cells were in the stationary phase at pH 7.2. Growth recovery and complete substrate utilization was observed, indicating that a factor inhibiting growth and substrate utilization was alleviated by adding acid to the system. Measured membrane potential and ATP pool size trends from the mid-log and stationary phases at pH 7.2 and 6.0 also supported the proposed hypothesis.

Lower membrane potentials in stationary phase than in mid-log and growth cessation with sufficient sugar left in the system indicated that not sugar but membrane potential driven PMF was limiting growth in stationary phase at pH 7.2. Growth recovery and higher membrane potential measurements upon adding acid indicated that the addition of protons to the medium lead to higher internal negative potentials that alleviated growth inhibition. Another noteworthy and important observation was that the elevated membrane potential post-acid addition was maintained even when the cells stopped growing in the stationary phase at pH 6. This showed that membrane potential was no longer limiting at this point and the growth cessation was due to some other factor, which was revealed to be sugar limitation by HPLC product profiles. The observation that adding more sugar to the system at this point resulted in cell growth recovery and the growth continued until the additional sugar supplied was consumed, further bolstered the inference above.

Measured ATP pools depicted that the pool sizes were smaller in the mid-log phases when the cells were actively growing at both pH 7.2 and 6.0. This is an expected trend as the cells are rapidly turning over ATP as they grow. Despite the active cell growth in log phases at both pH 7.2 and 6.0 the ATP pool at pH 6.0 was almost two times that of the pool size at pH 7.2. This observation supports our hypothesis that the cells would have higher ATP available due to PMF derived energy contribution. Essentially the higher internal negative potential brought about by the addition of protons to the media generates a proton gradient leading to PMF that drives ATP production. The cells can then utilize this additional ATP alongside the SLP derived-ATP to carry out various functions such as cell biomass production and active substrate transport. Transcriptomics data from our

batch reactors reveal genes from a sugar ABC transporter operon, as the most significantly overexpressed genes post-acid addition, suggesting that the additional ATP is being sequestered towards active substrate transport. This would understandably lead to better substrate utilization and higher product yields. Another trend noted with ATP pool measurements was that a bigger pool size was seen in the stationary phase at pH 7.2 in comparison to the mid-log phase. Whereas the opposite (i.e. bigger ATP pool size in mid-log and lower in stationary phase) was seen post-acid addition at pH 6.0. Looking back at the growth conditions during these phases, this trend can be explained with the sugar availability in the system, similar to what has been previously seen [195]. At pH 7.2 in stationary phase, the cell growth was inhibited by PMF limitation, but there was still sugar available in the system which was being metabolized by the cells. Therefore, as the cells were still accumulating ATP via ongoing metabolism at this point but not growing, thus the ATP pool size increased. On the other hand, in the stationary phase at pH 6.0, where the PMF limitation had been alleviated by added protons, the cells had stopped growing due to sugar limitation instead. As the growth at this point was limited by availability of the energy source, the ATP pool size rapidly declined, and growth stopped. Thus, the higher internal membrane potential upon acid addition and PMF driven additional ATP generation, resulting in improved cell yield and substrate internalization account for the phenotype observed in this study. These results suggest that *C. bescii* possesses an energy conservation system that is chemiosmotically coupled to the cellular metabolism. Such a system, if well understood, could be exploited to improve cellular productivity by only having to manipulate the external growth conditions making it an efficient and cost-effective strategy to be employed in industrial fermentations.

C. bescii being a hydrogen producing organism possesses two hydrogenases, a bifurcating [Fe-Fe] and a membrane bound [Ni-Fe] hydrogenase [182]. One of the ways the cell could be using the additional protons in the system, is via proton translocating ferredoxin oxidation, mode of energy conservation with the involvement of the membrane bound [Ni-Fe] hydrogenase. This can result in PMF and ATP generation as previously observed also with some anaerobes [196]. However, our results from using a *C. bescii* strain with an inactive [Ni-Fe] hydrogenase (JWCB038) indicate that although this hydrogenase may play some role in PMF generation in the organism as previously speculated [182], it is not critically involved in the generation of the phenotype observed in this study. Another way *C. bescii* could be harnessing the additional protons for energy is via a pyrophosphate fueled energy conservation system. It is known that in some bacteria inorganic pyrophosphates can serve as a source of energy [197, 198] and this ability to utilize intracellularly formed pyrophosphates aids energy conservation. In such a system the indispensable production of the pyrophosphates in an organism is coupled to their hydrolysis, which efficiently drives the reaction in the forward direction ensuring growth maintenance [199]. If the inorganic pyrophosphate (PPi) hydrolysis involves a membrane-bound proton translocating pyrophosphatase (H^+-PPase), the energy released by the hydrolysis could be used by the virtue of the established PMF [200, 201]. The proton gradient thus established can also be chemiosmotically coupled and used by ATP synthase to ultimately generate ATP [202]. *C. bescii*'s genome does encode for a membrane bound proton translocating pyrophosphatase (H^+-PPase): *ATHE_RS03785.* The presence and involvement of this would enable the above described mode of energy conservation in *C. bescii.* In a study with *C. saccharolyticus,* a close relative of *C. bescii*

it was also observed that the PPase enzyme activity measured from the membrane fraction was much higher than that of the cytosolic fraction, indicating an active membrane bound PPase [203]. Moreover, significantly higher phosphates were detected upon analysis of metabolomics data from our fermenters that were acid treated in comparison with the untreated controls. Therefore, based on our results and supporting information from genomics as well as literature, we speculate that the protons made available in the system (i.e. post-acid addition in this study) were sequestered via this mechanism to generate the phenotype observed. Future directions in line with the results from this study are discussed in detail in chapter 6.

Chapter 4.0

High Osmolarity inhibits the fermentation potential of *Caldicellulosiruptor bescii* on both liquid and solid substrates after PMF stress is relieved

4.1 Introduction

The capacity of a microorganism to respond to environmental fluctuations is an important physiological process that determines its ability to thrive in a variety of habitats [204, 205]. Although faced with changing surroundings throughout growth, extreme or extensive changes to the environment can be detrimental to growth. All microorganisms have a set of defined limits for growth beyond which they are unable to carry out their regular functions and growth declines. Microorganisms isolated from extreme or harsh conditions such as *C. bescii* also are prone to extreme changes in growth conditions and have a range of parameters for physiological functioning [205]. Some of the most important physical parameters for growth include temperature, pH and osmolarity [206]. When trying to maximize the performance as desired for biofuel producing microbe candidates, these vital conditions need to be optimal and more importantly maintained within physiological limits. Despite the ability of *C. bescii* and others to utilize a wide range of substrates, fermentation inhibition can occur while ample substrate is still available due to various factors, one of these factors is the presence of inhibitory compounds [29] that can result in unfavorable growth conditions and limit bioconversion to products. These

may be released from biomass, generated through pretreatment or as a product of metabolism. Several compounds released during fermentation such as carboxylic acids, furans, phenols can negatively affect cell membrane function, growth and glycolysis of the microorganisms [207, 208]. Other such potential inhibitory chemicals generated during the process can be sugar degradation products such as furfural and hydroxymethylfurfural (HMF); weak acids such as acetic, formic, and levulinic acids; lignin degradation products such as the substituted phenolics, vanillin and lignin monomers. Compounds like these not only cause cellular toxicity but can alter the overall growth environment. Physiological parameters such as osmolarity and pH that are crucial for the growth of an organism are also the conditions most prone to be easily affected.

Osmolarity is a well-known inhibitor of microbial growth and high salt or sugar concentrations are used as food preservatives [209, 210]. High solute concentrations like sugars, salts or ionic derivatives of inorganic salts, organic acids etc. can affect the osmolarity of the growth medium. Similarly, production and accumulation of organic acids, carboxylic acids and derivatives can have negative effects [211]. Previous studies in *C. bescii* have illustrated the potential for growth inhibition due to osmolarity. Development of the minimal growth media for *C. bescii*, LOD (low osmolarity defined) reported enhanced growth and attributed it to reduced solute load leading to lower osmolarity in comparison to the other rich media [180]. Growth inhibition with increasing amounts of NaCl in LOD medium has also been reported with complete inhibition at 50mM NaCl (~250 mOsm) [113]. Long lag phases indicative of stress have also been reported for *C. bescii* at high sugar loadings of 90 gL-1 glucose (>50h) and 150mM acetate [6]. Similar reports are also available for other thermophilic organisms, like severe growth inhibition

of *C. saccharolyticus* [212] at osmolarity values beyond 0.216 osm/kg, 200mM acetate concentrations and significantly decreased biohydrogen yields [213-215]. Use of CO_2 for stripping H_2 from the medium in a study led to a near 30% increase in the osmotic pressure of the culture, which was observed to inhibit the growth of *C. saccharolyticus* [213]. Thus, there is premise for investigating high osmolarity as yet another factor that can further limit the fermentation potential of *C. bescii.*

In chapter 3, the previously described study, lowered membrane potential was recognized as a factor limiting the growth and substrate utilization of *C. bescii*. Inhibition of growth, of microbes like *C. bescii* due to stress factors such as high osmolarity and low pH along the fermentation process can limit their ability to reach higher cell densities and in turn diminish product yields. This may be a major issue in reaching the goal for industrial biofuel production and needs increased attention. Thus, identifying such major inhibiting factors, and understanding the conditions under which they become growth limiting form potential immediate research targets. The aim of this study was two-fold including the objectives of investigating if membrane potential associated limitation also affects growth on solid substrate and whether osmolarity and/or organic acid accumulation are the next immediate or secondary factor(s) to limit the fermentation potential of *C. bescii* if PMF limitation was alleviated.

4.2 Materials and Methods

4.2.1 Strains, Media and Growth Conditions

Caldicellulosiruptor bescii strain DSM6725 (wild-type) was grown in liquid LOD medium [113] with 10g/L maltose (catalog no. M5895, Sigma) or 10g/L avicel (PH-101, catalog no. 11365, Sigma) as the carbon source to begin with, unless otherwise specified. Cells were grown anaerobically at 75 °C under batch reactor settings. Applikon bioreactors (Applikon Biotechnology, Inc.), 1.5 L were used with 1 L working volume and were seeded with a 5% inoculum, maintained at 75 °C, 300 rpm and sparged with nitrogen constantly at a low flow rate of ≥0.5 L/min. Seed cultures for fermenters were grown in anaerobic 50 mL serum bottles from a 0.5% inoculum and incubated at 75 °C, 200 rpm overnight. Liquid substrate fermenters with maltose were started and maintained at pH 6.0. The fermenters with the solid substrate, avicel, were started at pH 7.2 and then pH was lowered to 6.0 in the stationary phase using1 M H_2SO_4. In-depth details of the batch fermentations performed are individually specified for the experiments in results.

4.2.2 Growth Assessments

Samples from liquid substrate, maltose grown cultures were collected and growth was estimated by monitoring optical cell density. A spectrophotometer (Thermo Scientific™ GENESYS™ 20), was used for measuring absorbance at 680 nm. For samples from solid substrate, cultures grown on avicel, cell enumeration by direct counting of DAPI stained cells using microscopy was performed as previously described [183] using ImageJ software.

4.2.3 Quantification of Substrates and Products

Carbohydrate content of the dried (40°C) solids for ~24 hours (pre and post-fermentation avicel) was determined by the quantitative saccharification assay NREL/TP-510-42618 and high-performance liquid chromatography (HPLC) method NREL/TP-510-42623, similar to as previously described [216]. Briefly, biomass carbohydrates were acid-hydrolysed in 72% w/w H_2SO_4 (0.1 g solids/mL acid) for 1 h at 30 °C, followed by further oligomer breakdown in 4% w/w H_2SO_4 at 120 °C for 1 h. Calcium carbonate was used to neutralize the hydrolysed samples to pH ~7.0. Samples were then filtered before quantification of carbohydrate content (glucose, xylose, galactose, mannose and arabinose) by HPLC on a Shimadzu Prominence-i Series LC-2030C system (Shimadzu Scientific Instruments, Columbia, MD) against known standards. HPLC product separation was carried out using an Aminex™ HPX-87P column (Bio-Rad Laboratories Inc., Hercules, CA) at 0.6 mL/min flow rate of ultrapure water and column temperature of 80 °C, while signal was measured with a refractive index detector (model RID-20A) at 35 °C. Furfural and 5-hydroxymethyl furfural were also quantified (system-integrated UV–Vis detector) to confirm optimum acid solubilization with only trace amounts of monomeric sugar breakdown. Samples were analyzed in triplicates.

Fermentation products maltose, cellobiose, glucose, acetate and lactate, were quantified using high-performance liquid chromatography (HPLC) quantification using an Aminex™ HPX-87H column (Bio-Rad Laboratories Inc.) at 0.5 mL/min flow rate of 5 mm H_2SO_4 mobile phase and column temperature of 60 °C. The signal was quantified on a Shimadzu Prominence LC-20A Series system (Shimadzu Scientific Instruments, Columbia, MD) equipped with a refractive index detector (model RID-20A) at 35 °C. Integrated total peak

areas were compared to peak areas and retention times of known standards for each compound of interest.

4.2.4 Osmolarity Measurements

Samples were collected in 1.5 mL Eppendorf tubes and centrifuged at 16,873 x g for 10 min at room temperature. Supernatants were filtered using HPLC certified, Acrodisc® 13 mm syringe filters with 0.2 µm nylon membrane (Pall Laboratory) into fresh 1.5 mL Eppendorf tubes. Osmolarity values for each sample were measured three times independently on Wescor Vapro 5520 instrument by loading 10 µL of sample on Wescor Inc SS033 sample discs at room temperature.

4.3 Results

4.3.1 Bottle Growth Assessments with Sodium Chloride

Growth suppression was observed in *C. bescii* cultures (triplicates) at osmolarities in the range of ~200-300 mOsm obtained via addition of NaCl in 50 mL LOD medium (native osmolarity ~74 mOsm) at 12 and 24 hours post-inoculation (Figure 21a). On the other hand, no growth was observed at 12 and 24 hours post-inoculation in the medium at ~350 mOsm (138mM NaCl) (Figure 21a). These results are in accordance with what has been previously shown [1]. It was further observed that, upon continuing incubation past 24 hours to allow the cultures to grow over a span of 108 hours growth started after a long lag of ~30 hrs and cultures in higher osmolarity media (350 mOsm) where no growth was

previously seen reached similar final optical densities (OD 680nm) as the cultures in LOD medium (74 mOsm) without NaCl. Similar results were also seen in 10 mL cultures when repeated in triplicates in Hungate tubes for over 100 hours (Figure 21b).

4.3.2 High Liquid Substrate Loadings and Osmolarity at pH 6.0

In order to test if osmolarity and/or organic acid accumulation would be the secondary limiting factor(s) post-PMF limitation alleviation, *C. bescii* cultures grown in pH-controlled fermenters (1.5L Applikons) starting at pH 6.0 with 10 g/L maltose. Cultures were allowed to grow for over 250 hours and additional 10 g/L maltose with base salts equivalent of salts per liter were added at stationary phases determined by monitoring culture OD 680 nm and base addition to the fermenters (Figure 22). Samples for fermentation product profiling and osmolarity measurements were collected at several stages. The cells in the system reached a maximum OD of 1.0 and utilized ≥ 40g/L maltose to produce 102 mM total organic acids.

Osmolarity measurements from all three reactors with cultures growing on maltose at various stages of fermentation show a gradual increase from ~ 70 mOsm to 450 mOsm (Figure 22) in the medium osmolarity as products formed accumulate in the system. *C. bescii* was observed to experience growth suppression starting at total osmolarity values of ≥250 mOsm and severely inhibited as osmolarity approaches >350 mOsm and was unable to recover upon substrate or base salts addition (Figure 22).

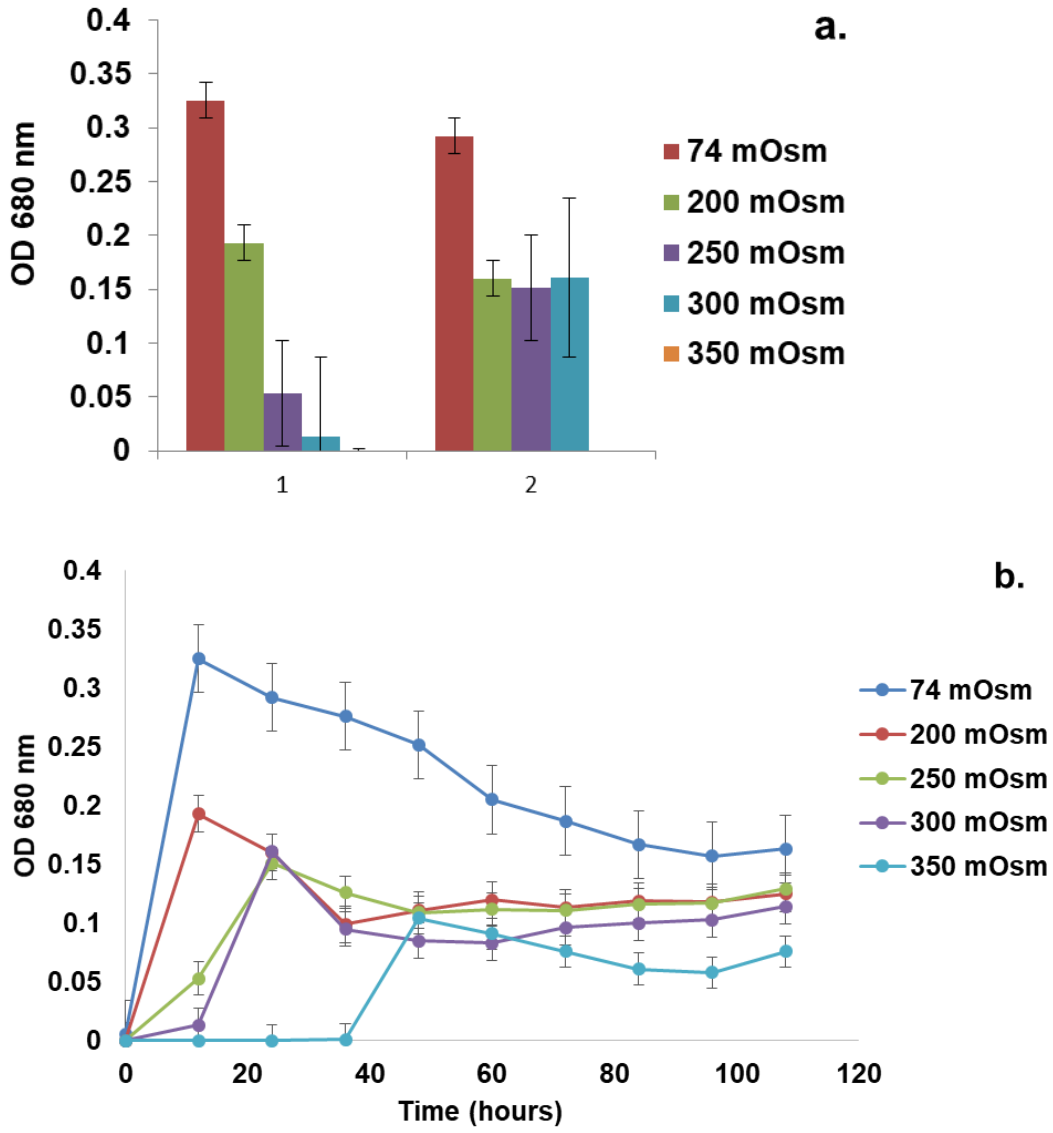


Figure 21: Effects of high osmolarities (200-350mOsm) on growth of *C. bescii* cultures due to added NaCl in LOD medium. Growth (triplicates) suppression observed at (a) 12 hours (1) and 24 hours (2) indicated completely inhibited growth at 350mOsm but when allowed to grow over 100 hours (b) cultures showed growth after a lag of ≥ 30 hour.

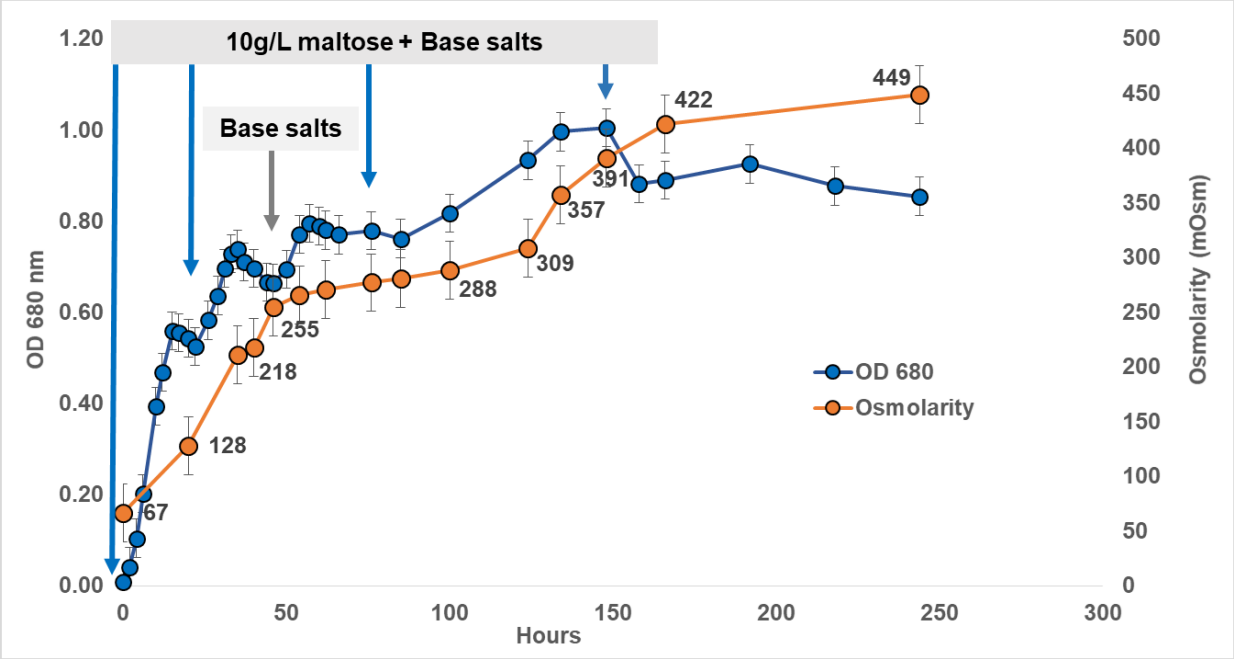


Figure 22: *C. bescii* cultures grown starting at pH 6.0 (controlled), with subsequent additions of substrate (maltose, 10 g/L) and base salts equivalent of per liter show growth limitation effects at high osmolarities. Blue arrows depict timepoints of sugar along with base salts/L addition and grey arrows indicate addition of base salts/L alone.

