METABOLIC ENGINEERING OF A CELLULOLYTIC THERMOPHILE FOR RENEWABLE PRODUCTION OF BIOFUELS

by

AMANDA MICHELE WILLIAMS-RHAESA

(Under the Direction of Michael W. W. Adams)

ABSTRACT

Sustainable generation of liquid transportation fuels is critical to mitigate the environmental impacts of fossil fuels. Lignocellulosic biomass is an attractive renewable resource but the natural recalcitrance of plant materials to enzymatic and microbial deconstruction must first be overcome. Costly pretreatment processing of the biomass and the addition of exogenous cellulase enzymes to release sugars are necessary. Microbes then utilize the sugars in standard industrial ethanol production. Consolidated bioprocessing (CBP) has the potential to decrease the costs associated with the use of lignocellulosic feedstocks by combining biomass deconstruction and microbial conversion of cellulose and hemicellulose into ethanol all in one fermentation vessel. Cellulolytic thermophiles have been explored for use in this technology as their high growth temperatures facilitate breakdown of the plant material, and decrease the risk of contamination of fermentation vessels. *Caldicellulosiruptor bescii* (T_{opt} 78°C) is the most thermophilic cellulolytic bacterium known and can utilize plant biomass without pretreatment. The recent development of a genetic system in this organism has greatly increased its potential utility for CBP. The goal of this research was to engineer *C. bescii* for maximum ethanol production. First, the limitations of growing wild-type *C. bescii* were explored

on high concentrations of both a model substrate, crystalline cellulose (Avicel), and the real-world substrate, switchgrass. This study revealed that substrate utilization was limited by nitrogen availability and the production of >150 mM organic acids. Generation of a neutral product, such as ethanol, would alleviate this second limitation. Second, improvements were made to the genetics system, including the development of a more stable genetic background and also of a native xylose-inducible promoter, thereby expanding the genetic toolkit for this organism. Third, the more stable background was used for heterologous expression of a cytoplasmic bifunctional alcohol dehydrogenase with and without a reduced ferredoxin NAD oxidoreductase. This six-subunit membrane-bound enzyme provides redox balance for ethanol production and allowed for the generation of a maximum of 75 mM ethanol from cellulose at 60°C. This is the highest ethanol production by *C. bescii* to date. These advancements bode well for the future use of *C. bescii* as a platform organism for biotechnology.

INDEX WORDS: *Caldicellulosiruptor bescii*, consolidated bioprocessing, biofuels, fermentation, metabolic engineering, thermophile, anaerobe, genome stability, transposon, inducible promoter, ethanol

METABOLIC ENGINEERING OF A CELLULOLYTIC THERMOPHILE FOR RENEWABLE PRODUCTION OF BIOFUELS

by

AMANDA MICHELE WILLIAMS-RHAESA

B.S., University of Georgia, 2012

A.B., University of Georgia, 2012

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2017

METABOLIC ENGINEERING OF A CELLULOLYTIC THERMOPHILE FOR RENEWABLE PRODUCTION OF BIOFUELS

by

AMANDA MICHELE WILLIAMS-RHAESA

Major Professor:	Michael W. W. Adams
Committee:	William N. Lanzilotta
	William B. Whitman
	Robert J. Maier
	Christopher M. West

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
December 2017

DEDICATION

To my best friend and husband, Jeff, you have helped me remember to enjoy life especially when facing challenges. Without your love and support none of this would have been possible.

To my mom, Julie, you have shaped me in so many ways and though you could not be here to see me graduate, you are in my heart always.

ACKNOWLEDGEMENTS

There are so many people that deserve thanks and acknowledgement for their contributions to my growth personally and as a scientist. First I would like to thank my major professor Dr. Adams for his exceptional guidance and support both as an undergraduate and a graduate student in Biochemistry and Molecular Biology Department. Thanks to my committee members, Dr. Lanzilotta, Dr. Maier, Dr. Whitman, and Dr. West, for their assistance. Special thanks to Dr. Puett who set me on the path to this accomplishment by taking the time to ask me about my research interests while enjoying a lovely breakfast in Rome during my undergraduate study abroad. His recommendation is what got my foot in the door with the Adams lab and started this entire journey.

I also want to thank all the members of the Adams lab, it has been a joy to work with all of you. Thank you for all the time you have spent sharing practical knowledge, technical support, training me, setting up jerry-rigged equipment, offering encouragement, taping cat sticky notes around my bench, and talking about science, politics or Georgia football. Without all of your support none of the work described in the following chapters would have been possible. In particular, I'd like to thank: Farris Poole, Gabe Rubinstein, Israel Scott, Mirko Basen, Gina Lipscomb, and Gerti Schut. Thanks to my collaborators and friends in Dr. Kelly's lab at North Carolina State University: Jonathan Conway, Laura Lee, Jeff Zurawski, and Piyum Khatibi. You all have been as much a part of this journey as members of the Adams lab and I greatly appreciate all of your support. Lastly, I would like to thank my friends and family, who played a vital role in keeping me motivated throughout this process.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Need for Sustainable Biofuels

Global warming and limits of fossil fuel supplies are driving the need for sustainable alternatives for fuels and chemicals. Fossil fuels are high energy carbon molecules produced through thermochemical processes occurring on ancient biomass on the geologic time scale of millions of years (1-3). For alternative energy sources to be sustainable, they need to rely on carbon neutral sources and processes that have faster production time scales than the geological time required by fossil fuels. Solar and wind energy have made extensive gains in recent years (**Fig. 1.1**) (4). However, these innovations focus primarily on electric power and will not be suitable for replacement of liquid fuels and commodity chemicals generated from petroleum. Major increases in biofuel consumption have occurred over the past twenty years, the majority of which can be attributed to fuel ethanol and biodiesel **(Fig. 1.2)** (4). While these gains help to push the U.S. towards using more renewable energy sources, they are not necessarily sustainable and major scientific and technological advancements are still needed.

It is useful to examine the metrics for fuel ethanol production as they inform the state of the field for each of the technologies discussed below. The two most important metrics are yield and titer (5-7). Yield can be thought of as the efficiency of the process where the majority of the feedstock is going to the desired product. For ethanol, the maximum theoretic yield from glucose is two, and processes with greater than 90% yield are highly desirable (5). Titer is perhaps the more important of the two metrics for industrial production of ethanol as the costs of ethanol

recovery decrease dramatically the higher the ethanol concentration. With mixtures below 4% w/w (42 g/L or 900 mM ethanol), recovery costs are too high, above 35% of the combustion energy of ethanol recovered, to be economically feasible (8). For all potential biofuels, the energy inputs for all stages of production must below the combustion energy recovered in the fuel for the process to be sustainable and for ethanol this requires the generation of high ethanol titers (8).

Feedstocks for Biofuel Production

Current ethanol production for biofuels relies primarily on food based feedstocks such as corn in the United States, sugarcane in Brazil and sugar beets in the European Union (2, 3, 9). While the use of these sugar based feedstocks for ethanol production are very mature industries, there are a number of economic and environmental issues related to the competition of food crops and the biofuel industry (2, 3, 10). These issues include increasing food and animal feed prices as corn crops are directed toward biofuel production, the need for fertilizers for these crops as well as limited availability of land suitable for farming (2, 10-13). While it is highly debated whether the use of food crops for ethanol production does increase food prices, it is clear they are tightly linked through the increase of demand for these crops (10, 11, 14-16). The use of food crops for ethanol production also requires intensive agriculture that increases strain on arable land and decreases the energy economy of these crops with corn ethanol generating just 25% more energy than invested in its production and reducing greenhouse gas emission by only 12% compared to fossil fuels (12). Clearly, other alternatives need to be explored to meet our needs for sustainable energy.

Non-food based feedstocks include small scale agriculture waste, waste generated during crop processing and manufacturing, municipal solid waste as well as dedicated biofuel crops which all serve a potential role in the sustainable fuel future. Small scale solutions include the use of

waste crops which are not suitable for market or waste materials generated in plant processing to help run the in-house equipment in a more sustainable fashion (2, 3, 17-19). In one study, *Saccharomyces cerevisiae* was used for the fermentation of cull peaches, rotten fruit not suitable for market, for the production of fuel ethanol on peach farms (19). The waste streams of rice milling, broken, unripe and discolored rice, rice bran and rice husk, have been explored as for their feedstock potential for bioethanol production by *S. cerevisiae* generating 50 g/L ethanol, above the required titers for an economical process, from both individual and mixed waste streams (17). Municipal waste as well as microalgae have been investigated as potential sources of renewable gases and biodiesel (15, 20, 21). Municipal wastewater is currently treated to remove the solid waste or sludge, which is then disposed of or reused as fertilizer. An additional potential use is for biofuel production (22). This sludge can either be placed in an anaerobic digester for biological methane production as a sustainable source or natural gas (21) or it can act as a growth substrate for microalgae, which are rich sources of hydrocarbons for biodiesel production (20, 23). These fuels all serve to meet the needs of niche fuel markets, but large scale solutions are also needed to meet sustainability goals.

Larger scale use of other agricultural wastes and dedicated biofuel crops have also been explored largely representing lignocellulosic feed stocks (24-29). Corn stover, stalks, leaves and cobs remaining in a corn field after harvest, is an example that has been explored in depth in the United States (24-29). As with any agricultural residue, many of the issues of land use and intensive agriculture requiring high fertilizer and pesticide applications remains. However, since these materials are products of food production and their use does not compete with the supply and demand for food, they make attractive potential feedstocks. One issue that is not often considered is the balance between the current uses of corn stover, to prevent erosion and soil

nutrient depletion, and the potential removal of some or all of the stover for biofuel production (29). Dedicated non-food biofuel crops, such as switchgrass and poplar, which can grow on more marginal lands, overcome many of these issues (12, 13, 16, 30, 31). Ideal biofuel crops would be perennial grasses or fast growing trees that require minimal inputs of fertilizers and ideally replenish the soil in which they grow (32). Switchgrass, for example, has been shown to increase the biodiversity of the environment compared to standard food crops, sequester carbon in the soil below tillage depths and improve the available phosphorus content of the soil (32-35). In fact, due to the lower inputs, dedicated fuel crops have the possibility of generating more than four times the energy needed to produce them (12). There are still differing opinions in the field about land use and the likelihood of implementation by farmers (10-13, 16, 30). As the disruption of forest land leads to the release of sequestered carbon in the soil, abandoned farm land, and the marginal edges of current farm land make excellent targets for potential regions to grow biofuel crops (30). Likely the solution will require a mixture of approaches, both small and large scale, to meet demand for sustainable liquid fuels.

Lignocellulose represents a major untapped feedstock for biofuel and chemical production as cellulose is the most abundant organic polymer on earth, however, the innate recalcitrance of these materials makes their use far more challenging (36). The major components of lignocellulose are cellulose, hemicellulose and lignin with the exact ratios varying between different plant species (37). Overall, the plant cell wall is a highly complex structure with the exact composition, linkages and properties still a highly rich area of research (38). A number of difficulties must be overcome before plant biomass can be used as a feedstock, including heterogeneity of plant cell walls, differences in composition between plant species and over different growing seasons, toxicity of feedstock components to fermentation hosts, and general recalcitrance of plant biomass to

breakdown (16, 36, 37, 39-44). Addressing these challenges requires work at all levels from process development to plant and microbial engineering.

Arguably the most difficult of these challenges to address is the recalcitrance of plant biomass. This issue is being tackled on two fronts. The first is engineering plants to improve them for use as biofuel feedstocks and the second is the development of pretreatment technologies (45, 46). Plants are recalcitrant to degradation largely due to the selective pressures shaping their evolution and resulting in plants that grow upright (45, 47). One way to try and overcome this recalcitrance is engineering plants for easier microbial and enzymatic breakdown while maintaining their structural integrity for growth in the field (45, 48). Alterations in biosynthetic pathways for lignin, pectin and hemicellulose have all resulted in plants that are more easily digested, but field trials are still ongoing and needed to confirm that these modified plants have robust growth and disease resistance (45, 48-51). In addition to modifying the plant cell wall through bioengineering, plants have been modified to produce cellulase enzymes from bacteria and fungi that are compartmentalized and activated when the plant material is processed for ethanol production (52, 53). Plant produced microbial-cellulases are currently not in use as pretreatment technologies for current industrial processing are too harsh for the enzymes to remain active.

Pretreatment technologies have been the focus of a large effort in the biofuels research community working to increasing the digestibility of feedstocks through thermochemical and solvent treatments (27, 54-57). The majority of pretreatments rely on temperature, pH extremes, or solvent interactions to remove specific components and linkages of the plant cell wall and increase accessibility of the cellulose to enzymatic breakdown (58). Dilute acid pretreatment using sulfuric acid in combination with high temperatures (>100°C) is very common and results in the solubilization of hemicellulose while the cellulose and lignin components of plant biomass remain

insoluble (55, 57). The acidic pretreatments have to be optimized in order to balance the removal of hemicellulose to increase accessibility to enzymatic hydrolysis and the generation of toxic byproducts such as furans (57). The addition of co-solvents, such as tetrahydrofuran, to dilute acid pretreatments allows for efficient removal of lignin from biomass, dramatically increasing the efficiency of downstream hydrolysis (56). Alkali pretreatments, primarily ammonia fiber expansion (AFEX), involve the cleavage of hemicellulose and lignin bonds (54, 57). The breakage of the key linkages as well as the mechanical stress of sudden pressure changes used in this treatment leads to an increase in the digestibility of the plant biomass (54, 57). Ionic liquid (IL) pretreatments involve the use of non-volatile ionic solvents with low melting points to fractionate plant biomass by competing for hydrogen bonding resulting in temporary solubilization of the lignocellulose (27, 55, 57). The cellulose, hemicellulose, and lignin are recovered when they precipitate as the IL is removed through dilution with water (27). The high cost of ILs is often prohibitive despite the efficiency of this treatment in generating easily digestible final materials. The use of ILs in conjunction with alkali treatments has been explored as a means to make their use more cost effective (57). The pairing of specific pretreatments with the specific fermentation organisms and downstream processing remains an under explored area.

Biological Conversion of Lignocellulose

A number of different technologies have been studied for the downstream processing of lignocellulosic biomass into fuel molecules and they can be separated by how many of the biologically mediated steps occur in one reaction vessel. Four biological steps have been identified in the production of ethanol from cellulosic biomass: hydrolase enzyme production, polysaccharide hydrolysis, hexose sugar fermentation, and pentose sugar fermentation (41). The first method for processing is termed separate hydrolysis and fermentation (SHF) and requires a

different reaction vessel for each step (**Fig 1.3**) (41, 59). The second method, simultaneous saccharification and fermentation (SSF), combines the polysaccharide hydrolysis and hexose, and more recently pentose, sugar fermentation steps and was shown to dramatically decrease costs (41, 59, 60). The final process combines all four biological steps in one microbial community or a single organism and was originally called direct microbial conversion (DMC), which was later renamed consolidated bioprocessing (CBP) to remove confusion generated in the literature (41). Even with this clarification, the definition in the literature of CBP has somewhat broadened to encompass any process that uses one organism to generate one product from a feedstock, often not cellulose based, such as starch, in a single step (20, 61-63). However this dissertation will focus solely on the original definition. CBP is the least developed of these technologies largely due to limitations discussed below, however, it has the potential to eliminate the need for dedicated cellulase production which could dramatically decrease costs (41, 42, 64).

Looking at the competitiveness of biological processes with oil and gasoline prices is useful for understanding how the field has moved in the past 30 years. SHF was projected to generate ethanol at a selling price of $3.60/gal in 1988, which was not competitive with oil prices at the time with gasoline at $0.62/gal (59, 60). In the same studies, SSF was predicted to generate ethanol with a selling price of between $1.22/gal and $1.78/gal, depending on how pentose sugars are processed (59, 60). These improvements in price were primarily due to increases in hydrolysis rates, as the free sugars are strong inhibitors to the cellulase enzymes, generated in a separate step in different organisms. In SSF, as soon as the sugars are released, they are fermented to ethanol, which has a much lower inhibitory effect on the enzymes (41, 59). Additionally, the likelihood of contamination by non-process microorganisms is also mitigated as there is less free sugars available in the broth and ethanol can act as a microbial growth inhibitor (60). Since these

technologies were first studied, a number of commercial plants began production of ethanol from a variety of cellulosic feedstocks using this method with projected prices ranging from $2.15/gal to $4.55/gal in 2016, with the price dependent largely on feedstock costs (65, 66). Only the lowest of these prices was competitive with gasoline prices in 2016, which were on average $2.15/gal in the United States (67). Further improvements are still needed to make cellulosic ethanol production competitive with sugar based ethanol production and CBP offers promising opportunities for future cost reduction.

Consolidated Bioprocessing

To achieve CBP, there are two potential starting points that have been pursued and each has different challenges (64, 68, 69). The first is starting from common microbial hosts with robust genetics systems and long histories of use at the industrial scale, capable of generating 123 g/L ethanol from sugar in the case of *S. cerevisiae*, and engineering them for the production and excretion of cellulase enzymes (70). This starting point comes with the challenge of transferring a complex phenotype of cellulose deconstruction, although any generation of cellulase by yeast has the potential to greatly decrease the cost and loadings of exogenous enzymes (46, 71-73). The second is starting from a natively cellulolytic organism and engineering it for robust production of a fuel. Major challenges faced when adopting this strategy include limited genetic tools available in cellulolytic organisms and limited understanding of their metabolisms and regulatory networks (73-75). Neither method for CBP has been demonstrated successfully at industrial scale to date despite many recent advancements.

For non-cellulolytic host organisms, the expression of carbohydrate-degrading enzymes is critical for potential use for CBP and for decreasing costs even if exogenous enzymes are still needed. Generally, three types of enzyme are needed to degrade cellulose: β-glucosidase, which

cleaves cellobiose from the non-reducing end of cellulose; endoglucanase, which cleaves non-terminal glycoside bonds; and cellobiohydrolase, which cleaves cellobiose into two glucose molecules (76). The degradation of plant biomass requires additional enzymes for the cleavage of hemicellulose linkages as well as a multitude of other bonds contributing to the recalcitrance of plant biomass, however, a minimum set of required genes has not yet been identified (52, 76-78).

Studies in yeast have led to strains with improved ethanol yields and reductions in enzyme loadings required for SSF (19, 76, 79-83). To process cull peaches, *S. cerevisiae* was engineered for the heterologous expression of the pectate lyase B from *Paenibacillus amylolyticus* and this allowed the generation of 82% of theoretical ethanol yield without the addition of exogenous pectinase (19). This strain would likely also be useful in processing other fruit waste rich in pectin. Another study examined the synergistic role of four strains of *S. cerevisiae* which individually expressed and hyper-secreted the cellobiohydrolase-1 from *Chaetomium thermophilium*, the cellobiohydrolase-2 from *Chrysosporium lucknowense*, the endoglucanse from *Trichoderma reesei*, and the β-glucosidase from *Saccharomycopsis fibuligera* (76). It was found that by using a mixture of strains each secreting one cellulase-related enzyme, enzyme loadings could be reduced by one third and still maintain the same performance in an SSF scheme (76).

It is not clear if these free enzyme are the most effective as there is evidence that attachment of cellulolytic organisms to the surface of plant materials and to cellulose play a synergistic role in cellulose degradation (84-86). In order to transfer this capability to non-cellulolytic organisms, two strategies have been employed, cell surface display and minicellulosome systems (87, 88). In one study, two cellulase genes from *Aspergillus oryzae* were co-expressed in one strain of yeast leading to the generation of 8 g/L ethanol from barley β-glucan with no exogenous enzymes (88). Successful use of a minicellulosome with genes from

Clostridium cellulolyticum in *S. cerevisiae* has been documented which makes use of the self-assembly of cellulase enzymes docked to a scaffold protein allowing for attachment of the host cell to cellulose for degradation, however only very low ethanol titer was achieved, 1.4 g/L from microcrystalline cellulose (87). It is clear from these studies that the use of non-cellulolytic organisms without the addition of cellulase enzymes will require a great deal of additional work.

Some of the innovations leading the way toward CBP include engineering organisms for co-utilization of both C_5 and C_6 sugars in addition to the heterologous expression of cellulase enzymes to improve solubilization. Since plant biomass typically contains comparable amounts of glucose- and xylose-based carbohydrates, the ideal organism will be able to utilize both simultaneously (89). To that aim, many different organisms, including *S. cerevisiae* and *Escherichia coli*, have been engineered to either overcome carbon catabolite repression or to utilize sugars they were unable to metabolize previously (89-91). Carbon catabolite repression is the transcriptional regulation of genes required for the utilization of different carbon sources in the presence of a preferred substrate. A strain of *E. coli* was engineered for the co-utilization of cellobiose and xylose, where xylose is normally the preferred substrate, through the constitutive expression of cellobiose transporters and β-glucosidase followed by strain evolution to improve the efficiency of this process (89). *S. cerevisiae* does not natively utilize xylose as a growth substrate, thus efforts have been made to generate strains expressing the xylose fermentation genes from *Scheffersomyces stipites*, which can co-utilize cellobiose and xylose, consuming both sugars equally and generating of 30 g/L ethanol from 40 g/L of each sugar (91).

Tolerance of inhibitors present in the native plant material as well as those generated during pretreatment is of major importance for potential CBP microbes. There has been work aimed at increasing tolerance to these microbial inhibitory compounds, including adaptive evolution of

strains (92), developing genetic tools to work in naturally tolerant organisms (93) and characterization of pathways allowing tolerance for transfer into other organisms (94). The goal of one study to evolve *S. cerevisiae* for increased tolerance to high pH found that duplications of whole chromosomes were common when pH changes were sudden, while different adaptations occurred when stress was more gradually applied (95). Adaptive evolution has also been used in the natively cellulolytic thermophile *Clostridium thermocellum*, which resulted in mutations in repair pathways for DNA and RNA (96). Many organisms, including a number of yeast species and members of the genus *Thermoanaerobacter*, are naturally tolerant of many of the inhibitory compounds found in the hydrolysates produced during pretreatment (93, 97).

In order to take advantage of these stress-tolerant organisms for use in CBP, genetic systems must be developed. To that aim, a transformation system was developed for the acid and thermotolerant yeast, *Issatchenkia orientalis*, allowing for the heterologous expression of the β-glucosidase from *Aspergillus aculeatus*, which resulted in the production of 29 g/L ethanol at 40°C without exogenous β-glucosidase in the cellulase mixture (93). Alternatively, as the tolerance mechanisms of these organisms are characterized, they can be transferred to the platform organisms currently in use or to natively cellulolytic microbes, but once more genetic systems are required. A common pathway that has been added to *S. cerevisiae*, *Caldicellulosiruptor bescii* and *C. thermocellum* involves the reduction of furfural and 5-hydroxymethylfurfural (5-HMF) in an NADH or NADPH dependent manner (94, 98, 99). The study in yeast examined four different alcohol dehydrogenase genes and characterized the cofactor preference and the activity with either furfural or 5-HMF. They found that a strain expressing all of these genes was able to grow in the presence of 12 mM furfural and 5-HMF, both of which completely inhibited growth in the parent strain (94). The expression of a thermostable butanol dehydrogenase from *Thermoanaerobacter*

pseudethanolicus 39E, which was shown previously to have activity with furfural and 5-HMF, increased the tolerance of both *C. bescii* and *C. thermocellum* to furfural and 5-HMF (98, 99).

The major breakthrough for the use of natively cellulolytic organisms for ethanol production has been the development of genetic systems. This has been achieved for cellulolytic organisms that natively produce ethanol, such as the cellulolytic fungus *Fusarium oxyporum* and the cellulolytic thermophile *C. thermocellum* (100-104). Both of these organisms produce ethanol below the yields and titers required for industry. Genetic systems have also been developed for other cellulolytic organisms that do not natively produce ethanol, such as *Caldicellulosiruptor bescii* (105-107). In either of these cases having genetic tools is absolutely critical. Cellulose and hemicellulose degradation has been improved in both *C. thermocellum* and *C. bescii* through the heterologous expression of additional cellulase enzymes from different organisms (108-112). This is particularly successful when the capabilities of the native organism are well characterized and the addition of a new enzyme targets a linkage not used by the endogenous enzymes.

Metabolic engineering for fuel production has been demonstrated for the natively cellulolytic fungi *Fusarium oxyporum*, *F. verticillioides* and *Acremonium zeae*. All of them natively produce ethanol but further engineering is required to minimize production of side products and improve xylose utilization (100-102, 113-117). Metabolomic data was used to identify bottle necks in the metabolism of *F. oxyporum*, and over expression of its native transaldolase and phosphoglucomutase doubled the ethanol productivity (101). However, the ethanol productivity of *S. cerevisiae* is still fourteen times higher even when grown on hydrolyzed fiber sludge, a pulp mill waste product containing glucose, xylose and mannose (118). Recombinant *F. verticillioides* and *A. zeae* fermented pretreated sugar cane bagasse to 4.6 g/L and 3.9 g/L ethanol, respectively, after six days (115). More work is needed to improve yields and

titers in these organisms as they are well below the required 42 g/L and the production times are far too long. In addition to ethanol, the use of endophytic fungi that are natively cellulolytic and naturally produce volatile organic compounds, mainly monoterpenes and sesquiterpenes, has been explored for biodiesel production (117). One major limitation of using these organisms is the slow growth rate, requiring long process times to reach critical mass before fermentation can begin (114). Cellulolytic fungi have therefore been the focus of a great deal of research and have a number of potential advances when used on the industrial scale (74, 119).

Thermophiles for Biofuel Production

Microorganisms that grow optimally at temperatures above 65°C have gained interest for use in the production of biofuels and other industrially relevant molecules in a sustainable manner (74, 119-122). These organisms have a number of advantages for use at the industrial scale, including decreased risk of contamination, less susceptible to phage attack, decreased cooling costs, and the potential for continuous recovery of volatile products directly from the fermentation broth (74). Additionally, many chemical reactions are more favorable at higher temperatures, such as the deconstruction of plant biomass, which is greatly improved with increased temperatures (74, 119). Finally, thermophilic organisms often have unique metabolic capabilities that are often difficult to transfer into more conventional industrial (74). In fact, thermophilic enzymes have been used commercially since the discovery of the *Thermus aquaticus* DNA polymerase, which has been used in polymerase chain reaction (PCR) since 1976 (123).

Application of thermophiles can involve either whole cell applications as with CBP or use of their enzymes. In many applications in the chemical, pharmaceutical, food, paper and textile industries, the thermostable enzymes alone are ideal (120-122). The vast majority of the work thus far has involved heterologous expression of thermophilic enzymes in mesophilic hosts, however,

their activities are often lower than natively purified enzymes due to incorrect folding (74, 120-122). In a number of cases, expression and purification of an enzyme of a thermophile from a mesophilic host is not successful and instead cell-dependent processes are needed for the maintenance of activity (122). CBP is an excellent example of an application that requires a whole cell process as the organisms are required to ferment the sugars to a fuel (74, 120, 122).

The major breakthrough in using these thermophilic organisms, instead of just their enzymes, has been the emergence of genetic systems allowing for metabolic engineering (74). Development of these systems has been challenging largely due to the extreme growth conditions and a lack of selectable markers under such extreme conditions (74). Antibiotic resistance is challenging as both the antibiotics themselves and the required resistance proteins suffer from low thermostability, however, successful use of antibiotics for selection in a few thermophiles has been documented (124-127). Kanamycin is a thermostable antibiotic and two separate studies have generated mutant versions of the kanamycin-resistance gene, resulting in thermostable gene products for use in *Thermus thermophilus* at 80°C and *Bacillus stearothermophilus* at 70°C (124, 128). These markers have been used in a number of genetic systems for thermophiles, including *Moorella thermoacetica* (T_{opt} 55°C) (129), *Thermoanaerobacter tengcongensis* (T_{opt} 60°C) (130), *Geobacillus thermoglucosidasius* (68°C) (131), *Thermus thermophilus* (T_{opt} 70°C) (132), *Caldicellulosiruptor bescii* (T_{opt} 78°C) (127), and *Thermotoga* spp. (T_{opt} 77°C) (133). Simvastatin has been a very useful antibiotic for use in archaea, which are not susceptible to most antibiotics used in bacterial genetics (74). Overexpression of the 3-hydroxy-3-methylglutaryl coenzyme A reductase gene allows for resistance to simvastatin and has been used in *Thermococcus kodakaraensis* (125) and *Pyrococcus furiosus* (126). Nutritional markers, particularly uracil

prototrophic selection first developed in yeast, are far more common as the basis for thermophile genetics and have been used with great success in both archaeal and bacterial hosts (74).

Thermophiles of particular interest for CBP include the cellulolytic *C. thermocellum* (134), xylanolytic members of the genus *Thermoanaerobacter* (135), and cellulolytic members of the genus *Caldicellulosiruptor* (107, 136). *C. thermocellum* is of high interest due to the high efficiency of the cellulosome in the degradation of crystalline cellulose and its native ability to generate ethanol among other carbon end products (103, 134). Many members of the genus *Thermoanaerobacter*, including *T. mathranii* and *T. pseudethanolicus*, are xylanolytic and metabolize both C_5 and C_6 sugars to generate primarily ethanol and acetate, and these organisms have been explored for use in co-cultures with cellulolytic organisms (103, 135). Members of the genus *Caldicellulosiruptor* are of great interest for CBP as well as biohydrogen production as they generate hydrogen at near the Thauer limit of four hydrogen molecules per glucose consumed (103, 137). Additionally, some of these species can grow up to 90°C and represent the most thermophilic cellulolytic organisms known. They also utilize multidomain free enzymes instead of the cellulosome to break down polysaccharides (107, 136). Advancements that have been made with these organisms will now be discussed.

Clostridium thermocellum

Within the field of cellulosic biofuel production by thermophiles, the majority of research has focused on the thermophile *C. thermocellum*, which grows optimally at 60°C. *C. thermocellum* can utilize a variety of substrates, including crystalline cellulose and plant biomass, but importantly it cannot grow on C_5 sugars present in xylan (138). The cellulose-degrading cellulosome of *C. thermocellum* has been extensively studied. This is a large (210 kDa) heterogeneous multienzyme complex that tethers *C. thermocellum* to the surface of insoluble growth substrates (139). One area

of focus has been on engineering designer cellulosomes with specific components as well as heterologous expression of the complex in other organisms to increase their cellulolytic capabilities (111, 112, 140-146). Designer cellulosomes can be generated by engineering specific dockerin and cohesion modules to generate homogenous and uniform assemblies of cellulosomes, which can be optimized for maximal deconstruction of a given plant material (112, 147, 148). Transferring functional cellulosomes to other organisms has been a challenge as this complex must be exported outside of the cell to function. Despite the challenges, it has been successfully demonstrated in *Thermoanaerobacterium saccharolyticum* (140), *Clostridium acetobutylicum* (141, 144, 146), *S. cerevisiae* (143), and *Bacillus subtilis* (145).

Use of *C. thermocellum* for CBP requires metabolic engineering as the wild-type strain generates ethanol at only low yields, 29% of theoretical maximum, and titers, 10 g/L, when grown on 60 g/L Avicel (149). *C. thermocellum* generates a wide range of fermentation products in addition to ethanol, including formate, acetate, lactate and hydrogen gas, and it also has an unusual glycolytic pathway that contains a malate shunt (103, 104, 150). *C. thermocellum* research has benefitted from the genetic system developed for this organism (151) and directed evolution has been used in an attempt to generate ethanol tolerant strains (152-155). Metabolic engineering in this organism has included the deletion of the lactate dehydrogenase gene (156), elimination of hydrogen production by deletion of the energy conserving (ECH) hydrogenase and the *hydG* responsible for maturation of the Fe-Fe hydrogenases (157), deletion of the pyruvate formate lyase genes responsible for formate production (158), and elimination of acetate production through the deletion of the phosphotransacetylase (159). Combining these engineering strategies has generated strains with more streamlined ethanol production leading to the generation of up to 480 mM (22 g/L) ethanol at 75% of theoretical maximum yield from cellobiose in the most advanced strain

(160). Very similar results were achieved, 415 mM (19 g/L) ethanol at 74% theoretical yield, from cellobiose for strains expressing the ethanol production pathway from *T. saccharolyticum* (161). While these results seem promising, *C. thermocellum* still suffers from major limitations, including its inability to utilize pentose sugars that are inhibitory to growth making this organism less than ideal for CBP of plant biomass where C_5 sugars are highly abundant (162).

Genus *Thermoanaerobacter*

Members of the genus *Thermoanaerobacter* utilize C_5 and C_6 sugars for the generation of ethanol, acetate, lactate and hydrogen with different product profiles for each of the different strains (135). Generally, these species have high ethanol and inhibitor tolerances making them of great interest for use in CBP (97, 163). Since *Thermoanaerobacter* species cannot hydrolyze cellulose, they are more frequently studied as potential co-culture partners to increase ethanol yield through the use of the C_5 sugars released in the hydrolysis of plant biomass (135, 164, 165). Comparative genomics of a variety of *Thermoanaerobacter* species showed that there are large differences in xylanolytic capacity with only strains *T. italicus*, *T. mathranii* and *T. thermohydrosulfuricus* possessing extracellular carbohydrate active enzymes (CAZy enzymes) (135). However, these three strains are not the highest ethanol producers in the genus and have not been evaluated in co-culture with *C. thermocellum*. Two strains in particular have been tested for their use with *C. thermocellum*, *T. pseudethanolicus* 39E and *T.* sp. X514 (164, 165). These studies found that both strains resulted in a 3-5 fold increase in ethanol production from crystalline cellulose, with strain X514 having a greater effect (164). The higher performance of strain X514 was later correlated with the ability to synthesize vitamin B12 de novo as it is an important co-factor for energy maintenance in *Thermoanaerobacter* species (165).

Genus *Caldicellulosiruptor*

Members of the genus *Caldicellulosiruptor* have also been the subject of interest for use in CBP as this genus contains the most thermophilic cellulose-degrading organisms known to date (**Table 1.1**) (107, 136). These gram-positive bacteria are strictly anaerobic thermophiles with optimal growth temperatures between 65°C and 80°C (107, 166). In contrast to *C. thermocellum*, members of the genus *Caldicellulosiruptor* are capable of using both C_5 and C_6 sugars and make use of free cellulase enzymes rather than a cellulosome to deconstruct cellulose. Moreover, a member of this genus, *C. bescii* (formerly *Anaerocellum thermophilum*) was the first thermophilic bacterium shown to degrade plant biomass without prior thermochemical treatment. (107, 136, 167-169). The extracellular plant biomass degrading enzymes utilized by the members of this genus have been of particular interest for industrial purposes as they are highly active even when compared to commercial enzyme preparations (170). The members of this genus encode a number of multidomain extracellular glycosyl hydrolase (GH) domain containing enzymes (**Table 1.1**). The genome of *C. bescii*, for example, encodes 68 GH domains, most of which can be found in two genomic regions termed the Glucan Degradation Locus (GDL) and Xylan Degradation Locus (XDL) (136). Additionally, their simplicity as free individual enzymes compared to the exceedingly complex cellulosome make them attractive for industrial use. The family 3 pectate lyase of *C. bescii* was purified and this led to the determination of its structure as well as characterization of the reaction mechanism of this enzyme (171-173). Another study on the most abundant secreted enzyme from *C. bescii*, CelA, yielded a crystal structure as well as the discovery of a unique mode of cellulose degradation involving the pitting of the surface of cellulose fibers (170, 174). The heterologous expression in *E. coli* of CAZy enzymes with S-layer homology

domains from *C. kronotskyensis* has led to the characterization of the biochemical properties of two of these multidomain enzymes including the individual roles of each domain (85).

In addition to the enzymes themselves, there is a lot to be learned from comparisons between the individual species of this genus. **Table 1.1** shows the different species that have been examined for their cellulose degradation capabilities. A number of these strains, including *C. bescii*, *C. kronotskyensis*, *C. obsidiansis*, and *C. saccharolyticus*, are highly efficient cellulose degraders capable of growth on even highly crystalline cellulose whereas other strains are less cellulolytic (107). *C. lactoaceticus* is moderately cellulolytic, capable of some growth on crystalline cellulose though it is not the preferred growth substrate (107). Other species were incapable of growth on crystalline cellulose providing a case study for the necessary machinery for this process (39). *C. hydrothermalis* is among the least cellulolytic of the *Caldicellulosiruptor* genus and has been purposed as a potential host for study of the minimal GH enzyme requirements for crystalline cellulose degradation (175). The pan-genome of *Caldicellulosiruptor,* which represents the genetic diversity of this genus, is still open such that the discovery of new members will likely lead to the discovery of new CAZy enzymes (136).

Caldicellulosiruptor bescii

Of the *Caldicellulosiruptor* genus members, *C. bescii* has been the most characterized for use as a potential CBP organism. Initial studies of *C. bescii* focused largely on the plant biomass degradation capabilities and physiology of this organism. Annotation of the genome showed that *C. bescii* utilizes a standard Embden-Meyerhof-Parnas glycolysis in addition to the ability to metabolize C_5 sugars through non-oxidative pentose phosphate pathway (**Fig 1.4** and **Fig 1.5**) (176, 177). The major redox recycling enzyme in *C. bescii* is the bifurcating hydrogenase (BF-H_2ase) which transfers electrons from both reduced ferredoxin and NADH to generate hydrogen

gas and represents ~5% of the cytoplasmic protein (unpublished data). In 2009, Yang *et. al.* showed that *C. bescii* was capable of utilizing even high lignin content plant biomass as a sole carbon source without pretreatment (168). Further studies characterized the residual plant material and found that *C. bescii* solubilizes plant biomass leaving the overall composition very similar to that of the initial substrate (106). This was also found to be true for *C. kronotskyensis* and *C. saccharolyticus* in a subsequent study examining all three organisms (178). Other studies have looked at *C. bescii* utilization of a variety of plant biomass sources including natural variants and engineered plants that have been shown to have decreased recalcitrance (168, 179, 180). The results of one study in particular found that the recalcitrance of untreated plant material was the limiting factor in utilization by *C. bescii* (179). Different cultivars of switchgrass having different bioavailability to *C. bescii* with carbohydrate solubilization between 7% for the most recalcitrant material to 36% for the caffeic acid O-methyltransferase (COMT) transgenic line, which has altered lignin biosynthesis and is less recalcitrant than its parent strain, and Cave-in-Rock cultivar, which is a low recalcitrant natural variant of switchgrass (179).

In the research reported herein, in an effort to understand some of the limitations of plant biomass utilization, *C. bescii* was grown on high concentrations, 50-200 g/L, of both crystalline cellulose and the model biomass switchgrass and the results are presented in Chapter 2. The major conclusion was that the physiological limitation of salt tolerance in *C. bescii* was the overall limit of cellulose degradation although switchgrass utilization was also affected by the recalcitrance of the substrate (105). In order to overcome the salt tolerance, *C. bescii* would need to generate a neutral end product, such as ethanol, to further utilize the model substrate cellulose. *C. bescii* does not natively produce ethanol and so a genetic system was required, and this became available (181) when the research reported herein was initiated.

C. bescii genetics

The development of a genetic system for *C. bescii* has been a critical step in the potential use of this organism for CBP. The *pyrF* and *pyrE* genes in *C. bescii* are required for uracil biosynthesis, however, they also lead to the toxicity of 5-fluoroorotic acid (5-FOA), making these genes useful targets for selectable and counter-selectable markers. The first iteration of the system in *C. bescii* relied on the selection of a spontaneous deletion in the *pyrFA* genes selected for by plating on 5-FOA containing medium (181, 182). The resultant strain has served as the parent strain for multiple studies and lead to an increase in understanding of *C. bescii* metabolism, biomass deconstruction, and metabolic engineering for ethanol production (85, 98, 108-110, 181, 183-191). A major improvement to the genetics system was the development of a kanamycin resistance gene for use in *C. bescii* (127). This breakthrough allowed for the development of a new and defined uracil auxotrophic parent strain, MACB1018, directly from the wild-type. In Chapter 3, I investigated the genetic stability of the two lineages to determine which should we used for optimizing ethanol production. I show that the MACB1018 lineage is more stable than the original uracil auxotrophic lineage, JWCB005, which was generated from a spontaneous mutant (192). The major difference in stability can be attributed to the activity of a transposable element, IS*Cbe4*, which has a dramatically increased copy number in the JWCB005 lineage leading to genome scale rearrangements in daughter strains and numerous deletion and insertion mutations (192, 193). While the use of the antibiotic resistance marker in *C. bescii* (127) was a major step forward, in order to further expand the genetic toolbox for this organism I design and developed an inducible promoter system based on the native xylose-inducible promoter, and this research is presented in Chapter 4.

Studies that have been made possible by the genetic tools in this organism have focused on its physiology and enzymology, as well as metabolic engineering. For example, the discovery of a putative transporter for tungstate, an element seldom used in biology, along with a novel oxidoreductase suggested that *C. bescii* could synthesize one or more tungstoenzymes. This was demonstrated by the heterologous expression in *C. bescii* of a known and well characterized tungstoenzyme (190). Similarly, the role of the glycosyl hydrolase (GH) enzymes used in *C. bescii* for plant biomass degradation has been examined through targeted deletion of each of the individual GHs (194). This work revealed that each GH affected growth on different substrates showing how the diversity of these enzymes allows *C. bescii* to metabolize a wide variety of carbon sources (194). In terms of metabolic engineering for biofuel production, heterologous expression of the bifunctional alcohol dehydrogenase, AdhE, from *C. thermocellum* in *C. bescii* led to the production of 12 mM ethanol from switchgrass (185). The overall goal of this research was therefore to use the knowledge learned from growing *C. bescii* on high concentrations of cellulose and switchgrass and use the new stable genetic background strain to significantly improve ethanol production by *C. bescii* by considering the primary redox metabolism of the cell. The results of that research are presented in Chapter 5.

Research Objectives

The overall goal of this research was to improve ethanol production by *C. bescii* by considering the redox metabolism of this organism and optimizing electron flow towards ethanol production. Prior to 2012, a genetic system was not available for this organism and research focused on understanding plant biomass deconstruction. Chapter 2 describes work aimed at understanding the limitations of fermentation of the model growth substrate crystalline cellulose and the real-world substrate switchgrass and highlights some of the physiological limitations of *C.*

bescii when grown on industrially relevant substrate loadings. With the development of the genetic system in *C. bescii,* CBP has become a feasible goal although continued advancements were necessary. Chapter 3 compares the two available background lineages of *C. bescii* in terms of overall genome stability and insertion element movement in these strains. Chapter 4 focuses on the development of a xylose inducible promoter that allows for conditional expression of genes that will be of utility in future studies of this organism. In Chapter 5, ethanol production strains were constructed in the new, stable MACB1018 parent and it was demonstrated that redox balancing is extremely important for ethanol production, leading to a five-fold increase in comparison to previous studies.

Tables and Figures

Table 1.1 Characterized members of the genus *Caldicellulosiruptor*.

Cellulose degradation capability is designated based on growth on crystalline cellulose as follows: high (robust growth), moderate (less than optimal growth, compared to glucose), low (no growth). The number of CAZy enymes was obtained from http://www.cazy.org/ (136).

Species	Cellulose Degradation Capability	Number of CAZy enzymes	Temperature Optimum (°C)
C. bescii	high	68	78
C. hydrothermalis	low	74	65
C. kristjanssonii	low	48	78
C. kronotskyensis	high	32	70
C. lactoaceticus	moderate	54	68
C. obsidiansis	high	59	78
C. owensensis	low	67	75
C. saccharolyticus	high	70	70

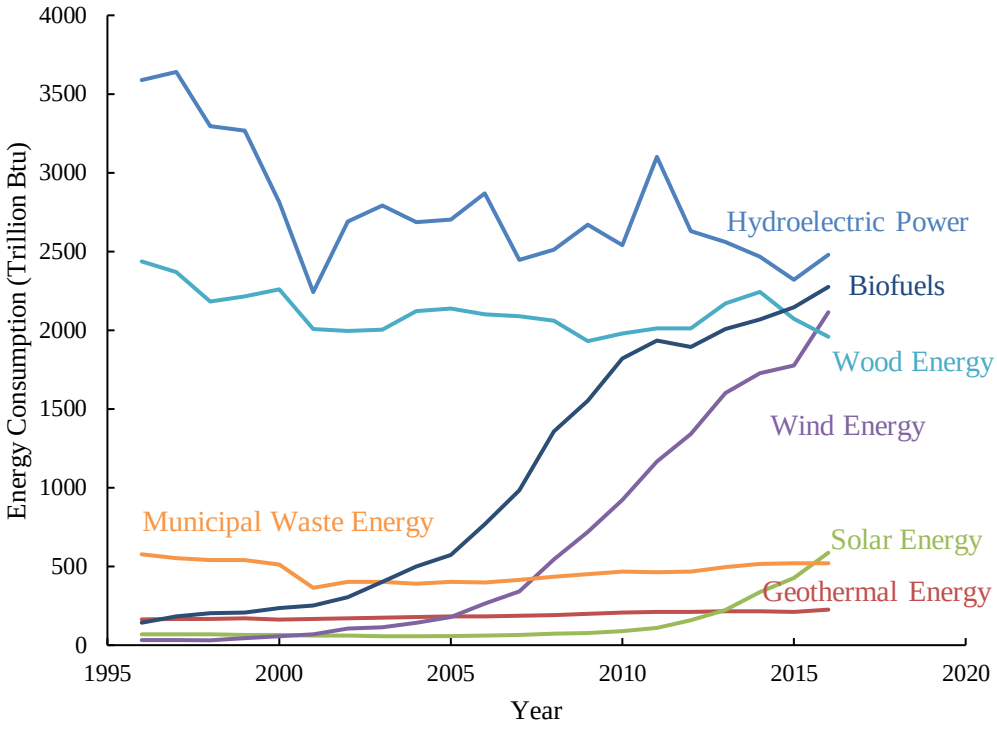


Figure 1.1 Renewable energy consumption in the United States.

Data for this figure is taken from the Monthly Energy Review by the U.S. Energy Information Administration (4). Biofuels constitutes a combination of fuel ethanol, biodiesel and other biofuel sources. Municipal waste energy includes solid waste, landfill gas and other biomass used for energy.

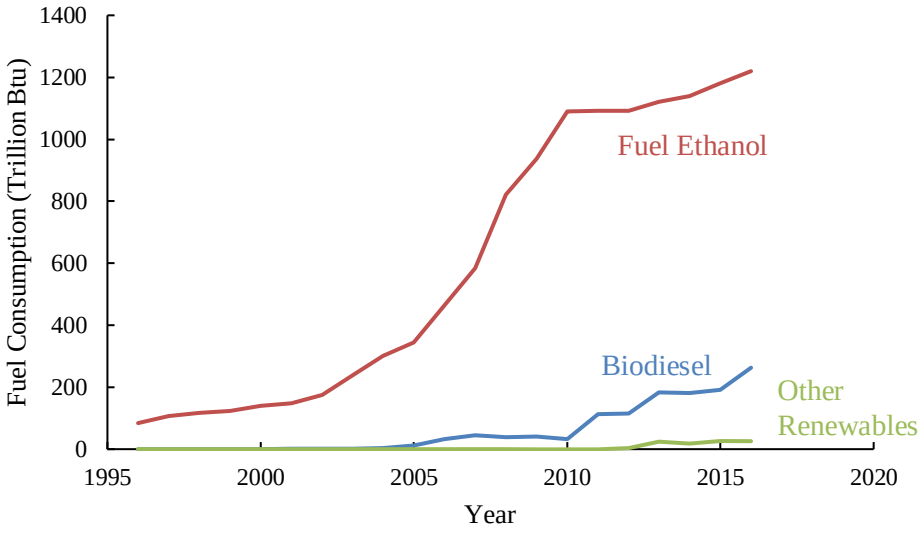


Figure 1.2 Biofuel Consumption in the U.S.

Data for this figure is taken from the Monthly Energy Review by the U.S. Energy Information Administration (4).

Figure 1.3 Vessels needed for each type of lignocellulose processing scheme.

Each box represents one vessel with text and color indicating the biological step which takes place in that vessel. Boxes with color gradients represent multiple biological steps being performed in the same fermentation vessel. The three processes represented here are separate hydrolysis and fermentation (SHF), simultaneous saccharification and fermentation (SSF), and consolidated bioprocessing (CBP).

Figure 1.3

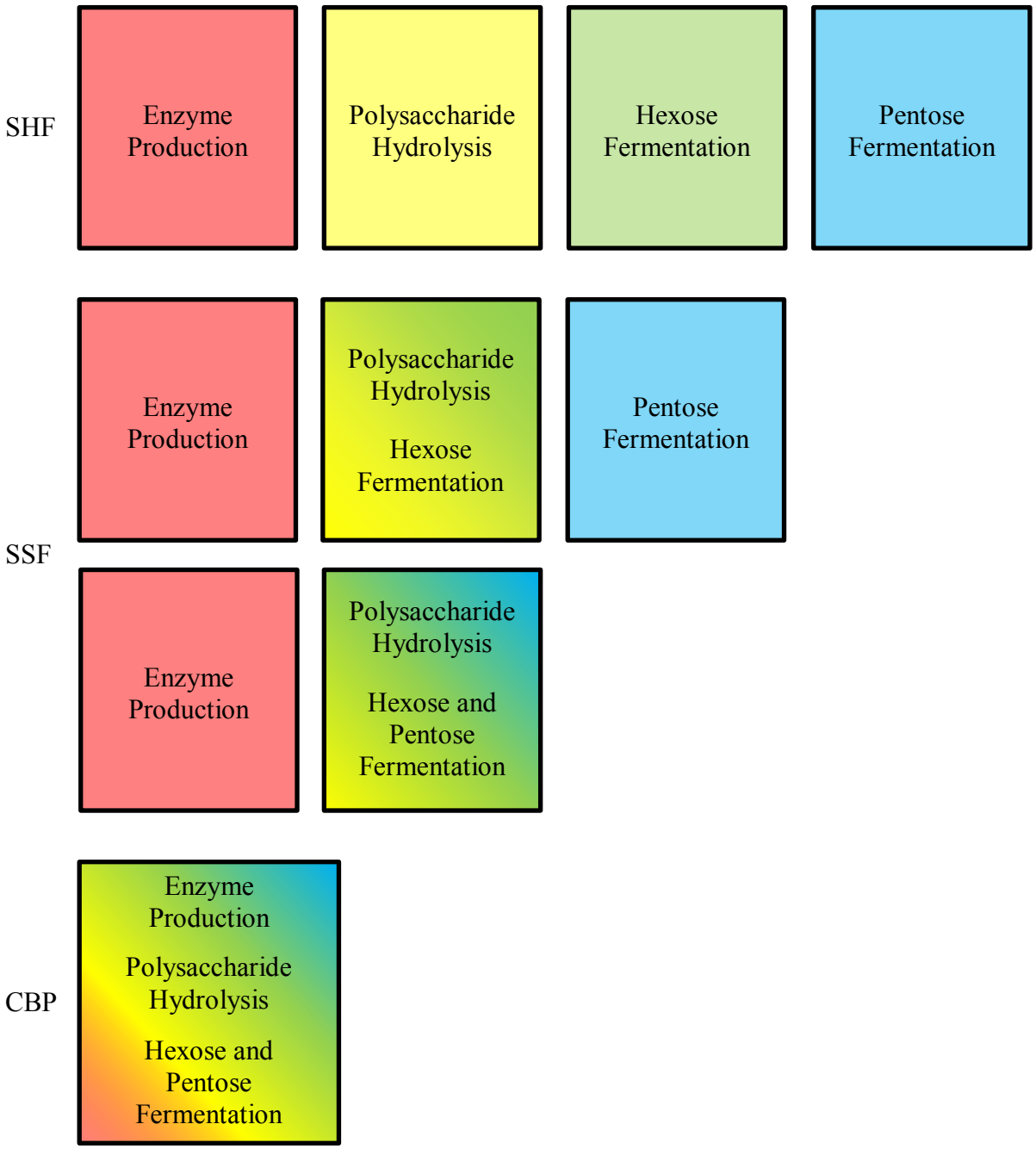
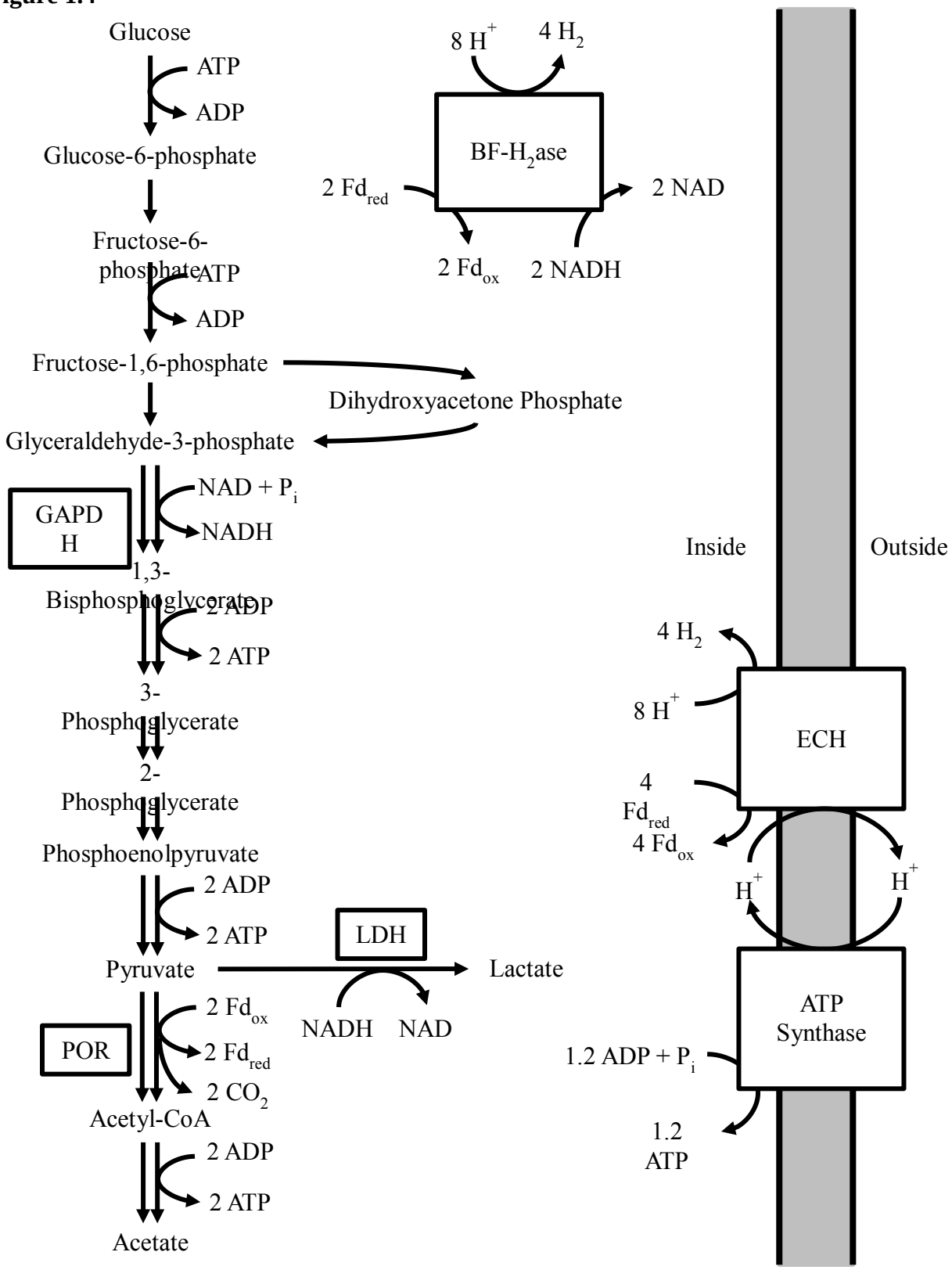

SHF
Enzyme Production
Polysaccharide Hydrolysis
Hexose Fermentation
Pentose Fermentation
SSF
Enzyme Production
Polysaccharide Hydrolysis
Hexose Fermentation
Pentose Fermentation
Enzyme Production
Polysaccharide Hydrolysis
Hexose and Pentose Fermentation
CBP
Enzyme Production
Polysaccharide Hydrolysis
Hexose and Pentose Fermentation

Figure 1.4 Glycolysis in *C. bescii*.

Glycolysis in *C. bescii* follows a standard Embden-Meyerhof-Parnas pathway. Glyceraldehyde-3-phosphate (GAP) is converted by the GAP dehydrogenase (GAPDH) generating NADH and 1,3-bisphosphoglycerate. The other key redox enzymes in glycolysis or involved in redox cofactor recycling are pyruvate oxidoreductase (POR), lactate dehydrogenase (LDH), bifurcating hydrogenase (BF-H_2ase), and energy conserving hydrogenase (ECH). Natively *C. bescii* generates acetate, lactate, carbon dioxide and hydrogen as the end products of glycolysis.

Figure 1.4


Glucose
ATP
ADP
Glucose-6-phosphate
Fructose-6-phosphate
ATP
ADP
Fructose-1,6-phosphate
Dihydroxyacetone Phosphate
Glyceraldehyde-3-phosphate
NAD + P_i
GAPDH
NADH
1,3-Bisphosphoglycerate
2 ADP
2 ATP
3-Phosphoglycerate
2-Phosphoglycerate
Phosphoenolpyruvate
2 ADP
2 ATP
Pyruvate
LDH
Lactate
NADH
NAD
2 Fd_ox
POR
2 Fd_red
2 CO_2
Acetyl-CoA
2 ADP
2 ATP
Acetate
8 H+
4 H_2
BF-H_2ase
2 Fd_red
2 NAD
2 Fd_ox
2 NADH
Inside
Outside
4 H_2
8 H+
ECH
4 Fd_red
4 Fd_ox
H+
H+
ATP Synthase
1.2 ADP + P_i
1.2 ATP

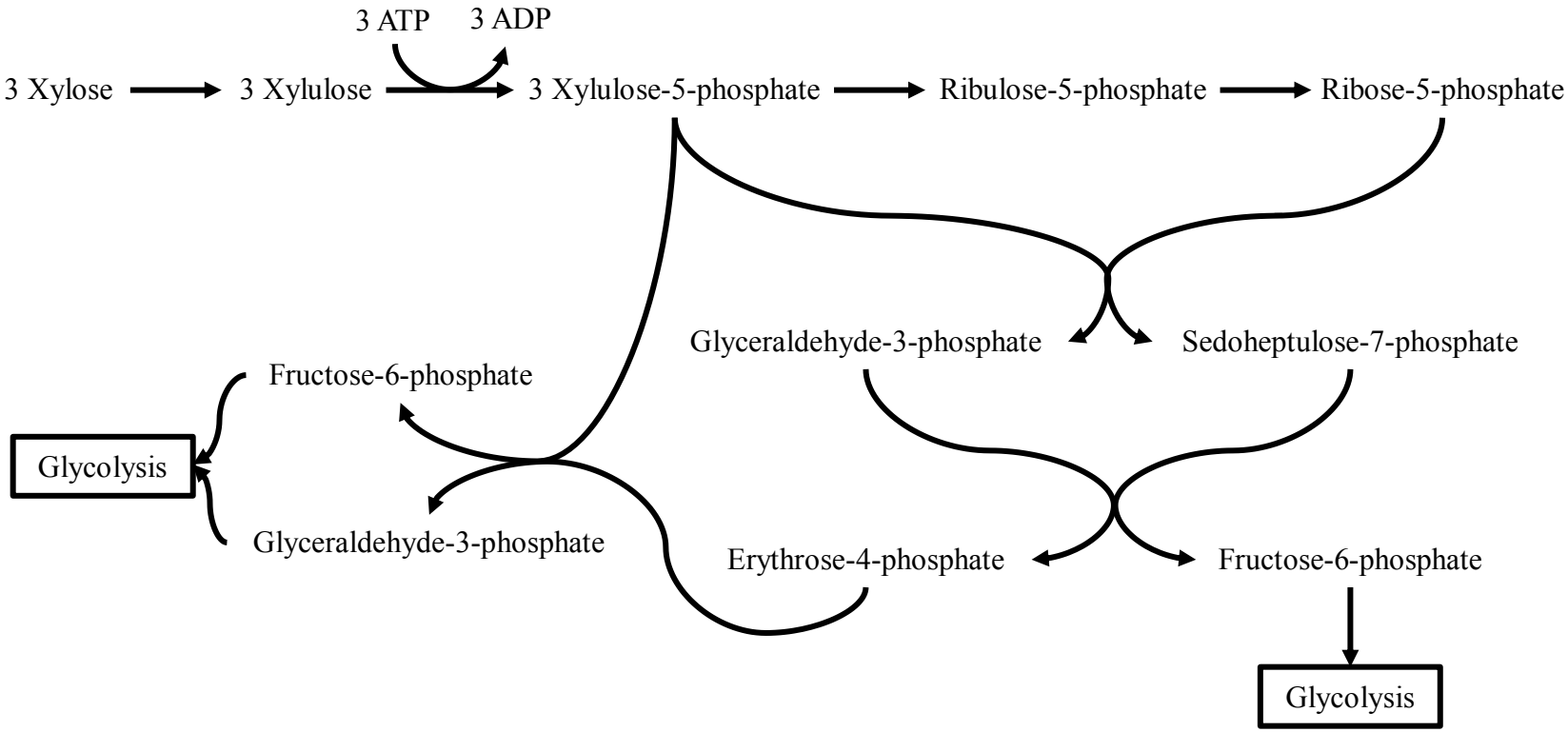


Figure 1.5 Pentose phosphate pathway in *C. bescii*.

In *C. bescii* xylose is funneled into glycolysis via the non-oxidative pentose phosphate pathway. First xylose is converted to xylulose by xylose isomerase and xylulose is then phosphorylated. Three molecules of xylulose-5-phosphate feed into the pentose phosphate pathway of sugar-phosphate interconversions resulting in the production of two molecules of fructose-6-phosphate and one molecule of glyceraldehyde-3-phosphate which feed directly into glycolysis.

CHAPTER 2

DEGRADATION OF HIGH LOADS OF CRYSTALLINE CELLULOSE AND OF UNPRETREATED PLANT BIOMASS BY THE THERMOPHILIC BACTERIUM *CALDICELLULOSIRUPTOR BESCII*[1]

---

[1]Rhaesa, A.M., Basen, M., Kataeva, I., Prybol, C.J., Scott, I.M., Poole, F.L., Adams, M.W.W. 2013. *Bioresource Technology.* 152:384-92.

Abstract

The thermophilic bacterium *Caldicellulosiruptor bescii* grows at 78°C on high concentrations (200 g L^{-1}) of both crystalline cellulose and unpretreated switchgrass, while low concentrations (<20 g L^{-1}) of acid-pretreated switchgrass inhibit growth. Degradation of crystalline cellulose, but not that of unpretreated switchgrass, was limited by nitrogen and vitamin (folate) availability. Under optimal conditions, *C. bescii* solubilized approximately 60% of the crystalline cellulose and 30% of the unpretreated switchgrass using initial substrate concentrations of 50 g L^{-1}. Further fermentation of crystalline cellulose and of switchgrass was inhibited by organic acid end-products and by a specific inhibitor of *C. bescii* growth that did not affect other thermophilic bacteria, respectively. Soluble mono- and oligosaccharides, organic acids, carbon dioxide and microbial biomass, quantitatively accounted for the crystalline cellulose and plant biomass carbon utilized. *C. bescii* therefore degrades industrially-relevant concentrations of lignocellulosic biomass that have not undergone pretreatment thereby demonstrating its potential utility in biomass conversion.

Introduction

Conversion of plant biomass into liquid biofuels is viewed as a potential sustainable and cost-effective alternative to fossil fuels (2). Plants have evolved numerous mechanisms to counteract microbial and enzymatic attacks. The major obstacle for their efficient and economic conversion to fuels is the recalcitrance of their three major components, the glycan polymers cellulose and hemicellulose and the polyaromatic lignin. Harsh thermochemical and physical pretreatments of the biomass are required to enable the cellulose and hemicellulose to be enzymatically hydrolyzed (195, 196). It has also been suggested to use microorganisms that produce cellulolytic and hemicellulolytic enzymes and subsequently ferment the sugars to biofuels in a process known as consolidated bioprocessing (CBP) (43). However, to date no microorganism is known that can at the same time ferment polysaccharides from plant biomass and produce a biofuel at the needed rates and efficiencies.

Advances have been made recently in genetically-modifying non-saccharolytic, ethanol-producing microorganisms, particularly *Saccharomyces cerevisiae*, to produce polysaccharide-degrading enzymes (44) but to do so in sufficient amounts for efficient biomass conversion is a major challenge. An alternative strategy is to use microorganisms that are naturally cellulolytic and hemicellulolytic and engineer them to produce the biofuel of choice. Among those relatively few microorganisms that have the capability to ferment crystalline cellulose, most studies have focussed on the moderately thermophilic bacterium *Clostridium* (*Cm.*) *thermocellum*. This organism grows optimally (T_{opt}) at 60°C and is a potential CBP organism. It degrades crystalline cellulose using a multienzyme complex known as the cellulosome and it also produces ethanol as fermentation product (43). *Cm. thermocellum* does not utilize hemicellulose, consequently it has been grown in co-culture with xylan- or xylose-utilizing, ethanol-producing bacteria such as

species of *Thermoanaerobacter* and *Thermoanaerobacterium* in order to increase ethanol yield and titer (197-199). As yet utilization of high loads of pretreated plant biomass (above 10 g L^{-1}) has not been reported for such co-cultures.

Some species of the thermophilic bacterial genus *Caldicellulosiruptor* utilize both crystalline cellulose and hemicellulose and metabolize both C_5 and C_6 sugars (136, 200). In addition, they grow optimally at higher temperatures (70−80°C) than species of *Clostridia* and *Thermoanaerobacter* (typically near 60°C). Advantages of fermentations at elevated temperatures include higher mixing rates, a lower risk of contamination and facilitated biofuel recovery (201). Moreover, genome sequencing of several *Caldicellulosiruptor* species has revealed that they do not contain cellulosomes. Rather, they represent the second paradigm for degrading crystalline cellulose that involves free enzymes rather than a large multienzyme complex. They contain a variety of glycosyl hydrolases and polysaccharide lyases that include many multidomain cellulases and xylanases (136, 202). With one exception (203), the *Caldicellulosiruptor* species produce only trace amounts of ethanol (166). Nevertheless, a genetic system has been recently developed for one species (204), so metabolic engineering of this group for biofuel production should be feasible in the near future.

A major development within the *Caldicellulosiruptor* genus was the demonstration that one species, *C. bescii* (T_{opt} 78°C), degraded switchgrass, a model plant for bioenergy production, that had not been subjected to thermochemical pretreatment (168). The organism also degraded switchgrass after it had simply been washed with hot water at the growth temperature of the microorganism (78°C). This showed that the microbe was not simply utilizing non-recalcitrant soluble sugars present in unpretreated biomass (106). However, these studies were performed at a relatively low biomass concentration (5 g L^{-1}), at least an order of magnitude lower that what

would be required for any industrial process. High concentrations of biomass require less water and energy for heating and cooling, therefore reducing processing costs (205), but the degradation of industrial loads of unpretreated plant biomass (>10 g L^{-1}) has not been previously studied. Surprisingly, there are also only a few reports on the microbial degradation of high loads (>10 g L^{-1}) of either thermochemically-pretreated lignocellulosic material or on crystalline cellulose (198, 203, 206, 207).

We show here that that *C. bescii* can grow on and degrade high concentrations of both unpretreated switchgrass and crystalline cellulose (up to 200 g L^{-1}), while it grows poorly on even low loads of thermochemically-pretreated switchgrass (>10 g L^{-1}). Unexpectedly, we also demonstrate that different physiological mechanisms are responsible for limiting growth of *C. bescii* on crystalline cellulose and on unpretreated plant biomass.

Materials and Methods

Growth experiments

Caldicellulosiruptor bescii DSM6725 was grown in the medium described previously (177). The carbon substrates used were cellobiose, crystalline cellulose (Avicel PH-101) (both obtained from Sigma, St. Louis, MO) or switchgrass (sieved 20/80-mesh fraction; provided by Dr. Brian Davison, Oak Ridge National Laboratory, Oak Ridge, TN). Samples of plant biomass were used without chemical or physical pretreatment other than washing for 18 h with water at 78°C. This material is referred to as unpretreated switchgrass. Switchgrass referred to as acid-pretreated was subjected to thermochemical treatment (0.05 g H_2SO_4 per g dry switchgrass, 190°C, 5 min) and was obtained from Dr. Jonathan Mielenz of Oak Ridge National Laboratory, TN. Before use, the acid-pretreated biomass was frozen, thawed, washed four times with four volumes of water to neutralize and to remove simple oligomeric sugars, and autoclaved (180) for 25 minutes at 121°C.

Vitamins and bicarbonate were added after autoclaving. Growth experiments were performed at 78°C, either as small scale closed cultures without pH control (50 mL volume, 75 mL headspace of N_2/CO_2 80/20; shaken at 150 rpm), or in pH-controlled, gas-flushed or gas sparged fermentors. These were either the DASGIP parallel bioreactor system (0.5 L, DASGIP, Jülich, Germany) or a 20 L custom fermentor. DASGIP bioreactors were equipped with heating jackets with a custom-made feed-back temperature control. The pH was maintained at 6.8 − 7.2 by the automated addition of $NaHCO_3$. Gas exchange in the fermentors was achieved by continuous flushing of the gas headspace with N_2/CO_2 (80/20 v/v) at a flow rate of 2 L per hour, or alternatively by sparging the culture at the same flow rate. To collect the spent media after microbial growth on crystalline cellulose or switchgrass, the DASGIP fermentors were transferred to an anaerobic chamber containing a N_2/H_2 (95/5, v/v) atmosphere. The insoluble substrate was allowed to settle and then separated from the medium by decanting. Cells were removed by centrifugation at 6,000 × g for 10 minutes under anaerobic conditions. In the anaerobic chamber, the cell free media was dispensed into 20 mL aliquots in 50 mL bottles, and the atmosphere was replaced by N_2/CO_2 (80/20, v/v), before it was used to support growth of new cultures. Cell densities were monitored using a Petroff-Hauser counting chamber. Before counting, a significant fraction of the insoluble switchgrass was allowed to settle and then removed. Total cell protein was determined using a standard Bradford assay.