HPLC product profiles show that the system is not substrate limited and shows sugar (glucose) accumulation representative of slowed metabolism at ≥350 mOsm (Table15). Growth rate (μ) and productivity (P) were estimated as below using measured OD and products over a period of 16-30 hours:

$$\mu = \frac{\ln(\frac{OD_2}{OD_1})}{T_2 - T_1}$$ Equation 2

where,

μ = Specific Growth Rate; OD_1 and OD_2 = Optical Density at time point 1 and 2; T= Time

$$P = \frac{Pdt_2 - Pdt_1}{T_2 - T_1}$$ Equation 3

where,

P = Productivity; Pdt_1 and Pdt_2 =Products (total organic acids concentration) at time point 1 and 2; T= Time

Estimations for productivity (P) at ≤150 mOsm, between 250mOsm - 300mOsm and >400mOsm used product measurements from 0 - 22 hours, 46 - 62 hours and 166-192 hours respectively (Table 15). It was observed that the productivity dropped from by 50% (0.09 g/L hr^{-1} to 0.045 g/L hr^{-1}) at system osmolarity > 250 mOsm and approaches zero

Table 15: HPLC product profiles and corresponding osmolarity values for maltose fermentations as depicted in figure 22. The table shows data for products accumulated, substrate additions along fermentation, sugar accumulation and utilization patterns along the growth curve. Values in “blue” highlight time points of sugar and base salts addition and “red” draw attention to accumulation of sugars observed.

Hours	Osm (mOsm)	Maltose (g/L)	Glucose (g/L)	Acetate (g/L)	Lactate (g/L)
0	67	**10.85**	0.00	0.00	0.00
22	143	3.75	1.77	1.33	0.72
26	*NA	**12.23**	1.75	1.23	0.65
46	255	2.54	8.15	2.39	1.09
54	265	**7.01**	5.37	2.30	0.76
62	271	0.28	9.50	3.05	1.15
134	358	**BDL	4.99	4.95	1.30
134		**8.78**	4.65	4.63	1.22
166	422	**BDL	11.67	4.64	1.26
192	*NA	**BDL	11.41	4.62	1.28
244	449	**BDL	11.23	4.66	1.31
244		**9.62**	10.93	4.51	1.22
362	487	0.25	17.85	4.49	1.26

*NA= Values “Not Available”

**BDL= Below Detection Limit

(0.00061 g/L hr^{-1}) at >400mOsm. Growth rate reduced to one fifth and one twenty fifth of the growth rate observed at medium osmolarity ≤150 mOsm ($\mu_{\leq 150}$ = 0.14 h^{-1}), at osmolarities > 250 mOsm ($\mu_{250\text{-}300}$= 0.026 h^{-1}) and >400mOsm ($\mu_{>400}$= 0.006 h^{-1}) respectively.

C. bescii displayed the ability to withstand high osmolarities of around 500mOsm as cell lysis was not observed when cells were viewed under the microscope (63X magnification). Cell viability was confirmed by sub-culturing cells extracted from fermenters at 400 hours. Moreover, upon replacing approximately half the volume of the fermenters with fresh media (osmolarity lowered to 270mOsm) the cells in the fermenters responded to medium dilution and grew back to reach densities similar to before dilution (OD ~ 0.5) and osmolarity increased to > 350 mOSm over the next 300 hours (Figure 23).

4.3.3 PMF limitation alleviation, Solid substrate conversion and Osmolarity

Cultures were grown in fermenters (triplicate) with additional 5 g/L avicel with base salts equivalent of salts per liter were added at stationary phases. System was monitored using base (alkali) addition to the fermenters, which is responsive to the rate of accumulation of products and representative of growth and metabolism (Figure 24). Monitoring base addition allows for a continuous, real time monitoring of the system as OD is not a reliable measurement on solid substrates such as avicel or switchgrass and cell count limit real time monitoring potential. Increase in the base (alkali) addition signifying growth recovery and/or product formation post stationary phase at pH 7.2 was observed within ~4-6 hours of lowering pH to 6.0 in all fermenters (triplicates) (Figure 24). This is similar to the

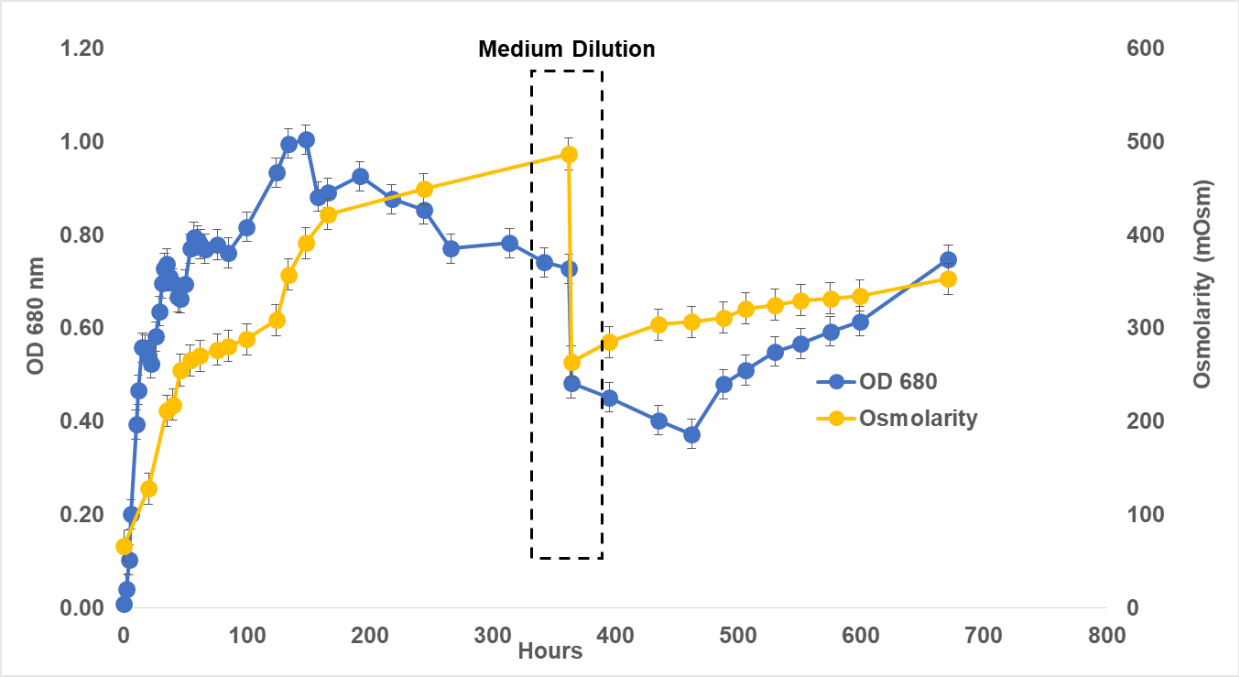


Figure 23: *C. bescii* depicts the ability to maintain cell viability at high osmolarities for long durations by slowly growing back to an OD of ~ 0.6 post-medium dilution.

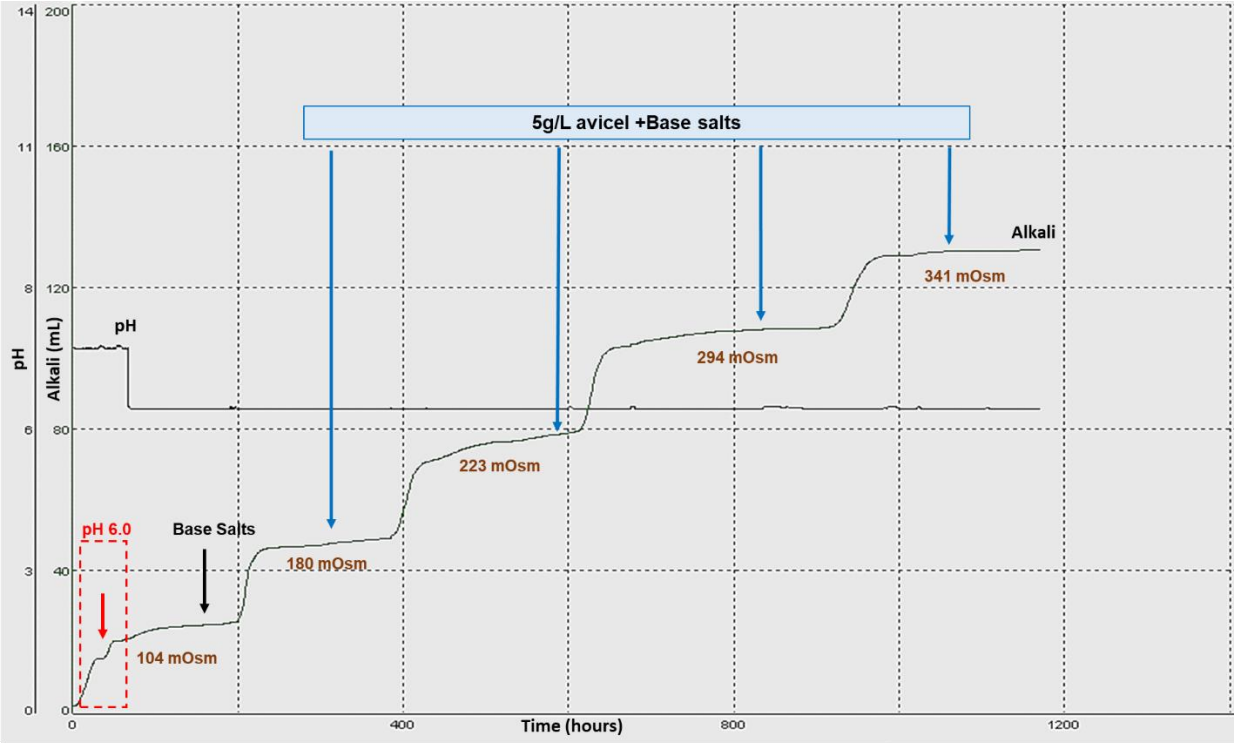


Figure 24*: Alkali addition curve from *C. bescii* fermentation representing growth of cultures starting at pH 7.2 on 10g/L Avicel grown over 1000 hours with subsequent addition of substrate and base salts (equivalent to per liter media). Growth depicts membrane potential limitation alleviation upon lowering pH to 6.0. The medium undergoes a gradual increase in osmolarity reaching ≥340mOsm with product accumulation (osmolarity values plotted are average of triplicate fermenters).

* Representative alkali addition curve from one of the triplicate fermenters is shown, all three fermenter curves showed similar effects. Osmolarity values labeled are average of measurements from all three fermenters.

response observed with growth on liquid substrate (maltose) due to alleviation of membrane potential limitation as previously described in chapter 3. Cell number increased from around $2x10^8$ to $6x10^8$ cells/mL (average of three replicates) in 6 hours after lowering pH, which is similar to the increase in OD seen when cells grown on liquid substrate, maltose. Samples for HPLC, osmolarity measurements and quantitative saccharification were collected throughout fermentation and continued till the system stopped taking any further base and even additional substrate or base salts did not affect the stationary phase. HPLC product profile and quantitative saccharification at the stationary stage (plateaued base curve) at pH 7.2 shows ample available solid substrate avicel and lower concentrations of products (Table 16). Post-acid addition stationary phase (after base salts replenished in the system) HPLC product profile and quantitative saccharification on the other hand shows significantly reduced avicel in the system and higher products (Table 14). Addition of 5 g/L avicel to the system in the post-acid addition stationary phase lead to elevation in base addition in all three biological replicates and increased products (Figure 24, Table 16). A total of 30 g/L avicel was added to the system in a fed-batch manner and 75.8% of the total was converted and utilized to produce a total of ~ 95 mM total organic acids (Acetate: ~90mM Lactate: ~5mM). Osmolarity measurements from various stages of this fermentation shows gradually increasing medium osmolarity in the system, which reaches high values of ≥340mOsm. A similar trend of product rate reduction with increasing osmolarity was observed with avicel as with maltose. Productivity (P) at >250 mOsm drops to almost one sixth of rate at osmolarities <200 mOsm (0.0078 $gL^{-1}hr^{-1}$).

Table 16: HPLC product profiles depict products accumulated, substrate additions along fermentation and utilization patterns. Values in “blue” highlight time points of 5g/L avicel and base salts addition.

Sample	Hours	Unutilized Avicel (mg glucose/g Avicel)	Acetate (g/L)	Lactate (g/L)	Glucose (g/L)	Cellobiose (g/L)
pH 7.2 stationary	**54**	**889.23**	0.28	0.01	0.09	0.15
pH 6 stationary (post-base salts addition)	**363**	**532.57**	2.52	0.16	1.38	0.40
1st addition		**938.63**	2.23	0.12	1.30	0.31
Pre-2nd addition	**585**	**522.93**	3.96	0.31	2.02	0.36
2nd addition		**938.00**	3.34	0.26	1.72	0.32
Pre-3rd addition	**849**	**547.77**	4.46	0.31	1.93	0.52
3rd addition		**952.00**	4.26	0.24	1.71	0.47
Pre-4th addition	**993**	**544.00**	4.67	0.32	1.92	0.59
4th addition		**926.43**	4.3	0.23	1.68	0.45
end-point	**1160**	**547.30**	5.32	0.39	2.09	0.37

4.4 Discussion

The study described in this chapter aimed to test whether *C. bescii* experiences a PMF related limitation while growing on the solid substrate avicel as was observed on maltose (previously described in chapter 3) and would the osmolarity and/or organic acid accumulation become the next immediate factor(s) limiting fermentation potential in *C. bescii*. The fermentation system used in the study ensured that the culture does not suffer from any other limitations (nutritional or nitrogen [77]) that can occur in a batch system over time. This was accomplished by addition of substrate and base salts during the stationary phase. Both metabolizable including maltose as well as sugars released from avicel and non-metabolizable ionic osmolyte sodium chloride (NaCl) were used and their effects were observed in this study. Avicel which does not contribute to the osmolarity of the system ("osmotically inert") until broken down into sugars, was used to differentiate and observe the effects of high loads of sugars such as maltose from accumulating organic acids and/or other by-products as major contributors to osmolarity. The results from growth of *C. bescii* on high osmolarities using NaCl showed growth suppression at ≥ 250 mOsm as previously described [180] but also added to the knowledge showing *C. bescii* remains viable at high osmolarities of 350mOsm and the growth is severely lagged (≥30 hours). *C. bescii* when grown on maltose with subsequent sugar and salts additions overtime and reaching high medium osmolarities also showed similar trends of growth suppression at ≥ 250mOsm and severe growth and metabolism suppression at ≥350 mOsm. Moreover, as the total organic acids accumulated in cultures with NaCl and on maltose were ≥160 mM as previously described [6, 77] as growth inhibitory concentration. Thus, these results signify the impact of 'total medium osmolarity' as a parameter and the

range over which the growth and metabolism of this organism are adversely affected irrespective of the use of these two different osmolytes. The maltose grown cells in fermenters also showed maintenance of cell viability at high osmolarities by responding to system dilution at >450mOsm. This supports that *C. bescii* possesses the ability to withstand high osmolarities and stay viable to resume growth when conditions become allowing.

Cultures grown in controlled pH fermenters at pH 6.0 maintained high cell densities throughout, reaching a maximum OD ~1 and resulted in the utilization of ≥ 40g/L maltose until growth and metabolism were limited by high medium osmolarities. This is by far the highest cell density and amount of liquid substrate utilized to be reported with *C. bescii* to our knowledge. These results also suggest some ways to most effectively employ *C. bescii* cultures to the production of desired products. The use of fed-batch systems may be one of the effective ways to overcome the effects of high loads of sugars and increase productivity. Another approach could be to monitor the system osmolarity which could be aided by real-time online osmolarity measurements using a specialized probe [217] and to dilute the system by replacing an amount of spent with fresh media, a similar approach has been preliminarily described in this study.

Results from avicel fermentations revealed that *C. bescii* experiences pmf limitation when growing on solid substrate as well, as suggested by the improved growth (3x), substrate utilization and product formation post-acid addition to lower pH to 6.0 as previously seen with maltose fermentations (described in chapter 3). Post PMF limitation alleviation and subsequent Avicel and base salts addition to the system (total of 30g/L avicel) showed overall high productivity resulting in breakdown and utilization of ~76% of the substrate

and accumulation of ~95mM products at pH 6.0 using additions of 1X base salts but without any additional ammonium chloride supplementation as has been previously reported [77]. Cultures on avicel also showed growth suppression patterns similar to when grown on maltose or with addition of NaCl. It was observed though that the total osmolarity values at severely suppressed fermentation stages with avicel were comparatively lower i.e. ~340mOsm than that of the maltose fermentations that approached ~450mOsm. Growth maintenance requirements on a complex solid substrate such as avicel are understandably higher than that on a liquid substrate like maltose as the cell would have additional energy costs for enzymes needed for breakdown of the substrate along with conversion for availability of enough substrate for physiological maintenance [218]. These higher maintenance requirements on Avicel are also reflected in the higher nitrogen requirements observed in such a system as seen in this study post acid addition and has also been previously described [77]**.** The pronounced effect of comparatively lower osmolarity values on the growth and metabolism of *C. bescii* growing on avicel in comparison to maltose could thus be attributed to this factor.

In conclusion it was observed that total osmolarity values of ≥ 350mOsm severely suppresses growth and metabolism in *C. bescii* on maltose as well as avicel as substrate or in the presence of NaCl leading to high osmolarity values of the growth medium. Moreover, as observed the degree of severity leading to growth lag, suppression or complete inhibition and thus the final osmolarities at end-point would vary based on presence of additional stress factors such as higher maintenance requirements on substrate or persisting PMF limitation. The results from this study would benefit and assist in designing fermentations that could be used to further study and overcome osmolarity

as a limitation as well as work directed towards evolving or engineering osmotolerant/robust *C. bescii* strains

Chapter 5.0

Investigating small RNA interference and its prospects in genome wide screening in *C. bescii*

5.1 Introduction

Extreme thermophiles with optimal growth temperatures of 70 to 100°C have found a new place in the realm of industrial biofuel production from biomass and currently attract the most attention in this area [219-221]. These organisms are often isolated from harsh environments where some of them survive on decaying organic matter like compost [29] and thus exhibit native metabolic capabilities for degradation and/or fermentation of cellulose as well as other complex polysaccharides. Some of the most studied organisms under this category include anaerobic bacteria from genera such as *Sulfolobus, Thermotoga, Thermococcus, Pyrococcus, Clostridium, Thermoanerobactor* and *Caldicellulosiruptor* [29, 30].

Caldicellulosiruptor bescii (*C.bescii*) from the genus *Caldicellulosiruptor* is one such organism with recently heightened interest for lignocellulosic biofuel production [29]. Interest in *C.bescii* and other species is motivated by their ability of possessing a combination of several useful properties together, including, the highest known temperature optima among described cellulolytic microbes, the ability to grow on a wide range of substrates and co-metabolize C5 and C6 sugars as well as the potential to solubilize un-pretreated biomass [2, 6]. Despite these attributes, *C. bescii* suffers from

several limitations such as inability of lignin degradation and fermentation inhibition under various stress conditions leading to incomplete substrate utilization and lowered product yields [2]. Moreover, to understand strictly anaerobic, thermophilic microorganisms their extreme conditions of growth need to be reconstructed, which can be a challenge [126, 127]. As a result, there are fewer tools developed or tailored for the exploration of anaerobic hyperthermophiles in comparison to vast diversity available for mesophiles which further impedes the growth of knowledge [128]. *C. bescii* and other similar microbes thus remain widely underexplored with the lack of variety (as available for many mesophilic microbes), in flexible genetic, high throughput tools being one of the major barriers to efficient production of industrially relevant amounts of biofuels, other products of interest and development of industrial variants. In the face of such challenges the need for tools that allow for deeper, wider and quicker high-throughput testing and manipulation of such organisms is generated.

Omics and next generation sequencing (NGS) are some of the currently used tools to study *C. bescii* [50, 97, 222] which can be used to identify changes in gene regulation and genetic variations such gene deletions/insertions, SNPs in an organism in response to a set of stimuli. In order to facilitate deeper and faster screening studies though, there is merit in tailoring techniques such as random transposon mutagenesis and RNA interference (RNAi) for use in these organisms. Such techniques can accelerate the process of population enrichment from a library of cells with inactive or attenuated genes (random/targeted mutagenesis) that might be a fitness advantage for survival in the conditions under study and provide relevant gene targets [223, 224]. Therefore, these can serve as tools for not only targeting genes of interest but also launching genome-

wide screening studies that can be made more efficient and high-throughput when used in combination with omics and/or NGS tools [225]. The recent development of a genetic system in *C. bescii* [113, 117] allowing for the basic manipulations such delivering reliable methods for inducing competence, genetic material transfer, gene expression and genome integration with relative ease now enables the development of these tools in this organism.

Some advantages of using RNAi mediated gene expression manipulation in comparison with other methods are that unlike gene knock-outs which results in loss of gene function and can cause lethality or a severe growth phenotype in the case of essential genes, RNAi allows for knock-downs (i.e. attenuation of the gene) and thus enables the study of such genes. Moreover, RNAi also could be used for tunable gene expression studies allowing development of strains with various degrees of gene function. RNA interference by antisense RNA (asRNA) molecules is widespread in nature and there are several examples of natural as well as synthetic RNA silencing or interference by small, antisense, non-coding RNA molecules in prokaryotes [226]. Antisense RNA sequences that hybridize to the target mRNA inhibit gene expression in several ways, but the mechanisms can be broadly grouped into translation repression and transcript degradation or a combination of both [226]. Masking of the ribosome binding site or stearic hindrance repelled ribosome attachment can lead translational repression of the transcript. On the other hand, a double stranded RNA due to hybridization of the asRNA to the target mRNA can set off a degradation or decay response by a number of RNases in bacteria [227]. RNaseP is one such example of a ribozyme that can be employed for the knock-down of target mRNAs.

The native function of RNaseP is cleavage of the 5'-leader sequence from pre-tRNAs[228], although over the years a wide range of other substrates have been identified in bacteria demonstrating that artificial substrates can be cleaved by RNase P [229]. Several studies have developed and used methods where either endogenous or exogenous RNaseP is employed to knockdown target mRNAs [230] using asRNA sequence called the external guide sequences [231, 232], which can target specific mRNAs. Guide sequences are designed with certain features to resemble the structure of the pre-tRNA substrates after hybridizing. Studies in *E. coli* have shown the use of an external guide sequence (EGS) that upon hybridization resembles the stem structure of the pre-tRNA [233] which is recognized by the endogenous RNaseP cleaving at the junction from single-stranded to double-stranded regions[234]. In these studies the following features were in-built in the EGS asRNAs such that it resembles the stem structure of pre- tRNA [230]. Sequence lengths between 13 and 16 base pairs (bp) are used as longer sequences have been shown to reduce efficiency [234]. These sequences from the targeted gene or region are fused to the RNaseP recognition sequence “RCCA” where ‘R’ is generally the purine adenine ‘A’) at the 3'-end, this sequence is present on natural pre-tRNA substrates and was found to be essential for the catalytic activity in *E. coli.* In this recognition 3'-RCCA sequence the cytidines are known to bound a GG sequence pair in the catalytic domain of RNaseP (RNA in the RNaseP complex). Other than the directed or targeted approaches mentioned above RNAi library bases screening tools have also been described in bacteria like *Staphylococcus aureus* where an RNAi library was made using randomly fragmented genome for the screening of essential genes in the organism [224].

RNA interference or silencing using antisense small RNA molecules to our knowledge have not been previously described in thermophiles. Such a technique would be an extremely valuable addition to the genetics tool box for organisms like *C. bescii* and calls for attempts at its development in such relatively underexplored organisms. This is whether to be used to target specific genes or developed into a genome wide screening tool. Here I describe and detail my attempts exploring the occurrence of RNA interference phenomenon in *C. bescii* and the possible attributes of designing an RNAi based genome-wide screening tool. As previously mentioned several strategies with asRNA have been tried, tested and optimized for obtaining RNA interference in other bacteria [227, 230]. In this study, I investigated two of these strategies namely, external guide sequence (EGS) RNAseP mediated transcript degradation [232] as “approach I” which uses gene of interest (GOI) asRNA fragment designs 01 – 03 (Figure 25) for RNA interference screening and expressed regular asRNA [227] in “approach II” for designing random genomic fragment RNAi library tool (Figure 25). With inbuilt parameters during the design of the antisense fragments the effect of ribosome binding site (RBS) sequence in the fragment and length of the fragment were also tested (Figure 25). As ultimately, the experiments and the approaches investigated in this study had limited success, I will also describe possible further approaches.

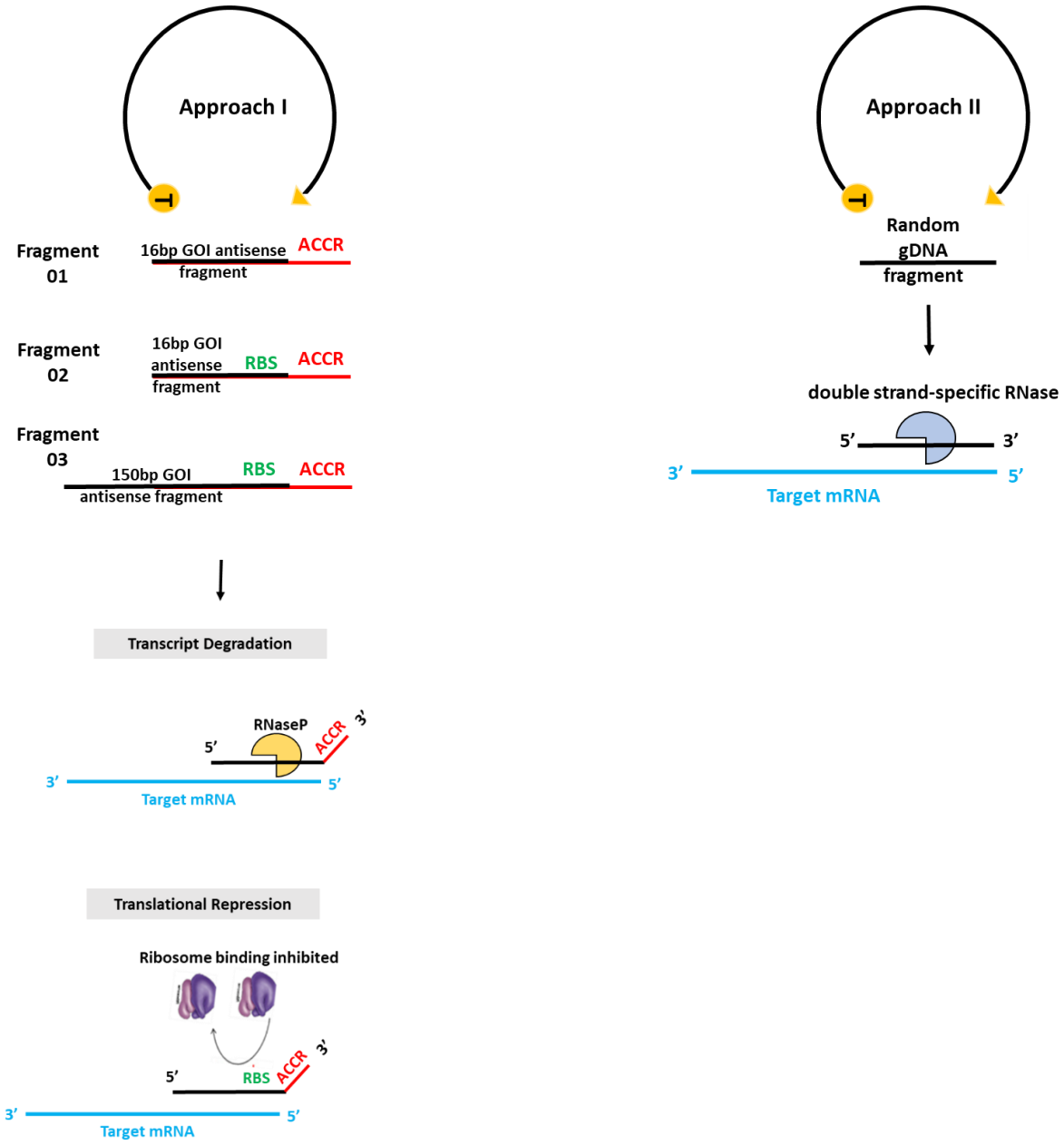


Figure 25: Overview of the antisense fragment designs depicting the in-built parameters tested in the two broad approaches (EGS, Approach I and random genomic RNAi library, Approach II) used in this study and the possible RNA interference causing mechanisms that may follow.

5.2 Materials and Methods

5.2.1 Strains, Media and Culture conditions

Experiments were performed using *C. bescii* strains JWCB049 (*ΔpyrFA Δldh CIS1*:: $P_{S\text{-}layer}$Teth39_0206) [112], JWCB018 (*ΔpyrFA ldh::ISCbe4 Δcbe1/(ura−*/5-FOA^R)) [6] and JWCB032 (*ΔpyrFA ldh::ISCbe4 Δcbe1*::PS-layer *Cthe-adhE2/(ura/*5-FOA^R)) [235]. All strains were grown anaerobically in liquid or on solid surface in LOD medium [113], pH was adjusted to 6.8 using 8M NaOH, with maltose [0.5% (wt/vol); catalog no. M5895, Sigma) as the carbon source unless otherwise noted. Liquid cultures were grown from a 0.2% inoculum or a single colony and incubated at 75 °C or 65 °C in anaerobic culture bottles or Hungate tubes. Media was reduced in the Coy anaerobic glove bag (COY, Grass Lake, Mich., USA). *E. coli* strain DH5α was used for plasmid DNA constructions and preparations. Standard techniques for *E. coli* were performed as previously described [236]. *E. coli* cells were grown in LB broth supplemented with apramycin (50 µg/mL) and plasmid DNA was isolated using a Qiagen Miniprep Kit. Chromosomal DNA from *C. bescii* strains was extracted using PowerLyzer® PowerSoil® DNA Isolation Kit (MO Bio Laboratories Inc.) according to the manufacturer's instructions.

5.2.2 Investigative Design and Methodology

Two approaches were used with the objectives of systematically investigating occurrence of the RNA interference phenomenon in *C. bescii* and designing as well as evaluating the plausibility of an RNAi based genome wide screening tool development were employed in this study (Figure 26a and 26b).

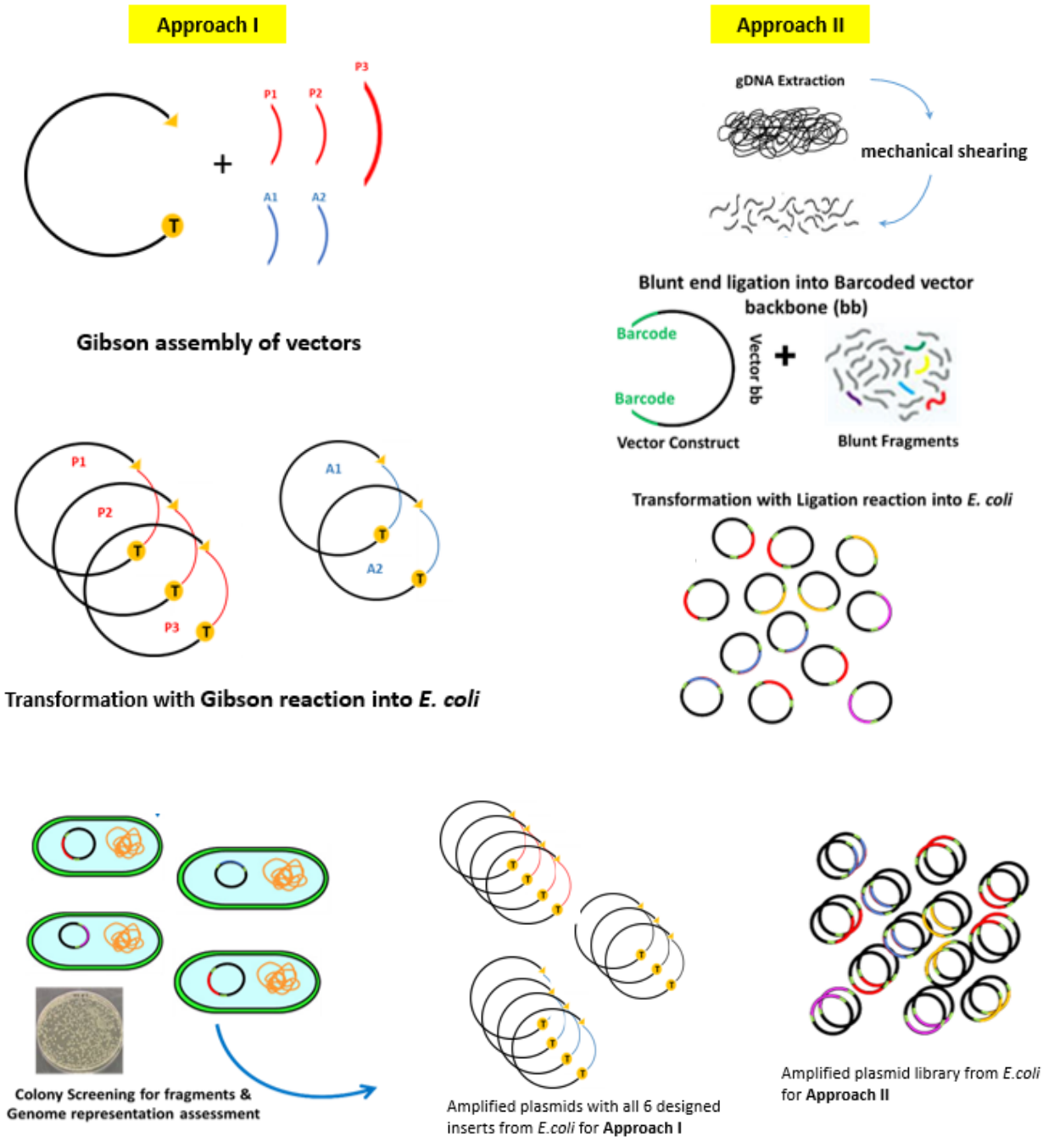


Figure 26 a: Design and methodology detailing the RNaseP directed transcript degradation (Approach I**) used in the study to screen *C. bescii* for RNAi effect, possible mechanism(s) and the random genome fragment-based RNAi library approach (Approach II) attempting development of RNAi based genome wide screening tool.

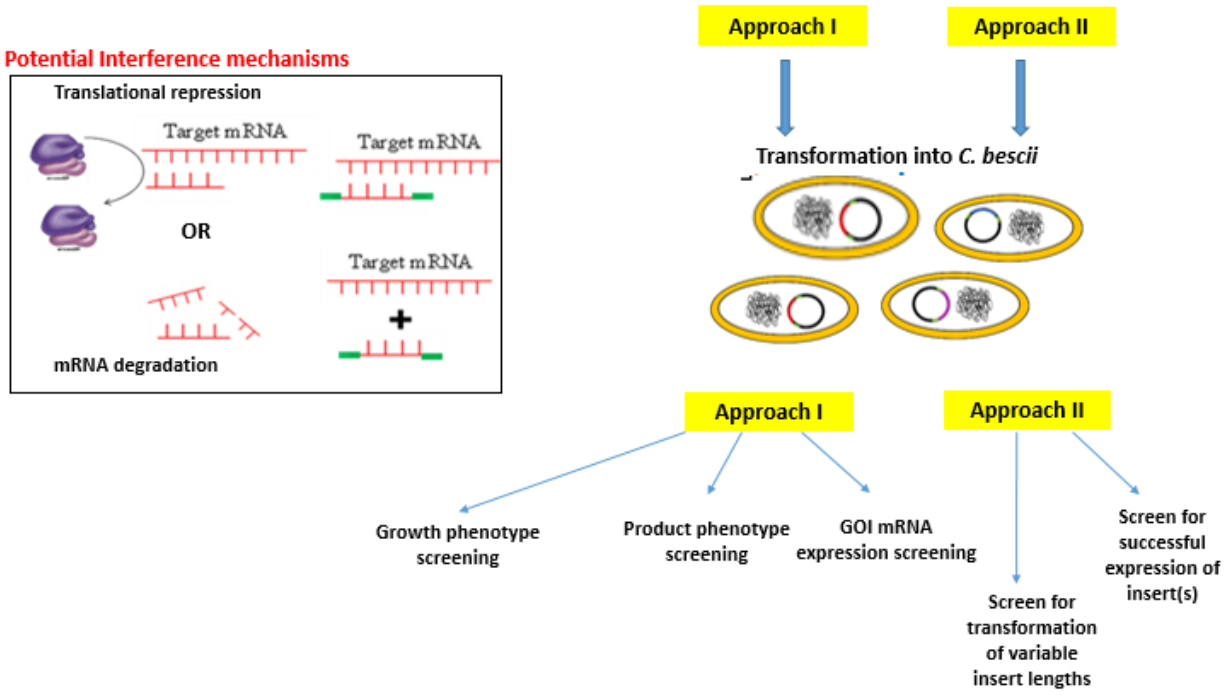


Figure 26b: Continuation of the design and methodology from figure 26a, depicting steps following transformation of the RNAi fragment carrying plasmids into *C. bescii*.

** In Figure 26a, P1, P2, P3, A1 and A2 are designed RNAi fragments (sequences are detailed in Figure 27a) and their corresponding plasmid constructs for the RNaseP directed approach (approach I). In the depiction, **"T"** denotes terminator and the Promoter is denoted as a triangle.

Approach I used a targeted RNAi scheme where antisense fragments for two genes of interest (GOIs) from the acetate production pathway namely phosphate acetyltransferase (pta), *Athe_1494* and acetate kinase (*ack*), *Athe_1493* were designed to be introduced in the plasmid backbone 'pMi'. DNA fragments were designed to test for the following mechanisms of RNA interference in bacteria as depicted in figure 27a. In the first design that used, RNaseP directed transcript degradation, 20 bp antisense fragments (sequences as listed in figure 27a) from the GOI were designed to have a four base pair RNaseP recognition sequence (ACCA) on the 3'-end of the antisense transcript as previously described as essential to process [230]. Such fragments were labeled GOI01 (pta01/ak01). The second type of fragment design tested for effects of ribosome binding site masking leading to translation inhibition which may or may not lead to transcript degradation along with the RNaseP directed degradation. These fragments were also 20 bp in length consisting of the RNaseP recognition sequence and 16 bp from the region upstream GOI till the start codon and labeled GOI02 (pta02/ak02). The third design took length of the interfering fragment into consideration and was designed longer i.e. to be 150 bp including 4 bp RNaseP recognition sequence, 46 bp from the upstream region of the GOI and 100 bp downstream. These were labeled GOI03 and only tested for pta (pta03). A heterogeneous sequence from a plasmid pRL27 [237] was inserted between the promoter and terminator as a unique identifier in the negative control vector, which could be used to confirm fragment expression from the designed plasmids. Post-transformation in *C. bescii* the transformants with each of these GOI constructs could be compared with the parent and negative control for growth and/or product phenotype and screened for GOI expression variation if any.

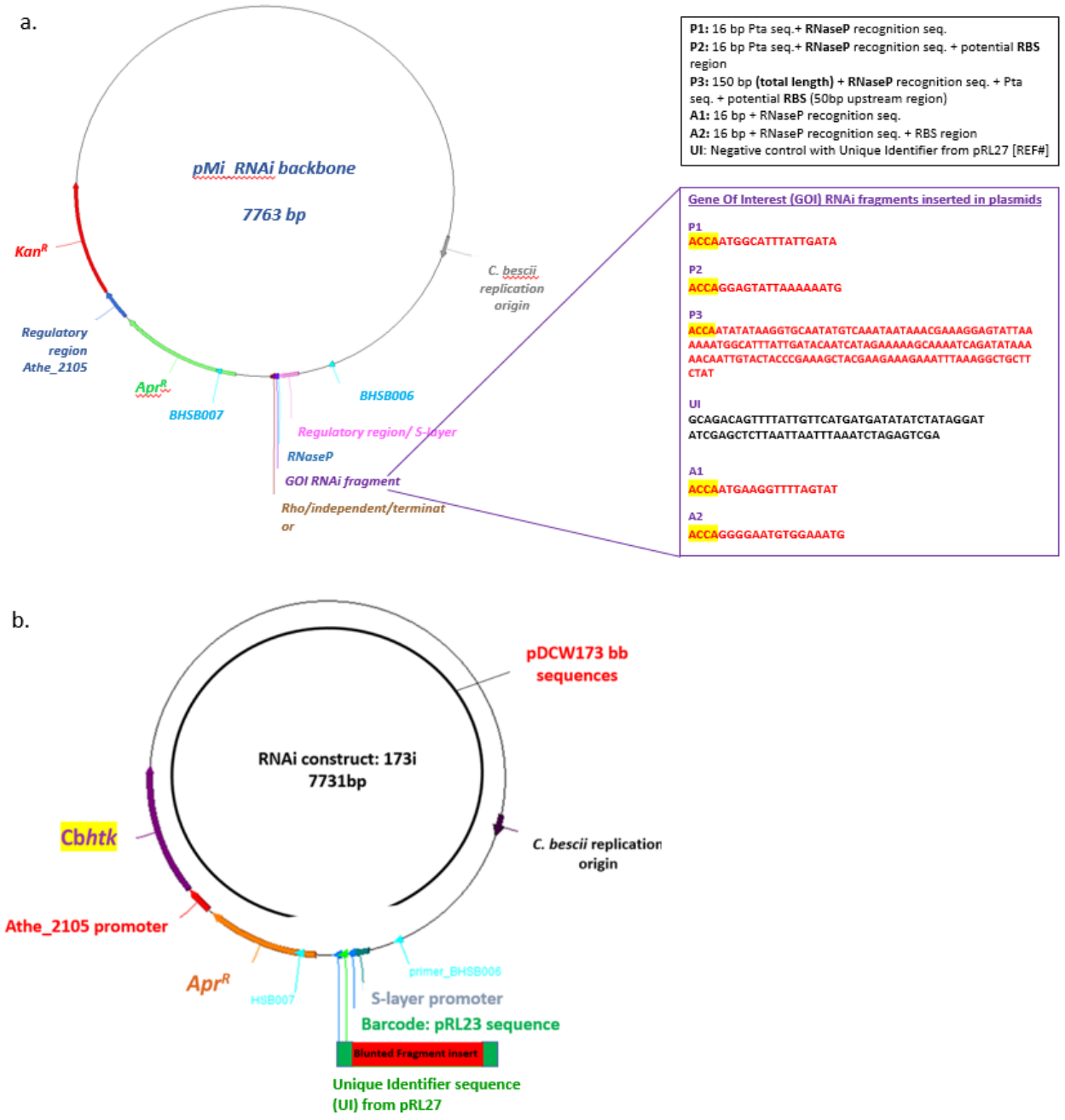


Figure 27: RNAi vector construct designs. Unmarked region of the constructs consists of backbone components from pDCW173. (a) Construct used in “Approach I”: Validation of RNA interference and modes of action using two genes of interest fragments Pta (Phosphotransacetylase) and AK (Acetate kinase). (b) Construct used in “Approach II”: Random genomic fragments based RNAi library development.