Quantification of substrates and products

The amount of residual insoluble substrates was determined as described previously (168). The products of the fermentations were determined after removal of cells and insoluble substrates by centrifugation. Total sugars in the culture supernatant were determined photometrically at 490 nm after hydrolysis in concentrated sulfuric acid and derivatization of the sugar with 5% phenol

(phenol-sulfuric acid method), using glucose as the standard. Acetate, formate, malate, butyrate, uracil, cellobiose, glucose and xylose were determined by high-performance liquid chromatography (HPLC) on a 2690 separations module (Waters, Milford, MA) equipped with an Aminex HPX-87H column (300 mm by 7.8 mm; Bio-Rad, Hercules, CA), a photodiode array detector (model 996; Waters) and a refractive index detector (model 410; Waters). The system was operated with 5 mM H_2SO_4 as the eluent at a flow rate of 0.5 ml min^{-1}. Samples for HPLC were acidified with 0.1 M H_2SO_4 and centrifuged before analysis to remove particles. Lactate and ethanol concentrations were determined by using assay kits from Megazyme (Wicklow, Ireland). Soluble protein was measured using the Bradford assay. Amino acids and free ammonium were determined by HPLC after phenyl isothiocyanate derivatization (PITC, Edman's reagent). Prior to derivatization, samples were filtered through a 0.2 µm filter. 5 µl of PITC and 10 µl of the filtered sample were added to 85 µl of a mixture of methanol and triethylamine (7/1 v/v), and incubated for 20 minutes in the dark. 400 µl of water were added to stop the derivatization, and precipitates were removed by centrifugation. Ammonium and amino acids in the soluble fraction were separated using the HPLC 2690 separations module (Waters, Milford, MA) equipped with a LiChrosorb C_8 reversed phase column (Supelco, Bellefonte, PA) in a linear gradient from 100% solvent A (0.05 % v/v trifluoroacetic acid) to 100% solvent B (100% acetonitrile). Ammonium was alternatively quantified by the salicylate method using the solutions of a test kit (Mars Fishcare, Chalfont, PA, USA) in a modified plate reader assay.

Carbon mass balances

Carbon mass balances were calculated using the amount of residual substrates and products as described previously (106). Uronic acids only contribute a minor fraction (<2%) to the total carbon in the cell wall and were therefore not included in the analysis. *C. bescii* does not use the

lignin from the cell walls as a carbon source or as an electron donor (106) and the lignin content was therefore not included in the carbon balance.

For the estimation of the concentration of carbon in total sugars (determined as described above), glucose was used as a standard and total sugars were expressed in glucose units. Carbon dioxide was not experimentally determined as the gas headspace from the microbial cultures was removed by flushing. It was assumed that carbon dioxide equaled the amount of acetate produced (via the pyruvate ferredoxin oxidoreductase, phosphotransacetylase and acetate kinase reactions). Microbial biomass formation was estimated from the amount of protein formed assuming that 50% of dry weight is protein and 50% of dry weight is carbon (208).

Results and Discussion

Fermentation of high concentrations of crystalline cellulose

The anaerobic thermophilic bacterium *C. bescii* has been shown to degrade crystalline cellulose and switchgrass that had not been chemically-pretreated (168, 176) but concentrations high enough to be industrially-relevant have not been previously tested. Here we show that *C. bescii* grows on concentrations of crystalline cellulose as high as 200 g L^{-1} (Fig. 2.1 A and Fig. 2.2), the highest yet reported for a microbial culture. Concentrations of cellulose above 200 g L^{-1} have been studied using enzymatic degradation, with conversion efficiencies decreasing from ~90% at 20 g L^{-1} to ~30% at 400 g L^{-1} (205). The highest concentrations using microbial conversion previously studied were 100 g L^{-1} using an unclassified *Caldicellulosiruptor* strain (203) and 92 g L^{-1} by a co-culture of a genetically-engineered *Cm. thermocellum* strain and *Thermoanaerobacterium saccharolyticum* (198). Crystalline cellulose has also been used in Solid Substrate/State Cultivation (SSC) where the substrate is soaked with media thereby maximizing

substrate and product concentrations. However, the conversion efficiency is very low and efficient mixing and control of pH are significant problems with SSC (209).

In closed bottle cultures with no pH control, *C. bescii* degraded a comparable amount of crystalline cellulose (2.1 - 2.6 g L^{-1}) after 48 h incubation at 78 °C independent of the initial crystalline cellulose concentration over the range 5 – 50 g/L^{-1} (Table 2.1). Accordingly, the concentrations of the primary end products, lactate and acetate, and the final pH values (4.8 ± 0.1) of the cultures, were similar independent of the initial crystalline cellulose concentration. These data suggested that the decrease in pH due to organic acid production and the build-up of the hydrogen gas generated by fermentation are likely limiting factors in degrading high loads of crystalline cellulose in closed cultures.

Therefore, *C. bescii* was grown in pH-controlled cultures with continuous flushing of the headspace or sparging of the culture medium with N_2/CO_2 (80/20 v/v). Such conditions had a marked effect on the microorganism's ability to degrade crystalline cellulose. With pH control and gas-flushing, *C. bescii* completely solubilized crystalline cellulose at substrate loads of both 5 and 10 g L^{-1} and there was approximately 20% conversion using a load of 50 g L^{-1} (see Table 2.2). The initial crystalline cellulose degradation rate was 96 mg h^{-1} (Fig. 2.3). When the medium was sparged with N_2/CO_2, rather than simply replacing the gas phase by flushing with N_2/CO_2, the growth rate also improved (from 0.19 h^{-1} to 0.24 h^{-1}), suggesting that efficient removal of hydrogen gas from the medium is not achieved simply by flushing (Fig. 2.4).

The main carbon end products of crystalline cellulose fermentation by *C. bescii* were acetate, lactate and CO_2 (Fig. 2.5). Using 50 g L^{-1} and with pH control and gas sparging, approximately 100 mM of these products were generated in the first 35 h (Fig. 2.6 AB). This is similar in terms of soluble carbon yield to what was reported previously with *Cm. thermocellum*,

where approximately 40 mM acetate and 35 mM ethanol were formed in 50 h, although the substrate load was lower (10 g L^{-1}) (210). For *C. bescii,* the carbon balances (substrates used *versus* products generated) were closed using both low and high crystalline cellulose loads (5 g L^{-1} and 50 g L^{-1}) showing that the major carbon products and microbial biomass accounted for all of the crystalline cellulose carbon consumed (Fig. 2.5). Minor products (ethanol, alanine and pyruvate) accounted for less than 2% of the total carbon. This is in contrast to crystalline cellulose conversion by *Cm. thermocellum,* where a significant number of minor end products were generated that together were equivalent to 11% of the crystalline cellulose utilized (210). Growth and fermentation of crystalline cellulose by *C. bescii* appeared complete after 55 h independent of the initial substrate concentration (Fig. 2.7). However, in the 50 g L^{-1} culture, glucose and cellobiose were still being generated after 100 h (Fig. 2.5) and represented more than 50% of the carbon products. Hence, at high loads, once *C. bescii* reaches a high cell density and ceases to grow, it continues to hydrolyze cellulose to glucose and cellobiose but does not ferment these sugars.

To investigate what limits growth of *C. bescii* on crystalline cellulose, attempts were made to grow the organism on the spent growth medium after cells had been previously grown for 160 h on 50 g L^{-1} crystalline cellulose (Fig. 2.7). The spent medium contained cellobiose (11 mM) and glucose (19 mM) so addition of carbon substrates was not necessary. However, as shown in Fig. 3A, growth was not observed when the spent medium was inoculated with a fresh culture of *C. bescii.* This was due to nitrogen limitation since the spent medium, which contained only 90 µM ammonium, supported growth to high cell densities (1.60×10^8 cells $mL^{-1,}$ Fig. 2.7 A) when supplemented with the concentration of ammonium (NH_4Cl) that was present in the standard medium (6.2 mM). *C. bescii* was therefore grown on the high crystalline cellulose load (50 g L^{-1})

in a pH-controlled, culture sparged with N_2/CO_2 (80/20, v/v) in media containing various ammonium concentrations. When the initial ammonium supply was increased from 6.2 to 12.3 mM, the cultures exhausted the nitrogen supply (< 100 µM NH_4^+) and acetate production ceased (near 80 mM at 35 h). Adding further ammonium (to 12.3 mM) stimulated growth and acetate production resumed, reaching a maximum of 140 mM at 130 h before plateauing (Fig. 2.7 B). Further degradation of the crystalline cellulose was not limited by nitrogen as more than sufficient ammonium (7 mM) remained.

Increasing the ammonium supply therefore had a dramatic effect on the conversion of crystalline cellulose (Fig. 2.7 CD). The amount degraded almost doubled to 60% (30 g L^{-1} degraded as determined by weight) compared to 36% conversion (18 g L^{-1} degraded) in the standard medium. This resulted in the production of approximately twice the amount of organic acids (135 − 165 mM, approx. 90% acetate and 10% lactate) and glucose (50 – 65 mM) compared to the standard medium (82 mM and 30 – 35 mM, respectively). In contrast, the amount of cellobiose produced (20 − 30 mM) was very similar in the different cultures (see Fig. 2.7 C and Fig. 2.6). The growth rate (0.3 h^{-1}) and the rate of product generation were also largely unchanged with the higher concentration of ammonium, indicating that the increased nitrogen supply enables *C. bescii* to continue crystalline cellulose degradation for a longer period (Fig. 2.6). These results indicate that on high substrate loads, *C. bescii* requires significantly higher amounts of nitrogen, presumably for cell maintenance, to more effectively utilize the substrate and ferment it to end products.

Fermentation of high concentrations of unpretreated lignocellulosic biomass

C. bescii also grew to high cell densities on unpretreated lignocellulosic biomass (switchgrass) at high concentrations up to 200 g L^{-1} (Fig. 2.1 and Fig. 2.2). The switchgrass

samples were first washed with hot water at the growth temperature of the organism (78°C for 18 h) to remove readily solubilized organic material (sugars, proteins, etc.) that might serve as growth substrates. Hence, the microorganism used the insoluble switchgrass biomass as its carbon source for growth. Previous investigations of microbial biomass degradation used only 5 g L^{-1} (168, 177). The conversion of high loads of thermochemically-treated biomass (50 g L^{-1} to 400 g L^{-1}) have been studied using enzymatic hydrolysis (205) and simultaneous saccharification and fermentation (SSF, 60 g L^{-1} to 115 g L^{-1}: (211)). Pretreated biomass in the form of ammonium-fibre extension (AFEX) corn stover was used at a concentration of 117 g L^{-1} (40 g L^{-1} glucan load) with the mesophilic bacterium *Cm. phytofermentans* (206). Hence, the finding that *C. bescii* grew on high loads of biomass that had not undergone thermochemical pretreatment at industrial relevant concentrations of 50 – 200 g L^{-1} and converted more than 20% to organic acids and sugars in five days bodes well for the concept of CBP in general. We estimate a threshold concentration for industrial fermentations of unpretreated switchgrass between 100 g L^{-1} and 150 g L^{-1} because at higher concentrations most of the liquid was absorbed into the substrate (Fig. 2.2).

Surprisingly, *C. bescii* degraded a constant fraction of the unpretreated switchgrass (27 − 33%) and this was independent of the initial concentration over the range 1 – 50 g L^{-1} (Table 2.3). This is in contrast to the situation with crystalline cellulose, where a constant amount of the substrate was degraded. In pH-controlled fermentor cultures, switchgrass degradation (50 g L^{-1}) resulted in the release of soluble sugars (~12 mM as determined by the phenol-sulfuric acid method; Fig. 2.8 B), which accounted for 20% of the product generated (on a carbon basis, Fig. 2.8 D), but less than 10% of these sugars were cellobiose (0.6 mM) and glucose (1.5 mM). The nature of the sugars released from switchgrass remains unknown as the method used measures the degradation products from all soluble carbohydrate. Compared to switchgrass, much more sugar

was released from crystalline cellulose (> 30 mM, Fig. 2.5 B), accounting for 50% of the carbon products, and in this case all of it was in the form of cellobiose and glucose (Fig. 2.5 D).

In further contrast to the situation with crystalline cellulose, the same products were generated from both low and high loads of unpretreated switchgrass (Fig. 2.8). We had previously shown that at the low load (5 g L^{-1}) the carbon balance between substrates utilized and products generated was approximately 100%, based on the C_6 and C_5 contents of switchgrass (0.45 g g^{-1} and 0.31 g g^{-1}, respectively) (106). For unknown reasons, a lower amount (86%) of the substrates was recovered as products for the high load case (50 g L^{-1}: Fig. 2.8 CD). However, in both cases, 3-4% of the substrate carbon was converted to lactate and 4-6% of it was found in the microbial biomass. The major difference between these two balances is in the amount of acetate formed, which represents 45% and 33% of the carbon product for 5 g L^{-1} and 50 g L^{-1}, respectively. The abiotic release of acetate from hemicellulose acetylation at 78°C was determined to be 11 mM from 50 g L^{-1} switchgrass. Therefore, up to ~20% of the acetate produced originated from switchgrass deacetylation (Fig. 2.8 CD). A low concentration of ethanol (2.5 mM) was also produced during degradation of the high switchgrass load representing close to 2% of carbon utilized (Fig. 2.8 D). While there are no previous reports on the degradation of unpretreated biomass by other microorganisms, *C. obsidiansis* (166) and an unclassified *Caldicellulosiruptor* species (203) fermented thermoacid-pretreated switchgrass (5 and 10 g L^{-1}, respectively) with the products accounting for 74% and up to 97%, respectively, of the carbon consumed. Herein we report the first carbon balance for a high load of unpretreated switchgrass (50 g L^{-1}) but in this case the recovery was only 86% and the nature of the 'missing' carbon is not known (although it is not ethanol, alanine or pyruvate). On the other hand, given the complex composition of plant material such as switchgrass, and the fact that more than 20% of the high load (50 g L^{-1}) was

solubilized and fermented by *C. bescii*, a recovery of close to 90% of biomass carbon is not unreasonable.

These data therefore demonstrate that *C. bescii* grows to high cell densities using high substrate loads (200 g L^{-1}) of unpretreated switchgrass without inhibition of growth with increasing substrate load. However, this was not the case with switchgrass that had undergone thermal acid-pretreatment (with sulfuric acid at 190°C for 5 min). As shown in Fig. 2.9, growth of *C. bescii*, as measured by the cell density obtained after 72 h incubation in closed cultures, dramatically decreased with increasing concentrations of acid-treated switchgrass from 10 to 50 g L^{-1}, such that there was virtually no growth at the highest biomass concentration tested. The pH at the end of the experiment (72 h) in the cultures with acid-pretreated switchgrass was 6.2 for loads of 10 and 20 g $L^{-1}$and 7.3 for 50 g L^{-1}, so growth of *C. bescii* was not inhibited by low pH. In contrast, under the same conditions, high cell densities (at least 2.3×10^8 cells mL^{-1}) were obtained for unpretreated switchgrass using the same biomass concentrations. Moreover, the addition of 50 g L^{-1} acid-pretreated switchgrass inhibited growth of *C. bescii* on 10 g L^{-1} unpretreated switchgrass (Fig. 2.9). Clearly, these results show that acid pretreatment of switchgrass releases inhibitors of bacterial growth and characterization of biomass degradation at high substrates loads is only possible with unpretreated switchgrass.

Thermochemical treatment of plant biomass is typically used to greatly reduce the lignin-associated recalcitrance of plant cell walls, rendering the sugar polymers (cellulose and hemicellulose) more accessible to the cellulolytic enzymes. On the other hand, it is one of the most expensive steps in converting plant biomass to biofuels (196). In addition, by-products of pretreatment like furan aldehydes, phenols and certain aliphatic acids can act as inhibitors of microbial fermentations or enzymatic hydrolysis (212). Pretreatment of biomass with dilute

sulfuric acid at high temperature (near 150°C) is the most commonly used method, and acid-pretreated switchgrass has been shown to support growth of *Cm. thermocellum* (180) and several *Caldicellulosiruptor* species (107, 166, 180), However, the highest concentration previously reported was only 10 g (dry weight) L^{-1} (180) . The results presented herein with *C. bescii* indicate that such a concentration may be close to the limit. Presumably, the higher concentrations of inhibitors that are released from higher concentrations of acid-pretreated switchgrass (above 10 g (dry weight) L^{-1}) affect growth and/or enzymatic hydrolysis of the biomass, an effect not observed with unpretreated switchgrass. Since *C. bescii* degrades unpretreated switchgrass (168), even at 200 g L^{-1}, conventional thermochemical pretreatment of the biomass is clearly unnecessary. This result gives a new perspective on the cost and efficiency of biofuel production using the CBP approach.

Factors limiting further crystalline cellulose and switchgrass fermentation

Degradation of high loads of crystalline cellulose by *C. bescii* under nitrogen-sufficient conditions resulted in the production of high concentrations of organic acids (~150 mM). Similar product concentrations (up to 150 mM) were achieved with an uncharacterized *Caldicellulosiruptor* species (203). As shown in Fig. 2.10, growth of *C. bescii* on the disaccharide cellobiose was affected by much lower organic acid concentrations. The addition of 75 mM sodium acetate to the standard growth medium resulted in a lag phase of about 6.5 h and a decreased growth rate (0.24 h^{-1} compared to a value of 0.40 h^{-1} for the control culture). The effect was even more pronounced at concentrations comparable to those seen during crystalline cellulose degradation. The presence of 150 mM resulted in a lag phase of over 24 h and a resulting growth rate of 0.13 h^{-1}. Nevertheless, the cell density eventually reached that seen in the control culture

(~4×10^8 cells mL^{-1}) albeit after 72 h rather than after 20 h. No growth was observed if 200 mM sodium acetate was added to fresh medium (data not shown).

To determine if the production of organic acids was the limiting factor for crystalline cellulose degradation in an N-sufficient medium, attempts were made to cultivate a fresh inoculum of *C. bescii* in the spent cell-free medium once cell growth on crystalline cellulose had ceased. The spent medium contained sufficient amounts of sugar for growth (50 mM glucose and 30 mM cellobiose) but also contained 135 mM of organic acids. *C. bescii* grew to a reasonably high cell density (1×10^8 cells mL^{-1}) after 48 h, but only when a mixture of vitamins and trace elements were added (at concentrations added to a fresh medium: Fig. 2.11). *C. bescii* requires only the thermally-labile vitamin folate for growth (213), which is in accordance with its lack of the folic acid biosynthetic pathway, and addition of only folate (rather than a vitamin mixture) stimulated growth of *C. bescii* on the spent medium. However, there was also a significant lag phase for growth of about 50 h (Fig. 2.11), which might be explained by the energetic burden to synthesize the other vitamins. Continuous addition of folate alone or the vitamin mixture and trace elements throughout the growth phase to fermentations of 50 g L^{-1} crystalline cellulose with sufficient ammonium supply (24 mM) did not lead to a higher concentration of end products (data not shown). Moreover, no growth was observed in the spent medium from a fermentor containing 165 mM of organic acids (acetate plus lactate; data not shown).

These results suggest that the supply of the thermolabile vitamin folate may be a limiting factor under prolonged fermentation conditions but that this is overridden by growth inhibition by the high concentrations of organic acids that accumulate. Complete inhibition of growth of *C. bescii* was reported with an osmolarity of 250 mOsmol (50 mM NaCl; (214)) and growth of its close relative, *C. saccharolyticus,* was affected by an osmolarity of 200-300 mOsmol (from sugar,

organic acids and NaCl) (215). However, we found that inhibition of fermentation and crystalline cellulose degradation is not due to osmotic pressure alone, since growth was observed even with 160 mM organic acids where the osmolarity of the medium is 520 mOsmol (440 mM ions). Additionally, we found that *C. bescii* was able to grow on medium containing 90 g L^{-1} glucose (initial osmolarity of >550 mOsmol), albeit with a long lag phase of >50 h (data not shown).

Organic acids such as acetic acid can also act to uncouple membrane function by dissipating ion gradients (216). Consequently, *C. bescii* did not grow in the presence of 200 mM acetate and exhibited a lag phase in the presence of concentrations of acetate comparable to those produced during fermentation (150 mM), although the cells reached cell densities comparable to those obtained in the absence of acetate. A similar concentration of acetate (175 mM) completely inhibited growth of *C. saccharolyticus* (215). It has also been reported that concentrations as low 3 mM cellobiose led to a significant decrease in the purified extracellular cellulase of *Cm. thermocellum* (217). With *C. bescii*, however, cellobiose and glucose production continued despite cellobiose accumulating to concentrations up to 30 mM during crystalline cellulose fermentation (Fig. 2.6 C). Hence, crystalline cellulose degradation by *C. bescii* might be much less sensitive to hydrolysis products than is *Cm. thermocellum.* Of course, the generation by *C. bescii* of ethanol or other alcohols instead of organic acids by metabolic engineering would alleviate issues of product inhibition, a scenario that now seems possible given the recent development of a transformation system for this organism (204).

In contrast to the situation with crystalline cellulose, the nitrogen supply did not limit growth of *C. bescii* on unpretreated switchgrass, presumably because additional nitrogen is available from the plant biomass. Similarly, the addition every 24 h of folate or a vitamin mixture (equivalent to that in the original medium) to a *C. bescii* culture growing on 50 g L^{-1} unpretreated

switchgrass did not affect the fraction of biomass that was solubilized (data not shown). When a fresh culture of *C. bescii* was added to the cell-free spent medium resulting from the fermentation of 50 g L^{-1} switchgrass, some growth was observed when the medium was supplemented with a carbon source (cellobiose, 5 g L^{-1}), but both the growth rate (0.02 h^{-1} *versus* 0.28 h^{-1}) and the final cell density (5 x 10^7 *versus* 3 x 10^8 cells ml^{-1}) were about an order of magnitude less that those measured with *C. bescii* in a fresh medium (Fig. 2.12). In contrast to the situation with crystalline cellulose, the addition of folate did not stimulate growth in the spent medium (data not shown).

These data suggest that deconstruction of plant biomass over prolonged periods by *C. bescii* eventually generates an inhibitor of microbial growth. It has been suggested that fermentation of both unpretreated (106) and acid-pretreated (180) switchgrass result in the release of lignin-derived phenolic compounds that potentially inhibit microbial growth (106). Either these phenolic compounds or furan aldehydes created by the acid-pretreatment could be responsible for inhibiting growth of *C. bescii* on the thermochemically-treated material at biomass concentrations above 10 g L^{-1}. However, this was not true with unpretreated biomass since *C. bescii* degraded more rather than less biomass with increasing loads (Table 2.3). Nevertheless, *C. bescii* did not grow well on the spent medium from a pH-controlled fermentation of unpretreated switchgrass (50 g L^{-1}) (Fig. 2.12), even though organic acids were at non-inhibitory concentrations (Table 2.3). Surprisingly, the spent medium supplemented with cellobiose did support vigorous growth of the anaerobic fermentative bacterium *Thermoanaerobacter mathranii* (which grows optimally at 70°C), to the same extent as a fresh culture medium. (Fig. 2.12). Hence if a growth inhibitor is involved, it appears to be specific for *C. bescii* and is presumably derived from the plant biomass as it is not generated when *C. bescii* grows on crystalline cellulose and generates similar

fermentation products. Further studies of the growth of *C. bescii* on unpretreated switchgrass are needed to identify the nature of this specific inhibitor.

At present there is no clear benchmark for the concentrations of unpretreated plant biomass that are relevant for industrial conversion processes. That is not the case for thermochemically-treated biomass since the results presented herein suggest that the limit is about 10 g L^{-1}, at least for *C. bescii*. One requirement for industrial biomass conversion seems to be the formation of at least 4% (w/v) ethanol (40 g L^{-1} or 868 mM) (206). Given the C_6 and C_5 content of the unpretreated switchgrass used in this study (106) and the maximum theoretical ethanol yields from the two sugar types (218), an initial concentration of 108 L^{-1} of unpretreated biomass would be sufficient to generate 4% (w/v) ethanol. The high load used herein with *C. bescii* (50 g L^{-1}) would theoretically generate close to 2 % (w/v) ethanol assuming optimal conversion efficiencies. Of course, as discussed above, at present this organism generates organic acids rather than alcohol, but clearly, in principle, a genetically-modified strain of *C. bescii* that generates the appropriate product could be used in industrial fermentations of extremely high concentrations of unpretreated plant biomass.

Conclusions

C. bescii does not contain cellulosomes but is able to degrade crystalline cellulose as efficiently as cellulosome-containing microbes, such as *Cm. thermocellum* and therefore represents the second paradigm for cellulose degradation. More importantly, *C. bescii* degrades unpretreated switchgrass almost as efficiently as it does crystalline cellulose and does so at substrate loads relevant to industrial conversion (up to 200 g L^{-1}). In contrast, growth of *C. bescii* was inhibited by much lower concentrations of thermochemically-pretreated switchgrass.

Biomass degradation at high substrates loads necessary for industrial processing, at least by this microorganism, is therefore only possible with unpretreated biomass.

Acknowledgements

This work was supported by a grant (DE-PS02-06ER64304) from the Bioenergy Science Center (BESC), Oak Ridge National Laboratory, a U.S. Department of Energy (DOE) Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. We thank Jonathan Mielenz and Miguel Rodriguez Jr. from Oak Ridge National Laboratory for supplying acid-pretreated switchgrass. We thank Gerrit J. Schut, Sanjeev Chandrayan, Puja Chandrayan, William Nixon and Paul R. Volny for technical assistance.

Tables and Figures

Table 2.1 Degradation of crystalline cellulose by *C. bescii* in flushed cultures without pH control.

Closed cultures (50 ml) were grown without pH control in three biological replicates (n=3 ± SD). End products and the pH were determined after 48 h of growth (see **Fig. 2.5** and **2.2**).

Initial ($g\ L^{-1}$)	**Converted ($g\ L^{-1}$)**	**Converted (%)**	**Acetate (mM)**	**Lactate (mM)**	**pH [end]**
Closed, no pH-control					
1	0.8	80	4.1 ± 0.9	1.8 ± 0.3	6.2
5	2.1	42	5.7 ± 0.9	4.8 ± 0.6	4.7
10	2.1	21	5.0 ± 1.0	4.4 ± 0.6	4.8
20	2.5	13	6.6 ± 0.2	4.3 ± 0.3	4.7
50	2.6	5	4.5 ± 0.5	5.3 ± 0.4	4.9
100	*n.d.*	*n.d.*	5.8 ± 0.3	4.8 ± 0.3	4.8
150	*n.d.*	*n.d.*	6.0 ± 0.3	4.3 ± 0.1	4.8
200	*n.d.*	*n.d.*	6.1 ± 0.3	4.0 ± 0.3	4.9

n.d. not determined

Table 2.2 Conversion of crystalline cellulose and unpretreated switchgrass by *C. bescii* in pH-controlled (pH 7.2), flushed 10 L fermentors.

	Initial ($g\ L^{-1}$)	Converted ($g\ L^{-1}$)	Acetate (mM)	Lactate (mM)	Glucose (mM)	Cellobiose (mM)	Total Sugars (mM)	Cell Protein ($\mu g\ mL^{-1}$)	Cell Density (mL^{-1})
Cellulose	5	5.0	32.9	23.1	0.0	0.2	*n.d.*	134	3.8×10^{8}
	10	9.9[a]	53.5	18.5	5.7	2.0	*n.d.*	269	5.5×10^{8}
	50	10.5[b]	35.4	23.1	32.9	2.4	*n.d.*	162	7.8×10^{8}
Switchgrass	5	1.8	12.1	0.6	0.0	0.0	10.0	44	3.2×10^{8}
	50	10.0	58.5	3.7	1.5	0.6	12.0	118	4.8×10^{8}

n.d. not determined; [a] rate of crystalline cellulose degradation 96 mg h^{-1} (**Fig. 2.3**); [b] Nitrogen limited (see Text and **Fig. 2.5**). With sufficient NH_4Cl 30 g L^{-1} were degraded (**Fig. 2.7**)

Table 2.3 Degradation of unpretreated switchgrass by *C. bescii* in closed and flushed cultures without pH control.

Closed cultures (50 ml) were grown without pH control in two biological replicates (n=2 ± SD). End products and the pH were determined after 48 h of growth (see **Fig. 2.5** and **2.2**).

Initial wSG (g L^{-1})	**Converted (g L^{-1})**	**Converted (%)**	**Acetate (mM)**	**Lactate (mM)**	**pH [end]**
Closed, no pH control					
1	0.3	32	4.7 ± 0.1	0.0 ± 0.0	7.8
5	1.6	33	4.7 ± 0.1	0.2 ± 0.1	7.2
10	3.1	31	7.8 ± 0.1	0.5 ± 0.1	5.7
20	5.8	29	9.4 ± 0.1	1.7 ± 0.9	5.6
50	13.6	27	9.7 ± 0.0	1.7 ± 1.0	5.5

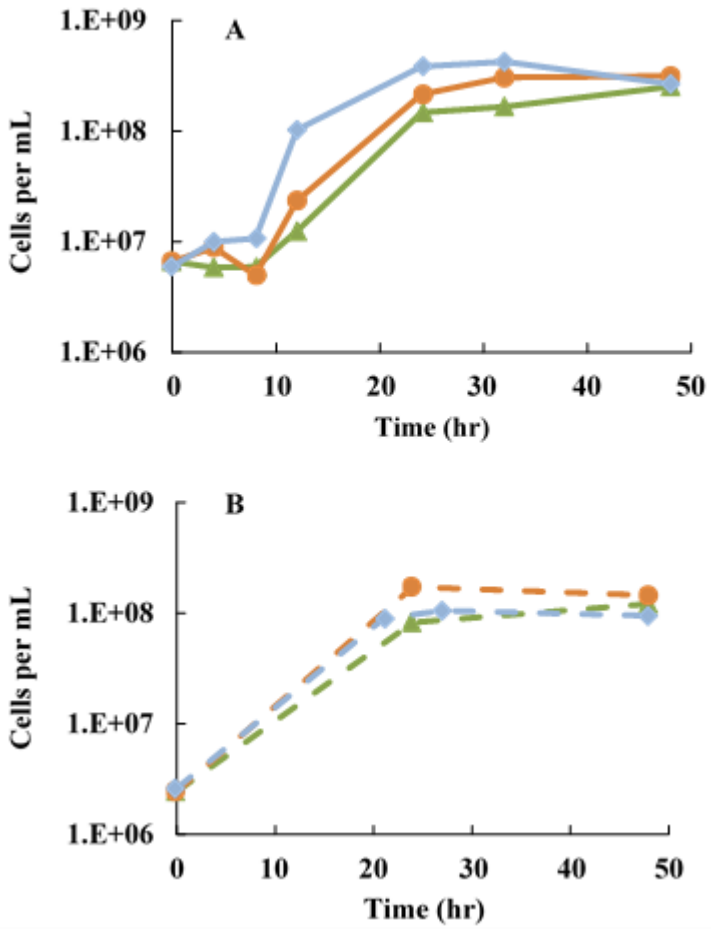


Figure 2.1 Growth of *C. bescii* on high biomass loads.

C. bescii was grown in closed cultures on 5 g L^{-1} (green triangles), 50 g L^{-1} (orange circles) or 200 g L^{-1} (blue diamonds) of crystalline cellulose (**A**) or unpretreated switchgrass (**B**). Values are averages of three (**A**) or two (**B**) biological replicates.

Figure 2.2 Growth of *C. bescii* on different loads of cellulose and unpretreated switchgrass. 5, 20 and 50 and 200 g L^{-1} of cellulose (A) and the same concentrations of unpretreated switchgrass (B) in closed cultures.

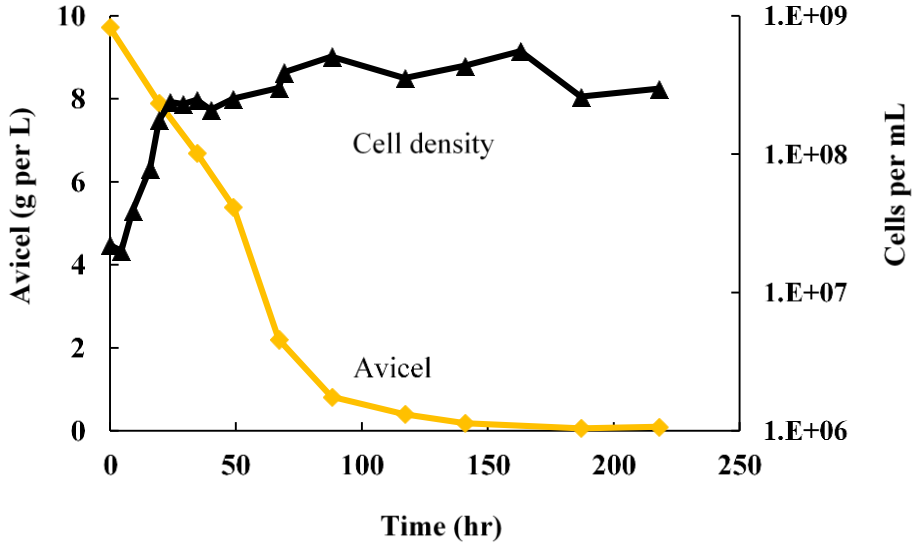


Figure 2.3 Degradation of 10 g L-1 of crystalline cellulose by C. bescii in a 10 L pH controlled fermenter.

Growth (cell density) and cellulose concentration (Avicel) are indicated.

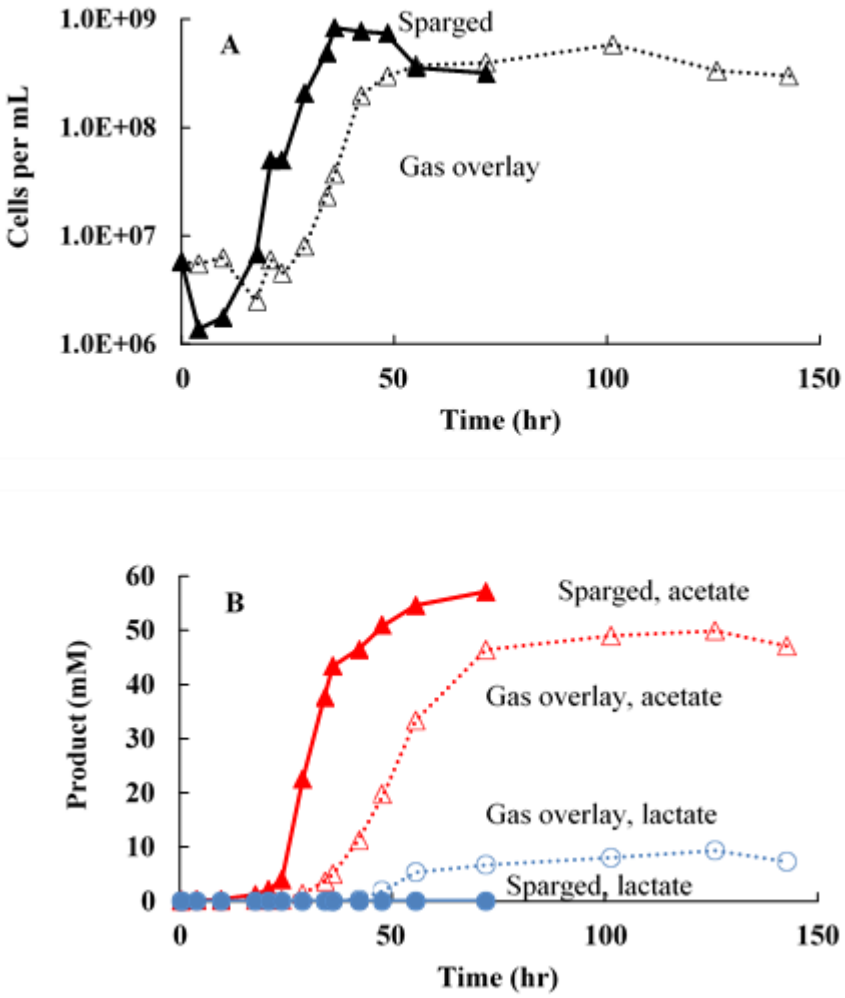


Figure 2.4 Effect of gas-sparging and gas-overlay on cellulose fermentation by C. bescii. Cells were grown in 0.5 L fermenters with gas removal at a flow rate of 2 L per h of N_2/CO_2 (80/20 v/v) either by gas-sparging (solid lines and symbols) or with continuous removal of the gaseous headspace (gas overlay; open symbols and dotted lines). Growth (cell density, black triangles) is shown in panel A, while the fermentation products (acetate, red triangles and lactate, blue circles) are shown in panel B.

Figure 2.5 Growth of *Caldicellulosiruptor bescii* on crystalline cellulose with gas-flushing and pH-control.

Cell density and fermentation products of *C. bescii* grown on 5 g L^{-1} (**A, C**) and 50 g L^{-1} (**B, D**) of crystalline cellulose in a 10 L pH-controlled fermentor with gas flushing. Time courses (**A, B**) of cell densities (black triangles), and concentrations of acetate (red triangles), lactate (blue circles) and glucose plus cellobiose (orange diamonds). The corresponding carbon balances (**C, D**) include crystalline cellulose (grey), sugars from yeast extract (black) on the substrate side; and glucose (yellow), cellobiose (orange), acetate (red), lactate (blue), carbon dioxide (pink) and carbon calculated from biomass formation (light green) on the product side.

Figure 2.5

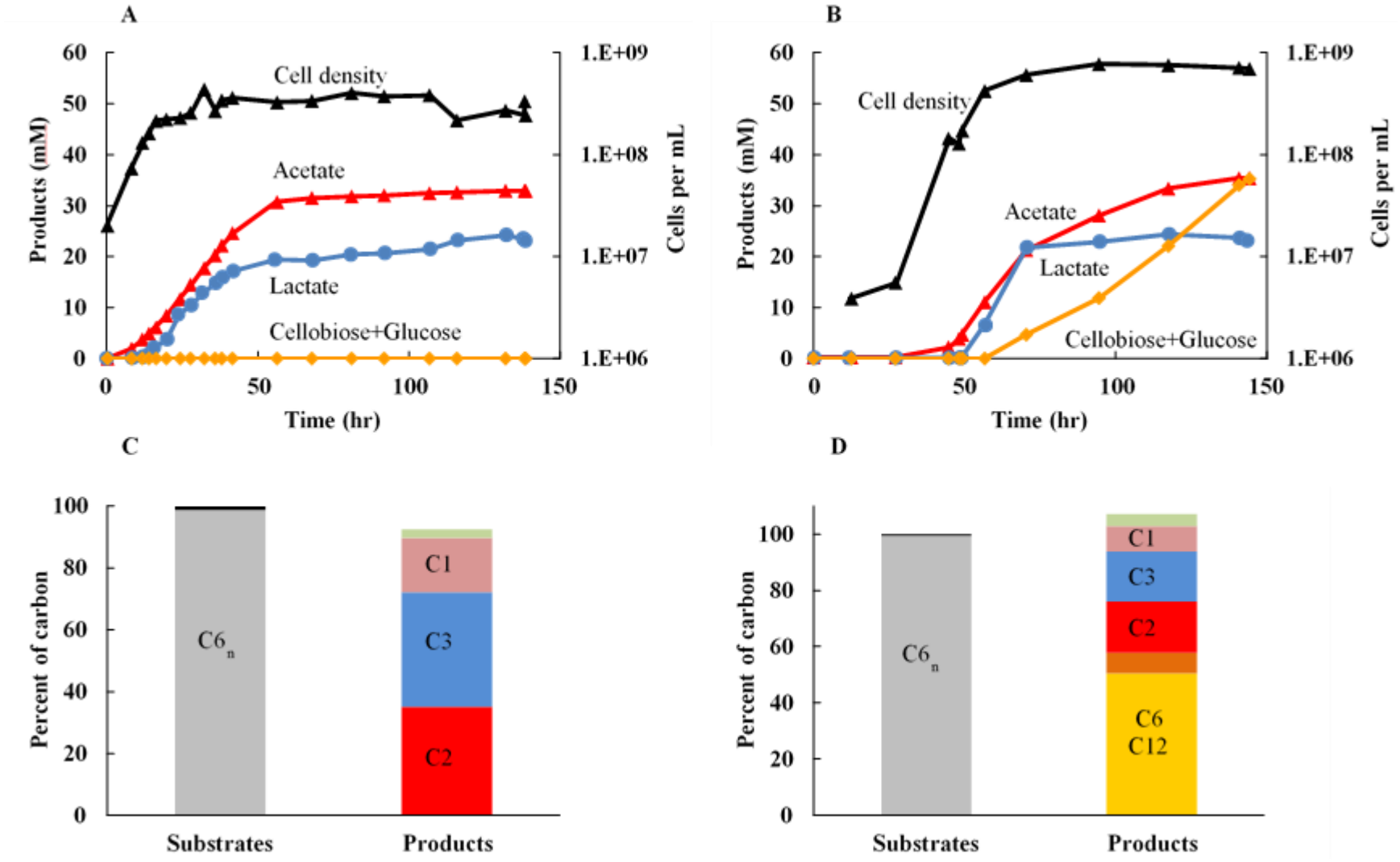

A
Products (mM)
Cells per mL
Time (hr)
Cell density
Acetate
Lactate
Cellobiose+Glucose
B
C
Percent of carbon
C6n
C1
C3
C2
Substrates
Products
D
C6
C12

Figure 2.6 Effect of ammonium on cellulose fermentation by *C. bescii*.

Growth of *C. bescii* in 0.5 L gas-sparged fermenters with pH control (see **Figure 2.7** B−D). Two cultures were initially supplied with twice the NH_4Cl concentration (12.3 mM; red squares) that was present in the standard medium, and supplied with an additional 12.3 mM NH_4Cl after it had been completely consumed. Two cultures were grown in the standard medium with 6.2 mM NH_4Cl (control, blue diamonds). Acetate (**A**), lactate (**B**), cellobiose (**C**) and glucose (**D**) concentrations shown.

Figure 2.6

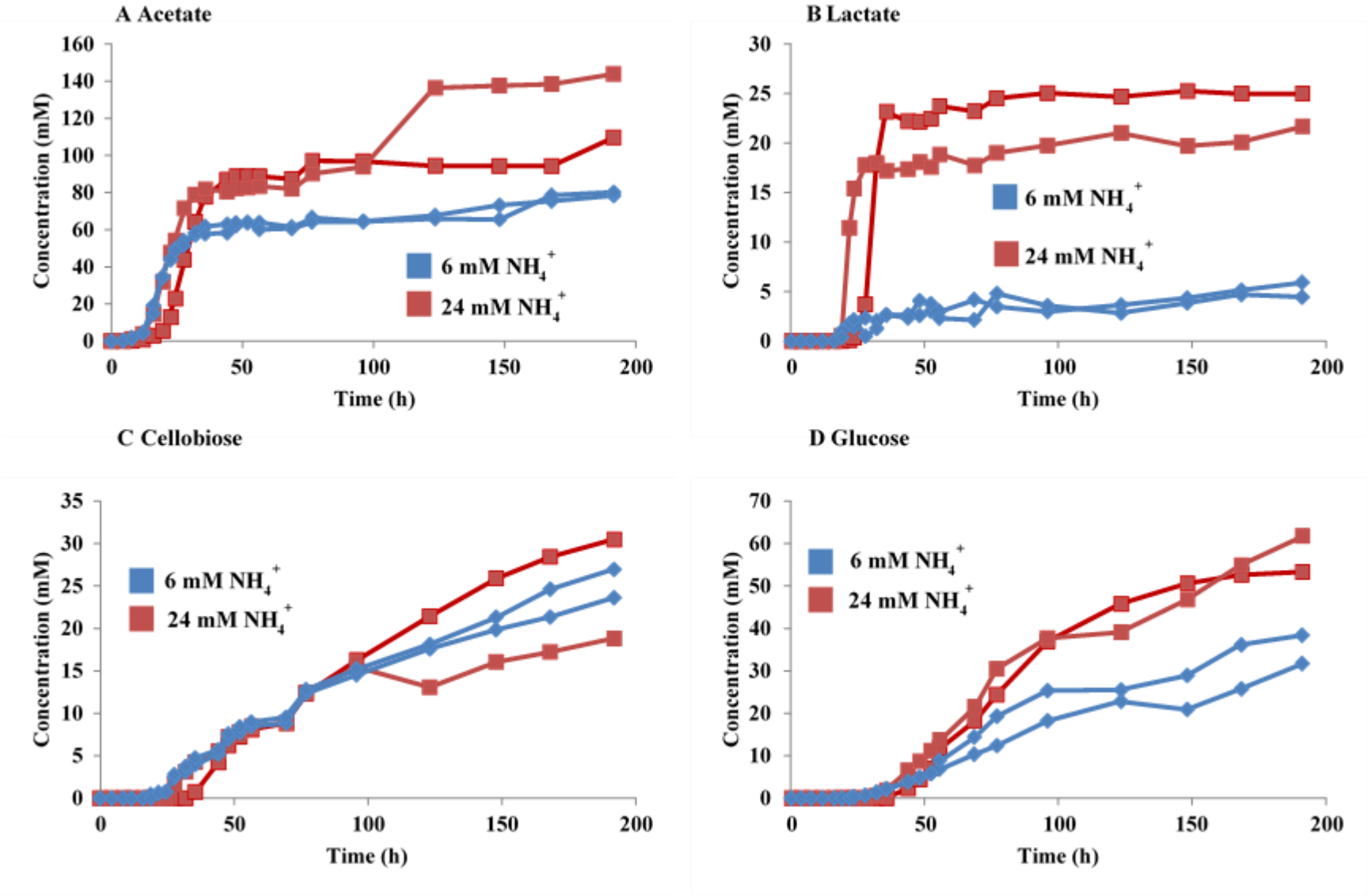

A Acetate
B Lactate
C Cellobiose
D Glucose
Concentration (mM)
Time (h)
6 mM NH4+
24 mM NH4+

Figure 2.7 Effect of ammonium addition on conversion of high loads of crystalline cellulose by *C. bescii*.

(A) Growth of *C. bescii* on 20 mL spent media in closed cultures without (blue diamonds) and with the addition of 6.2 mM NH_4Cl (n=3 ± SD). The media was obtained from a pH-controlled *C. bescii* fermentation of 50 g L^{-1} crystalline cellulose after *C. bescii* stopped producing acids. (**B**) Effect of ammonium addition on acetate production of *C. bescii* in-a pH controlled fermentation. (**C**) Fermentation products of *C. bescii* and (D crystalline cellulose degraded after 192 h by each two replicate pH-controlled, gas-sparged cultures of *C. bescii*. Two cultures were initially supplied with 12.3 mM NH_4Cl concentration (brown bars), twice the concentration that was present in the standard medium (6.2 mM blue bars). After ammonium depletion, the culture with the higher initial NH_4Cl concentration was supplied with another 12.3 mM NH_4Cl.

Figure 2.7

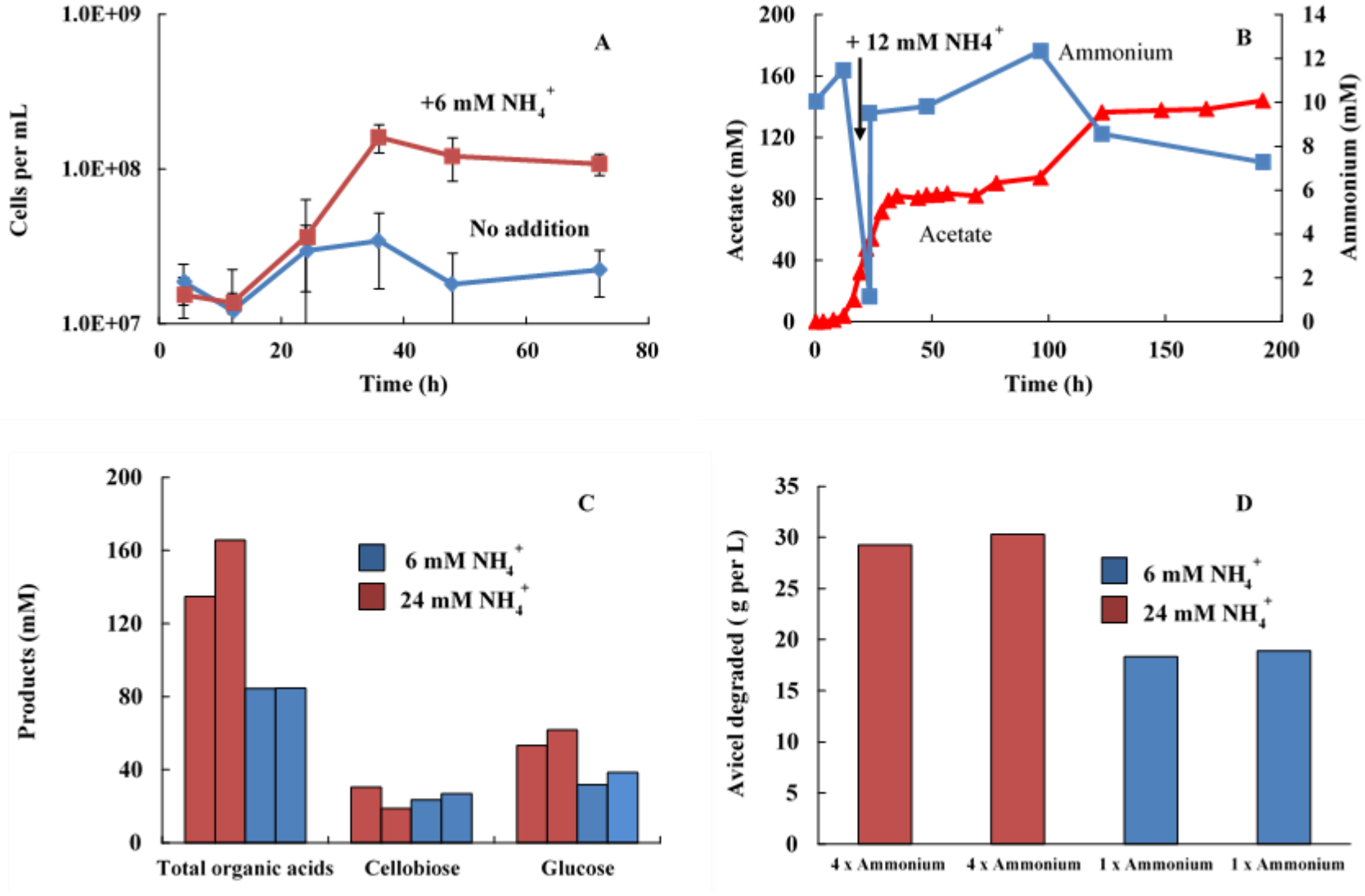
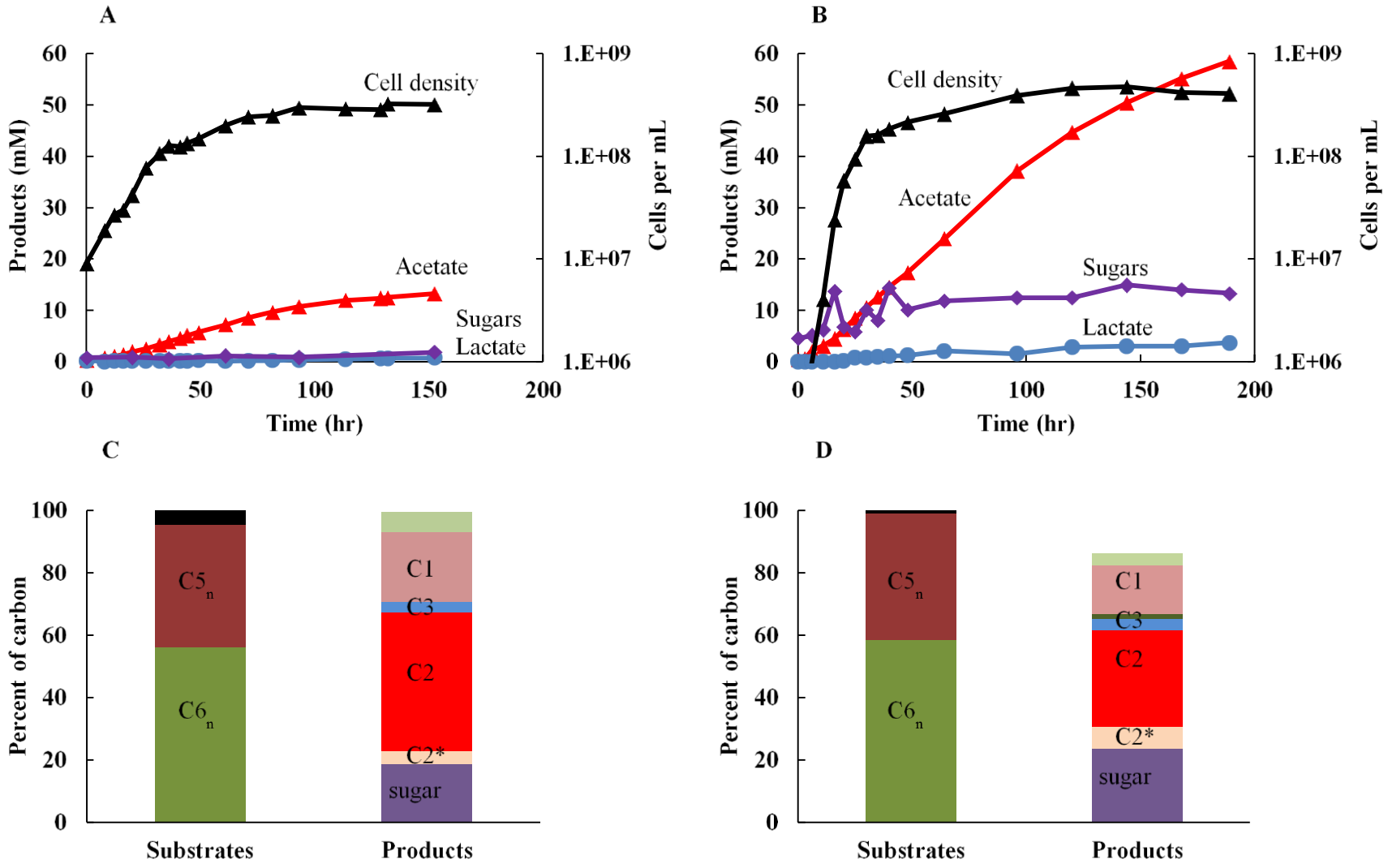

A
Cells per mL
1.0E+09
1.0E+08
1.0E+07
+6 mM NH4+
No addition
0
20
40
60
80
Time (h)
B
+ 12 mM NH4+
Ammonium
Acetate
Acetate (mM)
Ammonium (mM)
200
160
120
80
40
0
14
12
10
8
6
4
2
0
0
50
100
150
200
Time (h)
C
Products (mM)
6 mM NH4+
24 mM NH4+
Total organic acids
Cellobiose
Glucose
D
Avicel degraded (g per L)
35
30
25
20
15
10
5
0
6 mM NH4+
24 mM NH4+
4 x Ammonium
4 x Ammonium
1 x Ammonium
1 x Ammonium

Figure 2.8 Growth of *Caldicellulosiruptor bescii* on unpretreated switchgrass in a 10 L fermentor and fermentation products.

Cell density and fermentation products of *C. bescii* grown on 5 g L^{-1} (**A, C**) and 50 g L^{-1} (**B, D**) of unpretreated switchgrass. Time courses (**A, B**) of cell densities (black triangles), and concentrations of acetate (red triangles), lactate (blue circles) and total sugars (purple diamonds). The corresponding carbon balances (**C, D**) include C_6 sugars (dark green) and C_5 sugars calculated from switchgrass composition and consumption (as determined by weight) and sugars from yeast extract (black) on the substrate side. On the product side totals sugars (determined after acid hydrolysis; pink), acetate (red), acetate derived from hemicellulose acetylation (light pink), acetate (red), lactate (blue), ethanol (dark green), carbon dioxide (pink) and carbon calculated from biomass formation (light green) were quantified.

Figure 2.8


A
Products (mM)
Cells per mL
Time (hr)
Cell density
Acetate
Sugars
Lactate
B
C
Percent of carbon
C5n
C6n
C1
C3
C2
C2*
sugar
Substrates
Products
D

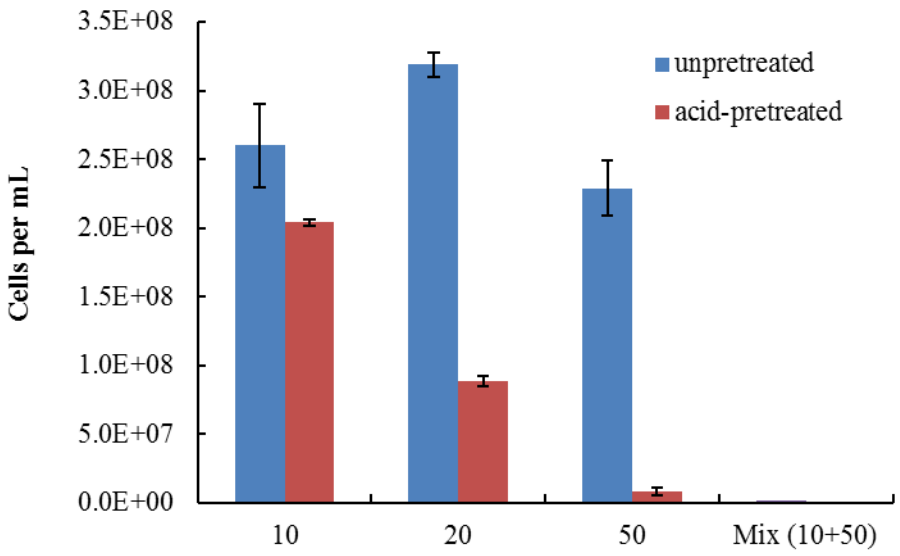


Figure 2.9 Growth of *C. bescii* on acid-pretreated and unpretreated switchgrass.

Cell densities in *C. bescii* cultures after 3 d of incubation at 78°C (stationary phase) in closed cultures on 10, 20 and 50 g L^{-1} of unpretreated switchgrass (blue bars) and acid-pretreated switchgrass (brown bars); and 10 g L^{-1} unpretreated plus 50 g L^{-1} acid-pretreated switchgrass (purple bars). The experiment was performed in three biological replicates (n=3 ± Standard Deviation).

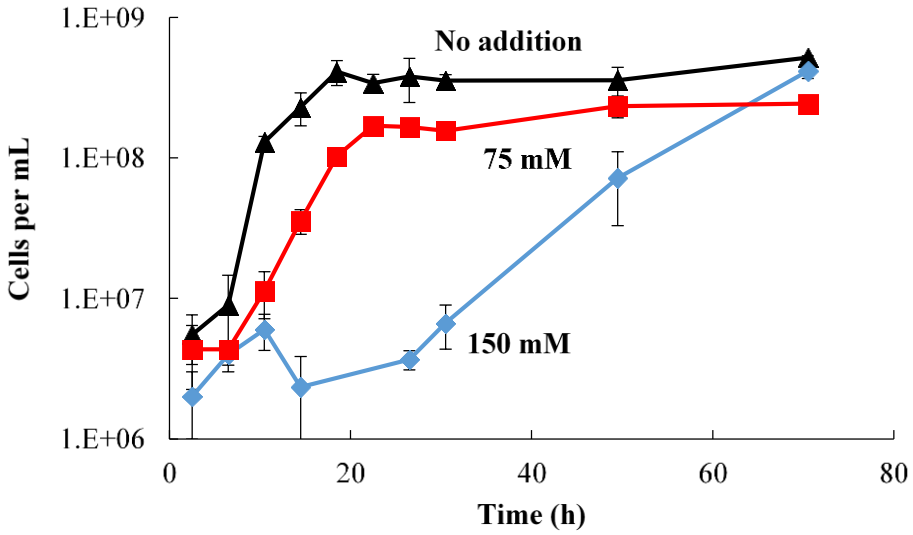


Figure 2.10 Effect of sodium acetate on *C. bescii* grown on 5 g L^{-1} cellobiose.

No (black triangles), 75 mM (red squares) or 150 mM sodium acetate (blue diamonds) were added to 50 mL cultures of *C. bescii* in closed bottles before inoculation. The experiment was performed in two (no addition; n=2 ± SD) or three biological replicates (n=3 ± SD).

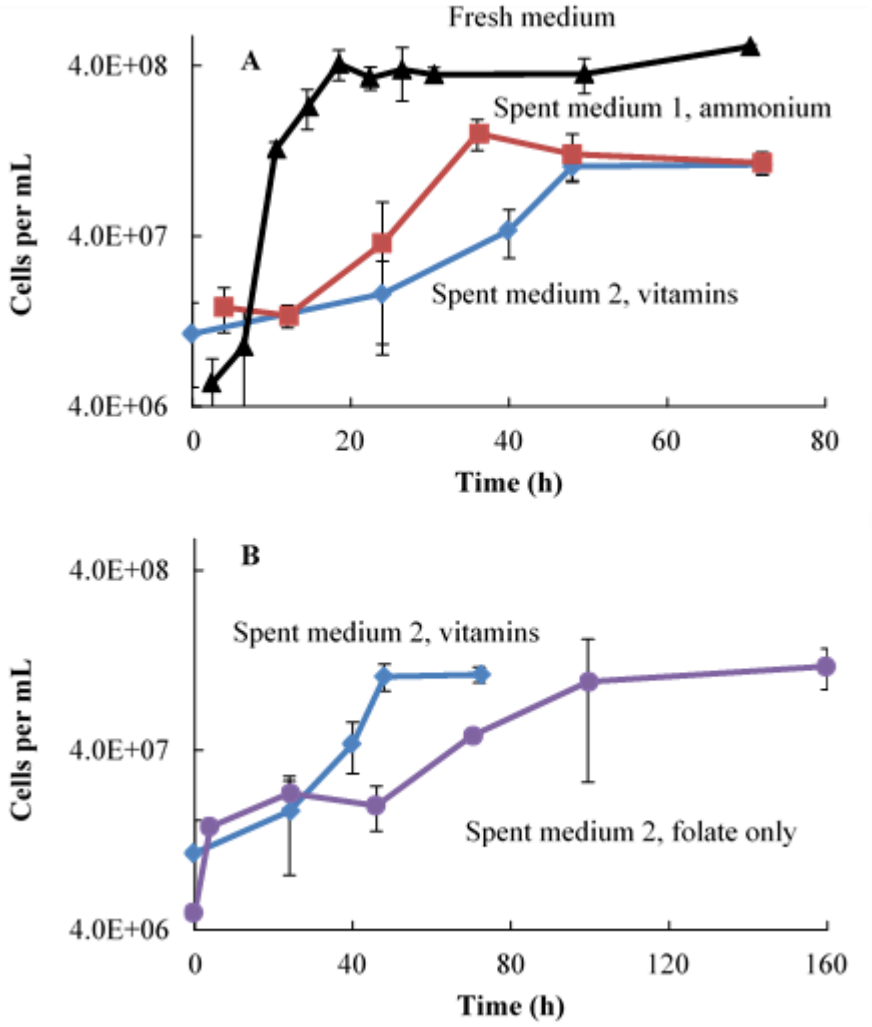


Figure 2.11 Addition of potential growth limiting factors to spent media from *C. bescii* grown on crystalline cellulose (50 g L^{-1}).

The spent media were collected anaerobically from fermenters after growth of *C. bescii* had ceased. Spent medium 1 (**A**) was collected after *C. bescii* had generated ~80 mM of organic acids and degraded ~15 g L^{-1} of cellulose. The medium was supplemented with 6 mM ammonium. Spent medium 2 (**A, B**) was collected from an N-sufficient fermentation that was not ammonium limited (see Fig. 6B − D). In this case, *C. bescii* produced 135 mM of organic acids and degraded 30 g L^{-1} cellulose. Spent media 2 (**A, B**) was supplemented with a vitamin mixture and trace elements at the same concentrations used in the standard medium, or with folate only (20 µg L^{-1}) (**B**).

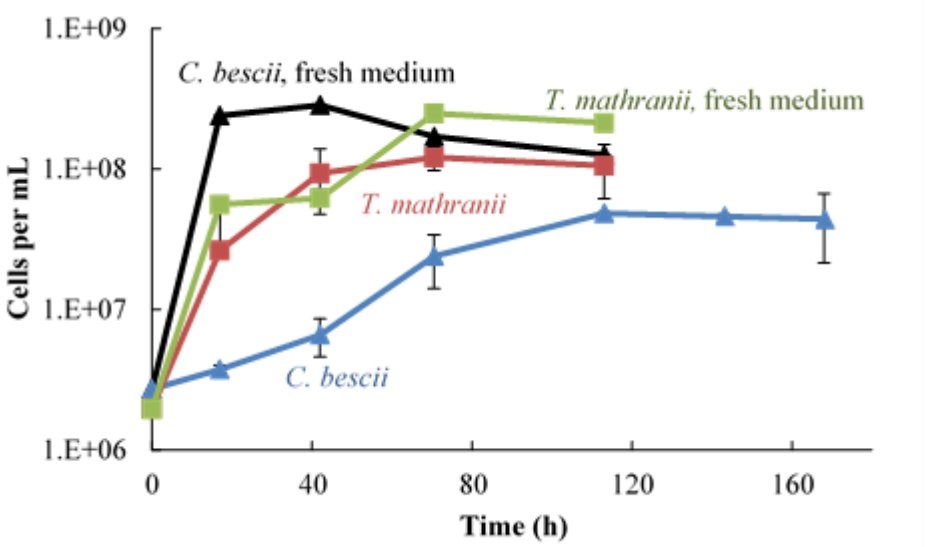


Figure 2.12 Growth of *C. bescii* on spent switchgrass media.

Growth of *C. bescii* (blue triangles) and *Thermoanaerobacter mathranii* DSM 11426 (brown squares) on 20 mL spent media in closed bottles. The media was obtained from a *C. bescii* fermentation (0.5 L volume, DASGIP system) of 50 g L^{-1} switchgrass after *C. bescii* stopped producing acids. The experiment was performed in three biological replicates (n=3 ± SD), except for the controls of *C. bescii* (black triangles; n=1) and *T. mathranii* (green squares; n=1) grown on their standard media.

CHAPTER 3

GENOME STABILITY IN ENGINEERED STRAINS OF THE EXTREMEMLY THERMOPHIILIC, LIGNOCELLULOSE DEGRADING BACTERIUM *CALDICELLULOSIRUPTOR BESCII*[1]

---

[1]Williams-Rhaesa, A.M., Poole, F.L., Dinsmore, J.T., Lipscomb, G.L., Rubinstein, G.M., Scott, I.M., Conway, J.M., Lee, L.L., Khatibi, P.A., Kelly, R.M., and Adams, M.W.W. 2017. *Applied and Environmental Microbiology*. 83(14):e00444-17.

Abstract

Caldicellulosiruptor bescii is the most thermophilic cellulose degrader known and is of great interest because of its ability to degrade nonpretreated plant biomass. For biotechnological applications, an efficient genetic system is required to engineer it to convert plant biomass into desired products. To date, two different genetically tractable lineages of *C. bescii* strains have been generated. The first (JWCB005) is based on a random deletion within the pyrimidine biosynthesis genes *pyrFA*, and the second (MACB1018) is based on the targeted deletion of *pyrE*, making use of a kanamycin resistance marker. Importantly, an active insertion element, IS*Cbe4*, was discovered in *C. bescii* when it disrupted the gene for lactate dehydrogenase (*ldh*) in strain JWCB018, constructed in the JWCB005 background. Additional instances of IS*Cbe4* movement in other strains of this lineage are presented herein. These observations raise concerns about the genetic stability of such strains and their use as metabolic engineering platforms. In order to investigate genome stability in engineered strains of *C. bescii* from the two lineages, genome sequencing and Southern blot analyses were performed. The evidence presented shows a dramatic increase in the number of single nucleotide polymorphisms, insertions/deletions, and IS*Cbe4* elements within the genome of JWCB005, leading to massive genome rearrangements in its daughter strain, JWCB018. Such dramatic effects were not evident in the newer MACB1018 lineage, indicating that JWCB005 and its daughter strains are not suitable for metabolic engineering purposes in *C. bescii*. Furthermore, a facile approach for assessing genomic stability in *C. bescii* has been established.

Importance

Caldicellulosiruptor bescii is a cellulolytic extremely thermophilic bacterium of great interest for metabolic engineering efforts geared toward lignocellulosic biofuel and bio-based

chemical production. Genetic technology in *C. bescii* has led to the development of two uracil auxotrophic genetic background strains for metabolic engineering. We show that strains derived from the genetic background containing a random deletion in uracil biosynthesis genes (*pyrFA*) have a dramatic increase in the number of single nucleotide polymorphisms, insertions/deletions, and IS*Cbe4* insertion elements in their genomes compared to the wild type. At least one daughter strain of this lineage also contains large-scale genome rearrangements that are flanked by these IS*Cbe4* elements. In contrast, strains developed from the second background strain developed using a targeted deletion strategy of the uracil biosynthetic gene *pyrE* have a stable genome structure, making them preferable for future metabolic engineering studies.

Introduction

Caldicellulosiruptor bescii is a strict anaerobe that grows optimally at 78°C and is the most thermophilic cellulose-degrader known. It produces acetate, lactate and hydrogen from the fermentation of a variety of sugars as well as from non-pretreated plant biomass (177). Its ability to deconstruct lignocellulosic biomass, combined with its high optimal growth temperature, makes *C. bescii* of great biotechnological interest for metabolic engineering efforts toward lignocellulosic bio-based fuels and chemicals production. To this end, a genetic system was developed for *C. bescii* utilizing a uracil auxotrophic mutant background strain and the counter-selectable marker *pyrF*, a gene required for biosynthesis of uracil that also confers sensitivity to 5-fluoroorotic acid (5-FOA) (181, 204). Because there was no method for direct selection of a targeted deletion of *pyrF* in wild-type *C. bescii*, the initial development of a genetic background strain relied on selection of random mutants containing deletions in uracil biosynthesis pathway genes (182). This method resulted in strain JWCB005, which has a partial deletion in both the *pyrF* and *pyrA* genes (182). A more recent development leading to the improvement of *C. bescii* genetic methodologies was the use of a high-temperature kanamycin resistance gene (*htk* codon optimized for *C. bescii* [Cb*htk*]) that enables antibiotic resistance to be utilized for gene insertion or deletion (127). This strategy allowed for the clean deletion of the *pyrE* gene from wild-type *C. bescii*, generating the uracil auxotrophic 5-FOA resistant strain MACB1018. This strain has also been used as a genetic background via utilization of Cb*htk* and kanamycin to select for transformants and 5-FOA resistance to select for loss of the *pyrE* marker (127).

To date, the JWCB005 genetic background strain has been the basis for the majority of the genetically engineered strains of *C. bescii* (**Fig. 3.1, Table 3.1**). Efforts to improve the genetic system included development of a replicating shuttle vector, as well as deletion of the *cbeI* gene

to generate strain JWCB018, allowing for transformation without prior methylation of the DNA (181). Other work has led to a better understanding of plant biomass degradation through deletion of the major cellulose-degrading enzymes CelA and pectate-lyase (187, 191). Additional studies included engineering improved biomass utilization via heterologous expression of cellulose-degrading enzymes from other members of the *Caldicellulosiruptor* genus or other thermophilic cellulolytic organisms (85, 108, 189). Detoxification of furan aldehydes found in pretreated plant materials has also been addressed (98). Utilizing this genetic system also led to the discovery of the unexpected ability of *C. bescii* to utilize tungsten, a metal seldom used in biology (190).

Of particular interest for industrial applications are strains where metabolism has been altered to engineer *C. bescii* for ethanol production. Deletion of the lactate dehydrogenase gene (*ldh*) in *C. bescii* eliminated lactate production and increased acetate and hydrogen production by 21% and 34%, respectively, compared to the parent strain (183). The deletion of the maturation genes required for the nickel-iron hydrogenase showed that this enzyme was not responsible for the majority of hydrogen production by *C. bescii* (184). The addition of a bifunctional alcohol dehydrogenase gene (*adhE*) from *Clostridium thermocellum*, resulting in strain JWCB032, allowed for the production of ethanol from plant biomass at 65°C (185), and production at 75°C was obtained by expressing the genes encoding AdhE and AdhB from *Thermoanaerobacter pseudethanolicus* 39E, although ethanol yield was much lower (186). JWCB032 is the best ethanol-producing strain of *C. bescii* to date, making it thus far the most promising strain for future industrial development.