Approach II on the other hand aimed at simultaneously establishing parameters for designing an RNAi based genome-wide screening tool. Here random genomic fragments were blunt-end ligated into the plasmid backbone flanked by a unique identifier sequence that can be used to selectively sequence the inserted fragments (Figure 26a) to create a random genomic fragment library which can be amplified by passaging through *E. coli* and then transformed in *C. bescii.* Direct ligation reaction mixture containing blunt-end ligated random genomic fragments to be used which results in lowered transformation efficiencies, and since transformation efficiencies in *C. bescii* are very low to begin with (10^4 cfu/µg) *E. coli* was used for amplification of blunt-end ligated plasmid library. This would generate a random fragment RNAi based library, the percentage genome representation in the library could be determined as a quality check parameter before the library is used in an enrichment study under different stress or environmental conditions of interest followed by sequencing to identify genetic elements that confer robustness under conditions tested. Approach II was set out to test and detail the logistics of the design of such a tool by screening for the following: representation of not a single but variety of fragments of different sizes and from regions of the genome, expression of the fragments inserted in plasmid, estimating the required number of *E.coli* and *C. bescii* transformations needed to get a 99.99% coverage of the genome with statistical confidence and if that were theoretically plausible going forward.

5.2.3 Construction of Vectors bearing RNAi Fragments

The plasmids for this study were generated using a previously designed shuttle vector for *C. bescii* pDCW173 [119] as backbone for majority of the construct. Regions of the pDCW173 containing components of protein CelA were removed, *pyrF* gene conferring uracil prototrophy as selective marker was replaced with kanamycin resistance gene, Cb*htk* (marked yellow in Figure 27b) [125, 238] under the control of *Athe_2105* promoter to get the vector (pDCW173i) used for vector backbone amplifications for this study. pDCW173i was achieved via a three piece Gibson assembly using primer pairs 173::Cbhtk_01 and 173::Cbhtk_02 (Table 1) for two segments amplified from pDCW173 and kanamycin resistance gene gblock fragment that was codon biased for *C. bescii* as previously described [125] further referred to as Cb*htk* in the manuscript. The plasmid pDCW173i (173i) was then further used to amplify the backbone for vector used in approach I (Figure 27a), followed by a two-piece Gibson assembly utilizing the primers as detailed in Table 17. This resulted in the generation of six plasmids: pMi_pta01 (P1), pMi_pta02 (P2), pMi_pta03 (P3), pMi_ak01 (A1), pMi_ak_02 (A2) and pMi_UI (UI) (primers are listed in Table 17). Using primers 173i_UI01 and 173i_UI02 the backbone vector for approach II was amplified and used for blunt-end ligations of random genomic fragments into the vector backbone Figure 26b.Random fragmentation of the genomic DNA extracted from JWCB049 was achieved using a Bioruptor® Plus (Diagenode) sonication device as listed in manufacturer's protocol to achieve 250bp fragments.

Table 17: List of primers used in the study (*gibson overlap regions are in written in red).

Primer	Sequence	Application
BHSB006 BHSB007	GCCGCATCTGAGAGTT CTACGGAAGGAGCTGTG	Screening primers used for checking presence of insert and sanger sequencing
003_01 003_2	GTGTCTGCTGATGGAAGGTATG GGGCTTGTCTGCGATGTTAT	Housekeeping gene RT-qPCR primers
UI_01 UI_02	GCAGACAGTTTTATTGTTCATGAT TCGACTCTAGATTTAAATTAATTAAGAGC	RT-qPCR primers for negative control vector with unique identifier sequence
Pta_01 Pta_02	GCAATAACCCCCACAATGTC TGCAGGACAAGCAAATGTTC	RT-qPCR primers for phosphate acetyltransferase
Ak_01 Ak_02	GCAAACCAATCCTGTCAA AAGGGGAATGTGGAAATG	RT-qPCR primers for acetate kinase
173::Cbhtk_01 173::Cbhtk_02	TCTTCTCTAGTCATAATTATAGGTCCTTTCA Tctcaccaaacctccttgtatg AGTAGATGTTAGCAAAAGAATACCATTTTG Aggaacgaaaactcacgttaag	Gibson assembly primers for construction of 173i
pMi_bb01_all genes pMi_UI07_bb02 pMi_pta01_bb02 pMi_pta02_bb02 pMi_pta03_bb02 pMi_ak01_bb02 pMi_ak02_bb02	TAAACTGTAATCAAACTGTcgatcgattcccatga gcccacgaacagtaa AGCTCTTAATTAATTTAAATCTAGAGTCGAg agggtgagattgattctca AGGTTTGGTGAGACCAATGGCATTTATTG ATAgagggtgagattgattctca AGGTTTGGTGAGACCAGGAGTATTAAAAA ATGgagggtgagattgattctca TACGAAGAAAGAAATTTAAAGGCTGCTTC TATgagggtgagattgattctca AGGTTTGGTGAGACCAATGAAGGTTTTAG TATgagggtgagattgattctca AGGTTTGGTGAGACCAGGGGAATGTGGA AATGgagggtgagattgattctca	Gibson asssembly primers (bb01_all genes with bb02 primer for each GOI) used for Approach I vector construct
173i_UI01 173i_UI02	**CTATAGATATATCATCATGAACAATAAAAC** **TGTCTGCctcaccaaacctccttgtatg** **GATATCGAGCTCTTAATTAATTTAAATCTA** **GAGTCGAggaacgaaaactcacgttaag**	Gibson asssembly primers for construction of vector used in Approach II with unique identifier overhangs

5.2.4 Transformation, Screening, Plasmid Purification, and Sequence Verification

Constructed plasmids, as described above, pMi_pta01 (P1), pMi_pta02 (P2), pMi_pta03 (P3), pMi_ak01 (A1), pMi_ak_02 (A2) and pMi_UI (UI) as well as the random genome fragments containing plasmid library post ligation were transformed in *E. coli* DH5α competent cells as per manufacturer's protocol from the NEB Gibson assembly and the NEB Quick Blunting and Ligation Kits. Colonies from the overnight LB media plates containing apramycin (50µg/mL) were PCR screened using primer pair BHSB006 and 007 using phusion polymerase HF master mix (NEB) to select for single colonies harboring the required plasmids. The primers used for screening and sequencing of all constructs made are listed in Table 17.

The screened single colonies were grown overnight in liquid selective media with apramycin (50µg/mL) and used to extract and purify plasmids using Qiagen Miniprep kit followed by Qiagen PCR purification kit. Using 1 µg of the purified plasmid DNA *C. bescii* strains JWCB032 (approach I: Targeted RNAi) JWCB018 (approach II: RNAi library) were electro-transformed as described previously [235]. Cells were then plated onto solid LOD medium and 40µM uracil and kanamycin (50µg/mL). Subsequently single colonies were picked and screened by PCR for the region of GOI or insertion of random sequence at the blunt-end ligation site between the unique identifier sequence. PCR products were Sanger sequenced to confirm gene of interest fragments in the single colonies of confirmed transformants.

5.2.5 Growth and Product Analysis of Mutants for Phenotypic Effects

Analysis of growth was conducted for cultures grown in Hungate tubes containing 10 mL LOD medium supplemented with 5 g/L Maltose catalog no. M5895, Sigma) and 40µM uracil. Triplicates were inoculated with a fresh 0.2% (vol/vol) inoculum and incubated at 65 °C and stationary for growth curves. Growth was estimated by monitoring optical cell density using a spectrophotometer (Thermo Scientific™ GENESYS™ 20), measuring absorbance at 680 nm. Fermentation products maltose, glucose, acetate, lactate and ethanol, were monitored and analyzed using high-performance liquid chromatography (HPLC) quantification with an Aminex™ HPX-87H column (Bio-Rad Laboratories Inc.) at 0.5 mL/min flow rate of 5 mm H_2SO_4 mobile phase and column temperature of 60 °C. The signal was quantified on a Shimadzu Prominence LC-20A Series system (Shimadzu Scientific Instruments, Columbia, MD) equipped with a refractive index detector (model RID-20A) at 35°C. Integrated total peak areas were compared to peak areas and retention times of known standards for each compound of interest.

5.2.6 RNA Extraction and Expression Quantification

High quality RNA was isolated from the parent strain (JWCB032) and transformants with the RNAi fragment containing vectors as well as the negative control vector (RIN > 7.2). Briefly, cell pellets from each sample were incubated in 10 mg/mL lysozyme (Sigma Aldrich part number L-7651, St. Louis, MO) resuspended in SET buffer (50 mM Tris-HCl pH 8.0 50 mM EDTA, 20% w/v Sucrose) followed by incubation at 37^{o}C for eight minutes, vertexing briefly every two minutes. After incubation, 900 µL of buffer RLT from a Qiagen RNeasy Kit (Qiagen, Hilden, Germany) was added to the lysozyme cell mixture and

vortexed. Then 650 µL of 100% ethanol was added to each aliquot and mixed by pipetting. Each sample was then applied to a RNeasy spin column (Qiagen, Hilden, Germany) and processed following the manufacturer's protocol included the optional on column Dnase I treatment. Samples were eluted in 35 µL RNAse free water (Qiagen, Hilden, Germany). The quantity and quality of RNA was assessed post-elution using a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE) and Agilent Bioanalyzer (Agilent, Santa Clara, CA, USA). First cDNA strand synthesis from the total RNA extracted was performed using the SuperScript™ III First-Strand Synthesis System (Invitrogen™) following manufacturer's protocol. Quantification of gene expression for comparison between parent and, negative control and RNAi fragment harboring single colony transformants was done using real-time quantitative PCR (RT-qPCR). Bio-Rad MyiQ2 Two-Color Real-Time PCR Detection System (Bio-Rad Laboratories, CA) and Roche FastStart SYBR Green Master (Roche Applied Science, IN) were used for this experiment. The primers targeting the genes in this experiment are listed in Table 1, DNA polymerase III subunit beta (*Athe_0003*) was used as the housekeeping gene.

5.3 Results

5.3.1 Growth and Product Profiles for Targeted RNAi

RNAi fragments from the two genes (*pta* and *ack*) were targeted to see a growth or metabolite (i.e. with actetate and ethanol) phenotype (Figure 28). Screened single colony isolates of confirmed RNAi transformants harboring the plasmids: pMi_pta01 (P1), pMi_pta02 (P2), pMi_pta03 (P3), pMi_ak01 (A1) and pMi_ak_02 (A2) and pMi_UI (UI,

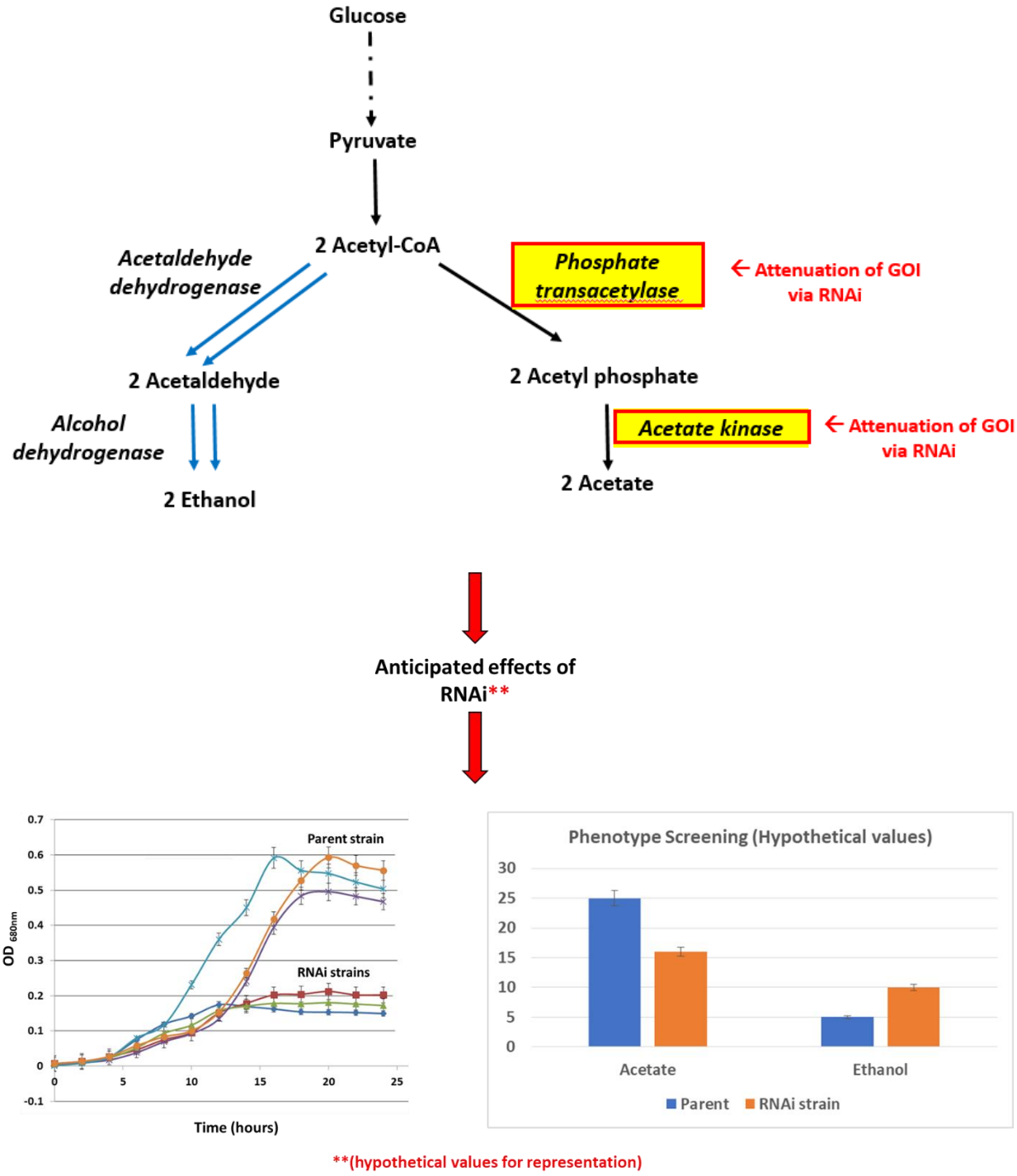


Figure 28: Effects of RNAi on target genes of interest *acetate kinase (ack)* and *phosphate trasacetylase (pta)* anticipated to result in reduced acetate and increased ethanol levels as depicted using hypothetical values (pathway modified from [239]).

negative control) when grown in triplicates at 65 ^{o}C, stationary and compared with the parent strain (JWCB032) did not show any significant growth phenotype (Figure 29a). The parent strain did reach a slightly higher max OD (0.26) in comparison to the other RNAi transformants (0.22) but the final ODs for all were similar. Product profile analysis of all the RNAi fragment containing transformants and the parent strain also indicated no significant differences upon comparison of the acetate and ethanol produced (Figure 29b). The cultures of all transformants from the growth curve showed the presence of the plasmids with the representative RNAi gene fragment upon PCR verification, post growth curve monitoring in all biological replicates (Figure 30). This shows that the RNAi fragment containing constructs were present and maintained during the growth curve.

5.3.2 Target Gene Expression Comparisons and Expression of Insert Fragments

RT-qPCR estimated expression values of targeted genes *pta* and *ack* normalized to the housekeeping gene (*Athe_0003*) in the RNAi transformants for all fragment designs. The comparison of relative expression values ($\Delta\Delta$Ct method) displayed similar expression of the target gene *pta* between the parent (JWCB032) and the negative control (UI) for all transformants with the three RNAi fragment designs (P1, P2, P3) (Figure 31a). The first RNAi fragment design (A1) containing transformant for the target gene *ack* also showed similar expression, but the transformant with the second type of RNAi fragment (A2) displayed between 19-23% mRNA expression decrease with for the gene (*ack*) when compared with the parent (JWCB032) and negative control (UI) using three individual biological replicates of each (values from each represent an average of three technical eplicates used in RT-qPCR) for the comparison and was tested to be statistically relevant

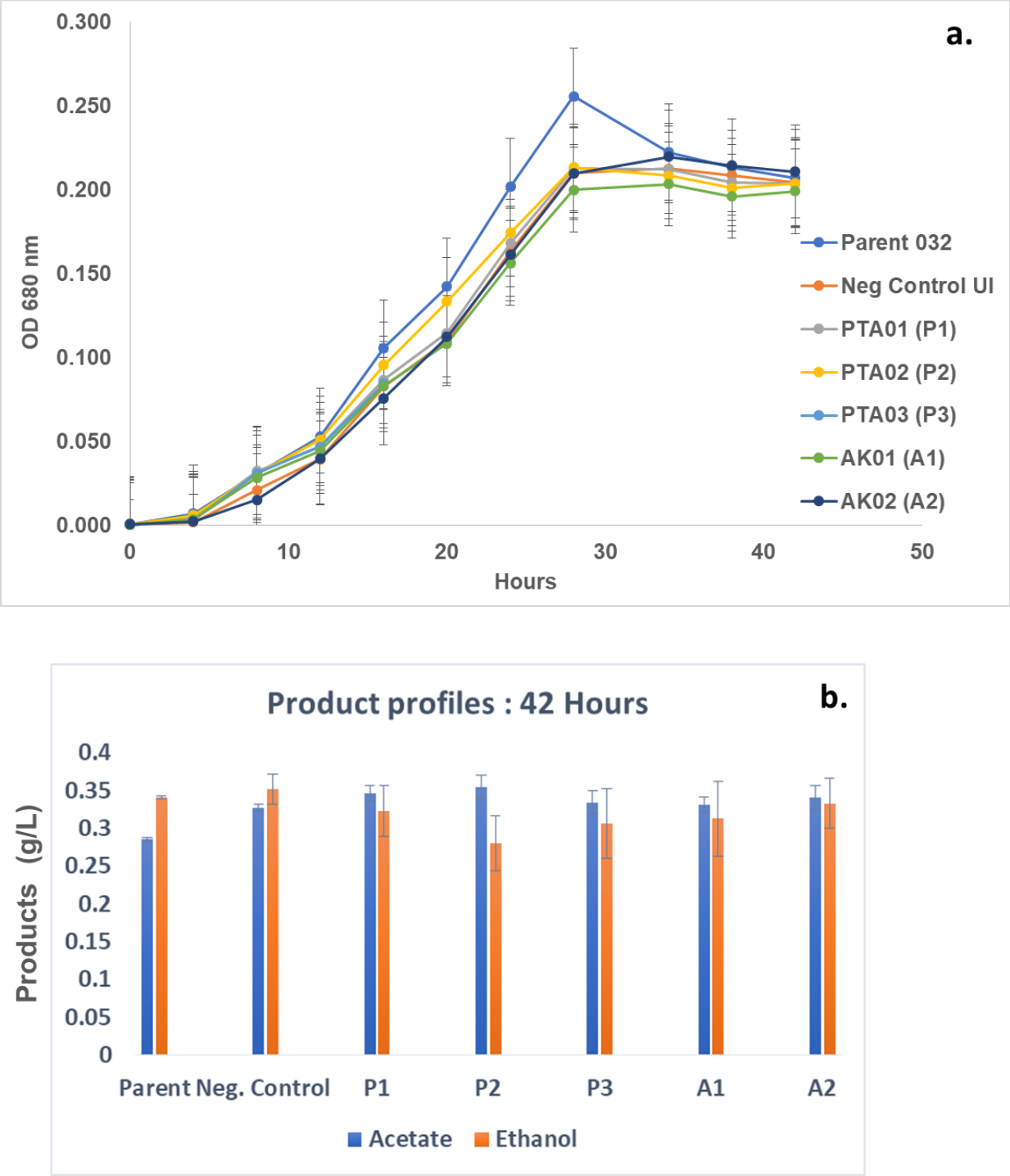


Figure 29: No significant growth (a) or product (b) phenotype was observed in the RNAi plasmid transformed single colony isolates (P1, P2, P3, A1, A2) in comparison to the negative control (UI) and parent (JWCB032).