Genetic stability of microbial strains is paramount for their industrial application. This is especially important as strains of relatively unstudied non-model microorganisms are developed for biotechnological applications. Insertion sequence (IS) elements can contribute to instability via

transposition into other parts of the chromosome, thereby potentially modifying or eliminating gene function or expression. However, little is known currently about IS elements and genome stability in *Clostridiales* and *Firmicutes.* In *C. bescii*, there are 45 annotated IS elements from 7 classes of transposon families, 25 of which are full-length and 20 of which are truncated and presumably inactive. One of these classes of IS elements, IS*Cbe4*, was shown to be active in *C. bescii.* It was observed in the generation of strain JWCB018 where its transposition into the *ldh* gene eliminated lactate production (193). Herein, we have investigated the mutations and IS element movement in strains within the two genetic background lineages, JWCB005 (Δ*pyrFA*) and MACB1018 (Δ*pyrE*), by genome sequencing and Southern blot analyses to assess their genomic stability for future studies. These analyses show that JWCB005 and its daughter strains have a significantly higher number of mutations and active IS*Cbe4* elements compared to wild type and that this contributes to genome instability that can result in large genome rearrangements. In contrast, the genetic background lineage with a clean deletion of *pyrE* (MACB1018) had no significant genome rearrangements and significantly fewer mutations.

Materials and Methods

Growth of C. bescii

C. bescii DSM 6725 was obtained from DSMZ German Collection of Microorganisms and Cell Cultures, and a glycerol stock of it was used to represent wild-type reference (WT-Ref) for assays. Another stock from 2015 that was used as the parent strain for MACB1017-MACB1021 was used for resequencing (WT-Parent) and assays. *C. bescii* strains JWCB005, JWCB018, and JWCB032 were obtained from J. Westpheling (University of Georgia). *C. bescii* strains MACB1013, MACB1017-MACB1021, MACB1032 and MACB1034 were generated as described previously (127). Strains of *C. bescii* were grown on the glucose-containing modified DSM 516

(CG516) medium containing 20 µM uracil, as previously described (127). For sequencing analysis, cultures were grown in 500 mL of cellobiose-containing low osmolarity complex (LOC) medium, as described previously (214). The 500-mL cultures for sequencing analysis were grown overnight statically at 70°C under anaerobic conditions All other cultures were grown overnight at 75°C with shaking at 150 rpm under anaerobic conditions. All strains were revived from glycerol stocks and transferred to fresh media before genomic DNA was extracted for Southern blot and/or DNA sequencing analysis.

Swimming motility assay

The swimming motility assay is based previous studies with mesophilic microorganisms (219). Cultures of each strain were grown overnight and then serially diluted in 1x low-osmolartiy defined (LOD) salts. In an anaerobic chamber containing N_2/H_2 (98%/2%, [vol/vol]), cells were distributed onto plates and grown while embedded in LOC media containing 0.3% Bacto agar. Plates were incubated at 65°C for 2 days.

Clumping assay

Glycerol stocks of WT-Parent, JWCB005, JWCB018, and MACB1018 were revived overnight in CG516 medium supplemented with uracil. Cell density was determined using a Petroff–Hauser counting chamber. Cultures were transferred to 50 mL glass serum bottles containing 20 mL CG516 medium with uracil at a starting cell density of 2×10^6 cells • mL^{-1} and grown at 78°C for 36 h with no shaking or disturbance. Cultures were then swirled gently just prior to being imaged.

Genomic DNA extraction

Cells from 50 mL and 500 mL cultures were harvested at 6000 x g for Southern blot and DNA sequencing analysis, respectively. A phenol chloroform/isoamyl alcohol extraction was

used, as described (136). Ethanol precipitation was performed to obtain highly pure DNA. The DNA concentration was determined using a Thermo Scientific NanoDrop 2000c spectrophotometer.

Strain Construction

pGR002 for transformation into strain JWCB018 was generated using Gibson Assembly from New England Biolabs (220). The structure of the plasmid is shown in **Figure 3.2** and was sequence verified. Competent cells were prepared by growing 500-mL cultures in LOD medium with amino acids (214) and washed with 10% (wt/vol) sucrose as described (127). Cells were mixed with 0.5 to 1µg of plasmid DNA and transferred to a 1-mm-gap electroporation cuvette, and electroporation was carried out as described (127). Recovery was into 20 mL of LOC medium at various intervals, 1mL samples of recovery cultures were centrifuged for one minute at 14,000 rpm and supernatant was removed to minimize uracil carry over to selective media. Cell pellets were resuspended in 0.2 mL of selective LOD medium without uracil supplementation and transferred to the selective medium (214). Transformation isolates were purified once on solid LOD medium without uracil and screened for complete plasmid insertion at the Δ*cbeI* locus. Counterselection was then performed on solid LOD medium with 4mM 5-FOA and 40 µM uracil for loss of the plasmid backbone or reversion to the parent strain. Colony isolates were once again screened and subjected to a final round of purification on solid LOD medium with 40 µM uracil and the insertion of Pf*aor* (*Pyrococcus furiosus* aldehyde ferredoxin oxidoreductase) and *adhA* were sequence confirmed. This strain was designated MACB1013.

Genome sequencing and assembly

All aspects of library construction and sequencing performed at the JGI can be found at http://www.jgi.doe.gov. The raw reads from each genome were *de novo* assembled using HGAP

(version: 2.3.0) (221) into a single chromosomal contig. However, the parameters were not optimal for the correct assembly of *C. bescii*'s two native plasmids and thus these data were not used for further analyses. Read coverage ranged from 72.6x to 499.1x. No genome polishing was performed beyond the PCR verifications listed in this study.

Genome annotation

Genome features and annotations were predicted using JGI's standard pipeline, but that information was not utilized by this study. Instead, the features and annotations from reference genome sequence were transferred to the draft genome sequences using the Rapid Annotation Transfer Tool (RATT, version: 1.0 using the "strain" settings profile) (222) for comparison with other analyses in this study. In the publicly available genome drafts, the genes were identified using Prodigal (223), followed by a round of manual curation using GenePRIMP (224). The predicted coding sequences (CDSs) were translated and used to search the National Center for Biotechnology Information (NCBI) non-redundant database, UniProt, TIGRFam, Pfam, KEGG, COG, and InterPro databases. The tRNAScanSE tool (225) was used to find tRNA genes, whereas rRNA genes were found by searches against models of the rRNA genes built from SILVA (226). Other noncoding RNAs, such as the RNA components of the protein secretion complex and the RNase P, were identified by searching the genome for the corresponding Rfam profiles using INFERNAL (227). Additional gene prediction analysis and manual functional annotation was performed within the Integrated Microbial Genomes (IMG) platform developed by the Joint Genome Institute, Walnut Creek, CA, USA (228).

Genome alignments and variations

The assembled draft genomes were downloaded from the JGI genome portal (229) and the plasmid contigs were removed due to assembly errors (data not shown). The reference genome

was wild-type *C. bescii* DSM 6725 (WT-Ref; accession CP001393.1) from NCBI's GenBank repository and the plasmids were removed for consistency in analyses (176). Whole genome alignments to this reference were created using progressiveMauve (version: 20150226 build 10) (230) with the following nondefault settings: default seed weigh, false; use seed families, true; match seed weight, 15; and min LCB weight, 3000. Based on an initial alignments of each draft genome sequences to the reference, the first base pair of each chromosomal contig was shifted and the reverse complement was generated, if required, using Geneious version: 8.1.8 (231) (see **Table 3.2** specific changes). The final alignments and all other analyses where based on these newly generated draft genome sequences. The locally collinear block (LCB) coordinates from the final progressiveMauve (version: 20150226 build 10) alignments were extracted and added to the genomes as GFF3 tracks. In addition, the final alignments produced the coordinates for SNPs and InDels (or gaps) identified by progressiveMauve. Those coordinates were mapped on the reference sequence features (e.g. genes, CDS, etc.) using a custom BioPerl (232) script and tabulated (**Tables 3.3 to 3.5**).

Southern blot analysis probe generation

Primers for insertion sequence elements IS*Cbe1*, IS*Cbe2*, IS*Cbe3*, IS*Cbe4* and IS*Cbe5* as identified by the ISfinder website (233) were designed using the wild-type strain *C. bescii* DSM 6725 (accession CP001393.1) from NCBI's GenBank repository (176). Probes were designed to generate 400- to 600-bp PCR products from *C. bescii* DSM 6725 genomic DNA using the primer pairs shown in **Table 3.6**. PrimeSTAR Max polymerase (Takara Bio) was used according to manufacturer's recommendations, and PCR products were then concentrated and purified using a DNA Clean & Concentrator kit (Zymo Research). Probes were labeled with digoxigenin (DIG) using the DIG High Prime DNA Labeling & detection Starter Kit I.

Southern blot analyses

Southern blots were generated with 3 µg of genomic DNA from each strain digested with the NsiI-high-fidelity (HF) restriction enzyme (New England BioLabs). The restriction fragments were separated by electrophoresis at 100 V for 2 hours in a 0.7% (wt/vol) agarose gel containing 0.5 µg mL^{-1} of ethidium bromide. The gels were subjected to depurination and denaturation followed by neutralization. Gels were then equilibrated in 1 x Tris-borated-EDTA (TBE) for 10 min. Restriction fragments from the gel were transferred onto positively charged membranes (Roche) using the Thermo Scientific Semidry Electroblotter at 120 mA for 45 min. The DNA fragments were fixed to the membrane by UV-crosslinking in a Stratagene UV Stratalinker 2400 using the auto-cross-link setting. After pre-hybridization, the nylon membrane with the restriction fragments was incubated with the probe while shaking at 45 ºC overnight in accordance with the DIG High Prime DNA Labeling & Detection Starter Kit I (Roche) protocol. Stringency washes were performed using 2x SSC (1x SSC is 0.15 M NaCl plus 0.015 M sodium citrate) and 0.1% SDS at 65°C for two 30-min washes. The immunological detection was performed according to the manufacturer's instructions and membranes were stripped and re-probed following the DIG Protocol.

***In silico* Southern blot analyses**

The reference and draft genome sequences were digitally digested with the NsiI restriction enzyme using Geneious version: 8.1.8, and the individual sequence fragments were assigned unique identification (ID) numbers. Theses fragments were added to a custom BLAST database queried with the IS*Cbe1*, IS*Cbe2*, IS*Cbe3*, IS*Cbe4* and IS*Cbe5* probes (as described above) using blastn (version: 2.2.29 with default settings, except an e-value cutoff of $1e^{-10}$, yet no hit was greater than $1e^{-100}$) (234). Then, only the probe hits were temporarily reassembled into a single pseudo-

contig for each genome using a custom BioPerl script. This pseudo-contig was digitally re-digested with NsiI using Geneious (version: 8.1.8), yielding the digital Southern blot images.

Insertion sequences analyses

The insertion sequence elements where predicted for reference and draft genome sequences using the semi-automatic annotation engine in the Web-based ISsaga 2 tool (http://issaga.biotoul.fr/). Predicted complete and partial IS elements were extracted and added to the genomes as GFF3 tracks.

Accession number(s)

The assembled level 3 improved high-quality draft genome sequences were generated for wild-type *C. bescii* DSM 6725 (WT-Parent; European Nucleotide Archive [ENA] accession no. FXXF01000001), JWCB005 (ENA accession no. FUZN01000002), JWCB018 (ENA accession no. FXXD01000001), MAC1017 (ENA accession no. FXXE01000001), MAC1018 (ENA accession no.FUZJ01000001), MAC1019 (ENA accession no. FUZL01000001), MAC1020 (ENA accession no. FXXC01000001), and MAC1021 (ENA accession no. FWDH01000001) by the U.S. Department of Energy's Joint Genome Institute (JGI) (30) using Pacific Biosciences (PacBio) SMRTbell libraries, with sequencing on the PacBio RS/RS II platform (31).

Results and Discussion

Phenotypic abnormalities and IS element movements in JWCB005 lineage strains

In 2012 the first genetically tractable strain of *C. bescii*, JWCB005, was reported, obtained via selection for a random mutation in the uracil biosynthetic genes (182)**.** JWCB005 contained a deletion in the *pyrFA* genes and growth without uracil was possible with only the addition of the *pyrF* gene (182)**.** Strain JWCB005 and its Δ*cbeI* daughter strain, JWCB018, have since been the basis for 30 of the 40 strains of *C. bescii* that have been developed (**Fig. 3.1**, **Table 3.1**) (181). A

second genetic background lineage of *C. bescii* was recently developed, based on strain MACB1018, which contains a targeted deletion of *pyrE* (2, 5) (**Fig. 3.1, Table 3.1**). So far MACB1018 has been used to develop only two strains, MACB1032 and MACB1034, which are described below. However, several phenotypic abnormalities in strains in the JWCB005 lineage have become apparent, including a decreased swimming motility, clumping when grown without shaking and the spontaneous production of lactate, despite the fact that the gene for *ldh* was assumed to be inactivated.

To further investigate the clumping phenotype in the JWCB005 lineage, a swimming motility assay was used to compare various strains (**Fig. 3.3**). We utilized a wild-type stock of the originally published *C. bescii* DSM 6725 strain (wild-type reference [WT-Ref]), JWCB005 (Δ*pyrFA*), JWCB018 (Δ*pyrFA* Δ*cbeI*), MACB1018 (Δ*pyrE*), and the wild-type parent to MACB1018 (WT-Parent). JWCB005 and WT-Parent both generated discrete colonies, indicative of a swimming motility defect while all other strains gave diffuse colonies (**Fig. 3.3**). Of these five strains, only JWCB018 showed evidence of clumping and settling in liquid growth media (**Fig. 3.4**). These results indicate that the motility and clumping phenotypes are likely not related, and, since they were not previously documented, no genetic basis for their existence is known. However, JWCB018 was reported to be unable to produce lactate and this phenotype was shown to be the result of an insertion sequence element IS*Cbe4* transposition into the lactate dehydrogenase (*ldh*) (193).

We have also observed two instances of IS*Cbe4* transposition during construction of strains using the JWCB005 lineage. The first occurred when a strain, MACB1002, was constructed to heterologously express the gene (*aor*) encoding the aldehyde oxidoreductase of *Pyrococcus furiosus*, driven by the *slp* promoter on a replicating shuttle vector (190). The replicating plasmid

was isolated from the recombinant strain, but the plasmid size differed from that of the original transformation plasmid by about 1.5 kb. PCR and sequence analysis confirmed the presence of an IS*Cbe4* element located 73 bases inside the 5' end of the 200 bp *slp* promoter sequence (**Fig. 3.5, Fig. 3.6**). A 10-base sequence of the promoter was repeated on each side of the IS element. Expression of the Pf*aor* gene did not appear to be negatively affected, as 127 bp of the 3' end of the promoter were still intact, and the expression level of Pf*aor* was higher than that of the native *slp* gene (190).

The second IS element movement that we observed in a JWCB005-derived strain was in strain MACB1013. This strain was constructed to express Pf*aor* and the *adhA* gene encoding the primary alcohol dehydrogenase of *Thermoanaerobacter* sp. strain X514 to investigate ethanol production via the novel AOR-AdhA pathway (235). JWCB018 was used as the parent strain, which contains an IS*Cbe4* insertion in the *ldh* locus, thereby eliminating lactate production (193). JWCB018 was transformed with pGR002 to insert the *aor-adhA* expression construct at the *cbeI* locus to generate strain MACB1013 (**Fig. 3.2**). However, MACB1013 was unexpectedly found to produce lactate. Sequencing of the *ldh* locus showed that the IS*Cbe4* element was no longer present in the *ldh* gene, which instead had the wild-type sequence (**Fig. 3.5)**. The mechanism of the transposase encoded within the IS*Cbe4* element has not been determined, and it is unknown whether loss of IS*Cbe4* in this instance was the result of a rare recombination event or if it was transposase-mediated (236, 237). All of these observations established the need to investigate IS element movements and genome stability within the JWCB005 lineage, as well as the newer MACB1018 (Δ*pyrE*-derived) lineage (**Fig. 3.1**).

Southern blot analyses of *C. bescii* strains

Given the phenotypic abnormalities and observed IS element movements in the Δ*pyrFA* mutant (JWCB005) lineage strains described above, several strains were selected to probe for IS element movement via Southern blot analyses: JWCB005, JWCB018, JWCB032 and MACB1013 from the Δ*pyrFA* mutant lineage, MACB1018, MACB1032, and MACB1034 from the Δ*pyrE* mutant lineage, and WT-Ref as a control. As shown in **Fig. 3.1**, JWCB032 and MACB1013 are daughters of JWCB018 that were engineered for ethanol production using two different pathways, with JWCB032 having the highest ethanol yield of any strain of *C. bescii* to date (185). MACB1018 is a recently developed alternative genetic background strain that was used for construction of MACB1032 and MACB1034, containing deletions of *ldh* and *cbeI*, respectively (127).

Southern blot analysis for IS*Cbe4* was performed on NsiI-digested genomic DNA from the various strains of *C. bescii* and results are shown in **Fig. 3.7**. The probe was designed to hybridize to a 444-base pair region of the transposase gene and contains homology to all intact copies of IS*Cbe4*. Wild-type *C. bescii*, as well as strains MACB1018, MACB1032 and MACB1034, all have the expected pattern of digestion and hybridization with the IS*Cbe4* probe with one additional band present for MACB1018 and its daughter strains. However, all strains from the JWCB005 lineage, including JWCB018, JWCB032, and MACB1013, have a clear increase in the instances of IS*Cbe4* in their genomes, although resolution of the exact number from the Southern blot alone is difficult, and more than one transposon may be present on a given restriction fragment. Furthermore, resolution of bands of similar size may not have occurred under the conditions used for this Southern blot analysis, which would result in a potential underestimation of the movements that have occurred. The same membrane used in **Fig. 3.7** was stripped and hybridized repeatedly with

probes designed for four of the other IS elements in *C. bescii*, IS*Cbe1*, IS*Cbe2*, IS*Cbe3* and IS*Cbe5* (**Fig. 3.8**). For these additional four probes, either limited or no movement was observed for each insertion element.

Sequencing of *C. bescii* strains

To better understand genome stability and IS element movement within the two *C. bescii* lineages, PacBio sequencing was performed on the following strains: JWCB005 (*ΔpyrFA* mutant) and JWCB018 (daughter of JWCB005), as well as MACB1018 (Δ*pyrE* mutant), its wild-type *C. bescii* DSM 6725 parent (WT-Parent), and four of its sister strains generated concurrently with MACB1018: MACB1017, MACB1019, MACB1020 and MACB1021. The sister strains of MACB1018 were included to gain insight into the variation that can occur among isolates within one round of genetic manipulation. PacBio sequencing was used in order to observe the transpositions of IS*Cbe4*, since the longer read length allows for greater accuracy in identification of IS element localization. The sequencing and assembly resulted in a single draft-quality contig for each the strains (see Materials and Methods for processing details). These eight draft genomes were compared to the previously-published level 6 finished Sanger/454-based reference genome of wild-type *C. bescii* DSM 6725 (WT-Ref; accession CP001393.1) (176).

IS element analyses

The Web-based ISsaga2 analysis tool (238) was used to predict the location and type of all known IS elements in all nine genomes (eight strains plus the WT parent). **Table 3.7** shows the number of all known complete and partial IS elements by family in the wild-type reference sequence and in the strains sequenced in this study. IS elements defined as complete by ISsaga are those that meet or exceed both global and local alignment thresholds for IS ends, inverted repeats, direct repeats and associated open reading frames (ORFs). Partial IS elements score below the

threshold and/or are missing components (e.g., inverted repeats). The numbers of complete and partial IS elements remain the same in each of the genomes with the notable exception of IS*Cbe4*, as well as one instance of IS*Cbe2* (**Table 3.7**). The only apparently active IS element is IS*Cbe4*, and the one IS*Cbe2* element that is lost from JWCB018 can be explained by a recombination event between two IS*Cbe4* elements that flank the lost IS*Cbe2* element in the parent strain (**Fig 3.9**).

The genome positions of each IS element were identified using the genome sequences upstream and downstream from them in order to examine the stability of individual elements from each parent strain to its daughter(s). With the exception of IS*Cbe4*, the positions of all the IS elements remain the same; however, as expected from the IS element counts in **Table 3.7**, the genome positions of IS*Cbe4* elements in the two lineages shown in **Figure 3.1** are very different . There was one instance that appeared to be an increase of one IS*Cbe3* element in MACB1018; however, further analysis indicated an assembly error as regions upstream and downstream of the element were also duplicated. PCR analysis confirmed that this region in MACB1018 was identical to the WT-Parent (**Fig 3.10**). This finding highlights the occasional errors that are present in the draft sequences and the need to verify results when questions arise.

In most cases where strains gained IS*Cbe4* elements, the existing IS*Cbe4* elements from the parent strains were retained at the same genome location in the daughter strains. As shown in **Table 3.7**, the sister strains MACB1017-1021 have different numbers of IS*Cbe4* elements. However, many of these insertion sites are the same, and there are only three new locations for IS*Cbe4*, as compared to the parent (WT-Parent). The new IS*Cbe4* in MACB1017 is shared by all five sister strains. The second new IS*Cbe4* in MACB1018 is shared by MACB1019 and MACB1020. The final IS*Cbe4* is unique to MACB1019. Examining the IS*Cbe4* elements in the JWCB005 lineage shows that most of the IS*Cbe4* elements are shared between JWCB005 and

JWCB018, except for those new to JWCB018. The individual elements were slightly more difficult to identify, as a number of elements had swapped flanking regions, indicating recombination had occurred at these IS*Cbe4* elements. There was one notable instance where one IS*Cbe4* element was lost between WT-Parent and JWCB005 .This loss did not result in the duplication of the direct repeats flanking IS*Cbe4*; instead, only one copy of the direct repeat remained. This observation is particularly important as it is the second instance of the loss of IS*Cbe4* without the duplication of the direct repeats. Interestingly, one IS*Cbe4* element in JWCB018 disrupted a gene previously identified as a cellulose-binding protein termed a tāpirin (Athe_1870) (84, 136), and this could potentially give rise to changes in attachment to and degradation of biomass substrates (**Table 3.8**). However, the cause of the swimming motility defect seen in strain WT-Parent and JWCB005 or the clumping phenotype in JWCB018 could not be attributed to movement of the known insertion elements.

While the mechanism of transposition of I*SCbe4* has not been biochemically characterized, the results presented here have implications for its mode of action. The increase in its copy number within the genomes of all strains that have undergone genetic manipulation without loss of the transposon at the wild-type or parent position suggests that this IS element may use a replicative method of transposition (237). A 'copy-out paste-in' mechanism, as described by Curcio and Derbyshire, seems the most likely mode of replication (237). Additionally, Guérillot et. al. reported circular forms of other IS*Lre2* family transposons of which IS*Cbe4* is a member (239). This mechanism would also explain the loss of IS*Cbe4* without duplication of the direct repeats, as this would be a rare event occurring during the repair of the single strand break caused during movement of the IS element, resulting in the clean loss of the element rather than duplication.

Comparison of published and re-sequenced wild type strains

To evaluate the quality of the draft contigs initially, a whole genome alignment of the published, wild-type reference genome (WT-Ref) to the PacBio re-sequenced wild-type contig (WT-Parent) was performed. This showed that there are four single nucleotide polymorphisms (SNPs) (three occurring within genes) between these two genome sequences, and seven deletions (three within genes) and 14 insertions (nine within genes) of two or more nucleotides in the reference genome **(Table 3.3)**. Upon closer inspection of the insertions and deletions in the alignment, most were located in homopolymer runs of four nucleotides or more and were on average less than one nucleotide long so that these may be attributed to sequencing or assembly artifacts. Overall, the two genome sequences generally agree but with only one notable exception, and this is at the so-called glucan degradation locus (GDL) (136). As shown in **Fig. 3.11**, these two genome assemblies have different arrangements within this locus, accounting for two of nine insertions into genes and two of three deletions within genes. Determining which orientation is correct is of biological relevance, since the GDL is critically important for biomass degradation in *C. bescii*. To confirm the correct orientation, we designed primers to bridge these regions and performed PCR (**Fig 3.11**). Genomic DNA from DSM 6725 was used as the template for the six reactions and the primers were paired to give products for either WT-Parent or WT-Ref (**Fig 3.11**). All expected products were generated for the orientation presented by the PacBio re-sequencing, while bands of the incorrect size were observed for the PCR products predicted by the orientation in the original Sanger/454 sequenced *C. bescii* genome. Hence, the PacBio assembly is the correct and biologically relevant sequence (**Fig 3.11**). This region of the *C. bescii* genome has stretches of identical nucleotides up to 2.5 kb in length, and the longer read length of PacBio sequencing is advantageous for assembling this highly repetitive region. Ultimately, there is no change to the

sequence of any particular gene in this region, however, the gene order is affected by these new results.

Genome rearrangements in *C. bescii* strains

To examine genome rearrangements in the two lineages, the JWCB005, JWCB018 and MACB1018 sequences were aligned to the WT-Parent sequence. The WT-Parent sequence was selected as the reference for these alignments because of the corrected GDL orientation. The alignment results show that dramatic rearrangements have occurred in JWCB018 (**Fig 3.9**). The overall organization of the WT-Parent, MACB1018 and JWCB005 genomes is the same, although JWCB005 has a notable increase in the number of IS*Cbe4* elements. JWCB018, however, has two large genome rearrangements and inversions (**Fig 3.9**). Intriguingly, each of these rearrangements is flanked by an IS*Cbe4* element, suggesting that the increased number of IS*Cbe4* elements is responsible for the rearrangements observed. There is also evidence for the loss of 10 kb of DNA encoding for 11 genes (Athe_0717 to Athe_0727) from JWCB018, which is located between two IS*Cbe4* elements in JWCB005. These genes include those encoding five *Leptospira* repeat proteins, two hypothetical proteins, a transglutaminase, a transpeptidase and one instance of IS*Cbe*2. These genes have not been studied in detail in *C. bescii* and clearly are not essential for growth under laboratory conditions. However, these dramatic changes to the genome, combined with the dramatic increase in the number of IS*Cbe4* elements, demonstrate the instability of JWCB005 and of its daughters. In contrast, the alignment of WT-Parent to MACB1018 or any of its sister strains (MACB1017, MACB1019, MACB1020, and MACB1021) show no genome re-arrangements (**Fig. 3.12**).

Mutations in *C. bescii* strains

While rearrangements can have major impacts on a genome's stability, so too can smaller variations. Thus, whole genome alignments of the eight non-wild-type genomes were performed against WT-Ref to investigate SNPs and indels. JWCB005 and JWCB018 had 50 and 56 SNPs in 13 and 16 unique genes, respectively, compared to MACB1018 and its sisters (MACB1017, MACB1019-1021), which ranged from six to nine SNPs in five to six unique genes (**Table 3.4**). As expected, each lineage contained SNPs that were highly conserved among related strains, indicating they are true mutations and are not sequencing or assembly errors. However, it is not clear that any of these SNPs can be attributed to the movements of IS*Cbe4* as none were in close proximity to the insertion elements.

Several SNPs are found in the 23S ribosomal RNA gene (Athe_R0035) of strains JWCB005 (3), JWCB018 (3), and MACB1017 (1). However, a disproportionately high number of the SNPs from the Δ*pyrFA* lineage strains (JWCB005 and JWCB018) were in an annotated pseudogene, Athe_2202. Yet, no SNPs were found in this pseudogene in the Δ*pyrE* lineage strains (MACB1017-MACB1021), suggesting it is not a sequencing or assembly error. Unfortunately, the swimming motility defect in WT-Parent and JWCB005 could not be easily attributed to any SNP. The clumping phenotype in JWCB018 could be related to a SNP in a gene that encodes a transketolase domain protein (Athe_2060). Athe_2060 is homolgous to a gene in *Mycoplasma genitalium* previously reported to cause cells to clump when disrupted (240). However, in *C. bescii*, this SNP cannot be the sole cause for the clumping phenotype because it is also found in JWCB005 which does not exhibit this phenotype.

JWCB005 and JWCB018 had 34 and 79 deletions (16 and 25 in unique genes) and 39 and 46 insertions (12 and 16 in unique genes) compared to the reference genome (**Table 3.5)**. By

contrast, the Δ*pyrE* lineage strains MACB1017-MACB1021, ranged from 15 to 19 deletions (10 to 12 in unique genes) and 9 to 14 insertions (5 to 7 in unique genes) compared to the WT-Ref genome (**Table 3.5)**. It should be noted that the error in WT-Ref GDL, as described above, does account for two of the gene insertions and two of the gene deletions in all eight strains. One of the deletions in JWCB005 caused a frameshift in the flagellar M-ring protein, *Flif* (Athe_2173), which is likely the source of the swimming motility defect based on previous studies in a different organism (219). Interestingly, the WT-Parent, but not the Δ*pyrE* daughter strains, also shares this mutation in a thymine -rich region. Therefore, it was further confirmed by PCR-amplification and sequencing. The results suggest that this region of the *Flif* gene seems to be highly unstable as the phenotype switches readily between generations of *C. bescii* strains.

Establishing a screen for *C. bescii* strains

Although genome sequencing analysis gives a much more precise picture of the movements of IS*Cbe4* and mutations, the approach is not high-throughput and is not feasible for every new *C. bescii* strain that is generated. To that aim, we sought to develop and validate a faster method to determine genome stability of *C. bescii* strains. For this analysis, we compared the Southern blot results shown in **Figure 3.7** with digital southern blots generated from the new sequencing results for those strains that were sequenced. This comparison shows that the Southern blot and sequencing results very closely related, allowing for the potential use of Southern blot analysis as a preliminary screen to estimate changes in the genomes of strains (**Fig 3.7, Fig. 3.13**). Since a goal of this study was to find a method to quickly evaluate the genomic stability of strains over time, it is noteworthy that IS*Cbe4* was the only IS element in *C. bescii* that changed significantly on the Southern blots, making it an ideal single marker for genome stability in these lineages. Since the Southern blots serve as a proxy for IS movements, we set out to calculate the

median number of mutations per IS*Cbe4* element so that we could estimate the total number of mutations in a new strain using only Southern blotting. As a relative control for sequencing and assembly errors, we subtracted the WT-Refseq total IS and total mutations from the total IS and total mutations for each of the newly sequenced strains. While not all IS elements and mutations are phenotypically equal, we assumed they were for ease of calculation. Next, we divided the number of mutations by the number of IS*Cbe4* elements per strain to get the ratio of mutation per IS*Cbe4*. This yielded a median of 8.2 mutations per IS*Cbe4* over the eight strains, with a range of 3 to 10 mutations per IS. This means that for every new band on a Southern blot, there are approximately 8 to 9 additional mutations accompanying it. While IS*Cbe4* elements provide a good estimation of the mutation rate, this method is still susceptible to the limitations of the Southern blotting, so that not all elements will be counted as separate bands when the fragments sizes overlap. However, since the IS elements predicted from the sequencing data are almost identical to those on the blot, this is not a major concern.

Conclusions

Strains constructed using the Δ*pyrFA* mutant lineage (JWCB005) generated by random mutagenesis have improved understanding of *C. bescii* metabolism and biomass deconstruction as well as expanded the possibilities for engineering biofuel production in this organism. However, moving forward, it will be important to utilize the most stable strains for metabolic engineering purposes. The phenotypic abnormalities, chromosomal rearrangement and genetic variations that we have documented in the Δ*pyrFA* mutant lineage (JWCB005) throughout this study are significant. In contrast, our analyses indicate that the newer Δ*pyrE* mutant lineage (MACB1018), constructed via the targeted deletion of *pyrE*, shows no major rearrangements, fewer IS element movements and significantly fewer mutations. Thus, as the *pyrE*-derived lineage is more similar

to wild type and appears to be significantly more stable, it will be a better choice for use in future metabolic engineering efforts. Additionally, with the aid of the long-reads generated from PacBio sequencing technology, we have shown the Glucan Degradation Locus (GDL) in the originally published *C. bescii* wild-type genome is in the incorrect orientation. In the future, Southern blot analysis can be used to initially examine the movement of IS elements in new strains, with particular attention given to IS*Cbe4*, and to provide an estimate of the number of genetic mutations. Genome sequencing will then be used to verify these results, where the time and cost deem it necessary.

Acknowledgements

We thank Daehwan Chung and Janet Westpheling for providing *C. bescii* strains JWCB005, JWCB018 and JWCB032.

This research was supported by a grant (DE-PS02-06ER64304) from the Bioenergy Science Center (BESC), Oak Ridge National Laboratory, a U.S. Department of Energy (DOE) Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Tables and Figures

Table 3.1 Genetically modified strains of *C. bescii* to date including genotype, parent strain and reference.

Strain	Genotype	Parent Strain	Reference
C. bescii DSM 6725		none	(177)
JWCB002	Δ*pyrBCF*	C. bescii DSM 6725	(204)
JWCB003	*pyrBCF* restored by marker replacement	JWCB002	(204)
JWCB005	Δ*pyrFA*	*C. bescii* DSM 6725	(182)
JWCB010	*ΔpyrFA ΔpecABCR*	JWCB005	(187)
JWCB011	transformed with pDCW89	JWCB005	(182)
JWCB014	transformed with pDCW129	JWCB005	(182)
JWCB017	Δ*pyrFA*, Δ*ldh*	JWCB005	(183)
JWCB018	Δ*pyrFA,* Δ*cbeI, ldh::ISCbe4*	JWCB005	(181)
JWCB021	transformed with pDCW89	JWCB018	(182)
JWCB029	Δ*pyrFA,* Δ*cbeI, ldh::ISCbe4, ΔcelA*	JWCB018	(191)
JWCB032	*ΔpyrFA, ldh::ISCbe4, Δcbe1,* $P_{S\text{-}layer}$ *Cthe-adhE*	JWCB018	(187)
JWCB033	*ΔpyrFA, ldh::ISCbe4, Δcbe1,* $P_{S\text{-}layer}$ *Cthe-adhE*(EA)*	JWCB018	(187)
JWCB036	*ΔpyrFA, Δldh, CIS1::*$P_{S\text{-}layer}$ *Cthe-adhE*	JWCB017	(184)
JWCB038	*ΔpyrFA, Δldh, CIS1::*$P_{S\text{-}layer}$ *Cthe-adhE, ΔhypADFCDE*	JWCB017	(184)
JWCB044	*ΔpyrFA, ldh::ISCbe4, ΔcbeI,* $P_{S\text{-}layer}$ *bdhA*	JWCB018	(98)
JWCB046	*transformed with pDCW173*	JWCB029	(188)

JWCB049	Δ*pyrFA,* Δ*ldh, CIS1::* $P_{S\text{-}layer}$ *adhE*(Teth39_0206)	JWCB017	(186)
JWCB052	*ΔpyrFA, ldh::ISCbe4, Δcbe1,* $P_{S\text{-}layer}$ *E1*	JWCB018	(108)
JWCB054	Δ*pyrFA,* Δ*ldh, CIS1::* $P_{S\text{-}layer}$ *adhB*(Teth39_0218)	JWCB017	(186)
JWCB069	transformed with pSKW06	JWCB029	(109)
JWCB070	transformed with pSKW07	JWCB029	(109)
JWCB071	transformed with pSKW09	JWCB029	(109)
JWCB073	transformed with pJGW07 containing *C. thermocellum pyrF*	JWCB052	(189)
JWCB074	transformed with pSKW10 containing $P_{S\text{-}layer}$ *Acel_0180*	JWCB052	(189)
JWCB075	transformed with pSKW11 containing $P_{S\text{-}layer}$ *Acel_0372*	JWCB052	(189)
JWCB079	transformed with pSKW14	JWCB029	(109)
JWCB080	transformed with pSKW15	JWCB029	(109)
JWCB081	transformed with pSKW16	JWCB029	(109)
RKCB103	Δ*pyrFA,* Δ*cbeI, ldh::ISCbe4,* pJMC009 ($P_{S\text{-}layer}$ Calkro_0402)	JWCB018	(85)
RKCB106	Δ*pyrFA,* Δ*cbeI, ldh::ISCbe4,* $P_{S\text{-}layer}$ *Cb*htk	JWCB018	(127)
MACB1002	Δ*pyrFA,* Δ*cbeI, ldh::ISCbe4,* pIMSPfAOR ($P_{S\text{-}layer}$-his6-PF0346)	JWCB018	(190)
MACB1013	Δ*pyrFA,* Δ*cbeI,* $P_{S\text{-}layer}$ *aor adhA* (PF0346, Teth514_0564)	JWCB018	This study
MACB1015	transformed with pSBS4	*C. bescii* DSM 6725	(127)
MACB1017	Δ*pyrE*	*C. bescii* DSM 6725	(127)
MACB1018	Δ*pyrE*	*C. bescii* DSM 6725	(127)
MACB1019	Δ*pyrE*	*C. bescii* DSM 6725	(127)

MACB1020	Δ*pyrE*	*C. bescii* DSM 6725	(127)
MACB1021	Δ*pyrE*	*C. bescii* DSM 6725	(127)
MACB1032	Δ*pyrE*, Δ*cbeI*	MACB1018	(127)
MACB1034	Δ*pyrE*, Δ*ldh*	MACB1018	(127)

Table 3.2 The chromosomal NCBI/JGI contig name, length and public accession number for each strain.

Plus, the adjustments made to each contig to match the reference (WT-Ref) including the reverse complement and/or the shift in start nucleotide.

Strain Name	NCBI/JGI Contig Name	Contig Length (bp)	Change from Ref. (bp)	Reverse Compl.	Shift to 1 in Ref. (bps)	ENA Accession Numbers
WT-Ref	GCA_000022325.1_ASM2232v1_genomic	2919718	0	No	N/A	CP001393.1 (NCBI)
WT-Parent	Ga0101194_11	2919647	-71	No	264482	FXXF01000001
JWCB005	Ga0070649_12	2948278	28,560	No	94489	FUZN01000002
JWCB018	Ga0070650_12	2930037	10,319	No	675606	FXXD01000001
MACB1017	Ga0100407_11	2932284	12,566	No	271738	FXXE01000001
MACB1018	Ga0100408_11	2937804	18,086	Yes	2555409	FUZJ01000001
MACB1019	Ga0100409_11	2945647	25,929	No	2074887	FUZL01000001
MACB1020	Ga0100410_11	2915698	-4,020	Yes	2608936	FXXC01000001
MACB1021	Ga0100411_11	2917586	-2,132	No	2083044	FWDH01000001

Table 3.3 Summary tables of the single nucleotide polymorphisms (SNPs) (A) and insertions/deletions (InDels) (B) in the eight PacBio-sequenced strains as compared to the reference sequence (WT-Ref).

The number in parenthesis is the number of unique genes. The InDels table only accounts for two of the eleven missing genes from JWCB018. The genes at the ends of the glucan degradation locus (GDL) are also counted as two insertions and two deletions for each strain due to the error in the WT-Ref. The pyrFA and pyrE deletions are accounted for in these tables. The asterisks means the deletion spans more than one gene, see InDels table for more details (**Table 3.5**).

A

Summary of Single Nucleotide Polymorphisms (SNPs)

Location	WT-Parent	JWCB005	JWCB018	MACB1017	MACB1018	MACB1019	MACB1020	MACB1021
Gene	3(3)	36(13)	39(16)	5(5)	6(6)	5(5)	6(5)	5(5)
CRISPRs	-	5	7	-	-	-	-	-
Intergenic	1	9	10	2	3	1	1	2
Total	**4**	**50**	**56**	**7**	**9**	**6**	**7**	**7**

B

Summary of Insertions and Deletions (InDels)

Location	**WT-Parent**		**JWCB005**		**JWCB018**		**MACB1017**	
	Deleted	**Inserted**	**Deleted**	**Inserted**	**Deleted**	**Inserted**	**Deleted**	**Inserted**
Gene	9(9)	3(3)	16(16)	12(12)	38(25)	18(16)	12*(12)	5(5)
CRISPRs	1	1	7	8	7	7	1	1
Intergenic	4	3	11	19	34	21	6	3
Total	**14**	**7**	**34**	**39**	**79**	**46**	**19**	**9**

Location	**MACB1018**		**MACB1019**		**MACB1020**		**MACB1021**	
	Deleted	**Inserted**	**Deleted**	**Inserted**	**Deleted**	**Inserted**	**Deleted**	**Inserted**
Gene	10*(10)	5(5)	10*(10)	5(5)	12*(11)	7(7)	10*(10)	5(5)
CRISPRs	1	1	-	-	1	1	1	1
Intergenic	6	7	5	8	6	6	6	4
Total	**17**	**13**	**15**	**13**	**19**	**14**	**17**	**10**

Table 3.4 A detailed table of the single nucleotide polymorphisms (SNPs) in the eight PacBio-sequenced strains as compared to the reference sequence (WT-Ref).

The table provides the base change, the position of the change in the reference, and the location of the area affected by each SNP in each strain.

Strain	Change from Reference (from base, to base)	Start Nucleotide in Reference	Location of Change in Reference	Reference Accession	Reference Locus Tag
WT-Parent	ga	743778	Gene	ACM59783.1	Athe_0659
	tc	1099916	Gene	ACM60103.1	Athe_1002
	ga	1472117	Gene	ACM60487.1	Athe_1387
	tc	1476058	Intergenic		
JWCB005	ct	1023064	Intergenic		
	tc	1099916	Gene	ACM60103.1	Athe_1002
	ct	1386789	Gene	ACM60397.1	Athe_1297
	tc	1476058	Intergenic		
	ga	1483203	Gene	ACM60500.1	Athe_1402
	ct	1647432	Gene	ACM60654.1	Athe_1556
	ga	1654621	Gene	ACM60661.1	Athe_1563
	ct	1694416	Gene	ACM60693.1	Athe_1599
	ct	1794365	Gene		Athe_R0035
	tg	1794366	Gene		Athe_R0035
	ag	1794369	Gene		Athe_R0035
	ga	1800964	Gene	ACM60804.1	Athe_1710
	ct	1947124	Gene	ACM60947.1	Athe_1859
	tc	2178091	Gene	ACM61137.1	Athe_2060
	ag	2331504	Gene		Athe_2202
	ca	2331577	Gene		Athe_2202
	ta	2331605	Gene		Athe_2202
	at	2331617	Gene		Athe_2202
	at	2331638	Gene		Athe_2202
	ag	2331641	Gene		Athe_2202
	ag	2331779	Gene		Athe_2202
	ct	2331780	Gene		Athe_2202
	tc	2331787	Gene		Athe_2202
	ct	2331813	Gene		Athe_2202
	tc	2331815	Gene		Athe_2202
	ac	2331816	Gene		Athe_2202
	gc	2331818	Gene		Athe_2202
	ct	2331819	Gene		Athe_2202
	ct	2331823	Gene		Athe_2202

	tc	2331824	Gene		Athe_2202
	ct	2331825	Gene		Athe_2202
	tc	2331831	Gene		Athe_2202
	at	2331840	Gene		Athe_2202
	tc	2331872	Gene		Athe_2202
	ac	2332150	Gene		Athe_2202
	at	2332151	Gene		Athe_2202
	tc	2341043	Intergenic		
	ct	2448916	Gene	ACM61380.1	Athe_2310
	ga	2586532	Gene	ACM61511.1	Athe_2443
	tc	2705921	Intergenic		
	gc	2705930	Intergenic		
	at	2705931	Intergenic		
	ct	2705932	Intergenic		
	ta	2705933	Intergenic		
	ga	2705934	Intergenic		
	ta	2866454	CRISPRs		
	ag	2866516	CRISPRs		
	tc	2866517	CRISPRs		
	at	2866519	CRISPRs		
	at	2866520	CRISPRs		
JWCB018	cg	167795	Gene	ACM59289.1	Athe_0134
	ga	904605	Gene	ACM59924.1	Athe_0809
	ct	1023064	Intergenic		
	tc	1099916	Gene	ACM60103.1	Athe_1002
	ga	1199166	Intergenic		
	ct	1386789	Gene	ACM60397.1	Athe_1297
	tc	1476058	Intergenic		
	ct	1480969	Gene	ACM60497.1	Athe_1399
	ga	1483203	Gene	ACM60500.1	Athe_1402
	ct	1647432	Gene	ACM60654.1	Athe_1556
	ga	1654621	Gene	ACM60661.1	Athe_1563
	ct	1694416	Gene	ACM60693.1	Athe_1599
	ct	1794365	Gene		Athe_R0035
	tg	1794366	Gene		Athe_R0035
	ag	1794368	Gene		Athe_R0035
	ga	1800964	Gene	ACM60804.1	Athe_1710
	tc	2178091	Gene	ACM61137.1	Athe_2060
	ag	2246191	Gene	ACM61191.1	Athe_2116
	ag	2331504	Gene		Athe_2202
	ca	2331577	Gene		Athe_2202
	ta	2331605	Gene		Athe_2202
	at	2331617	Gene		Athe_2202
	at	2331638	Gene		Athe_2202
	ag	2331641	Gene		Athe_2202

	ag	2331779	Gene		Athe_2202
	ct	2331780	Gene		Athe_2202
	tc	2331787	Gene		Athe_2202
	ct	2331813	Gene		Athe_2202
	tc	2331815	Gene		Athe_2202
	ac	2331816	Gene		Athe_2202
	gc	2331818	Gene		Athe_2202
	ct	2331819	Gene		Athe_2202
	ct	2331823	Gene		Athe_2202
	tc	2331824	Gene		Athe_2202
	ct	2331825	Gene		Athe_2202
	tc	2331831	Gene		Athe_2202
	at	2331840	Gene		Athe_2202
	tc	2331872	Gene		Athe_2202
	ac	2332150	Gene		Athe_2202
	at	2332151	Gene		Athe_2202
	tc	2341043	Intergenic		
	ct	2448916	Gene	ACM61380.1	Athe_2310
	ga	2586532	Gene	ACM61511.1	Athe_2443
	tc	2705921	Intergenic		
	gc	2705930	Intergenic		
	at	2705931	Intergenic		
	ct	2705932	Intergenic		
	ta	2705933	Intergenic		
	ga	2705934	Intergenic		
	tc	2866512	CRISPRs		
	at	2866513	CRISPRs		
	at	2866516	CRISPRs		
	tc	2866517	CRISPRs		
	ct	2866518	CRISPRs		
	at	2866519	CRISPRs		
	ag	2866520	CRISPRs		
MACB1017	gt	971622	Gene	ACM59982.1	Athe_0872
	tc	1099916	Gene	ACM60103.1	Athe_1002
	ct	1196449	Intergenic		
	ga	1205042	Gene	ACM60211.1	Athe_1110
	tc	1476058	Intergenic		
	ac	1597066	Gene	ACM60609.1	Athe_1511
	gc	1792303	Gene		Athe_R0035
MACB1018	ta	261516	Gene	ACM59366.1	Athe_0215
	ct	552007	Intergenic		
	gt	971622	Gene	ACM59982.1	Athe_0872
	tc	1099916	Gene	ACM60103.1	Athe_1002
	gt	1277027	Gene	ACM60282.1	Athe_1182
	tc	1476058	Intergenic		

	ac	1597066	Gene	ACM60609.1	Athe_1511
	gt	2529893	Gene	ACM61447.1	Athe_2378
	ag	2813985	Intergenic		
MACB1019	gt	971622	Gene	ACM59982.1	Athe_0872
	tc	1099916	Gene	ACM60103.1	Athe_1002
	ct	1299689	Gene	ACM60309.1	Athe_1209
	tc	1476058	Intergenic		
	ac	1597066	Gene	ACM60609.1	Athe_1511
	ta	1994378	Gene	ACM60973.1	Athe_1885
MACB1020	ag	303966	Gene	ACM59401.1	Athe_0251
	ag	303970	Gene	ACM59401.1	Athe_0251
	gt	971622	Gene	ACM59982.1	Athe_0872
	tc	1099916	Gene	ACM60103.1	Athe_1002
	tc	1476058	Intergenic		
	ac	1597066	Gene	ACM60609.1	Athe_1511
	ga	2254075	Gene	ACM61195.1	Athe_2120
MACB1021	gt	971622	Gene	ACM59982.1	Athe_0872
	gt	1082027	Gene	ACM60082.1	Athe_0980
	tc	1099916	Gene	ACM60103.1	Athe_1002
	tc	1476058	Intergenic		
	ac	1597066	Gene	ACM60609.1	Athe_1511
	ga	1991244	Gene	ACM60971.1	Athe_1883
	ca	2084015	Intergenic		

Table 3.5 A detailed table of the insertions/deletions (InDels) in the eight PacBio-sequenced strains as compared to the reference sequence (WT-Ref).

Deletions are represented by grey regions and insertions by white regions. The table provided the size of the change in base pairs, the position of the change in the reference, and the location of the area affected by each InDel in each strain. The table only accounts for two of the eleven missing genes from JWCB018. The genes at the ends of the glucan degradation locus (GDL) are also counted as two insertions and two deletions for each strain due to the error in the WT-Ref. The pyrFA and pyrE deletions are accounted for in these tables. The blank values in the "Start Nucleotide in Reference" column means that the proceeding deletion spans more than one gene (i.e. one deletion affects two genes).

Strain	Change from Reference	Size of Change (bp)	Start Nucleotide in Reference	Location of Change in Reference	Reference Accession
WT-Parent	Deleted	6	371975	Intergenic	
		1	471092	Intergenic	
		1	581142	Gene	ACM59637.1
		1	1074360	Gene	
		1	1083492	Gene	
		1	1465972	Gene	ACM60479.1
		1	1939565	Gene	ACM60945.1
		877	1948071	Gene	ACM60948.1
		2	1969680	Gene	ACM60955.1
		1	2302460	Gene	ACM61248.1
		2	2372589	Intergenic	
		1	2654955	Intergenic	
		1	2693807	Gene	ACM61637.1
		68	2867839	CRISPRs	
WT-Parent	Inserted	8	647457	Intergenic	
		1	1491795	Gene	ACM60506.1
		1	1598746	Intergenic	
		1	1805271	Intergenic	
		309	1939565	Gene	ACM60945.1
		571	1948071	Gene	ACM60948.1
		2	2867907	CRISPRs	

JWCB005	Deleted	4	39084	Gene	
		4	73715	Gene	ACM59218.1
		4	157089	CRISPRs	
		4	585909	Intergenic	
		4	729778	Intergenic	
		3	807815	Gene	ACM59832.1
		1	1074360	Gene	
		1	1083492	Gene	
		880	1461844	Gene	ACM60477.1
		1401	1472541	Intergenic	
		2	1598731	Intergenic	
		4	1702170	Gene	ACM60697.1
		4	1759853	Gene	ACM60759.1
		1623	1771412	Intergenic	
		1	1794367	Gene	
		4	1903977	Gene	ACM60922.1
		4	1919956	Intergenic	
		1	1939565	Gene	ACM60945.1
		877	1948071	Gene	ACM60948.1
		2	1969680	Gene	ACM60955.1
		3	2017659	Gene	ACM61001.1
		1	2302460	Gene	ACM61248.1
		2	2372585	Intergenic	
		4	2449207	Intergenic	
		1	2675894	Gene	ACM61624.1
		29	2705881	Intergenic	
		9	2705940	Intergenic	
		29	2759683	Intergenic	
		4	2829308	CRISPRs	
		35	2866353	CRISPRs	
		32	2866419	CRISPRs	
		30	2866483	CRISPRs	
		34	2866551	CRISPRs	
		202	2867431	CRISPRs	
JWCB005	Inserted	1623	39088	Gene	
		1622	73719	Gene	ACM59218.1
		70	157093	CRISPRs	
		2	316961	Intergenic	
		1	568604	Intergenic	
		1623	585913	Intergenic	
		6	647457	Intergenic	
		1623	729782	Intergenic	
		1	753553	Intergenic	
		1621	807818	Gene	ACM59832.1
		1	993722	Intergenic	

		1	1004397	Intergenic	
		1	1199112	Intergenic	
		1	1348516	Intergenic	
		2	1462724	Gene	ACM60478.1
		4	1473942	Intergenic	
		1	1491795	Gene	ACM60506.1
		2	1674697	Intergenic	
		1622	1702174	Gene	ACM60697.1
		1622	1759857	Gene	ACM60759.1
		4	1773035	Intergenic	
		1	1805271	Intergenic	
		1623	1903981	Gene	ACM60922.1
		1622	1919960	Intergenic	
		309	1939565	Gene	ACM60945.1
		571	1948071	Gene	ACM60948.1
		1622	2017662	Gene	ACM61001.1
		2	2331599	Gene	
		1623	2449211	Intergenic	
		196	2705910	Intergenic	
		4	2705919	Intergenic	
		3266	2759712	Intergenic	
		10579	2829312	CRISPRs	
		10	2866331	CRISPRs	
		296	2866388	CRISPRs	
		227	2866451	CRISPRs	
		97	2866513	CRISPRs	
		299	2866585	CRISPRs	
		3	2867633	CRISPRs	
		4	39078	Gene	
		10741	39083	Gene	
		4	73728	Gene	ACM59218.1
		4	157123	CRISPRs	
		1	191715	Gene	ACM59306.1
		1	344569	Intergenic	
		1	568616	Intergenic	
		4	585923	Intergenic	
JWCB018	Deleted	1	641562	Gene	ACM59694.1
		7	647482	Intergenic	
		1	663300	Intergenic	
		4	729792	Intergenic	
		1	807663	Gene	ACM59832.1
		1	1071111	Intergenic	
		1	1074360	Gene	
		1	1083492	Gene	
		4	1267954	Intergenic	

		880	1461844	Gene	ACM60477.1
		1403	1472539	Intergenic	
		2	1565432	Intergenic	
		5	1565436	Intergenic	
		2	1598757	Intergenic	
		4	1702175	Gene	ACM60697.1
		4	1759868	Gene	ACM60759.1
		1622	1773045	Intergenic	
		4	1783909	Intergenic	
		1	1794370	Gene	
		4	1919956	Intergenic	
		9285	1939663	Gene	ACM60945.1
		3	1969679	Gene	ACM60955.1
		1	1978392	Intergenic	
		2365	1980233	Gene	ACM60958.1
		3	2017659	Gene	ACM61001.1
		2	2023877	Gene	ACM61006.1
		1	2092992	Intergenic	
		1	2243370	Gene	ACM61189.1
		1	2243615	Intergenic	
		1	2243805	Intergenic	
		1	2244092	Intergenic	
		1	2244154	Intergenic	
		1	2244387	Intergenic	
		1	2244786	Gene	ACM61190.1
		1	2245320	Gene	ACM61190.1
		1	2245703	Gene	ACM61191.1
		1	2245796	Gene	ACM61191.1
		1	2245923	Gene	ACM61191.1
		1	2246074	Gene	ACM61191.1
		1	2246165	Gene	ACM61191.1
		2	2246188	Gene	ACM61191.1
		1	2246194	Gene	ACM61191.1
		1	2246309	Gene	ACM61191.1
		2	2246752	Gene	ACM61191.1
		1	2247085	Gene	ACM61192.1
		1	2247145	Gene	ACM61192.1
		1	2247617	Gene	ACM61193.1
		1	2247848	Gene	ACM61193.1
		1	2248789	Gene	ACM61194.1
		2	2248815	Gene	ACM61194.1
		1	2249349	Gene	ACM61195.1
		1	2271465	Intergenic	
		4	2315676	Intergenic	
		4	2372607	Intergenic	

		1	2413735	Intergenic	
		4	2449207	Intergenic	
		1	2517436	Intergenic	
		1044	2581097	Intergenic	
		1	2620206	Gene	ACM61547.1
		1	2675894	Gene	ACM61624.1
		29	2705881	Intergenic	
		9	2705940	Intergenic	
		1	2730739	Intergenic	
		29	2759683	Intergenic	
		136	2807172	Intergenic	
		3	2829278	CRISPRs	
		39	2866353	CRISPRs	
		38	2866418	CRISPRs	
		22	2866484	CRISPRs	
		37	2866548	CRISPRs	
		202	2867431	CRISPRs	
JWCB018	Inserted	1625	39078	Gene	
		1622	73723	Gene	ACM59218.1
		70	157118	CRISPRs	
		2	316967	Intergenic	
		1623	585918	Intergenic	
		1623	729787	Intergenic	
		1621	807825	Gene	ACM59832.1
		1	1044437	Intergenic	
		1622	1267958	Intergenic	
		1	1348516	Intergenic	
		2	1462724	Gene	ACM60478.1
		6	1473942	Intergenic	
		1	1491795	Gene	ACM60506.1
		1618	1565432	Intergenic	
		1	1674715	Intergenic	
		1622	1702170	Gene	ACM60697.1
		1622	1759863	Gene	ACM60759.1
		3	1771422	Intergenic	
		1623	1783904	Intergenic	
		1	1805279	Intergenic	
		7	1903981	Gene	ACM60922.1
		1622	1919960	Intergenic	
		211	1948948	Gene	ACM60948.1
		9077	1969682	Gene	ACM60955.1
		3	1982598	Gene	ACM60959.1
		1622	2017662	Gene	ACM61001.1
		1620	2023879	Gene	ACM61006.1
		1	2246335	Gene	ACM61191.1

		1	2246453	Gene	ACM61191.1
		1	2246872	Gene	ACM61191.1
		1	2247386	Gene	ACM61193.1
		1	2254613	Intergenic	
		1622	2315680	Intergenic	
		2	2331599	Gene	
		1623	2449211	Intergenic	
		4	2582141	Intergenic	
		196	2705910	Intergenic	
		4	2705919	Intergenic	
		3266	2759712	Intergenic	
		4	2807308	Intergenic	
		200	2829281	CRISPRs	
		496	2866392	CRISPRs	
		104	2866456	CRISPRs	
		88	2866506	CRISPRs	
		236	2866585	CRISPRs	
		3	2867633	CRISPRs	
		1	296422	Gene	ACM59398.1
		6	371975	Intergenic	
		1	471092	Intergenic	
		4	647457	Intergenic	
		4	875003	Intergenic	
		1	1074360	Gene	
		1	1083492	Gene	
		569	1468034	Gene	ACM60481.1
				Gene	ACM60482.1
MACB1017	Deleted	843	1471558	Gene	ACM60486.1
		1	1674697	Intergenic	
		1	1939565	Gene	ACM60945.1
		877	1948071	Gene	ACM60948.1
		2	1969680	Gene	ACM60955.1
		1	2039956	Gene	
		4	2372585	Intergenic	
		3188	2651776	Gene	ACM61596.1
		1	2712154	Gene	ACM61640.1
		69	2866587	CRISPRs	
		1622	875007	Intergenic	
		4	1468603	Gene	ACM60482.1
		4	1472401	Gene	ACM60487.1
MACB1017	Inserted	1	1491795	Gene	ACM60506.1
		1	1805271	Intergenic	
		309	1939565	Gene	ACM60945.1
		571	1948071	Gene	ACM60948.1
		15624	2654964	Intergenic	

		4	2866656	CRISPRs	
MACB1018	Deleted	69	157939	CRISPRs	
		1	296420	Gene	ACM59398.1
		4	371981	Intergenic	
		3	647457	Intergenic	
		4	875003	Intergenic	
		1	1074360	Gene	
		1	1083492	Gene	
		1	1465972	Gene	ACM60479.1
		569	1468034	Gene	ACM60481.1
				Gene	ACM60482.1
		843	1471558	Gene	ACM60486.1
		1	1939565	Gene	ACM60945.1
		877	1948071	Gene	ACM60948.1
		2	1969680	Gene	ACM60955.1
		2	2372595	Intergenic	
		4	2449202	Intergenic	
		1	2564195	Intergenic	
MACB1018	Inserted	4	158008	CRISPRs	
		16327	371985	Intergenic	
		1622	875007	Intergenic	
		1	1044437	Intergenic	
		4	1468603	Gene	ACM60482.1
		4	1472401	Gene	ACM60487.1
		1	1491795	Gene	ACM60506.1
		1	1598731	Intergenic	
		1	1674697	Intergenic	
		1	1805271	Intergenic	
		309	1939565	Gene	ACM60945.1
		571	1948071	Gene	ACM60948.1
		1623	2449206	Intergenic	
MACB1019	Deleted	1	296420	Gene	ACM59398.1
		1	302270	Intergenic	
		6	371976	Intergenic	
		1	471092	Intergenic	
		1161	855636	Gene	ACM59881.1
		4	875003	Intergenic	
		1	1074360	Gene	
		1	1083492	Gene	
		569	1468034	Gene	ACM60481.1
				Gene	ACM60482.1
		843	1471559	Gene	ACM60486.1
		1	1939565	Gene	ACM60945.1
		877	1948071	Gene	ACM60948.1
		2	1969680	Gene	ACM60955.1

		4	2449202	Intergenic	
MACB1019	Inserted	14	647457	Intergenic	
		25249	856797	Intergenic	
		1622	875007	Intergenic	
		1	1004397	Intergenic	
		1	1199112	Intergenic	
		4	1468603	Gene	ACM60482.1
		4	1472402	Gene	ACM60487.1
		1	1491795	Gene	ACM60506.1
		2	1674697	Intergenic	
		1	1805271	Intergenic	
		309	1939565	Gene	ACM60945.1
		571	1948071	Gene	ACM60948.1
		1622	2449206	Intergenic	
MACB1020	Deleted	135	155976	CRISPRs	
		1	296420	Gene	ACM59398.1
		3	344552	Intergenic	
		1	344829	Intergenic	
		6	371975	Intergenic	
		1	405754	Gene	ACM59487.1
		4	875003	Intergenic	
		1	1074360	Gene	
		1	1083492	Gene	
		569	1468034	Gene	ACM60481.1
				Gene	ACM60482.1
		843	1471558	Gene	ACM60486.1
		1	1939565	Gene	ACM60945.1
		877	1948071	Gene	ACM60948.1
		2	1969680	Gene	ACM60955.1
		1	2372597	Intergenic	
		4	2449202	Intergenic	
		68	2687762	Gene	ACM61633.1
		7013	2687864	Gene	ACM61633.1
MACB1020	Inserted	4	156111	CRISPRs	
		15	647459	Intergenic	
		1622	875007	Intergenic	
		1	1013658	Intergenic	
		4	1468603	Gene	ACM60482.1
		4	1472401	Gene	ACM60487.1
		1	1491795	Gene	ACM60506.1
		1	1674697	Intergenic	
		1	1805271	Intergenic	
		309	1939565	Gene	ACM60945.1
		571	1948071	Gene	ACM60948.1
		1623	2449206	Intergenic	

		68	2687830	Gene	ACM61633.1
		1287	2694877	Gene	ACM61637.1
		1	17250	Gene	ACM59178.1
		1	296420	Gene	ACM59398.1
		6	371975	Intergenic	
		230	451681	Intergenic	
		1	471092	Intergenic	
		4	875003	Intergenic	
		1	1074360	Gene	
		1	1083492	Gene	
MACB1021	Deleted	569	1468034	Gene	ACM60481.1
				Gene	ACM60482.1
		843	1471558	Gene	ACM60486.1
		1	1598731	Intergenic	
		1	1674697	Intergenic	
		1	1939565	Gene	ACM60945.1
		877	1948071	Gene	ACM60948.1
		2	1969680	Gene	ACM60955.1
		2185	2828752	CRISPRs	

Table 3.6 Primers used in this study including those for to amplify IS element probes for Southern Blot analysis, GDL arrangement confirmation and IS*Cbe3* duplication confirmation.

Target	Primer name	Sequence (5'-3')
IS*Cbe4*	**JD001**	GGAGCGATAATAGAGAGAGTAGTAGAC
	JD002	CTTTATCCAATTAGCTCCATCTCC
IS*Cbe1*	**JD003**	GCATACTTCAACATCCCTATCACAG
	JD004	GTACGTATATGCGACGATACAAGC
IS*Cbe2*	**JD005**	CTGGCTTCAGATACAGACG
	JD006	GCTCTTGGTGGAATAGGATTTAGTTC
IS*Cbe3*	**JD007**	CAGTCAAGGGTGTTTATGAG
	JD008	CACCCTGATATGGCAGTATC
IS*Cbe6*	**JD009**	CTCAACTGGTGGATTATACATC
	JD010	CCAAGAGACAGAGAAGGTG
GDL A	**AR073**	CCACTTGGTGCCACATAAATAGC
GDL B	**AR075**	GCAAGAAGGTTAGGTGGAAACAG
	AR076	CTGTTTCCACCTAACCTTCTTGC
GDL C	**AR077**	AAGAAGAAATTCAATCAAAGTTGATG
	AR074	TCACGTATGACGATTGAAGC
GDL D	**AR078**	GAGGTTAGAGATTTATGAAGCGTTACAG
IS*Cbe3* region A	**AR082**	TAAAAGCTGTATCGCACCACC
	AR084	AATTGAAGCAGAGTGTGGAGC

IS*Cbe3* region B	**AR083**	CTCAGCTTATTCAAGGACGAC
	AR081	CACTTCTCAGTGGAGTAGAGTC

Table 3.7 The number of all known insertion elements in each of the re-sequenced strains of *C. bescii*.

	Strains								
ISType	**WT-Ref**	**WT-Parent**	**JWCB005**	**JWCB018**	**MACB1017**	**MACB1018**	**MACB1019**	**MACB1020**	**MACB1021**
IS*Cbe1*	4(2)	4(2)	4(2)	4(2)	4(2)	4(2)	4(2)	4(2)	4(2)
IS*Cbe2*	5(2)	5(2)	5(2)	4(2)	5(2)	5(2)	5(2)	5(2)	5(2)
IS*Cbe3*	4(2)	4(2)	4(2)	4(2)	4(2)	4(2)	4(2)	4(2)	4(2)
IS*Cbe4*	7(5)	7(5)	19(5)	23(5)	8(5)	9(5)	10(5)	9(5)	8(5)
IS*Cbe6*	4(3)	4(3)	4(3)	4(3)	4(3)	4(3)	4(3)	4(3)	4(3)
IS*Csa1*	0(2)	0(2)	0(2)	0(2)	0(2)	0(2)	0(2)	0(2)	0(2)
IS*Csa9*	1(4)	1(4)	1(4)	1(4)	1(4)	1(4)	1(4)	1(4)	1(4)

The number of partial elements is given in parentheses.

Table 3.8 Genes likely disrupted by IS elements in sequenced strains.

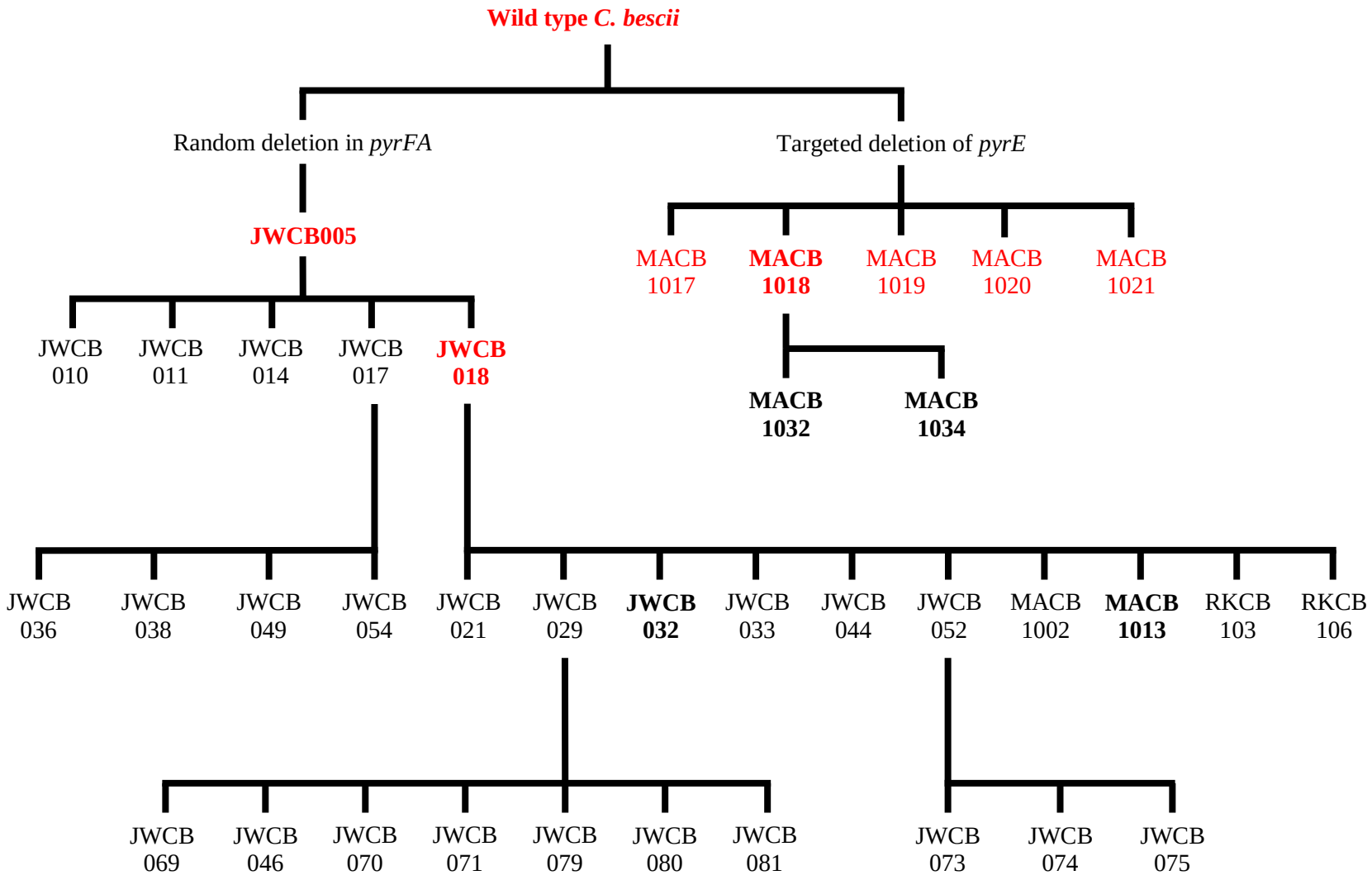
Genes were determined to be disrupted if an IS element was inserted either within the gene itself or within 100bp upstream of the start site potentially disrupting the promoter.

Strains	Gene Number	Annotation
JWCB005, JWCB018	Athe_0031	tryptophan synthase TrpB
JWCB005, JWCB018	Athe_0056	dTDP-glucose 4,6-dehydratase
JWCB005, JWCB018	Athe_0717	hypothetical protein: DUF4163, DUF3298
MACB1019	Athe_0769	cell division protein MraZ
JWCB018	Athe_1484	phosphate transport regulator
JWCB005, JWCB018	Athe_1603	UvrD/REP helicase
JWCB005, JWCB018	Athe_1665	hypothetical protein
JWCB005, JWCB018	Athe_1829	polysaccharide deacetylase
JWCB005, JWCB018	Athe_1844	adenylate/ guanylate cyclase
JWCB018	Athe_1870	tāpirin
JWCB018	Athe_1918	lactate dehydrogenase
JWCB018	Athe_2185	radical SAM domain protein
JWCB018	Athe_2186	ArsR family transcriptional regulator

Figure 3.1 Tree of genetically modified strains of *C. bescii* originating from two genetic background lineages.