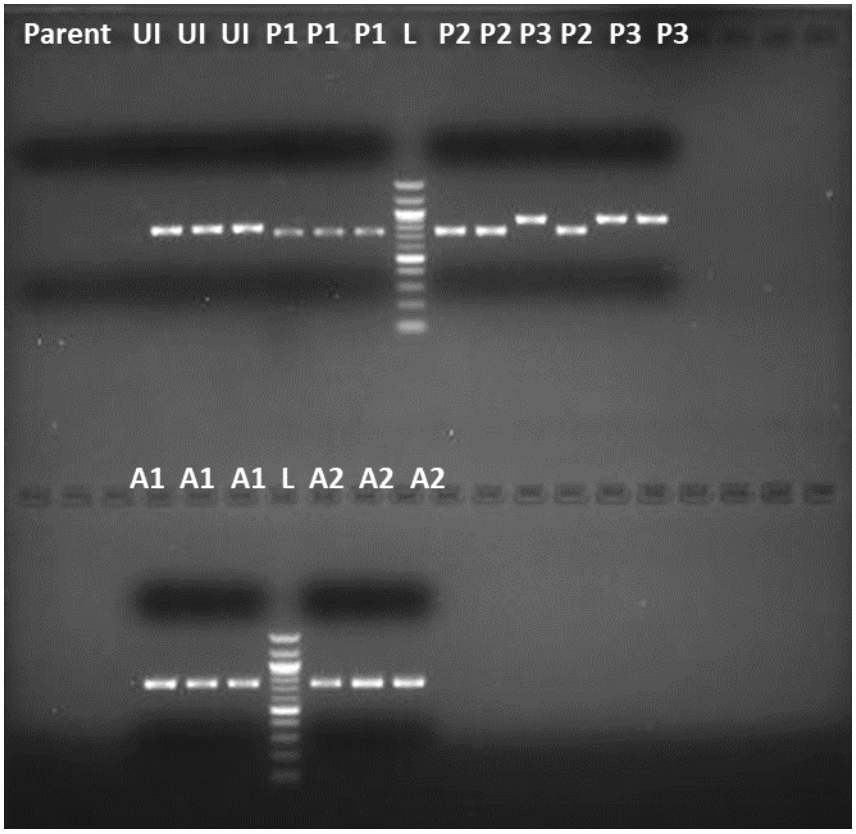


Figure 30: All replicates of the single colony transformants from the growth curves with RNAi fragments for targeted genes *pta* (P1, P2, P3) and *ack* (A1, A2) show presence of stable RNAI fragments upon PCR screen using primer pair BHSB006/ BHSB007 and NEB 100 bp ladder (L).

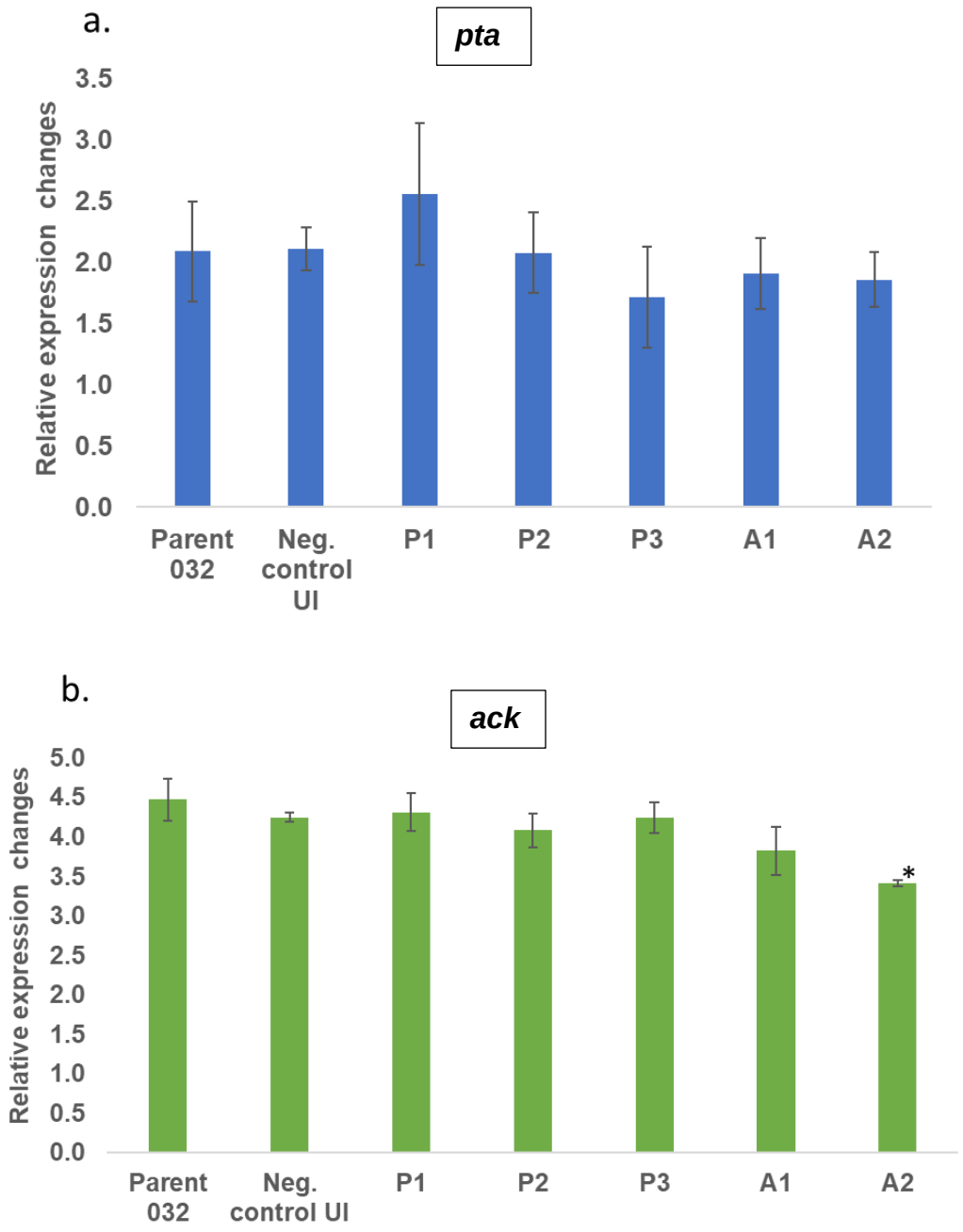


Figure 31: qRT-PCR estimated mRNA expression levels of target genes *pta* (a) *ack* (b) showed no significant expression changes for the gene *pta* but transformant A2 for *ack* depicted significant mRNA reduction (~20%) in comparison to both parent and negative control.

using student's T-test (two -tailed, paired) (Figure 31b). The expression of the heterogeneous fragment (UI) from pRL27, inserted between the S-layer promoter and terminator in plasmid was confirmed via PCR amplification (primers UI_01 and UI_02) of the fragment from the cDNA obtained from the reverse transcribed total RNA of the negative control (Figure 32).

5.3.3 Attributes of RNAi based Genome-wide Screening Tool Development

High quality genomic DNA (Figure 33a) which was extracted from *C. bescii* strain JWCB049 (*ΔpyrFA Δldh CIS1:: PS-layerTeth39_0206*) and fragmented using bioruptor sonication, generated ~150-700bp fragments (Figure 33b). Blunted genomic fragments ligated into the backbone vector 173i were successfully transformed into *E. coli* cells with a transformation efficiency with ligation reaction ~ 1.5 x 10^4 cfu/µg (using ligation reaction). All colonies from the plate were inoculated in LB media (100mL) with apramycin overnight and plasmids were extracted and concentrated followed by transformation of 1 µg of DNA in *C. bescii* by electroporation. Transformation efficiency (TE) of 2.25 x 10^5 cfu/µg (qualifies as high TE for *C. bescii*) was achieved with the empty RNAi vector in *C. bescii*. A 50% insert to backbone ratio was observed upon screening 50 colonies from two recovery cultures plated (Figure 34) for fragments using flanking unique identifier sequence specific primers.

Using the following formula as previously described [224, 240] the required number of clones needed to ensure the representation of the entire genome in the library with >99.99% confidence was calculated to be 4.7 x 10^5 clones (using equation 4.0). Furthermore, taking into consideration the above required number of clones for as high

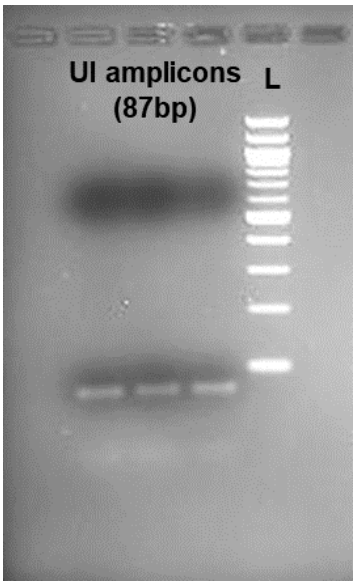


Figure 32: Unique identifier sequence (87 bp) amplified using cDNA (primer pair UI_01 and UI_02) from three biological replicates of the single colony isolate and run on agarose gel against 100 bp ladder (L). Amplification of the UI sequence (Sanger sequenced for confirmation) from the cDNA of the negative control plasmid containing indicates expression from the plasmid.

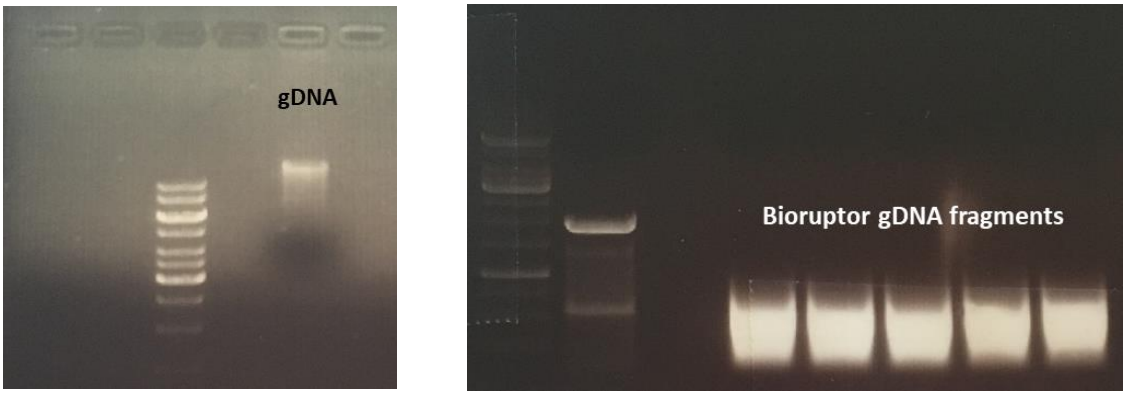


Figure 33: Extracted gDNA (a) on 0.8% and fragments generated via bioruptor sonication method (b) on 1 % agarose gel.

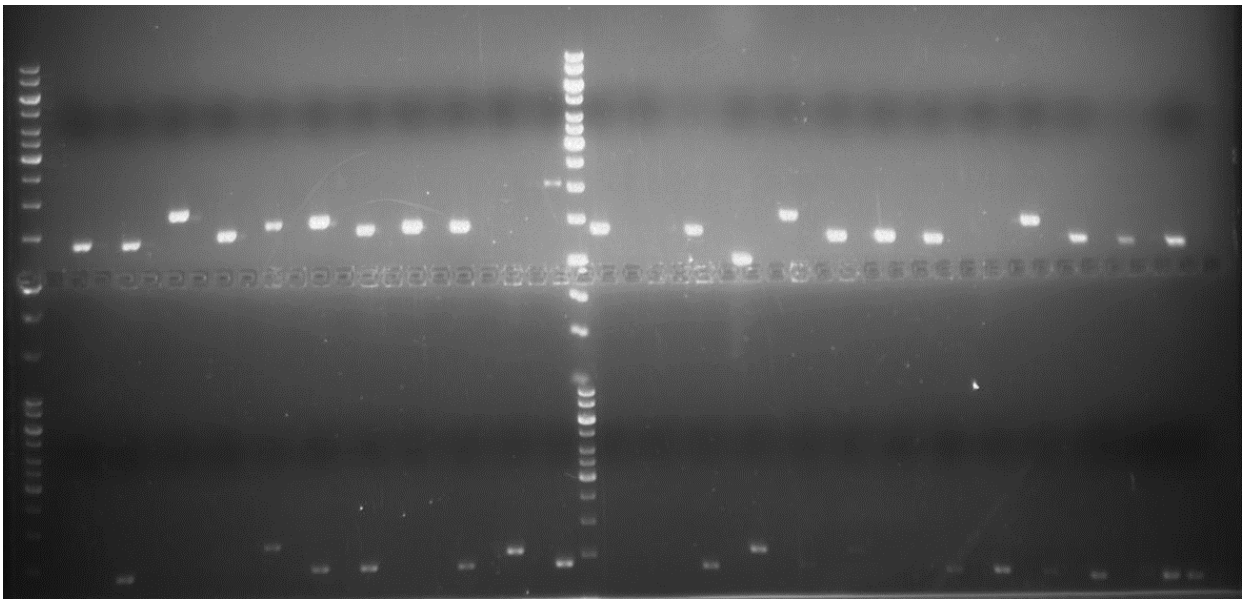


Figure 34: Fifty transformed *C. bescii* colonies screened for inserted genomic fragment between unique identifier sequence overhangs in the plasmid construct. Insert to backbone ratio was estimated based on band sizes. Bands ≥ 250-750 bp compared with 1kb (o'gene ladder, used to evaluate inserts and ≤ 100 bp for backbone.

as 99.99% (theoretical) coverage of genome in the library, transformation efficiencies from *E. coli* and *C. bescii* as well as the 50% insert to fragment ratio, it was estimated that for the complete RNAi library construction: ~50 *E. coli* and ~6 *C. bescii* transformations will be required.

$$N = \left(\frac{\ln(1 - P^{f})}{\ln(1 - f)}\right) \qquad \text{[Equation 4.0]}$$

where, $f = \left(\frac{a/b}{2}\right)$; a= Insert fragment size; b= Genome size; P= Probability of finding a given insert in the representative library of clones; N= Required number of clones in library.

Fragment expression and insert sequence were validated using three single colony isolates (#11,16 and 21) with inserts and one with backbone colony (# 35) chosen based presence of fragment bands of varying sizes (Figure 35). Sanger sequencing across unique identifier sequence for colonies 11, 16 and 21 mapped to *anthranilate ribosylphophotransferase* (*Athe_1694*) and two hypothetical genes (*Athe_2621* and *Athe_2747*) respectively upon BLAST analysis. No fragment sequence was found between barcodes for backbone suspected colony isolate #35.

Amplification of fragment using UI sequence specific primers from colonies 16 (with insert) and 35 (backbone – no insert) as indicated by amplification curves of RT-qPCR from colony isolated cDNA confirmed transcript expression from the plasmid housed in each colony. Melt curves for the two colonies confirm difference in transcript sizes as indicated by higher melting temperature for colony #16 where fragment was present and

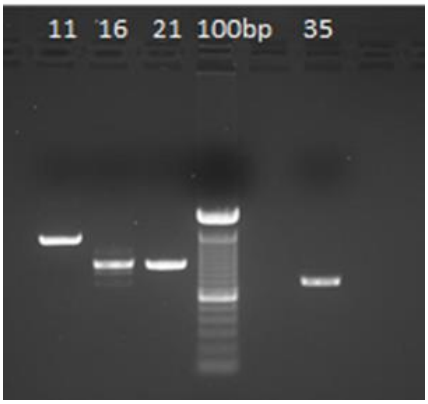


Figure 35: PCR bands for four single colony isolates #11, 16, 21 and 35 (primer pair UI01 and UI_02), Sanger sequenced. Sanger sequencing fragment sizes 663 bp, 243 bp, 228 bp and 74 for # 11, 16, 21 and 35 respectively.

expressed in comparison to lower melting temperature of the amplicon from colony #35 which lacked an insert and only had the UI sequence present in the “empty” vector. (Figure 36). Ct values for amplicon from colony #16 shows fragment expression under S-layer promoter to be similar to the house-keeping gene (denoted as: DNA pol_003) (Figure 36).

5.4 Discussion

In this study the attempts at investigating RNA interference in the anaerobic thermophilic organism *C. bescii* using asRNA-mediated sequence specific recognition approaches are described. Results from the RNaseP mediated approach, using EGS fragments (Approach I) in this study showed for both the targeted genes (*ack and pta*) the plasmids containing the designed asRNA fragments were transformed, stable and expressed in cultures growing at 65^0C. Observable growth or product phenotype were not seen for any of the fragments designed for either of the two targeted genes. To explain this, I note that the mRNA levels and changes in an organism do not directly and quantitatively correlate to the protein level changes, thus if the mRNA levels for a gene drop the protein levels may not drop by the same degree. Parallel proteomics analysis might provide a deeper understanding. Also, if the gene being targeted in a metabolic pathway is not the rate limiting gene, the mRNA level reduction may not change the overall pathway flux to the end-product, unless the decrease in the expression level is low enough to result in that step then becoming the limiting step. In a pathway leading to a major end product with the contribution from several genes sometimes one gene may compensate for the

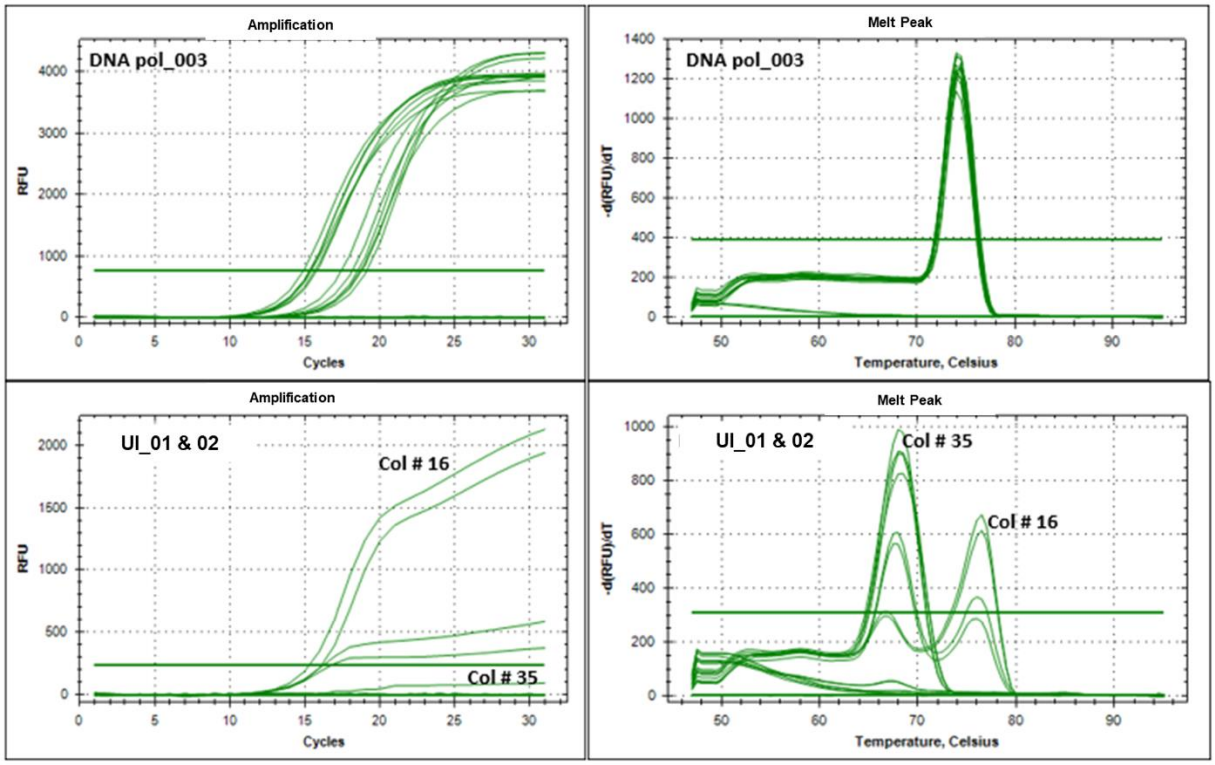


Figure 36: Fragment expression and insert sequence length differences validated using RT-qPCR amplification and melt curves respectively. Single colony isolates chosen based on preliminary PCR screens for presence of insert fragment (colony #16) and absence of insert fragment (colony #35) i.e. backbone between unique identifier sequence (UI).

function of another, again keeping the metabolic flux unchanged or there may be more than one copies of the targeted gene to subdue the effects of repression by RNAi. Reasons such as these can account for an absence of a phenotype in RNAi experiments. Upon comparing gene transcript levels for the target genes between parent and RNAi fragment harboring strains, a transcript reduction of ~20% was observed for one fragments ("A2" -16bp with RNaseP recognition and RBS sequence) for *acetate kinase* (*ack*) in three biological replicates. This provides some evidence for successful RNA interference and suggests promise in the development of this technology in *C. bescii* with further experimentation and troubleshooting. Protein estimations and comparisons between parent and RNAi fragment containing strains may be of value as a future experiment to further the results of this study and provide more insight into the changes in the protein levels. Additionally, it may be of importance to test the native RNaseP expression and/or activity in *C. bescii* as approach I relied on the presence of a native RNaseP (Athe_R0019).

RNA interference is a complex phenomenon [227] and there are several factors that can contribute to the successful reduction of mRNA expression. Some of these factors are: target accessibility that can depend on the target mRNA structure as well as turnover duration [241, 242]; length and structure of asRNA fragment [243]; binding kinetics [244-246], intracellular concentration of RNAi fragment [247, 248], degradation resistance of the chosen antisense molecule[227]. Taking these factors into consideration while troubleshooting or re-designing the RNAi study will be important for further research. In addition, to the factors that affect RNAi efficiency there are also many different methods/approaches/mechanisms by which transcript expression can be altered such as

RNaseP mediated EGS [230], expressed regular antisense[227], RBS blocking to block translation, stearic hinderence of translation using RNA with large molecules such as PMO (phosphorodiamidate morholino oligomer) [249, 250], catalytic RNA with attached ribozyme [251-253] and CRISPR mediated [254]. Since there are so many factors, mechanisms and parameters that need to be tested and can affect the efficiency of interference differently in different organisms [227], RNA interference establishment in a new organism requires empirically testing the known methods as well as the various parameters in combinations to determine the set of conditions that provide the best outcome - which in this case would be the most efficient target mRNA expression reduction. The following are approaches to be empirically tested in future for determining the most efficient strategy for establishing RNA interference in *C. bescii*:

- Development and use of a high copy number plasmid in *C. bescii* to clone RNAi fragments may be one way to achieve a high copy number of the asRNA. Another way could be to clone multiple copies of the EGS or targeting multiple sections of the target mRNA by using multiple fragments separated by hammerhead ribozyme sequence. All of these approaches have shown to improve the RNAi inhibition efficacy [255] [256, 257] and since a decent transcript reduction was observed with a RBS sequence containing fragment, but no phenotype was seen, approaches that may enhance the effect would be a good place to start.,

- Use of RNAi methodologies independent of the native RNaseP such as the attaching the guide sequence to the catalytic RNaseP RNA subunit from *C. bescii* or another thermophile or test the M1 RNA from *E.coli* [230], and

- Targeting genes with lesser metabolic or energy consequences unlike acetate metabolism genes used in this study. These genes may offer less resistance to the approach. For example, an antibiotic marker gene like kanamycin resistance gene *(cbhtk)* could be inserted in the genome and then targeted via RNAi. This would also ensure presence of only single copy of the gene as inserted making the approach less complex with minimal confounding effects and easier to optimize.