Strains included in the Southern blot analyses are shown in bold text. Strains sequenced using PacBio technology are shown in red. Strain genotypes and publications in which they are described can be found in **Table 3.1**.

Figure 3.1


Wild type C. bescii
Random deletion in pyrFA
Targeted deletion of pyrE
JWCB005
MACB 1017
MACB 1018
MACB 1019
MACB 1020
MACB 1021
JWCB 010
JWCB 011
JWCB 014
JWCB 017
JWCB 018
MACB 1032
MACB 1034
JWCB 036
JWCB 038
JWCB 049
JWCB 054
JWCB 021
JWCB 029
JWCB 032
JWCB 033
JWCB 044
JWCB 052
MACB 1002
MACB 1013
RKCB 103
RKCB 106
JWCB 069
JWCB 046
JWCB 070
JWCB 071
JWCB 079
JWCB 080
JWCB 081
JWCB 073
JWCB 074
JWCB 075

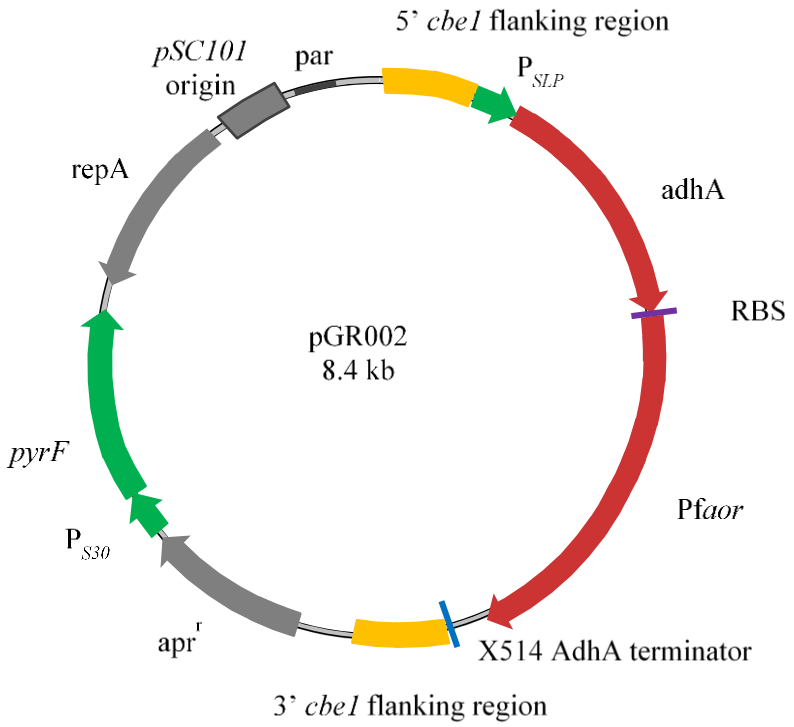


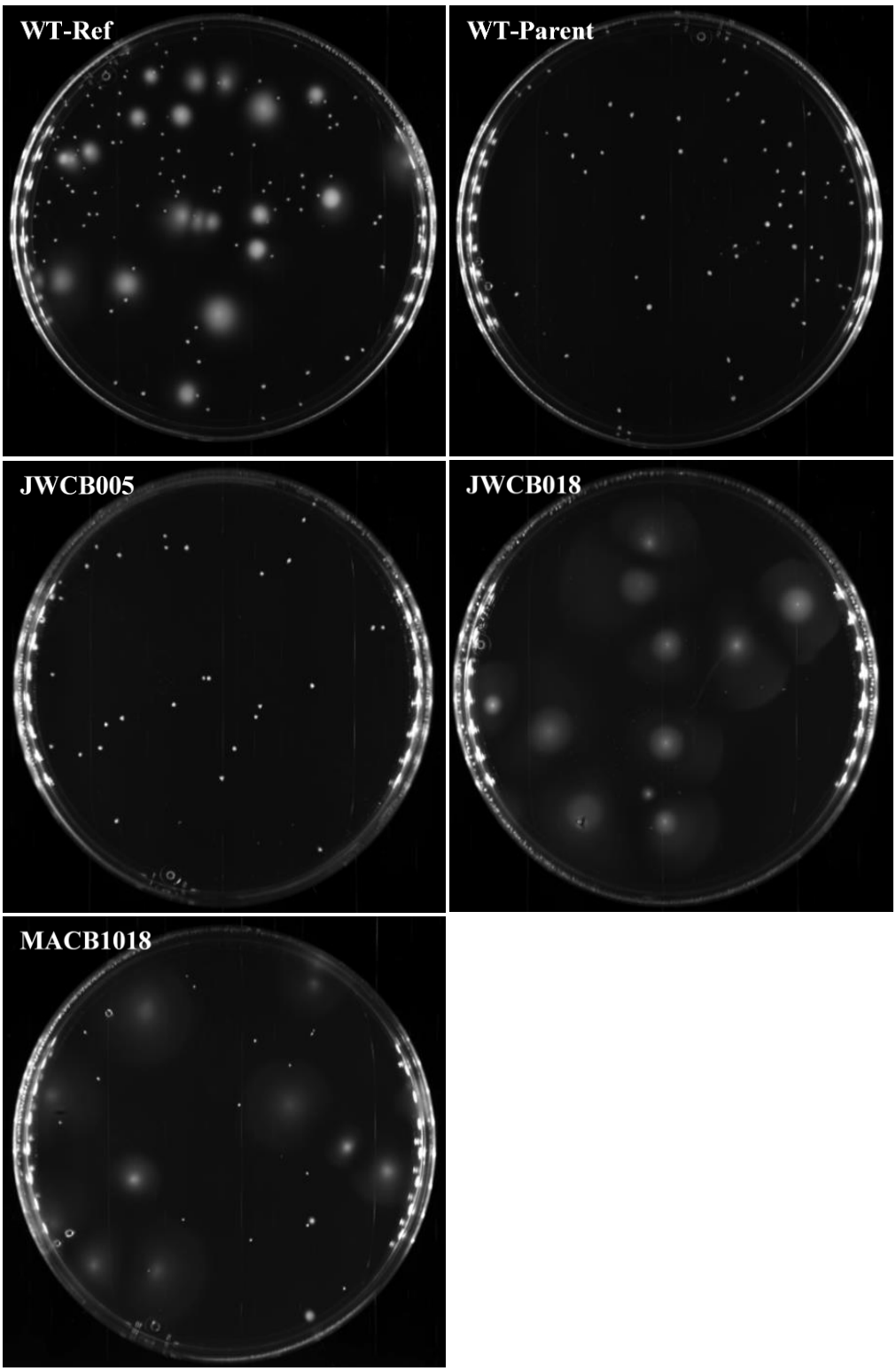
Figure 3.2 Plasmid pGR002 used in construction of MACB1013.

The plasmid is targeted for integration at the Δ*cbeI* locus utilizing uracil protrophy and 5FOA resistance as selections for plasmid integration and loss respectively.

Figure 3.3 Motility test using 0.03% bactoagar plates. Mobile cells generate diffuse colonies and non-motile cells generate discrete colonies.

WT-Ref, JWCB018, and MACB1018 are motile while WT-Parent and JWCB005 lose their motility.

Figure 3.3



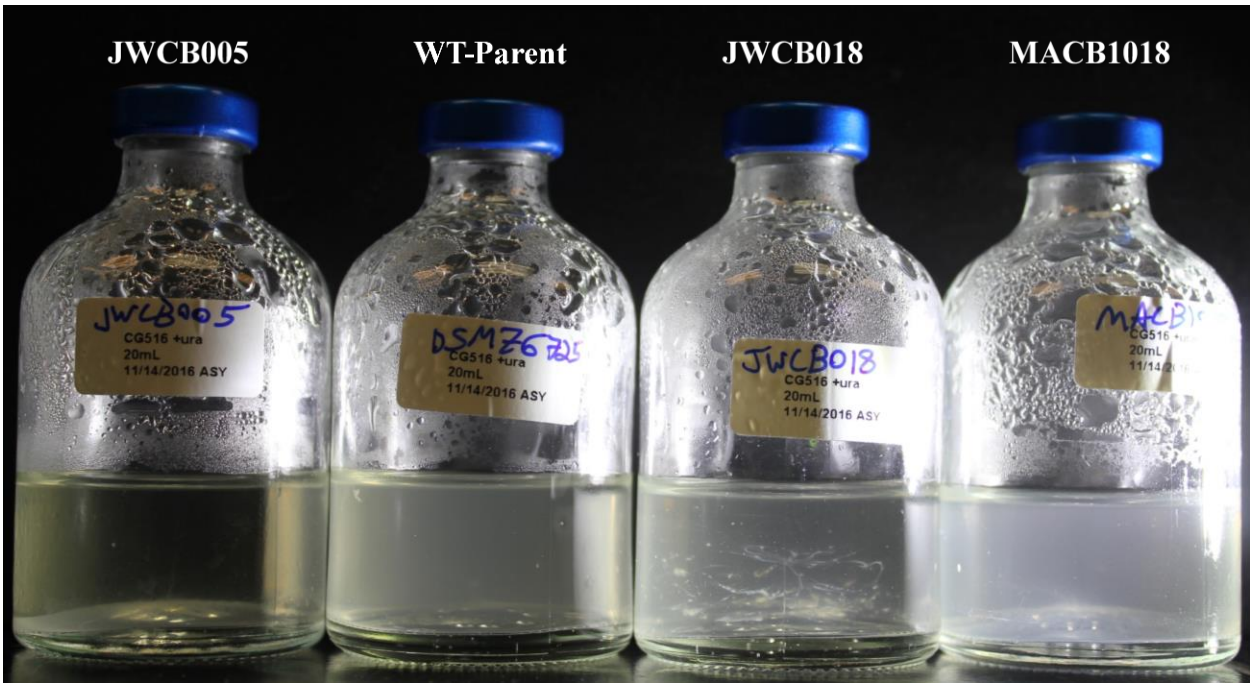


Figure 3.4 Cultures of strains JWCB005, WT-Parent, JWCB018, and MACB1018 were all started at a density of 2 x 10^6 cells/mL and allowed to grow for 36 hours of growth at 78°C with no shaking.

Cultures were gently swirled prior to being photographed. WT-Ref exhibits similar growth and phenotype to WT-Parent, JWCB005 and MACB1018 (data not shown)

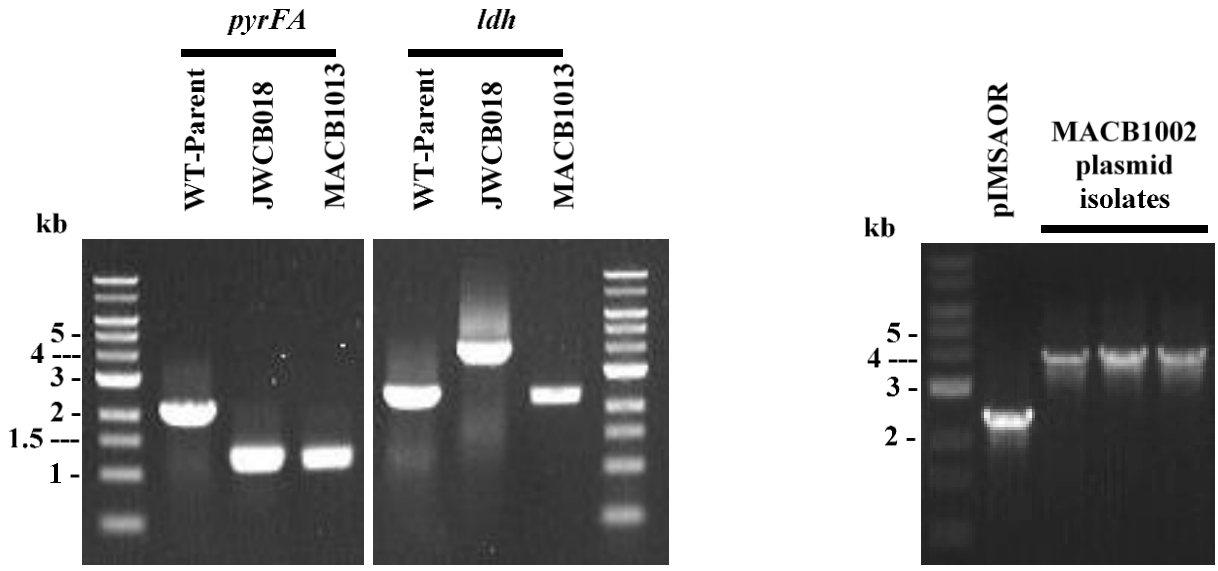


Figure 3.5 Movement of IS*Cbe4* IS elements observed in strains derived from JWCB005.

A. PCR verification of *pyrFA* and *ldh* loci of MACB1013. MACB1013 has the Δ*pyrFA* mutation matching its parent strain JWCB018 (1.1kb) and a wild-type *ldh* gene (2.3kb). **B.** PCR products showing IS*Cbe4* element insertion in plasmids originating from strain MACB1002 (4.1 kb) compared to the control plasmid used for constructing the strain pIMSAOR (2.4 kb).

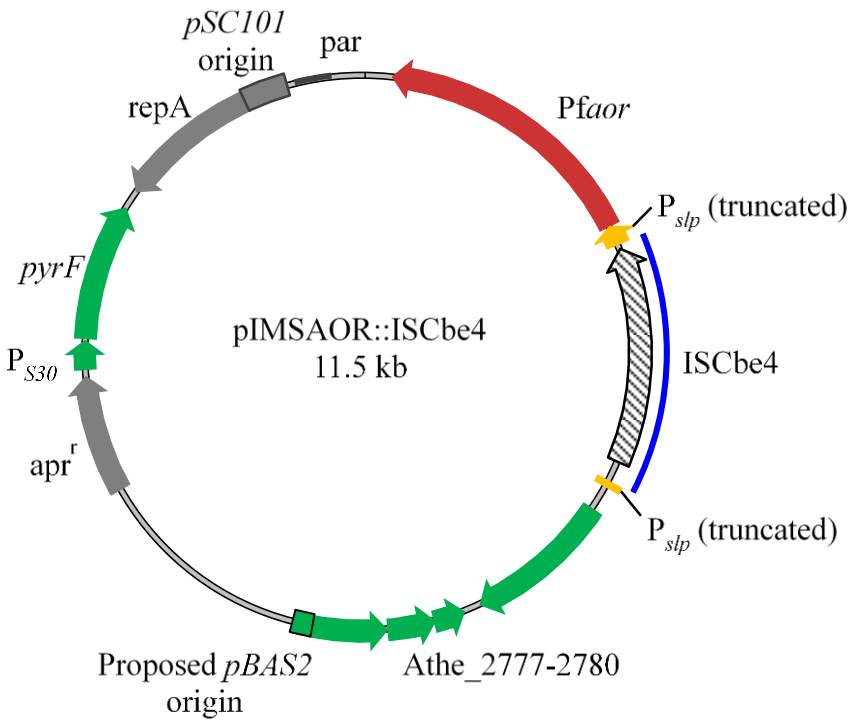


Figure 3.6 IS*Cbe4* element insertion into a replicating shuttle vector.

pIMSAOR plasmid showing insertion of IS*Cbe4* within the 5' end of P_{slp}. Annotations are as follows: *C. bescii* genes and plasmid elements (green), Pf*aor* (red), *slp* promoter (yellow), IS*Cbe4* insertion region (blue line), transposon coding region (black stripes), and *E. coli* plasmid elements (grey).

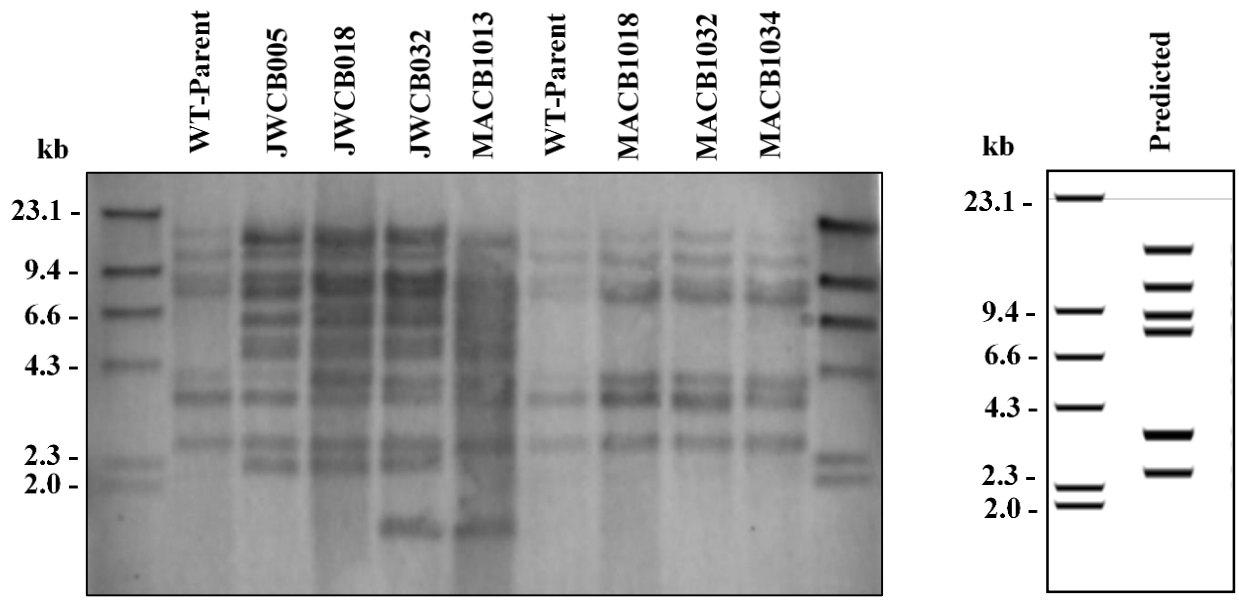


Figure 3.7 Southern blot probed for IS*Cbe4* element.

The predicted banding pattern is shown at the right with a digital blot generated using the published WT-Ref sequence. WT-Parent gives the expected banding pattern based on sequencing results. There is only one additional band observable for MACB1018, MACB1032 and MACB1034, the strains from the targeted deletion of *pyrE* lineage. JWCB005, JWCB018, JWCB032 and MACB1013 (which was derived from JWCB018) all have a dramatic increase in IS*Cbe4* elements, the exact number of which is difficult to determine from the Southern blot alone.

Figure 3.8 Southern blots probed for four other IS elements in *C. bescii*.

The same membrane as used in Figure 3 was used for all blots. Overall the banding patterns as expected by the Pacbio genome sequence was observed. Unexpected bands can be due to incomplete digestion, or mutations in a NsiI site. Additionally, quality of the blot decreases with each stripping and reprobing limiting the number of IS elements that could be probed

Figure 3.8

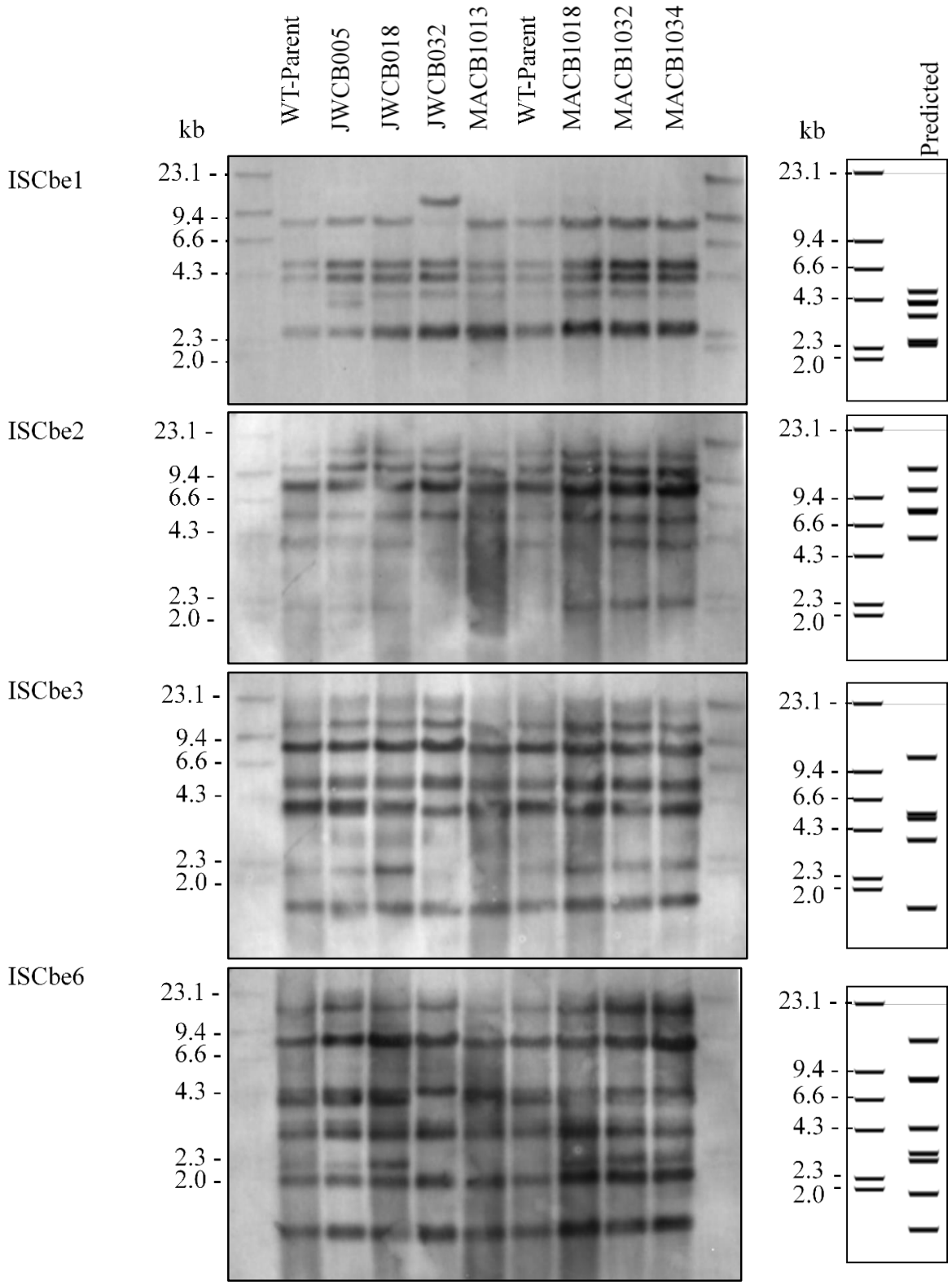
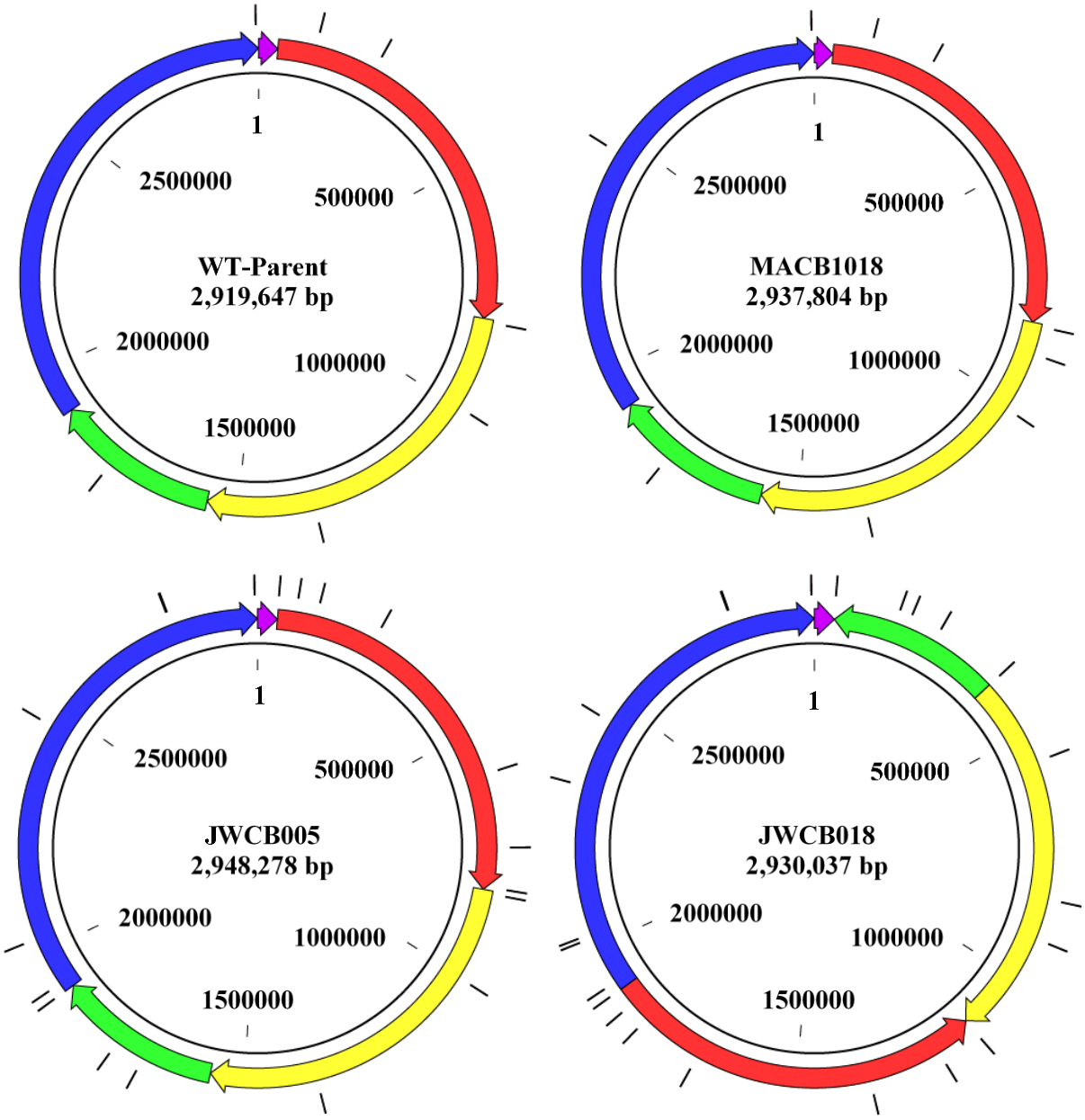

WT-Parent
JWCB005
JWCB018
JWCB032
MACB1013
WT-Parent
MACB1018
MACB1032
MACB1034
kb
Predicted
ISCbe1
ISCbe2
ISCbe3
ISCbe6
23.1
9.4
6.6
4.3
2.3
2.0

Figure 3.9 Overall genome arrangements for the WT-Parent, JWCB005, MACB1018, and JWCB018.

IS*Cbe4* elements are shown in black lines along the outside of the circular genome diagrams. There are no large scale rearrangements in JWCB005 and MACB1018 when compared to wild type. There are two large rearrangements and inversions in JWCB018 (green and red regions) when compared to wild type and its parent JWCB005.

Figure 3.9


1
500000
1000000
1500000
2000000
2500000
WT-Parent
2,919,647 bp
1
500000
1000000
1500000
2000000
2500000
MACB1018
2,937,804 bp
1
500000
1000000
1500000
2000000
2500000
JWCB005
2,948,278 bp
1
500000
1000000
1500000
2000000
2500000
JWCB018
2,930,037 bp

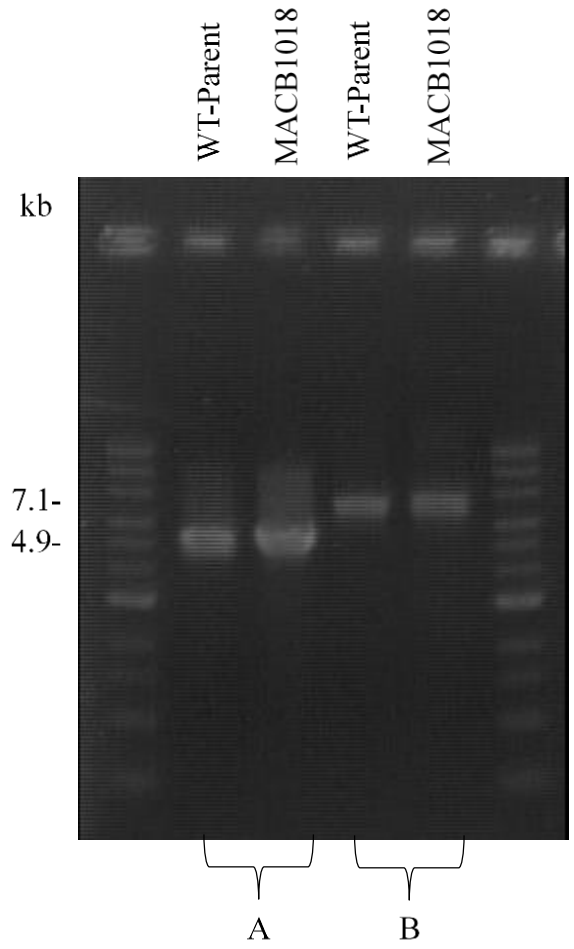
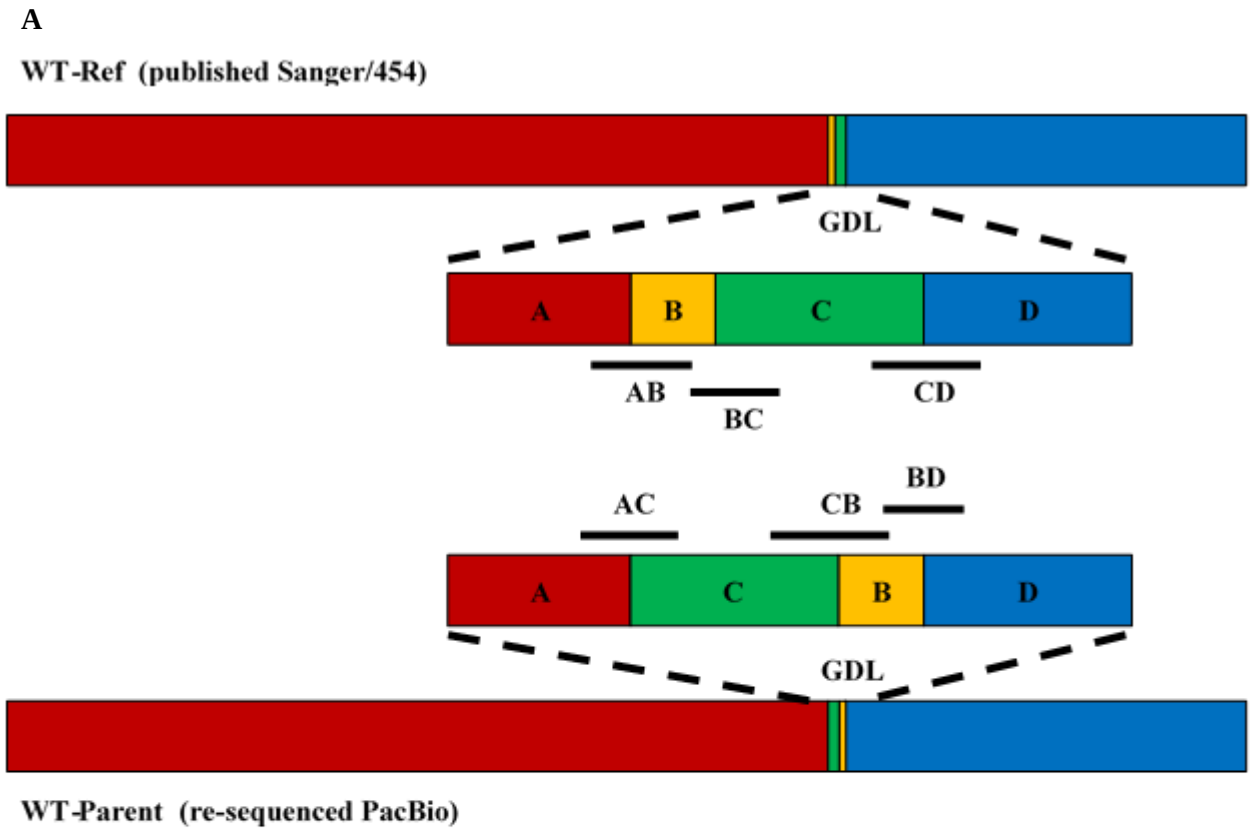


Figure 3.10 Gel for PCR results demonstrating that IS*Cbe3* is not duplicated in MACB1018. Three primers used in each reaction so that if a duplication had occurred a double band would have been observed for MACB1018. WT-Parent genomic DNA was used as a control. For product A, the expected products are 4.9kb for wild type and MACB1018 to have an additional band at 6.6kb which was not observed. For product B, the expected products are 7.1kb for wild type and MACB1018 to have an additional band at 9.0kb which was also not observed.

Figure 3.11 Differences in genome arrangement between previously published genome sequence (WT-Ref) and resequencing results (WT-Parent).

A. The overall arrangement of the wild-type genome sequenced by the two methods. There is only one major disagreement at the glucan degradation locus (GDL). The black lines at each region show the location of the PCR products used to validate the biologically relevant genome region. **B.** PCR validation of the arrangement at the GDL. Expected product sizes for the resequenced genome are: AC, 10.0kb; CB, 12.0kb; and BD, 7.9kb. Product sizes for the published genome are: AB, 10.5kb; BC, 9.0kb; and CD, 10.6kb. The expected products were attained for the orientation reported by the PacBio resequencing. For reactions targeting the orientation in the published genome, one larger than expected product was observed for reaction AB, and all other bands were faint or smaller than expected, indicative of non-specific PCR products.