Additionally, results from the second approach (Approach II) in the study tested and described various attributes for designing an RNAi based genome-wide screening tools. It showed that random genomic fragments could be blunt-end ligated into a suitable plasmid backbone towards generating a library and fragments with varying sizes, mapping to different genes stably transformed as well as expressed into *C. bescii* using this method. The estimated design parameters revealed that fifty *E. coli* and only six *C. bescii* transformations may be needed to generate a complete library of clones with random genomic fragments that would statistically represent the entire genome with 99.99% probability. The work from this approach and these findings provide a design scaffold for an RNAi based genome wide screening tool to be developed in *C. bescii*. One example of building over this scaffold could include testing ways to improve library quality such as estimating and accounting for or avoiding any potential bias that may occur due

to passaging the library through *E. coli* and attempting conjugation/mating approach instead to directly transform the library into *C. bescii.* Moreover, the components for the tool development such as the plasmid vector with UI sequence on each end, (designed to facilitate reductions in cost and time by enabling sequencing of only the inserts to screen the library for genetic elements that confer robustness under the studied condition post-enrichment) could also be used with subtle variations to tailor to the needs of RNAi method or parameters that improve inhibitory efficiency resulting in RNA interference upon further experimentation.

Chapter 6.0

Conclusions, Significance and Future Directions

The work described in this dissertation was performed with the major objective of enhancing the growth and overall productivity of the anaerobic extreme thermophile *Caldicellulosiruptor bescii*. The studies were designed to take a wholesome approach at the described aim by first identifying the major factors that can limit the fermentation potential of the organism in question by optimizing and using a range of tools such as integrated omics. This was followed by understanding the nature of limitations in order to devise strategies to overcome the limitation and by finally designing and attempting development of new high throughput tools to aid further exploration of the organism as a platform microbe.

The first study (chapter 2) aimed at understanding and describing the parameters as well as logistics required to obtain statistically reliable and globally comparable data analysis using RNA-Seq, a powerful next generation sequencing based transcriptomics tool. It was concluded that designing an experiment with as little as three good replicates (high correlation coefficients, indicating reproducibility across samples) and at least 5–10 million reads per sample would be the most efficient and cost-effective design for a microbial transcriptomics study. The depth and complexity of the RNA-seq data from *Bacillus thuringiensis* used for this study will be useful to others for a range of studies such as for insights into *Bacillus* physiology and for further developments in the field of

bioinformatics for microbial transcriptomics. Moreover, the data processing and analysis pipeline thus developed as a part of this study as well as the described parameters were further used in other following studies with *C. bescii* during generation and analysis of RNA-seq data.

The following two studies (chapters 3 and 4) were designed to identify the factor(s) contributing to the growth and substrate utilization limitation in *C. bescii.* The study described in chapter 3 focused on understanding and characterizing the physiological and systems level responses of *C. bescii* at acidic pH using integrated omics data analysis from growth under controlled conditions. The results from this study concluded that lowered membrane potential is one of the major factors limiting the growth and substrate utilization in *C. bescii* and alleviation of lowered membrane potential upon lowering pH to 6.0 from 7.2 overcomes this limitation by dramatically increasing cell and product yields as well as substrate utilizations. By looking deeper into the mechanism, it was concluded that the dramatic increase in productivity upon alleviation of PMF stress resulted from higher energy availability. The protons made available in the system, post-acid addition, were speculated to most likely be sequestered via the membrane bound H^+-translocation PPase present in *C. bescii* and it was shown that the [Ni-Fe] hydrogenase in this organism was not critically involved in the generation of the improved growth phenotype observed in this study. This work serves to demonstrate the need to better understand the coupling of membrane potential with cellular energy and metabolism in such organisms. Since anaerobes generally suffer from low ATP yields, it forms an important limitation to be targeted to advance their use in industry.

Future endeavors from this point onwards could include:

Identifying and characterizing other immediate limiting factors post- PMF stress alleviation in C. bescii: While still working with *C. bescii* it is now of significance to further explore the possible secondary limitations that could further limit its productivity. Total osmolarity, organic acid accumulation and/or acetate toxicity could be the next potential targets.

Identifying other organisms that display a similar form of limitation: Other organisms could also be experiencing PMF-limiting conditions affecting their growth, substrate utilization and in turn product yields. It seems relevant to investigate organisms that show similar inhibition or growth cessation patterns as seen with *C. bescii,* for membrane potential changes. For instance, there are reports of growth improvement in *Clostridium thermocellum (C. thermocellum)*, another potential CBP organism, upon addition of organic acids such as acetate and formate [258-261]. Although, these growth phenotypes to date have been entirely attributed to carbon flux balance rationalizations, our work offers PMF limitation as a possible alternate hypothesis towards the same effect. *C. thermocellum* may have an analogous energy conservation system resulting in improved or rescued growth phenotype post organic acid addition even in mutants producing lower acetate. Qualitative carbon flux analysis alone thus may not be enough for such studies and an overall ATP balance under such conditions may help elucidate the involvement of PMF-derived energy contribution. Similar growth enhancement effects have also been observed with citrate addition in *Lactococcus lactis* where increased ATP generation was also noted among the positive effects of citrate on the growth of lactic acid bacteria [262]. This also hints at presence of a PMF linked energy conservation system.

Developing a deeper understanding of the underlying mechanism in C. bescii and other similar anaerobic fermentative bacteria: A *C. bescii* mutant strain with an inactive or deleted PPase, if engineered in future, could be used to test and confirm its role in utilizing the added protons in this study to establish a PMF and its link to non-SLP derived ATP generation. Similarly, using such approaches as described with *C. bescii* in this study, directed towards identifying analogous systems or players in other organisms such as *C. thermocellum* could be the very next step.

This study carries the novel aspects of estimating membrane potentials as well as showing the ability to maintain growth at pH as low as 5.5 in *C, bescii*, both of which to our knowledge have not been descried before. It brings proton motive force to light, which is not an unknown cellular parameter but one that has not been directly studied, especially with anaerobes. PMF is introduced as an important physiological/ metabolic metric here, which can be used to improve cellular productivity in microbes. Designing growth environments that aid optimum PMF maintenance, leading to higher growth and product titers in a cost-effective manner, would be extremely desirable and widely applicable in any organism-based, product driven industry.

Further based on the results of the above-mentioned study, once membrane potential/PMF limitation was alleviated, the system showed the ability to completely utilize the available sugar and growth recovery post stationary phase upon adding more sugars to the system. Following this line of evidence for utilization of higher amounts of sugars

seen on liquid substrate maltose, the next study as described in chapter 4 was designed with a two-fold objective of investigating if membrane potential associated limitation also affects growth on solid substrate and whether osmolarity and/or organic acid accumulation would be the secondary factor(s) post-PMF limitation alleviation to limit the fermentation potential of *C. bescii*. It was concluded from the results of this study that *C. bescii* also experiences PMF limitation when growing on solid substrate avicel and that total osmolarities ≥ 350mOsm severely suppresses growth as well as metabolism before total organic acids could reach inhibitory levels, becoming the next immediate factor limiting productivity of *C. bescii* on both liquid as well as solid substrates post-PMF limitation. It was also concluded based on different substrates that the degree of severity of the effect of these high osmolarities would vary from growth lag, suppression or complete inhibition depending on presence of additional factors stressing the system such as higher maintenance requirements on substrate or persisting PMF limitation.

The identification and understanding of the factors limiting the growth and productivity of *C. bescii* as with identification of the PMF followed by high total osmolarity, would facilitate designing better fermentations with parameters supporting enhanced growth and product yields. Some ways to most effectively employ *C. bescii* cultures to the production of desired products based the factors identified in this work include, growth of cultured at acidic pH of 6.0; use of fed-batch systems to overcome the effects of high single loads of sugars and increase productivity; real-time online osmolarity monitoring of the fermentation system aided by specialized probe [217] and using it to estimate time-points for system dilution by replacing an amount of spent media with accumulated osmolytes

with fresh media. The advantage to studying the physiological and systems level responses of the organism to various stress factors identified, is that the knowledge gained could be further applied towards growth on real world substrates such as switchgrass and guide most efficient and productive fermentations resulting in maximum product formation and substrate utilization. Apart from fermentation system improvisations, further work would also entail development of robust strains, either via evolution to withstand higher osmotic pressures or by identifying target genes to improve osmotolerance and genetically engineering such strains.

The dissertation also includes the novel attempt at trying to investigate asRNA mediated RNA interference (chapter 5) which has not been previously described in *C. bescii.* The study also described and detailed work on the attributes as well as prospects of developing a genome wide screening tool based on RNAi. The results from the approaches applied, provide a basis for further studies to build upon the technology and its high- throughput application in future as it concludes that RNA interference technology for *C. bescii* antisense inhibition still needs further development and optimization. There is need for much experimentation with many approaches and parameters that can affect RNA interference efficiency to empirically arrive at the most effective strategy tailored for *C. bescii.* as also discussed as a part of this study. Once an effective RNAi strategy is established for *C. bescii* the tool development design and components (e.g. plasmid backbone) used here could be further developed with subtle variations to fit the new strategy.

Overall, the growth and substrate utilization in *C. bescii* were dramatically improved by identifying and alleviating lowered membrane potential limitation. Moreover, the work also emphasizes that *C. bescii* still displays further the potential for improved productivity, which can be achieved by overcoming successive limiting factors such as high medium osmolarity and others that might be identified and this work would aid the future studies towards this aim. Knowledge from the studies here would facilitate designing better fermentations with parameters supporting enhanced growth and product yields. The initial work and efforts described here show promise with RNAi and would inspire future studies in developing this technology and development of high-throughput genomic tools based on it. RNAi technology would be a valuable addition to the genetic engineering tool box for *C. bescii* and other similar thermophiles as it would facilitate not only genome wide fitness assays across various stress conditions such as high osmolarity or with specific inhibitors and also aid several genome-wide screening studies. The genes/genetic elements identified as a result could then be further used as targets for metabolically engineering robust *C. bescii* strains in the future.

References

1. Lu, X., *Biofuels*. 2014: Caister Academic Press.
2. Keller, M., et al., *Production of lignofuels and electrofuels by extremely thermophilic microbes.* Biofuels, 2014. **5**(5): p. 499-515.
3. Taylor, K., et al., *Atmospheric CO_2 over the last 1000 years: A high-resolution record from the West Antarctic Ice Sheet (WAIS) Divide ice core.* 2012.
4. Bessou, C., et al., *Biofuels, greenhouse gases and climate change. A review.* Agronomy for Sustainable Development, 2011. **31**(1): p. 1.
5. Liao, J.C., et al., *Fuelling the future: microbial engineering for the production of sustainable biofuels.* Nat Rev Microbiol, 2016. **14**(5): p. 288-304.
6. Chung, D., et al., *Direct conversion of plant biomass to ethanol by engineered Caldicellulosiruptor bescii.* Proc Natl Acad Sci U S A, 2014. **111**(24): p. 8931-6.
7. Mika, L.s.T., E. Cséfalvay, and A.r. Németh, *Catalytic conversion of carbohydrates to initial platform chemicals: chemistry and sustainability.* Chemical reviews, 2017. **118**(2): p. 505-613.
8. Owusu, P.A. and S. Asumadu-Sarkodie, *A review of renewable energy sources, sustainability issues and climate change mitigation.* Cogent Engineering, 2016. **3**(1): p. 1167990.
9. Demirbas, M.F., M. Balat, and H. Balat, *Biowastes-to-biofuels.* Energy Conversion and Management, 2011. **52**(4): p. 1815-1828.
10. Singhania, R.R., et al., *Waste to Wealth*. 2017: Springer.
11. Lynd, L.R., *The grand challenge of cellulosic biofuels.* Nature Biotechnology, 2017. **35**: p. 912.
12. Chisti, Y., *Biodiesel from microalgae.* Biotechnology advances, 2007. **25**(3): p. 294-306.
13. Li, Y., et al., *Biofuels from microalgae.* Biotechnology progress, 2008. **24**(4): p. 815-820.
14. Mathews, J.A., *Carbon-negative biofuels.* Energy policy, 2008. **36**(3): p. 940-945.
15. Speight, J.G., *The biofuels handbook*. 2011: Royal Society of Chemistry.
16. Solomon, B.D., *Biofuels and sustainability.* Annals of the New York Academy of Sciences, 2010. **1185**(1): p. 119-134.
17. Schmer, M.R., et al., *Energy potential and greenhouse gas emissions from bioenergy cropping systems on marginally productive cropland.* PloS one, 2014. **9**(3): p. e89501.
18. Carroll, A. and C. Somerville, *Cellulosic Biofuels.* Annual Review of Plant Biology, 2009. **60**(1): p. 165-182.
19. Lynd, L.R., et al., *Cellulosic ethanol: status and innovation.* Current Opinion in Biotechnology, 2017. **45**: p. 202-211.
20. Robertson, G.P., et al., *Cellulosic biofuel contributions to a sustainable energy future: Choices and outcomes.* Science, 2017. **356**(6345): p. eaal2324.
21. Himmel, M.E., et al., *Biomass recalcitrance: engineering plants and enzymes for biofuels production.* science, 2007. **315**(5813): p. 804-807.
22. Lynd, L.R., et al., *How biotech can transform biofuels.* Nat Biotechnol, 2008. **26**(2): p. 169-72.
23. DOE, U., *Lignocellulosic biomass for advanced biofuels and bioproducts*. 2016, DOE/SC-1070. Available at: http://genomicscience. energy. gov/biofuels/lignocellulose/. Accessed: April 29th.

24. Cosgrove, D.J., *Growth of the plant cell wall.* Nature reviews molecular cell biology, 2005. **6**(11): p. 850.
25. Iiyama, K., T.B.-T. Lam, and B.A. Stone, *Covalent cross-links in the cell wall.* Plant physiology, 1994. **104**(2): p. 315.
26. Wyman, C.E., et al., *Coordinated development of leading biomass pretreatment technologies.* Bioresource technology, 2005. **96**(18): p. 1959-1966.
27. Alvira, P., et al., *Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: a review.* Bioresource technology, 2010. **101**(13): p. 4851-4861.
28. Yang, B. and C.E. Wyman, *Pretreatment: the key to unlocking low-cost cellulosic ethanol.* Biofuels, Bioproducts and Biorefining, 2008. **2**(1): p. 26-40.
29. Blumer-Schuette, S.E., et al., *Thermophilic lignocellulose deconstruction.* FEMS Microbiol Rev, 2014. **38**(3): p. 393-448.
30. Olson, D.G., et al., *Recent progress in consolidated bioprocessing.* Current Opinion in Biotechnology, 2012. **23**(3): p. 396-405.
31. Mbaneme-Smith, V. and M.S. Chinn, *Consolidated bioprocessing for biofuel production: recent advances.* Energy and Emission Control Technologies, 2015. **3**: p. 23-44.
32. Lynd, L.R., et al., *Microbial cellulose utilization: fundamentals and biotechnology.* Microbiology and molecular biology reviews, 2002. **66**(3): p. 506-577.
33. Hu, W.-J., et al., *Repression of lignin biosynthesis promotes cellulose accumulation and growth in transgenic trees.* Nature biotechnology, 1999. **17**(8): p. 808.
34. Chen, F. and R.A. Dixon, *Lignin modification improves fermentable sugar yields for biofuel production.* Nature biotechnology, 2007. **25**(7): p. 759.
35. Fu, C., et al., *Downregulation of cinnamyl alcohol dehydrogenase (CAD) leads to improved saccharification efficiency in switchgrass.* BioEnergy Research, 2011. **4**(3): p. 153-164.
36. Coleman, H.D., et al., *Perturbed lignification impacts tree growth in hybrid poplar—a function of sink strength, vascular integrity, and photosynthetic assimilation.* Plant Physiology, 2008. **148**(3): p. 1229-1237.
37. Mansfield, S.D., K.Y. Kang, and C. Chapple, *Designed for deconstruction–poplar trees altered in cell wall lignification improve the efficacy of bioethanol production.* New Phytologist, 2012. **194**(1): p. 91-101.
38. Biswal, A.K., et al., *Sugar release and growth of biofuel crops are improved by downregulation of pectin biosynthesis.* Nature biotechnology, 2018. **36**(3): p. 249.
39. Fu, C., et al., *Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass.* Proceedings of the National Academy of Sciences, 2011. **108**(9): p. 3803-3808.
40. Fu, C., et al., *Overexpression of miR156 in switchgrass (Panicum virgatum L.) results in various morphological alterations and leads to improved biomass production.* Plant biotechnology journal, 2012. **10**(4): p. 443-452.
41. Shen, H., et al., *Functional characterization of the switchgrass (Panicum virgatum) R2R3-MYB transcription factor PvMYB4 for improvement of lignocellulosic feedstocks.* New Phytologist, 2012. **193**(1): p. 121-136.

42. Yoo, C.G., et al., *Insights of biomass recalcitrance in natural Populus trichocarpa variants for biomass conversion.* Green Chemistry, 2017. **19**(22): p. 5467-5478.
43. Dumitrache, A., et al., *Transgenic switchgrass (Panicum virgatum L.) targeted for reduced recalcitrance to bioconversion: a 2-year comparative analysis of field-grown lines modified for target gene or genetic element expression.* Plant biotechnology journal, 2017. **15**(6): p. 688-697.
44. Liu, G., J. Zhang, and J. Bao, *Cost evaluation of cellulase enzyme for industrial-scale cellulosic ethanol production based on rigorous Aspen Plus modeling.* Bioprocess and biosystems engineering, 2016. **39**(1): p. 133-140.
45. Klein-Marcuschamer, D., et al., *The challenge of enzyme cost in the production of lignocellulosic biofuels.* Biotechnology and bioengineering, 2012. **109**(4): p. 1083-1087.
46. Olson, D.G., et al., *Recent progress in consolidated bioprocessing.* Curr Opin Biotechnol, 2012. **23**(3): p. 396-405.
47. Himmel, M.E., et al., *Microbial enzyme systems for biomass conversion: emerging paradigms.* Biofuels, 2010. **1**(2): p. 323-341.
48. Barabote, R.D., et al., *Complete genome of the cellulolytic thermophile Acidothermus cellulolyticus 11B provides insights into its ecophysiological and evolutionary adaptations.* Genome research, 2009. **19**(6): p. 1033-1043.
49. Berka, R.M., et al., *Comparative genomic analysis of the thermophilic biomass-degrading fungi Myceliophthora thermophila and Thielavia terrestris.* Nature biotechnology, 2011. **29**(10): p. 922.
50. Blumer-Schuette, S.E., et al., *Caldicellulosiruptor core and pangenomes reveal determinants for noncellulosomal thermophilic deconstruction of plant biomass.* Journal of bacteriology, 2012. **194**(15): p. 4015-4028.
51. Dam, P., et al., *Insights into plant biomass conversion from the genome of the anaerobic thermophilic bacterium Caldicellulosiruptor bescii DSM 6725.* Nucleic acids research, 2011. **39**(8): p. 3240-3254.
52. Lykidis, A., et al., *Genome sequence and analysis of the soil cellulolytic actinomycete Thermobifida fusca YX.* Journal of bacteriology, 2007. **189**(6): p. 2477-2486.
53. Suzuki, H., et al., *Comparative genomics of the white-rot fungi, Phanerochaete carnosa and P. chrysosporium, to elucidate the genetic basis of the distinct wood types they colonize.* BMC genomics, 2012. **13**(1): p. 444.
54. Raman, B., et al., *Impact of pretreated switchgrass and biomass carbohydrates on Clostridium thermocellum ATCC 27405 cellulosome composition: a quantitative proteomic analysis.* PloS one, 2009. **4**(4): p. e5271.
55. Martinez, D., et al., *Genome sequencing and analysis of the biomass-degrading fungus Trichoderma reesei (syn. Hypocrea jecorina).* Nature biotechnology, 2008. **26**(5): p. 553.
56. Lynd, L.R., et al., *Microbial cellulose utilization: fundamentals and biotechnology.* Microbiol Mol Biol Rev, 2002. **66**(3): p. 506-77, table of contents.
57. Lynd, L.R., et al., *Consolidated bioprocessing of cellulosic biomass: an update.* Current opinion in biotechnology, 2005. **16**(5): p. 577-583.

58. La Grange, D.C., R. Den Haan, and W.H. Van Zyl, *Engineering cellulolytic ability into bioprocessing organisms.* Applied microbiology and biotechnology, 2010. **87**(4): p. 1195-1208.
59. Bokinsky, G., et al., *Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered Escherichia coli.* Proceedings of the National Academy of Sciences, 2011. **108**(50): p. 19949-19954.
60. Nakamura, C.E. and G.M. Whited, *Metabolic engineering for the microbial production of 1, 3-propanediol.* Current opinion in biotechnology, 2003. **14**(5): p. 454-459.
61. Zhang, X.-Z., et al., *One-step production of lactate from cellulose as the sole carbon source without any other organic nutrient by recombinant cellulolytic Bacillus subtilis.* Metabolic engineering, 2011. **13**(4): p. 364-372.
62. Thammasittirong, S.N.-R., et al., *Improvement of ethanol production by ethanol-tolerant Saccharomyces cerevisiae UVNR56.* SpringerPlus, 2013. **2**(1): p. 583.
63. Tsai, S.-L., G. Goyal, and W. Chen, *Surface display of a functional minicellulosome by intracellular complementation using a synthetic yeast consortium and its application to cellulose hydrolysis and ethanol production.* Applied and environmental microbiology, 2010. **76**(22): p. 7514-7520.
64. Den Haan, R., et al., *Hydrolysis and fermentation of amorphous cellulose by recombinant Saccharomyces cerevisiae.* Metabolic engineering, 2007. **9**(1): p. 87-94.
65. Yamada, R., et al., *Direct and efficient ethanol production from high-yielding rice using a Saccharomyces cerevisiae strain that express amylases.* Enzyme and microbial technology, 2011. **48**(4-5): p. 393-396.
66. Zhang, M., et al., *Creation of an ethanol-tolerant Saccharomyces cerevisiae strain by 266 nm laser radiation and repetitive cultivation.* J Biosci Bioeng, 2014.
67. Chang, J.J., et al., *Assembling a cellulase cocktail and a cellodextrin transporter into a yeast host for CBP ethanol production.* Biotechnol Biofuels, 2013. **6**(1): p. 19.
68. Egorova, K. and G. Antranikian, *Industrial relevance of thermophilic Archaea.* Current opinion in microbiology, 2005. **8**(6): p. 649-655.
69. Cysewski, G.R. and C.R. Wilke, *Rapid ethanol fermentations using vacuum and cell recycle.* Biotechnology and Bioengineering, 1977. **19**(8): p. 1125-1143.
70. Shabtai, Y., et al., *Continuous ethanol production by immobilized yeast reactor coupled with membrane pervaporation unit.* Biotechnology and bioengineering, 1991. **38**(8): p. 869-876.
71. Taylor, F., et al., *Continuous fermentation and stripping of ethanol.* Biotechnology progress, 1995. **11**(6): p. 693-698.
72. Unsworth, L.D., J. van der Oost, and S. Koutsopoulos, *Hyperthermophilic enzymes− stability, activity and implementation strategies for high temperature applications.* The FEBS journal, 2007. **274**(16): p. 4044-4056.
73. Shaw, A.J., et al., *Metabolic engineering of a thermophilic bacterium to produce ethanol at high yield.* Proceedings of the National Academy of Sciences, 2008. **105**(37): p. 13769-13774.
74. Ou, M.S., et al., *Thermophilic Bacillus coagulans requires less cellulases for simultaneous saccharification and fermentation of cellulose to products than*

mesophilic microbial biocatalysts. Applied biochemistry and biotechnology, 2009. **155**(1-3): p. 76-82.
75. Aksenova, H.Y., et al., *Spirochaeta thermophila sp. nov., an obligately anaerobic, polysaccharolytic, extremely thermophilic bacterium.* International Journal of Systematic and Evolutionary Microbiology, 1992. **42**(1): p. 175-177.
76. Rainey, F.A., et al., *Isolation and characterization of an obligately anaerobic, polysaccharolytic, extremely thermophilic member of the genus Spirochaeta.* Archives of microbiology, 1991. **155**(4): p. 396-401.
77. Basen, M., et al., *Degradation of high loads of crystalline cellulose and of unpretreated plant biomass by the thermophilic bacterium Caldicellulosiruptor bescii.* Bioresour Technol, 2014. **152**: p. 384-92.
78. Fontes, C.M. and H.J. Gilbert, *Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates.* Annual review of biochemistry, 2010. **79**: p. 655-681.
79. Demain, A.L., M. Newcomb, and J.D. Wu, *Cellulase, clostridia, and ethanol.* Microbiology and molecular biology reviews, 2005. **69**(1): p. 124-154.
80. Nataf, Y., et al., *Cellodextrin and laminaribiose ABC transporters in Clostridium thermocellum.* Journal of bacteriology, 2009. **191**(1): p. 203-209.
81. Blumer-Schuette, S.E., et al., *Extremely thermophilic microorganisms for biomass conversion: status and prospects.* Curr Opin Biotechnol, 2008. **19**(3): p. 210-7.
82. Svetlitchnyi, V.A., et al., *Single-step ethanol production from lignocellulose using novel extremely thermophilic bacteria.* Biotechnology for biofuels, 2013. **6**(1): p. 31.
83. Yang, S.-J., et al., *Efficient degradation of lignocellulosic plant biomass, without pretreatment, by the thermophilic anaerobe "Anaerocellum thermophilum" DSM 6725.* Applied and environmental microbiology, 2009. **75**(14): p. 4762-4769.
84. Blumer-Schuette, S.E., D.L. Lewis, and R.M. Kelly, *Phylogenetic, microbiological, and glycoside hydrolase diversities within the extremely thermophilic, plant biomass-degrading genus Caldicellulosiruptor.* Appl Environ Microbiol, 2010. **76**(24): p. 8084-92.
85. Rainey, F.A., et al., *Description of Caldicellulosiruptor saccharolyticus gen. nov., sp. nov: an obligately anaerobic, extremely thermophilic, cellulolytic bacterium.* FEMS Microbiol Lett, 1994. **120**(3): p. 263-6.
86. Mladenovska, Z., I.M. Mathrani, and B.K. Ahring, *Isolation and characterization of Caldicellulosiruptor lactoaceticus sp. nov., an extremely thermophilic, cellulolytic, anaerobic bacterium.* Archives of microbiology, 1995. **163**(3): p. 223-230.
87. Huang, C.-Y., et al., *Caldicellulosiruptor owensensis sp. nov., an anaerobic, extremely thermophilic, xylanolytic bacterium.* International Journal of Systematic and Evolutionary Microbiology, 1998. **48**(1): p. 91-97.
88. Bredholt, S., et al., *Caldicellulosiruptor kristjanssonii sp. nov., a cellulolytic, extremely thermophilic, anaerobic bacterium.* International Journal of Systematic and Evolutionary Microbiology, 1999. **49**(3): p. 991-996.
89. Onyenwoke, R.U., et al., *Reclassification of Thermoanaerobium acetigenum as Caldicellulosiruptor acetigenus comb. nov. and emendation of the genus*

description. International journal of systematic and evolutionary microbiology, 2006. **56**(6): p. 1391-1395.
90. Miroshnichenko, M.L., et al., *Caldicellulosiruptor kronotskyensis sp. nov. and Caldicellulosiruptor hydrothermalis sp. nov., two extremely thermophilic, cellulolytic, anaerobic bacteria from Kamchatka thermal springs.* International journal of systematic and evolutionary microbiology, 2008. **58**(6): p. 1492-1496.
91. Yang, S.-J., et al., *Classification of 'Anaerocellum thermophilum'strain DSM 6725 as Caldicellulosiruptor bescii sp. nov.* International journal of systematic and evolutionary microbiology, 2010. **60**(9): p. 2011-2015.
92. Hamilton-Brehm, S.D., et al., *Caldicellulosiruptor obsidiansis sp. nov., an anaerobic, extremely thermophilic, cellulolytic bacterium isolated from Obsidian Pool, Yellowstone National Park.* Applied and environmental microbiology, 2010. **76**(4): p. 1014-1020.
93. Bing, W., et al., *Caldicellulosiruptor changbaiensis sp. nov., a cellulolytic and hydrogen-producing bacterium from a hot spring.* International journal of systematic and evolutionary microbiology, 2015. **65**(1): p. 293-297.
94. Gibbs, M.D., et al., *Sequencing, cloning and expression of a β-1, 4-mannanase gene, manA, from the extremely thermophilic anaerobic bacterium, Caldicellulosiruptor Rt8B. 4.* FEMS microbiology letters, 1996. **141**(1): p. 37-43.
95. van de Werken, H.J., et al., *Hydrogenomics of the extremely thermophilic bacterium Caldicellulosiruptor saccharolyticus.* Appl Environ Microbiol, 2008. **74**(21): p. 6720-9.
96. Zurawski, J.V., et al., *The extremely thermophilic genus Caldicellulosiruptor: physiological and genomic characteristics for complex carbohydrate conversion to molecular hydrogen*, in *Microbial bioenergy: hydrogen production*. 2014, Springer. p. 177-195.
97. Lochner, A., et al., *Use of label-free quantitative proteomics to distinguish the secreted cellulolytic systems of Caldicellulosiruptor bescii and Caldicellulosiruptor obsidiansis.* Applied and environmental microbiology, 2011. **77**(12): p. 4042-4054.
98. Gibbs, M.D., et al., *Multidomain and multifunctional glycosyl hydrolases from the extreme thermophile Caldicellulosiruptor isolate Tok7B. 1.* Current microbiology, 2000. **40**(5): p. 333-340.
99. Saul, D., et al., *celB, a gene coding for a bifunctional cellulase from the extreme thermophile" Caldocellum saccharolyticum".* Applied and environmental microbiology, 1990. **56**(10): p. 3117-3124.
100. Te'o, V., D. Saul, and P. Bergquist, *celA, another gene coding for a multidomain cellulase from the extreme thermophile Caldocellum saccharolyticum.* Applied microbiology and biotechnology, 1995. **43**(2): p. 291-296.
101. Gibbs, M., et al., *The beta-mannanase from" Caldocellum saccharolyticum" is part of a multidomain enzyme.* Applied and environmental microbiology, 1992. **58**(12): p. 3864-3867.
102. Brunecky, R., et al., *Revealing nature's cellulase diversity: the digestion mechanism of Caldicellulosiruptor bescii CelA.* Science, 2013. **342**(6165): p. 1513-1516.

103. Izquierdo, J.A., M.V. Sizova, and L.R. Lynd, *Diversity of bacteria and glycosyl hydrolase family 48 genes in cellulolytic consortia enriched from thermophilic biocompost.* Applied and environmental microbiology, 2010. **76**(11): p. 3545-3553.
104. Kanafusa-Shinkai, S., et al., *Degradation of microcrystalline cellulose and non-pretreated plant biomass by a cell-free extracellular cellulase/hemicellulase system from the extreme thermophilic bacterium Caldicellulosiruptor bescii.* J Biosci Bioeng, 2013. **115**(1): p. 64-70.
105. Cherry, J.R. and A.L. Fidantsef, *Directed evolution of industrial enzymes: an update.* Curr Opin Biotechnol, 2003. **14**(4): p. 438-43.
106. Martinez, D., et al., *Genome sequencing and analysis of the biomass-degrading fungus Trichoderma reesei (syn. Hypocrea jecorina).* Nat Biotechnol, 2008. **26**(5): p. 553-60.
107. Kataeva, I., et al., *Carbohydrate and lignin are simultaneously solubilized from unpretreated switchgrass by microbial action at high temperature.* Energy & Environmental Science, 2013. **6**(7): p. 2186-2195.
108. Carere, C.R., et al., *Linking genome content to biofuel production yields: a meta-analysis of major catabolic pathways among select H 2 and ethanol-producing bacteria.* BMC microbiology, 2012. **12**(1): p. 295.
109. Schut, G.J. and M.W. Adams, *The iron-hydrogenase of Thermotoga maritima utilizes ferredoxin and NADH synergistically: a new perspective on anaerobic hydrogen production.* Journal of bacteriology, 2009. **191**(13): p. 4451-4457.
110. Willquist, K., A.A. Zeidan, and E.W. van Niel, *Physiological characteristics of the extreme thermophile Caldicellulosiruptor saccharolyticus: an efficient hydrogen cell factory.* Microbial Cell Factories, 2010. **9**(1): p. 89.
111. Klinke, H.B., A.B. Thomsen, and B.K. Ahring, *Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass.* Appl Microbiol Biotechnol, 2004. **66**(1): p. 10-26.
112. Chung, D., et al., *Cellulosic ethanol production via consolidated bioprocessing at 75 degrees C by engineered Caldicellulosiruptor bescii.* Biotechnol Biofuels, 2015. **8**: p. 163.
113. Farkas, J., et al., *Improved growth media and culture techniques for genetic analysis and assessment of biomass utilization by Caldicellulosiruptor bescii.* J Ind Microbiol Biotechnol, 2013. **40**(1): p. 41-9.
114. Dien, B., M. Cotta, and T. Jeffries, *Bacteria engineered for fuel ethanol production: current status.* Applied microbiology and biotechnology, 2003. **63**(3): p. 258-266.
115. Jin, M., et al., *Consolidated bioprocessing (CBP) of AFEX™-pretreated corn stover for ethanol production using Clostridium phytofermentans at a high solids loading.* Biotechnology and bioengineering, 2012. **109**(8): p. 1929-1936.
116. Kridelbaugh, D.M., et al., *Nitrogen and sulfur requirements for Clostridium thermocellum and Caldicellulosiruptor bescii on cellulosic substrates in minimal nutrient media.* Bioresource technology, 2013. **130**: p. 125-135.
117. Chung, D., et al., *Construction of a stable replicating shuttle vector for Caldicellulosiruptor species: use for extending genetic methodologies to other members of this genus.* PloS one, 2013. **8**(5): p. e62881.

118. Cha, M., et al., *Metabolic engineering of Caldicellulosiruptor bescii yields increased hydrogen production from lignocellulosic biomass.* Biotechnology for biofuels, 2013. **6**(1): p. 85.
119. Chung, D., et al., *Homologous Expression of the Caldicellulosiruptor bescii CelA Reveals that the Extracellular Protein Is Glycosylated.* PLoS ONE, 2015. **10**(3): p. e0119508.
120. Williams-Rhaesa, A.M., et al., *Genome stability in engineered strains of the extremely thermophilic lignocellulose-degrading bacterium Caldicellulosiruptor bescii.* Applied and environmental microbiology, 2017. **83**(14): p. e00444-17.
121. Cha, M., D. Chung, and J. Westpheling, *Deletion of a gene cluster for [Ni-Fe] hydrogenase maturation in the anaerobic hyperthermophilic bacterium Caldicellulosiruptor bescii identifies its role in hydrogen metabolism.* Applied microbiology and biotechnology, 2016. **100**(4): p. 1823-1831.
122. Chung, D., et al., *Methylation by a Unique a-class N4-Cytosine Methyltransferase Is Required for DNA.* 2012.
123. Chung, D.-H., et al., *Identification and characterization of CbeI, a novel thermostable restriction enzyme from Caldicellulosiruptor bescii DSM 6725 and a member of a new subfamily of HaeIII-like enzymes.* Journal of industrial microbiology & biotechnology, 2011. **38**(11): p. 1867.
124. Yang, S.J., et al., *Efficient degradation of lignocellulosic plant biomass, without pretreatment, by the thermophilic anaerobe "Anaerocellum thermophilum" DSM 6725.* Appl Environ Microbiol, 2009. **75**(14): p. 4762-9.
125. Lipscomb, G.L., et al., *A highly thermostable kanamycin resistance marker expands the tool kit for genetic manipulation of Caldicellulosiruptor bescii.* Applied and environmental microbiology, 2016. **82**(14): p. 4421-4428.
126. Teske, A., *Grand Challenges in Extreme Microbiology.* Frontiers in Microbiology, 2010. **1**(111).
127. Mori, K. and Y. Kamagata, *The Challenges of Studying the Anaerobic Microbial World.* Microbes and Environments, 2014. **29**(4): p. 335-337.
128. Taylor, M.P., et al., *Genetic tool development underpins recent advances in thermophilic whole-cell biocatalysts.* Microbial biotechnology, 2011. **4**(4): p. 438-448.
129. Croucher, N.J. and N.R. Thomson, *Studying bacterial transcriptomes using RNA-seq.* Curr. Opin. Microbiol., 2010. **13**(5): p. 619-624.
130. Marguerat, S. and J. Bahler, *RNA-seq: from technology to biology.* Cell Mol Life Sci, 2010. **67**(4): p. 569-79.
131. Williams, A.G., et al., *RNA-seq data: Challenges in and recommendations for experimental design and analysis.* Curr. Protoc. Hum. Genet., 2014. **83**: p. 11.13.1-11.13.20.
132. Martin, J.A. and Z. Wang, *Next-generation transcriptome assembly.* Nat. Rev. Genetics, 2011. **12**(10): p. 671-682.
133. McGettigan, P.A., *Transcriptomics in the RNA-seq era.* Curr. Opin. Chem. Biol., 2013. **17**(1): p. 4-11.
134. Mutz, K.-O., et al., *Transcriptome analysis using next-generation sequencing.* Curr. Opin. Biotechnol., 2013. **24**(1): p. 22-30.
135. Ozsolak, F., et al., *Direct RNA sequencing.* Nature, 2009. **461**(7265): p. 814-818.

136. Auer, P.L. and R.W. Doerge, *Statistical design and analysis of RNA sequencing data.* Genetics, 2010. **185**(2): p. 405-16.
137. Oshlack, A., M.D. Robinson, and M.D. Young, *From RNA-seq reads to differential expression results.* Genome Biol, 2010. **11**(12): p. 220.
138. Peixoto, L., et al., *How data analysis affects power, reproducibility and biological insight of RNA-seq studies in complex datasets.* Nucleic Acids Res., 2015. **in press**: p. doi: 10.1093/nar/gkv736.
139. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics.* Nat Rev Genet, 2009. **10**(1): p. 57-63.
140. Pinto, A.C., et al., *Application of RNA-seq to reveal the transcript profile in bacteria.* Genetics Mol. Res., 2011. **10**(3): p. 1707-1718.
141. Fang, Z. and X. Cui, *Design and validation issues in RNA-seq experiments.* Brief. Bioinformatics, 2011. **12**(3): p. 280-287.
142. Dillies, M.A., et al., *A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis.* Brief Bioinform, 2013. **14**(6): p. 671-83.
143. Robinson, M.D. and A. Oshlack, *A scaling normalization method for differential expression analysis of RNA-seq data.* Genome Biol, 2010. **11**(3): p. R25.
144. Anders, S. and W. Huber, *Differential expression analysis for sequence count data.* Genome Biol, 2010. **11**(10): p. R106.
145. Anders, S., et al., *Count-based differential expression analysis of RNA sequencing data using R and Bioconductor.* Nat Protoc, 2013. **8**(9): p. 1765-86.
146. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.
147. Soneson, C. and M. Delorenzi, *A comparison of methods for differential expression analysis of RNA-seq data.* BMC Bioinformatics, 2013. **14**: p. 91.
148. Li, J. and R. Tibshirani, *Finding consistent patterns: a nonparametric approach for identifying differential expression in RNA-Seq data.* Stat Methods Med Res, 2013. **22**(5): p. 519-36.
149. Miller, C.A., et al., *ReadDepth: a parallel R package for detecting copy number alterations from short sequencing reads.* PLoS One, 2011. **6**(1): p. e16327.
150. Mi, G., Y. Di, and D.W. Schafer, *Goodness-of-Fit Tests and Model Diagnostics for Negative Binomial Regression of RNA Sequencing Data.* PLoS One, 2015. **10**(3): p. e0119254.
151. Gierliński, M., et al., *Statistical models for RNA-seq data derived from a two-condition 48-replicate experiment.* Bioinformatics, 2015. **31**(22): p. 3625-3630.
152. Liu, Y., J. Zhou, and K.P. White, *RNA-seq differential expression studies: more sequence or more replication?* Bioinformatics, 2014. **30**(3): p. 301-4.
153. Fonseca, N.A., et al., *Tools for mapping high-throughput sequencing data.* Bioinformatics, 2012. **28**(24): p. 3169-77.
154. Sims, D., et al., *Sequencing depth and coverage: key considerations in genomic analyses.* Nat Rev Genet, 2014. **15**(2): p. 121-32.
155. Joung, K.B. and J.C. Cote, *Phylogenetic analysis of Bacillus thuringiensis serovars based on 16S rRNA gene restriction fragment length polymorphisms.* J Appl Microbiol, 2001. **90**(1): p. 115-22.

156. Baxter, S.W., et al., *Parallel evolution of Bacillus thuringiensis toxin resistance in lepidoptera.* Genetics, 2011. **189**(2): p. 675-9.
157. Bravo, A., et al., *Bacillus thuringiensis: A story of a successful bioinsecticide.* Insect Biochem Mol Biol, 2011. **41**(7): p. 423-31.
158. Gassmann, A.J., et al., *Field-evolved resistance by western corn rootworm to multiple Bacillus thuringiensis toxins in transgenic maize.* Proc Natl Acad Sci U S A, 2014. **111**(14): p. 5141-6.
159. He, J., et al., *Complete genome sequence of Bacillus thuringiensis subsp. chinensis strain CT-43.* J. Bacteriol., 2011. **193**(13): p. 3407-3408.
160. Johnson, S.L., et al., *Complete genome sequences for 35 biothreat assay-relevant Bacillus species.* Genome Announcements, 2015. **3**(2): p. e00151-15.
161. Wilson, C.M., et al., *Global transcriptome analysis of Clostridium thermocellum ATCC 27405 during growth on dilute acid pretreated Populus and switchgrass.* Biotechnol Biofuels, 2013. **6**(1): p. 179.
162. Zhang, F., et al., *Kinetic analysis and modeling of oleate and ethanol stimulated uranium (VI) bio-reduction in contaminated sediments under sulfate reduction conditions.* J Hazard Mater, 2010. **183**(1-3): p. 482-9.
163. Marioni, J.C., et al., *RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays.* Genome research, 2008. **18**(9): p. 1509-1517.
164. Robasky, K., N.E. Lewis, and G.M. Church, *The role of replicates for error mitigation in next-generation sequencing.* Nature Reviews Genetics, 2014. **15**(1): p. 56.
165. Todd, E.V., M.A. Black, and N.J. Gemmell, *The power and promise of RNA-seq in ecology and evolution.* Molecular ecology, 2016. **25**(6): p. 1224-1241.
166. Churchill, G.A., *Fundamentals of experimental design for cDNA microarrays.* Nature genetics, 2002. **32**: p. 490.
167. Yang, Y.H. and T. Speed, *Design issues for cDNA microarray experiments.* Nature Reviews Genetics, 2002. **3**(8): p. 579.
168. Haas, B.J., et al., *How deep is deep enough for RNA-Seq profiling of bacterial transcriptomes?* BMC Genomics, 2012. **13**: p. 734.
169. Yang, S., et al., *Clostridium thermocellum ATCC27405 transcriptomic, metabolomic and proteomic profiles after ethanol stress.* BMC Genomics, 2012. **13**: p. 336.
170. Blair, J.M., et al., *Choice of bacterial growth medium alters the transcriptome and phenotype of Salmonella enterica Serovar Typhimurium.* PLoS One, 2013. **8**(5): p. e63912.
171. Alam, S.I., et al., *Characterization of an environmental strain of Bacillus thuringiensis from a hot spring in Western Himalayas.* Curr Microbiol, 2011. **62**(2): p. 547-56.
172. Bishop, A.H. and C.V. Robinson, *Bacillus thuringiensis HD-1 Cry- : development of a safe, non-insecticidal simulant for Bacillus anthracis.* J Appl Microbiol, 2014. **117**(3): p. 654-62.
173. Aronson, A.I., W. Beckman, and P. Dunn, *Bacillus thuringiensis and related insect pathogens.* Microbiological reviews, 1986. **50**(1): p. 1.

174. Seyednasrollah, F., A. Laiho, and L.L. Elo, *Comparison of software packages for detecting differential expression in RNA-seq studies.* Briefings in bioinformatics, 2013. **16**(1): p. 59-70.
175. Medina, I., et al., *Highly sensitive and ultrafast read mapping for RNA-seq analysis.* DNA Research, 2016. **23**(2): p. 93-100.
176. Moulos, P. and P. Hatzis, *Systematic integration of RNA-Seq statistical algorithms for accurate detection of differential gene expression patterns.* Nucleic Acids Res, 2015. **43**(4): p. e25.
177. Suen, T., *Growth and regulatory metabolism of cellulose-utilizing thermophilic Caldicellulosiruptor species*. 2015, University of Waterloo.
178. Reginatto, V. and R.V. Antonio, *Fermentative hydrogen production from agroindustrial lignocellulosic substrates.* Braz J Microbiol, 2015. **46**(2): p. 323-35.
179. Chung, D., et al., *Expression of a heat-stable NADPH-dependent alcohol dehydrogenase in Caldicellulosiruptor bescii results in furan aldehyde detoxification.* Biotechnol Biofuels, 2015. **8**: p. 102.
180. Farkas, J., et al., *Improved growth media and culture techniques for genetic analysis and assessment of biomass utilization by Caldicellulosiruptor bescii.* Journal of industrial microbiology & biotechnology, 2013. **40**(1): p. 41-49.
181. Pawar, S.S., et al., *Biofilm formation by designed co-cultures of Caldicellulosiruptor species as a means to improve hydrogen productivity.* Biotechnology for Biofuels, 2015. **8**(1): p. 19.
182. Cha, M., D. Chung, and J. Westpheling, *Deletion of a gene cluster for [Ni-Fe] hydrogenase maturation in the anaerobic hyperthermophilic bacterium Caldicellulosiruptor bescii identifies its role in hydrogen metabolism.* Appl Microbiol Biotechnol, 2016. **100**(4): p. 1823-31.
183. Dumitrache, A., et al., *Tracking the cellulolytic activity of Clostridium thermocellum biofilms.* Biotechnol Biofuels, 2013. **6**(1): p. 175.
184. Manga, P., et al., *Replicates, Read Numbers, and Other Important Experimental Design Considerations for Microbial RNA-seq Identified Using Bacillus thuringiensis Datasets.* Frontiers in Microbiology, 2016. **7**(794).
185. Kataeva, I.A., et al., *Genome sequence of the anaerobic, thermophilic, and cellulolytic bacterium "Anaerocellum thermophilum" DSM 6725.* J Bacteriol, 2009. **191**(11): p. 3760-1.
186. Clarkson, S.M., et al., *Construction and Optimization of a Heterologous Pathway for Protocatechuate Catabolism in Escherichia coli Enables Bioconversion of Model Aromatic Compounds.* Appl Environ Microbiol, 2017. **83**(18).
187. Tabb, D.L., C.G. Fernando, and M.C. Chambers, *MyriMatch: highly accurate tandem mass spectral peptide identification by multivariate hypergeometric analysis.* J Proteome Res, 2007. **6**(2): p. 654-61.
188. Ma, Z.Q., et al., *IDPicker 2.0: Improved protein assembly with high discrimination peptide identification filtering.* J Proteome Res, 2009. **8**(8): p. 3872-81.
189. Edgar, R.C., *Search and clustering orders of magnitude faster than BLAST.* Bioinformatics, 2010. **26**(19): p. 2460-1.
190. Tschaplinski, T.J., et al., *Down-regulation of the caffeic acid O-methyltransferase gene in switchgrass reveals a novel monolignol analog.* Biotechnol Biofuels, 2012. **5**(1): p. 71.

191. Novo, D., et al., *Accurate flow cytometric membrane potential measurement in bacteria using diethyloxacarbocyanine and a ratiometric technique.* Cytometry, 1999. **35**(1): p. 55-63.
192. Yang, N.C., et al., *A convenient one-step extraction of cellular ATP using boiling water for the luciferin-luciferase assay of ATP.* Anal Biochem, 2002. **306**(2): p. 323-7.
193. Kridelbaugh, D.M., et al., *Nitrogen and sulfur requirements for Clostridium thermocellum and Caldicellulosiruptor bescii on cellulosic substrates in minimal nutrient media.* Bioresour Technol, 2013. **130**: p. 125-35.
194. Maloney, P.C., E.R. Kashket, and T.H. Wilson, *A protonmotive force drives ATP synthesis in bacteria.* Proc Natl Acad Sci U S A, 1974. **71**(10): p. 3896-900.
195. Forrest, W.W., *Adenosine triphosphate pool during the growth cycle in Streptococcus faecalis.* J Bacteriol, 1965. **90**(4): p. 1013-8.
196. Buckel, W. and R.K. Thauer, *Energy conservation via electron bifurcating ferredoxin reduction and proton/Na(+) translocating ferredoxin oxidation.* Biochim Biophys Acta, 2013. **1827**(2): p. 94-113.
197. Baltscheffsky, M., *Inorganic pyrophosphate and ATP as energy donors in chromatophores from Rhodospirillum rubrum.* Nature, 1967. **216**(5112): p. 241-3.
198. Baykov, A.A., et al., *Pyrophosphate-fueled Na+ and H+ transport in prokaryotes.* Microbiol Mol Biol Rev, 2013. **77**(2): p. 267-76.
199. Chen, J., et al., *Pyrophosphatase is essential for growth of Escherichia coli.* J Bacteriol, 1990. **172**(10): p. 5686-9.
200. McIntosh, M.T. and A.B. Vaidya, *Vacuolar type H+ pumping pyrophosphatases of parasitic protozoa.* Int J Parasitol, 2002. **32**(1): p. 1-14.
201. Serrano, A., et al., *Proton-pumping inorganic pyrophosphatases in some archaea and other extremophilic prokaryotes.* J Bioenerg Biomembr, 2004. **36**(1): p. 127-33.
202. Schocke, L. and B. Schink, *Membrane-bound proton-translocating pyrophosphatase of Syntrophus gentianae, a syntrophically benzoate-degrading fermenting bacterium.* Eur J Biochem, 1998. **256**(3): p. 589-94.
203. Bielen, A.A., et al., *Pyrophosphate as a central energy carrier in the hydrogen-producing extremely thermophilic Caldicellulosiruptor saccharolyticus.* FEMS Microbiol Lett, 2010. **307**(1): p. 48-54.
204. Csonka, L.N., *Physiological and genetic responses of bacteria to osmotic stress.* Microbiological Reviews, 1989. **53**(1): p. 121-147.
205. Wood, J.M., *Bacterial responses to osmotic challenges.* The Journal of General Physiology, 2015. **145**(5): p. 381-388.
206. Beales, N., *Adaptation of microorganisms to cold temperatures, weak acid preservatives, low pH, and osmotic stress: a review.* Comprehensive reviews in food science and food safety, 2004. **3**(1): p. 1-20.
207. Klinke, H.B., A. Thomsen, and B.K. Ahring, *Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass.* Applied microbiology and biotechnology, 2004. **66**(1): p. 10-26.
208. Almeida, J.R., et al., *Increased tolerance and conversion of inhibitors in lignocellulosic hydrolysates by Saccharomyces cerevisiae.* Journal of chemical technology and biotechnology, 2007. **82**(4): p. 340-349.

209. Martorell, P., et al., *Physiological characterization of spoilage strains of Zygosaccharomyces bailii and Zygosaccharomyces rouxii isolated from high sugar environments.* International journal of food microbiology, 2007. **114**(2): p. 234-242.
210. Ingram, M. and A. Kitchell, *Salt as a preservative for foods.* International Journal of Food Science & Technology, 1967. **2**(1): p. 1-15.
211. Roe, A.J., et al., *Perturbation of anion balance during inhibition of growth of Escherichia coli by weak acids.* Journal of bacteriology, 1998. **180**(4): p. 767-772.
212. Ljunggren, M., et al., *A kinetic model for quantitative evaluation of the effect of hydrogen and osmolarity on hydrogen production by Caldicellulosiruptor saccharolyticus.* Biotechnology for Biofuels, 2011. **4**(1): p. 31.
213. Willquist, K., P.A. Claassen, and E.W. Van Niel, *Evaluation of the influence of CO 2 on hydrogen production by Caldicellulosiruptor saccharolyticus.* International Journal of Hydrogen Energy, 2009. **34**(11): p. 4718-4726.
214. Liu, D., B. Min, and I. Angelidaki, *Biohydrogen production from household solid waste (HSW) at extreme-thermophilic temperature (70 °C) – Influence of pH and acetate concentration.* International Journal of Hydrogen Energy, 2008. **33**(23): p. 6985-6992.
215. Willquist, K. and E.W. van Niel, *Lactate formation in Caldicellulosiruptor saccharolyticus is regulated by the energy carriers pyrophosphate and ATP.* Metabolic engineering, 2010. **12**(3): p. 282-290.
216. Dumitrache, A., et al., *Consolidated bioprocessing of Populus using Clostridium (Ruminiclostridium) thermocellum: a case study on the impact of lignin composition and structure.* Biotechnology for Biofuels, 2016. **9**(1): p. 31.
217. Mattes, R., et al., *Real-time bioreactor monitoring of osmolality and pH using near-infrared spectroscopy.* BioProcess Int, 2009. **7**(4).
218. Russell, J.B. and G.M. Cook, *Energetics of bacterial growth: balance of anabolic and catabolic reactions.* Microbiological Reviews, 1995. **59**(1): p. 48-62.
219. Zeldes, B.M., et al., *Extremely thermophilic microorganisms as metabolic engineering platforms for production of fuels and industrial chemicals.* Frontiers in Microbiology, 2015. **6**: p. 1209.
220. M. Zeldes, B., et al., *Extremely Thermophilic Microorganisms as Metabolic Engineering Platforms for Production of Fuels and Industrial Chemicals*. Vol. 6. 2015.
221. Olson, D.G., R. Sparling, and L.R. Lynd, *Ethanol production by engineered thermophiles.* Current opinion in biotechnology, 2015. **33**: p. 130-141.
222. DeCastro, M.-E., E. Rodríguez-Belmonte, and M.-I. González-Siso, *Metagenomics of Thermophiles with a Focus on Discovery of Novel Thermozymes.* Frontiers in Microbiology, 2016. **7**(1521).
223. Lin, T., et al., *Transposon mutagenesis as an approach to improved understanding of Borrelia pathogenesis and biology.* Frontiers in Cellular and Infection Microbiology, 2014. **4**: p. 63.
224. Forsyth, R.A., et al., *A genome-wide strategy for the identification of essential genes in Staphylococcus aureus.* Mol Microbiol, 2002. **43**(6): p. 1387-400.

225. Wetmore, K.M., et al., *Rapid Quantification of Mutant Fitness in Diverse Bacteria by Sequencing Randomly Bar-Coded Transposons.* mBio, 2015. **6**(3): p. e00306-15.
226. Good, L. and J.E.M. Stach, *Synthetic RNA Silencing in Bacteria – Antimicrobial Discovery and Resistance Breaking.* Frontiers in Microbiology, 2011. **2**: p. 185.
227. Rasmussen, L.C.V., H.U. Sperling-Petersen, and K.K. Mortensen, *Hitting bacteria at the heart of the central dogma: sequence-specific inhibition.* Microbial Cell Factories, 2007. **6**: p. 24-24.
228. Robertson, H.D., S. Altman, and J.D. Smith, *Purification and properties of a specific Escherichia coli ribonuclease which cleaves a tyrosine transfer ribonucleic acid presursor.* J Biol Chem, 1972. **247**(16): p. 5243-51.
229. Forster, A.C. and S. Altman, *External guide sequences for an RNA enzyme.* Science, 1990. **249**(4970): p. 783-6.
230. Derksen, M., V. Mertens, and G.J.M. Pruijn, *RNase P-Mediated Sequence-Specific Cleavage of RNA by Engineered External Guide Sequences.* Biomolecules, 2015. **5**(4): p. 3029-3050.
231. Shen, N., et al., *Inactivation of expression of several genes in a variety of bacterial species by EGS technology.* Proceedings of the National Academy of Sciences of the United States of America, 2009. **106**(20): p. 8163-8168.
232. McKinney, J., et al., *Inhibition of Escherichia coli viability by external guide sequences complementary to two essential genes.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(12): p. 6605-6610.
233. Li, Y., C. Guerrier-Takada, and S. Altman, *Targeted cleavage of mRNA in vitro by RNase P from Escherichia coli.* Proceedings of the National Academy of Sciences of the United States of America, 1992. **89**(8): p. 3185-3189.
234. Guerrier-Takada, C., Y. Li, and S. Altman, *Artificial regulation of gene expression in Escherichia coli by RNase P.* Proceedings of the National Academy of Sciences of the United States of America, 1995. **92**(24): p. 11115-11119.
235. Chung, D., J. Farkas, and J. Westpheling, *Overcoming restriction as a barrier to DNA transformation in Caldicellulosiruptor species results in efficient marker replacement.* Biotechnology for biofuels, 2013. **6**(1): p. 82.
236. Maniatis, T., E.F. Fritsch, and J. Sambrook, *Molecular cloning: a laboratory manual*. Vol. 545. 1982: Cold Spring harbor laboratory Cold Spring Harbor, NY.
237. Larsen, R.A., et al., *Genetic analysis of pigment biosynthesis in Xanthobacter autotrophicus Py2 using a new, highly efficient transposon mutagenesis system that is functional in a wide variety of bacteria.* Arch Microbiol, 2002. **178**(3): p. 193-201.
238. Hoseki, J., et al., *Directed evolution of thermostable kanamycin-resistance gene: a convenient selection marker for Thermus thermophilus.* J Biochem, 1999. **126**(5): p. 951-6.
239. Cha, M., J. Westpheling, and D. Chung, *Recombinant Caldicellulosiruptor bescii and methods of use*. 2016, Google Patents.
240. Zilsel, J., P.H. Ma, and J.T. Beatty, *Derivation of a mathematical expression useful for the construction of complete genomic libraries.* Gene, 1992. **120**(1): p. 89-92.

241. Sczakiel, G. and R. Far, *The role of target accessibility for antisense inhibition.* Current opinion in molecular therapeutics, 2002. **4**(2): p. 149-153.
242. Engdahl, H.M., T.Å. Hjalt, and E.G.H. Wagner, *A two unit antisense RNA cassette test system for silencing of target genes.* Nucleic acids research, 1997. **25**(16): p. 3218-3227.
243. Campbell, T.B., C.K. McDonald, and M. Hagen, *The effect of structure in a long target RNA on ribozyme cleavage efficiency.* Nucleic acids research, 1997. **25**(24): p. 4985-4993.
244. Kronenwett, R., R. Haas, and G. Sczakiel, *Kinetic selectivity of complementary nucleic acids: bcr-abl-directed antisense RNA and ribozymes.* J Mol Biol, 1996. **259**(4): p. 632-44.
245. Persson, C., E.G. Wagner, and K. Nordstrom, *Control of replication of plasmid R1: kinetics of in vitro interaction between the antisense RNA, CopA, and its target, CopT.* EMBO J, 1988. **7**(10): p. 3279-88.
246. Rittner, K., C. Burmester, and G. Sczakiel, *In vitro selection of fast-hybridizing and effective antisense RNAs directed against the human immunodeficiency virus type 1.* Nucleic Acids Res, 1993. **21**(6): p. 1381-7.
247. Engdahl, H.M., M. Lindell, and E.G.H. Wagner, *Introduction of an RNA stability element at the 5'-end of an antisense RNA cassette increases the inhibition of target RNA translation.* Antisense and Nucleic Acid Drug Development, 2001. **11**(1): p. 29-40.
248. Nakashima, N., T. Tamura, and L. Good, *Paired termini stabilize antisense RNAs and enhance conditional gene silencing in Escherichia coli.* Nucleic Acids Research, 2006. **34**(20): p. e138-e138.
249. Geller, B.L., et al., *Inhibition of Gene Expression in Escherichia coli by Antisense Phosphorodiamidate Morpholino Oligomers.* Antimicrobial Agents and Chemotherapy, 2003. **47**(10): p. 3233-3239.
250. Deere, J., P. Iversen, and B.L. Geller, *Antisense Phosphorodiamidate Morpholino Oligomer Length and Target Position Effects on Gene-Specific Inhibition in Escherichia coli.* Antimicrobial Agents and Chemotherapy, 2005. **49**(1): p. 249-255.
251. Chen, H., G. Ferbeyre, and R. Cedergren, *Efficient hammerhead ribozyme and antisense RNA targeting in a slow ribosome Escherichia colimutant.* Nature biotechnology, 1997. **15**(5): p. 432.
252. Uhlenbeck, O.C., *A small catalytic oligoribonucleotide.* Nature, 1987. **328**(6131): p. 596.
253. Tabler, M., et al., *A three-nucleotide helix I is sufficient for full activity of a hammerhead ribozyme: advantages of an asymmetric design.* Nucleic Acids Research, 1994. **22**(19): p. 3958-3965.
254. Vigouroux, A., et al., *Tuning dCas9's ability to block transcription enables robust, noiseless knockdown of bacterial genes.* Molecular systems biology, 2018. **14**(3): p. e7899.
255. Cruz, C. and J. Houseley, *Endogenous RNA interference is driven by copy number.* eLife, 2014. **3**: p. e01581.

256. Tatout, C., E. Gauthier, and H. Pinon, *Rapid evaluation in Escherichia coli of antisense RNAs and ribozymes.* Letters in applied microbiology, 1998. **27**(5): p. 297-301.
257. Guerrier-Takada, C., R. Salavati, and S. Altman, *Phenotypic conversion of drug-resistant bacteria to drug sensitivity.* Proceedings of the National Academy of Sciences, 1997. **94**(16): p. 8468-8472.
258. Biswas, R., et al., *Improved growth rate in Clostridium thermocellum hydrogenase mutant via perturbed sulfur metabolism.* Biotechnol Biofuels, 2017. **10**: p. 6.
259. Biswas, R., et al., *Elimination of hydrogenase active site assembly blocks H2 production and increases ethanol yield in Clostridium thermocellum.* Biotechnol Biofuels, 2015. **8**: p. 20.
260. Rydzak, T., et al., *End-product induced metabolic shifts in Clostridium thermocellum ATCC 27405.* Appl Microbiol Biotechnol, 2011. **92**(1): p. 199-209.
261. Rydzak, T., L.R. Lynd, and A.M. Guss, *Elimination of formate production in Clostridium thermocellum.* J Ind Microbiol Biotechnol, 2015. **42**(9): p. 1263-72.
262. Sanchez, C., et al., *Contribution of citrate metabolism to the growth of Lactococcus lactis CRL264 at low pH.* Appl Environ Microbiol, 2008. **74**(4): p. 1136-44.

Appendix

1. https://www.google.com/search?q=Biofuels+issues+and+trends+%282012%29%2C+US+EIA&ie=utf-8&oe=utf-8&client=firefox-b-1-ab
2. https://www.frontiersin.org/article/10.3389/fmicb.2016.00794
3. https://www.frontiersin.org/article/10.3389/fmicb.2016.00794
4. https://www.frontiersin.org/article/10.3389/fmicb.2016.00794

Vita

Punita Manga was born in Jodhpur, Rajasthan, India on July 13th, 1989, the daughter of Dr. Vinod Kumar Manga and Mrs. Sudershan Manga. After completing her high school studies from Delhi Public School in Jodhpur, Rajasthan, she received her Bachelor of Technology in Biotechnology at SRM University, Chennai, Tamil Nadu, India. She spent a year in the Master of Technology in Biotech program at Birla Institute of Technology, Hyderabad, India before joining as a graduate student in the Genome Science and Technology Graduate Program at The University of Tennessee, Knoxville in the fall of 2012. She received her Ph.D. in Life Sciences in 2018.