Figure 3.11


A
WT-Ref (published Sanger/454)
GDL
A
B
C
D
AB
BC
CD
AC
CB
BD
A
C
B
D
GDL
WT-Parent (re-sequenced PacBio)

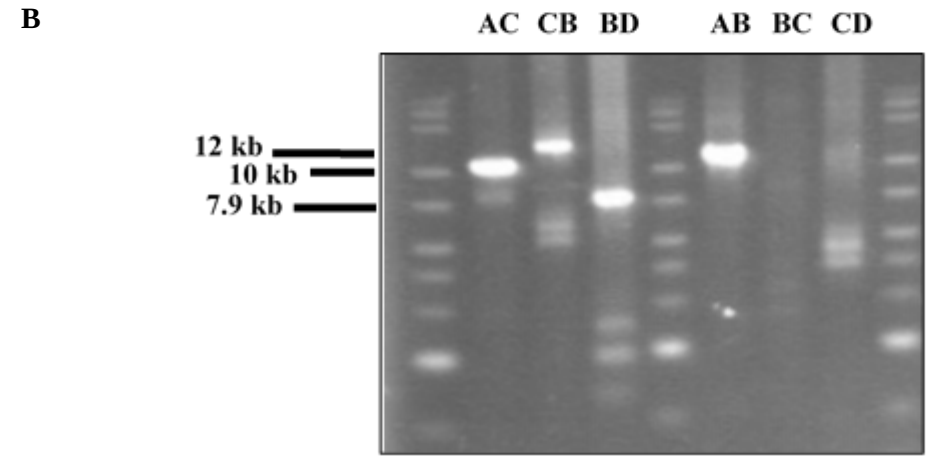
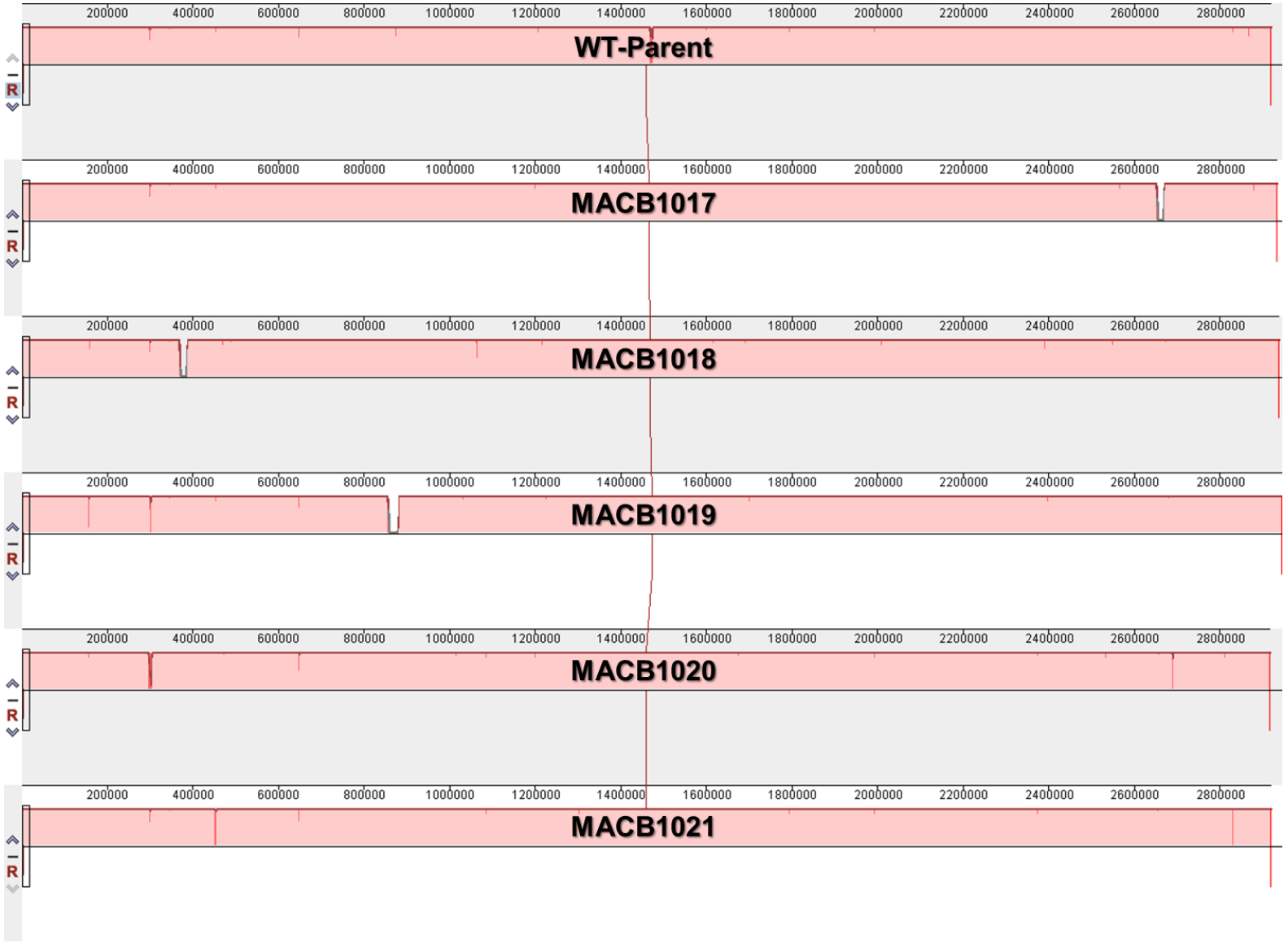

B
AC CB BD AB BC CD
12 kb
10 kb
7.9 kb

Figure 3.12 Overall genome arrangements from progressiveMauve for the WT-Parent, MACB1017, MACB1019, MACB1020 and MACB1021.

There are no large-scale rearrangements compared the WT-Parent which only varies significantly from the WT-Ref at the GDL as shown in Fig. 5.

Figure 3.12


200000
400000
600000
800000
1000000
1200000
1400000
1600000
1800000
2000000
2200000
2400000
2600000
2800000
WT-Parent
MACB1017
MACB1018
MACB1019
MACB1020
MACB1021
R

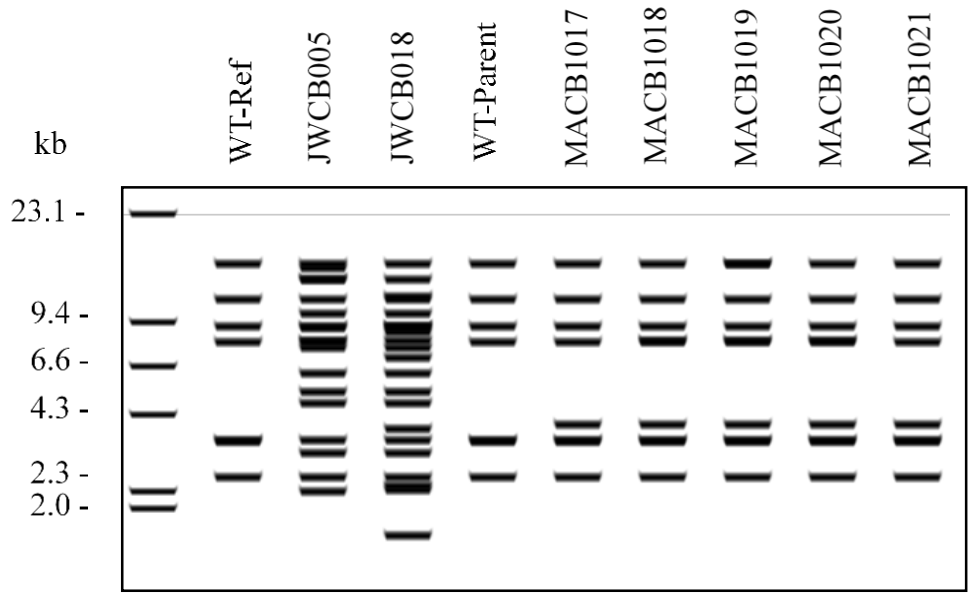


Figure 3.13 Digital Southern blot for IS*Cbe4* generated for sequenced strains of *C. bescii*.

CHAPTER 4

NATIVE XYLOSE INDUCIBLE PROMOTER EXPANDS THE GENETIC TOOLS FOR THE BIOMASS-DEGRADING, EXTREMELY THERMOPHILIC BACTERIUM *CALDICELLULOSIRUPTOR BESCII*[1]

[1]Williams-Rhaesa, A.M., Awuku, N.K., Lipscomb, G.L., Poole, F.L., Rubinstein, G.M., Conway, J.M., Kelly, R.M., and Adams, M.W.W. Submitted to *Extremophiles*, 12/04/2017.

Abstract

Regulated control of both homologous and heterologous gene expression is essential for precise genetic manipulation and metabolic engineering of target microorganisms. However, there are often no options available for inducible promoters when working with non-model microorganisms. These include extremely thermophilic, cellulolytic bacteria that are of great interest for renewable lignocellulosic conversion to biofuel and useful chemicals. In fact, improvements to the genetic systems in these organisms often ceases once transformation is achieved. This present study expands the tools available for genetically-engineering *Caldicellulosiruptor bescii,* the most thermophilic cellulose-degrader known growing up to 90°C on unpretreated plant biomass. A native xylose-inducible (P_{xi}) promoter was utilized to control the expression of the reporter gene (*ldh*) encoding lactate dehydrogenase. The P_{xi}-*ldh* construct resulted in a greater than 20-fold increase in expression of *ldh* in the presence of xylose compared to the expression level under glucose. Accordingly, lactate dehydrogenase activity was 32-fold higher in extracts of cells grown with xylose as the growth substrate instead of glucose. Finally, lactate production during growth of the recombinant *C. bescii* strain was proportional on the initial xylose concentration provided in the medium, showing tunable expression of genes are now possible using this xylose-inducible system. This study represents a major step in the use of *C. bescii* as a potential platform micoorganism for biotechnological applications using renewable biomass.

Importance

The thermophilic cellulolytic bacterium *Caldicellulosiruptor bescii* is of great interest for biotechnological applications, particularly consolidated bioprocessing for biofuel production since the organism can grow on unpretreated renewable plant biomass. The current genetics system in

this organism enables gene deletions and gene insertions using both chromosomal mutations and replicating shuttle vectors. To date, only two promoters have been demonstrated for use in *C. bescii*, the s-layer protein promoter and the the ribosomal protein S30EA promoter, which are both high expression constitutive promoters. Here we describe how a native xylose-inducible promoter can be used for the reglated control of heterologous gene expression in this organism. The ability to control gene expression considerably expands the potential for use of *C. bescii* as a platform organism for biotechnological applications.

Introduction

Caldicellulosiruptor bescii is the most thermophilic cellulose-degrading bacterium known, growing optimally at 78°C. It is capable of utilizing plant biomass as the sole carbon source without prior pretreatment (168). These attributes make *C. bescii* a candidate organism for lignocellulosic-based biotechnology via consolidated bioprocessing (CBP) (74, 119). CBP is an economically attractive alternative for generating bio-based fuels and chemicals from renewable plant biomass without the need for expensive exogenous cellulase enzymes (42, 44). However, metabolic engineering of *C. bescii* is required since the organism does not naturally produce industrial products (177). Thus, a robust and stable genetic system is required if *C. bescii* is to become a platform organism for CBP (181, 204).

The first genetic background of *C. bescii* was based on the uracil auxotrophic strain, JWCB005, which was selected by growth on 5-flouroortoric acid (5-FOA) (204). 5-FOA is toxic to uracil prototrophic strains and thus can be used for counter-selection for loss of plasmid backbones when making chromosomal manipulations in *C. bescii* (204). The majority of *C. bescii* strains have been developed in this background strain leading to increased understanding of plant biomass deconstruction, native metabolism and metabolic engineering (85, 98, 108, 183-188, 190, 194). However, recent work has shown that this genetic background is relatively unstable due to an increase in the insertion element IS*Cbe4* in this background (192). A uracil auxotrophic strain was, therefore, developed by making use of a kanamycin resistance strategy to allow for a targeted deletion of the *pyrE* gene to generate strain MACB1018 (127). IS*Cbe4* is still active in this strain, albeit at a lower copy number, making strain MACB1018 more stable when compared to the original background strain, JWCB005 (192).

While major improvements have been made in the *C. bescii* genetic system, genetic tools are still limited, especially for controlling both homologous and heterologous gene expression in this organism. To date, only two promoters have been deployed and both are high expression, constitutive promoters in *C. bescii* that are responsible for the expression of genes encoding the S-layer protein and the ribosomal protein S30EA (127). While these are useful in some cases, they do not allow for targeted control of gene expression or decreased expression of essential genes within *C. bescii*. To address this, we describe here the use of a xylose-inducible promoter in *C. bescii* based on the gene encoding lactate dehydrogenase (LDH) (Athe_1918, *ldh*) as a reporter. In the recombinant strain, the *ldh* mRNA transcript level, LDH activity and lactate production were all increased in xylose-containing media compared to glucose-containing media. Additionally, the amount of lactate produced was directly dependent on the initial xylose concentration in the medium, demonstrating titratable expression using this inducible promoter system.

Materials and Methods

Growth of *C. bescii*

Strains were grown under anaerobic conditions at 75°C. All strains used in this study are described in **Table 4.1**. For genetic manipulation, *C. bescii* competent cells were grown in the low osmolarity defined (LOD) medium, containing amino acids (LOD-AA) as described (214) with high purity argon in the headspace. Recovery, selection, and growth experiments were performed in modified DSMZ 516 medium having the following composition per liter: 1x salt solution, 1x vitamin solution, 1x trace element solution, 0.16 µM sodium tungstate, 0.25 mg resazurin, 5 g glucose (or 5 g xylose where described), 0.5 g yeast extract, 1 g cysteine hydrochloride, 1 g sodium bicarbonate and 1 mM potassium phosphate buffer, pH 7.2. The 50x stock salt solution contained the following per liter: 16.5 g NH_4Cl, 16.5 g KCl, 16.5 g $MgCl_2 \cdot 6H_2O$ and 7 g $CaCl_2 \cdot 2H_2O$. Stock

solutions of 200x vitamins and 1000x trace elements were prepared, as previously described (177). All media were filter-sterilized using a Millipore 0.22 µm filter. The complex medium containing glucose as the carbon source will be referred to as CG516 and the same medium containing xylose as the carbon source will be referred to as CX516. Both CG516 and CX516 use 20% carbon dioxide and 80% nitrogen in the culture headspace. For cultures with varying initial xylose concentration below 33.3 mM, 27.8 mM glucose was used as a supplemental carbon source, except when 16.6 mM xylose was used which had 13.9 mM glucose. All uracil auxotrophic strains were grown with 20 µM uracil in the growth medium. Kanamycin and 5-fluoroorotic acid (5-FOA) were used at concentrations of 50 µg/mL and 4 mM, respectively, when appropriate for selection (127).

The solid medium for genetic manipulation was prepared by mixing equal volumes of pre-heated 2X CG516 medium with 6% (wt/vol) agar. One hundred microliters (100 µL) of culture dilution were pipetted into an empty petri dish and the mixed medium was poured directly into the plate while swirling to distribute cells. Plates were allowed to solidify by cooling at room temperature for 20 minutes, inverted and placed into anaerobic canisters, as described (127), and flushed with 20% carbon dioxide and 80% nitrogen gas prior to incubation at 70°C for 2 days. Colonies were picked using toothpicks pressed into the agar and inoculated into 4-5 mL medium in screw-cap Hungate tubes.

Vector construction

The plasmid, pAR012, was constructed for insertion of the lactate dehydrogenase gene (Athe_1918, *ldh*) under the control of the xylose isomerase promoter (P_{xi} promoter). It was assembled from PCR products using the NEBuilder® HiFi DNA assembly (New England Biosciences). Primers used in this study can be found in **Table 4.2**. The *ldh* gene, the 199 bp upstream of Athe_0603 constituting the P_{xi} promoter region (note that *C. bescii* genes are

designated Athe), as well as the 1.2 kb upstream and 1.1 kb downstream flanking regions were amplified from *C. bescii* strain DSMZ 6725 genomic DNA. The plasmid backbone was generated by amplifying the colE1 origin of replication, the P_{S30} promoter, Cb*htk* and *pyrE* regions from pGL100 (127). All PCR products were generated using Takara PrimeSTAR Max polymerase, following manufacturer protocols. The plasmid was sequence-verified and found to have a silent mutation in the Cb*htk* gene after passage through *E. coli* changing the codon for T44 from ACA to ACG; however this did not affect kanamycin resistance in *C. bescii* strains transformed with this plasmid (data not shown).

Plasmid DNA methylation

Plasmid was methylated with recombinant M. CbeI as described (204). The *E. coli* strain, JW284, contains the M. CbeI expression plasmid, pDCW73, used for heterologous expression of M. CbeI that can be purified from cell lysate using the His-Spin Protein Miniprep kit (Zymo Research) (204).

***C. bescii* transformation and strain construction**

Competent cells of strain MACB1034 were inoculated to a density of 6-8 x 10^5 cells/mL into 500 mL of LOD-AA medium (214) and grown at 70°C to an $OD_{680}$of 0.06-0.07. 500 mL Cultures were cooled for 10-15 min in a room temperature water bath and cells were harvested by centrifugation at 6000 g for 10 min. Cell pellets were washed twice in 250 mL 10% (wt/vol) sucrose and again centrifuged at 6000 g for 10 min. Cells were gently re-suspended in 50 mL 10% sucrose and transferred to 50 mL falcon tubes, where they were once again pelleted by centrifugation at 6000 g for 10 min and again re-suspended in 1 mL of 10% sucrose. They were then transferred to a 1.5 mL microcentrifuge tube and centrifuged from 30 sec at 18,000 g. Most of the supernatant was removed, leaving 150-180 µL into which pellets were gently suspended by

pipetting. Plasmid DNA (0.5-1 µg) was added to 50 µL aliquots of competent cells. These were then transferred to 1 mm gap electroporation cuvettes. Cells were electroporated in a Bio-Rad Gene Pulser with a voltage of 1.8-2.2 kV, a resistance of 400-200 Ω and capacitance of 25 µF. Cells were immediately transferred to 20 mL CG516 medium (127), preheated to 70°C for recovery and incubated from 1-2h after electroporation. Recovery culture (5 ml) was periodically removed and transferred into CG516 medium containing 50 µg/mL kanamycin for selection. Transformants were screened, as described below, and colony-purified at least once on solid CG516 medium containing 50 µg/mL kanamycin. Counter selection for plasmid loss was performed on CG516 solid medium containing 20 µM uracil and 4 mM 5-FOA. Isolates were further purified at least once on solid CG516 medium containing 20 µM uracil. Insertion of the lactate dehydrogenase (*ldh*) gene was confirmed by sequencing of strain MACB1066.

For preliminary screening of transformants and colonies, culture PCR was used: 1 mL of liquid culture was pelleted and supernatant was removed. The cell pellet was suspended in 100 µL DNase/RNase free water (Invitrogen) and sonicated briefly to lyse cells. Cell debris was pelleted again by centrifugation at 18,000 g for 10 min. The supernatant was then used as template for PCR screening reactions by using 1 µL in a 10 µL reaction with SpeedSTAR polymerase (Takara). Genomic DNA for final strain verification was purified using the ZymoBead™ Genomic DNA Kit (Zymo Research), with a brief sonication step to improve cell lysis.

Quantitative reverse transcription PCR

RNA was extracted from 50 mL cultures of *C. bescii* strains in mid to late exponential phase by phenol:chloroform extraction, followed by precipitation with isopropanol. Genomic DNA was removed by digestion with Turbo DNase (Ambion) and subjected to an additional phenol:chloroform extraction. cDNA was synthesized from the purified RNA using the

Stratagene/Agilent Affinity Script QPCR cDNA synthesis kit. RNA was quantified using a Thermoscientific NANODROP 2000c spectrophotometer. Quantitative reverse transcription PCR, (qPCR) was performed using the Brilliant III Ultra-Fast SYBR® Green QRT-PCR Master Mix run on an Agilent MX3000p qPCR instrument and analysis was performed using the MxPro software (Agilent). Controls not subjected to RT were verified to have no amplification prior to use of cDNA for experimental quantification of mRNA. All samples were analyzed in technical duplicates.

Lactate dehydrogenase activity

Cells were grown to late exponential phase in 500 mL of both CG516 and CX516 media at 75°C. Cell pellets were collected by centrifugation at 6,000 g for 10 min, flash frozen with liquid nitrogen, and stored at -80°C prior to preparation of cell lysates. Cell lysates were prepared anaerobically in a COY chamber by suspending *C. bescii* cells in 0.5 mL of 25 mM phosphate buffer pH 7.0 with 2 mM dithionite, 1% trehalose and 1% 6-aminohaxanoc acid. Cells were disrupted by sonication for 3 x 30 sec at a maximum of 40W with 2 min between each round of sonication. Lactate dehydrogenase (LDH) activity was determined by the pyruvate-dependent oxidation of NADH. The assay was performed under anaerobic conditions in glass cuvettes at 75°C, containing 50 mM MOPS buffer, pH 7.0, and 0.5 mM NADH. The reaction was started by the addition of 5 mM pyruvate and an extinction coefficient of 6.22 $mM^{-1}cm^{-1}$ was used for NADH at 340 nm. In cases where no LDH activity was observed, the integrity of the protein extract was confirmed by assaying for pyruvate oxidoreductase (POR) activity by the pyruvate-dependent reduction of methyl viologen (MV). The assay solution contained 0.2 mM CoASH, 0.4 mM thiamine pyrophosphate, 2.0 mM $MgCl_2$, 2.0 mM MV. Anaerobic conditions were maintained by the addition of trace amounts (< 2 mM) of sodium dithionite. The reaction was started by the

addition of 10 mM pyruvate; an extinction coefficient of 13.0 $mM^{-1}cm^{-1}$ was used for reduced MV at 600 nm (241).

Lactate, glucose and xylose quantification

Lactate, glucose and xylose were measured at the end of growth by high-performance liquid chromatography (HPLC) on a 2690 separations module (Waters, Milford, MA), using a BioRad fermentation column, a photodiode array detector (model 996; Waters) and a refractive index detector (model 410; Waters). The mobile phase was 5 mM sulfuric acid at a flow rate of 0.5 mL per min. Samples were prepared by centrifugation at 18,000 g for 5 min to remove cells and particulate, followed by acidification with a final concentration of 0.1 M sulfuric acid.

Results and Discussion

Selection of the xylose isomerase promoter

Based on work in the related organism, *Caldicellulosiruptor saccharolyticus*, where the gene encoding an annotated xylose isomerase (XI) was shown to be differentially-expressed under growth of a variety of sugars (167), we determined if the homologous gene in *C. bescii* (Athe_0603) was differentially expressed using quantitative reverse transcription PCR (qPCR) under conditions where glucose or xylose was the carbon source (**Fig. 4.1**). The strains tested were the *Δldh* strain (MACB1034) and its parent strain (MACB1018). In both strains, expression under glucose conditions for the XI gene was one-tenth of the level of that encoding the glycolytic enzyme, glyceraldehyde-3-phosphate dehydrogenase (GAPDH). Conversely, under xylose conditions, expression of the XI gene was increased to levels more comparable to that of GAPDH. These results suggested that the xylose isomerase promoter (P_{xi}) could be used as an inducible promoter in *C. bescii*.

Strain generation

To test use of P_{xi} as an inducible promoter, we selected the gene (Athe_1918) encoding lactate dehydrogenase from *C. bescii* as the reporter. This gene had been previously used in *Pyrococcus furiosus* as a reporter gene for testing a cold inducible promoter (242). Additionally, lactate dehydrogenase deletion strains (*Δldh*) of *C. bescii* have been developed in multiple genetic lineages, allowing for the use of *ldh* as a reporter in a *ldh* null background (127, 183). The *Δldh* strain MACB1034 was used as the parent strain for this study, since it contains the deletion in the more stable background strain MACB1018 (127, 192). Plasmid pAR012 was constructed (**Fig. 4.2A**) containing the *colE1* origin for cloning in *E. coli*, upstream and downstream flanking regions of ~1 kb for targeting plasmid integration at the intergenic region between Athe_0949 and Athe_0950, the *Cbhtk* gene for selection of kanamycin resistance in both *E. coli* and *C. bescii*, and the *pyrE* gene for counter selection of plasmid loss from *C. bescii* (**Fig. 4.2A**). Between the flanking regions is the *C. bescii ldh* gene (Athe_1918) under the control of a 199 bp region from directly upstream of the xylose isomerase gene, Athe_0603. The plasmid, pAR012, was used to transform *C. bescii* strain MACB1034. After final strain development, MACB1066 was confirmed to encode the P_{xi} *ldh* gene construct (**Fig. 4.2B**)

Xylose-dependence of *ldh* transcription and LDH activity

In order to test the responsiveness of cells containing the P_{xi} *ldh* gene construct to xylose, we grew strains MACB1018, MACB1034 and MACB1066 (see **Table 4.1**) in media containing either glucose or xylose as the carbon source. RNA was collected from exponentially growing cells and quantitative reverse transcription PCR (qPCR) was performed. In **Figure 4.3A,** we show the change in relative expression for the XI gene and *ldh* between the same strain grown on xylose as compared to glucose. No amplification product could be detected for the *ldh* gene from strain

MACB1034, which is the Δ*ldh* parent strain (**Fig. 4.3**). All three strains showed between a 10- and 60-fold increase in XI expression under xylose conditions and only strain MACB1066 showed a similar increase, 20-fold, in the *ldh* transcript from the glucose to xylose growth conditions (**Fig. 4.3**). **Figure 4.3B** shows the expression of *ldh* relative to *gapdh* for strains MACB1018, MACB1034 and MACB1066 when grown on glucose and xylose. While *ldh* under the control of the native promoter in MACB1018 is transcribed more highly than *ldh* under the control of the P_{xi} promoter in MACB1066 (**Fig. 4.3B**), the P_{xi} promoter clearly allows for inducible transcription on xylose in MACB1066, which is not observed with the native ldh promoter in MACB1018 (**Fig. 4.3A**). Furthermore, the expression level of *ldh* relative to *gapdh* under the control of P_{xi} (**Fig. 4.3B**) is similar to the expression level of the XI gene under the control of the native P_{xi} promoter relative to *gapdh* in parent strains MACB1018 and MACB1034 under both glucose and xylose growth conditions (**Fig. 4.1**), as would be expected. These results indicate that the 199 bp selected for the XI promoter region is sufficient for regulating the expression of a downstream reporter gene, *ldh* in strain MACB1066

Having analyzed the transciptomic levels of the *ldh* reporter gene in MACB1066, the next goal was to determine if there was a measurable difference in LDH activity in this strain when it was grown on xylose, 33.3 mM, or glucose, 27.75 mM, as the carbon source. As shown in **Figure 4.4**, LDH specific activity in the cell-free extract increases 32-fold in MACB1066 when it is grown on xylose compared to glucose. There was no detectable LDH activity in the *Δldh* strain MACB1034, as expected, and the LDH specific activity for the original parent strain MACB1018 showed no significant response to changes in carbon source (**Fig. 4.4**). Interestingly, the LDH specific activity in the xylose-inducible MACB1066 strain was 3-times higher than that in the original parent strain MACB1018 (**Fig. 4.4**), even though the measured *ldh* expression levels were

significantly lower (**Fig. 4.3B**). This suggests that, while the P_{xi} promoter drives lower RNA expression levels than the native *ldh* promoter, factors in the transcribed P_{xi} RNA sequence upstream of the gene (e.g., ribosomal binding site, riboswitch, etc.) cause increased expression of LDH, at the translational level, under xylose growth conditions.

Xylose-dependent lactate production

The extent to which the production of lactate by xylose-inducible strain MACB1066 was dependent upon the initial xylose concentration in the growth medium was determined using xylose concentrations up to 33.3 mM. The original parent strain MACB1018, where its *ldh* is under the control of the native promoter, when grown on glucose produces approximately 3 mM lactate; this amount is independent of whether xylose is also present in the growth medium (**Fig. 4.5**). As expected, the *Δldh* parent strain, MACB1034, did not produce lactate when grown on glucose regardless of whether xylose was also present (data not shown). However, the MACB1066 strain did exhibit xylose-dependent lactate generation (**Fig. 4.5**). Indeed, the responsiveness to the initial xylose concentration was very specific, such that the amount of lactate produced increased almost proportionally with xylose. A very low concentration of lactate (~ 0.25 mM) was produced in the absence of added xylose, presumably due to the production of the P_{xi} inducer, which is assumed to be xylose, during the normal metabolism of *C. bescii*. This increased to ~ 5 mM lactate when 33 mM xylose was present, which is almost twice the lactate produced by the original parent strain MACB1018 in which *ldh* expression is under native control. In general, these data correlated with the approximately 3-fold higher LDH specific activity when MACB1066 was grown in the presence of 33.3 mM xylose compared to the parent strain (**Fig. 4.4**).

Under all conditions, the strains consumed similar total amounts of sugar, 12.5-16.5 mM total, with 50-56% of the total substrate remaining in excess at the end of growth (**Fig. 4.6**). For

cultures with initial xylose concentrations of 16.6 and 8 mM, 6.3-7.0 mM and 7.0-7.2 mM xylose was consumed, respectively, leaving 9-10 mM and 1 mM xylose in excess (**Fig. 4.6B**). This would explain the similar results for these two conditions with both resulting in 2.8-3.0 mM lactate for strain MACB1066 (**Fig. 4.5**). For both 4 mM and 1 mM initial xylose concentrations, no xylose was detected at the end of growth for all three strains (**Fig. 4.6B**). In cultures with both glucose and xylose present, slightly less glucose was consumed overall. However, there is no evidence of preferential consumption of either sugar, as neither was fully depleted prior to utilization of the other except when xylose concentrations were very low (i.e., 4mM and 1 mM initial xylose) (**Fig. 4.6**). This result is in agreement with the study in *C. saccharolyticus* which showed simultaneous utilization of six monosaccharides (167). The decrease in lactate production for both 16.6 and 8mM initial xylose, where xylose was still in excess, compared to 33.3 mM xylose, where no glucose was present, suggests that the presence of glucose may result in slight down regulation of P_{xi} (**Fig. 4.5** and **Fig. 4.6**). Further work is needed to characterize the mechanisms of regulation of this promoter. It is also possible that this effect is due to inefficient sugar transport, which is not well characterized for any member of this genus. However, there is evidence that for *C. saccharolyticus* glucose and xylose are transported by the same uptake systems (167, 200).

Conclusions

Here, we present a tunable xylose-inducible promoter for use in *C. bescii*. We used the lactate dehydrogenase gene, *ldh*, native to this organism as the reporter gene and demonstrated the xylose-dependent transcription of *ldh* mRNA, expression of LDH protein (as measured by specific activity), and lactate product formation in the recombinant strain where *ldh* expression is under control of P_{xi}. This demonstrated use of this promoter opens the door to inducible and regulated control of genes within this organism and greatly expands the molecular genetics tool kit available.

Future use of this promoter will allow conditional deletions to confirm gene essentiality as well as controlled inducible expression of genes for heterologous pathways in *C. bescii*. Likely candidates that will need to be tested include the bifurcating hydrogenase, shown to be the major hydrogen production enzyme in *C. bescii*, and other genes involved in glycolysis (184). Future studies to characterize the molecular mechanisms responsible for induction of this promoter will also increase our understanding of *C. bescii* and its metabolism.

Acknowledgements

This research was supported by a grant (DE-PS02-06ER64304) from the Bioenergy Science Center (BESC), Oak Ridge National Laboratory, a U.S. Department of Energy (DOE) Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

Tables and Figures

Table 4.1 Strains of *C. bescii* used and constructed in this study.

Strain	Genotype	Parent Strain	Reference
MACB1018	Δ*pyrE*	*C. bescii* DSM 6725	(127)
MACB1034	Δ*pyrE*, Δ*ldh*	MACB1018	(127)
MACB1066	Δ*pyrE*, Δ*ldh*, P_{xi}-*ldh*	MACB1034	This study

Table 4.2 Primers used in this study.

Target	Primer	Sequence
Plasmid Construction		
Plasmid backbone	AR015	ATCTAAGAGGTATGATTAAACAAAATAAAAGAGG
	GLCB086	GCATGTGAGCAAAAGGCC
5' flanking region	AR067	TTTATTGCATATTGCTGTAAGTGCGGCCGCGTTTAAACGGCGCGCCGCATGCTCACCAAACCTCCTTGTATG
	AR066	ATTTTGTTTAATCATACCTCTTAGATTGCAAGTTTAATAGTAGTTTTAATACTG
Xylose isomerase promoter (P_{xi})	NA08	AAGCCTCCTTTTTGGTAATCTATGCTTGCAAGTTGAACTGACAATCTGCC
	NA04	CAATTTTACCCGGTTTTCTCATTTTATAACTCACTCTCCTTCTATAATAAATCAG
Lactate dehydrogenase (*ldh*)	NA06	ATGAGAAAACCGGGTAAAATTGTAATTATTGG
	NA10	GAGTCAATCTCCCATTGATGTTATAGTTTTAAAGACTCTATCACACTTTTTATTACC
Stem loop terminator	AR069	CATCAATGGGAGATTGACTC
	AR068	ACTTACAGCAATATGCAATAAAGGC
3' flanking region	AR072	TTGCTGGCCTTTTGCTCACATGCTTTGAGAAAATCTAAGGTATATTGAGAG
	NA09	AGCATAGATTACCAAAAAGGAGGC
Strain Validation		
pyrE	GLCB024	AAGGCGGAATGGTCTTTGGTAAAG
	GLCB025	CGTTAATTGGTGCAAACGGTGC
ldh	GLCB003	CTGCAATTAAAGGACCCAAATACG
	GLCB004	GTTTGTGCAACCTTCTATGCC
Athe_0949 genome region	AR031	CTTGCTCTGCTGTTTTGTCTC
	AR032	GAGTATCATTTGCTGCTAATTTATATCC
qPCR		
gapdh	Athe_1406qF	GCTGCAGAAGGCGAATTAAAG
	Athe_1406qR	GGTCTGCAACTCTGTTGGAATA

ldh	AR102	GGCGTAATATGTTGCTCC
	AR103	TGAAATTGCTGCCTGGTCTC
Xylose isomerase	AR106	GGATACAGATCAGTTCCCAACAG
	AR108	CAATGTGACCAATGACCAAGTCTTC

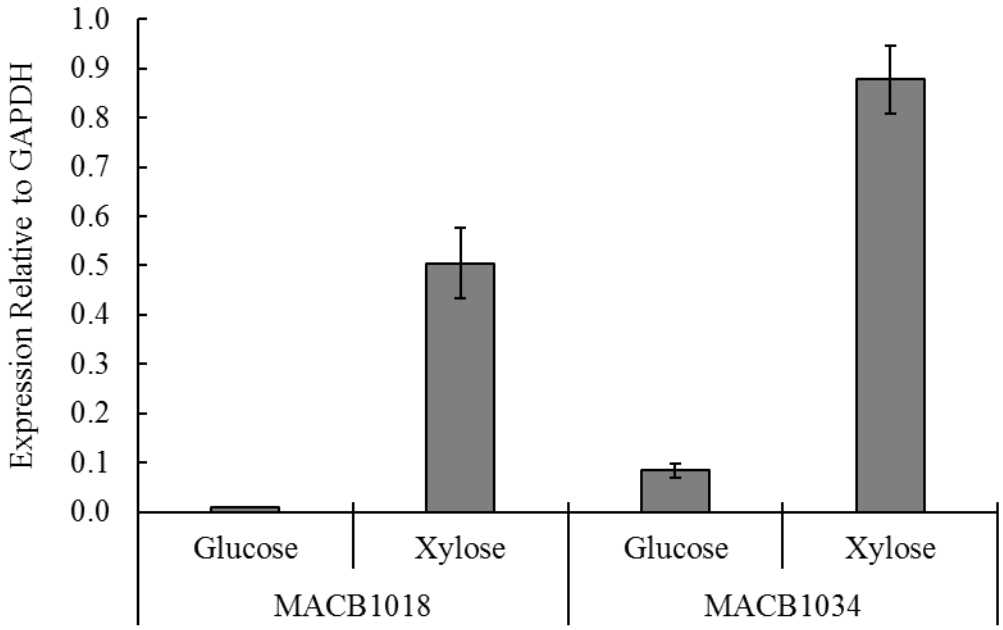


Figure 4.1 Xylose dependent expression of the native xylose isomerase gene in two genetically tractable strains of *C. bescii*.

Figure 4.2 Verification for P_{xi}-*ldh* strain.

A) Integrating vector pAR012 encoding the selection cassette (black) containing the Cb*htk* gene for selection of transformants via kanamycin resistance and the *pyrE* gene for counter-selection of plasmid loss for marker-less strain generation, elements used for cloning in *E. coli* (striped), the P_{xi} *ldh* cassette (grey), and flanking regions for targeted insertion into the genome (white). **B)** Gel electrophoresis results showing the *ldh*, *pyrE* and the P_{xi} *ldh* loci in strains MACB1018, MACB1034 and MACB1066. All strains give the same 2.4 kb product for the *pyrE* locus indicating that they are Δ*pyrE*. For the *ldh* locus, MACB1018 gives the 3.2 kb product indicating the presence of the native *ldh* gene while MACB1034 and MACB1066 both give the 2.3 kb product indicating that *ldh* is knocked out. Strain MACB1066 gives a larger, 3.9 kb, product for the Athe_0949 locus indicating insertion of the P_{xi}-*ldh* gene construct at this location in the genome.

Figure 4.2

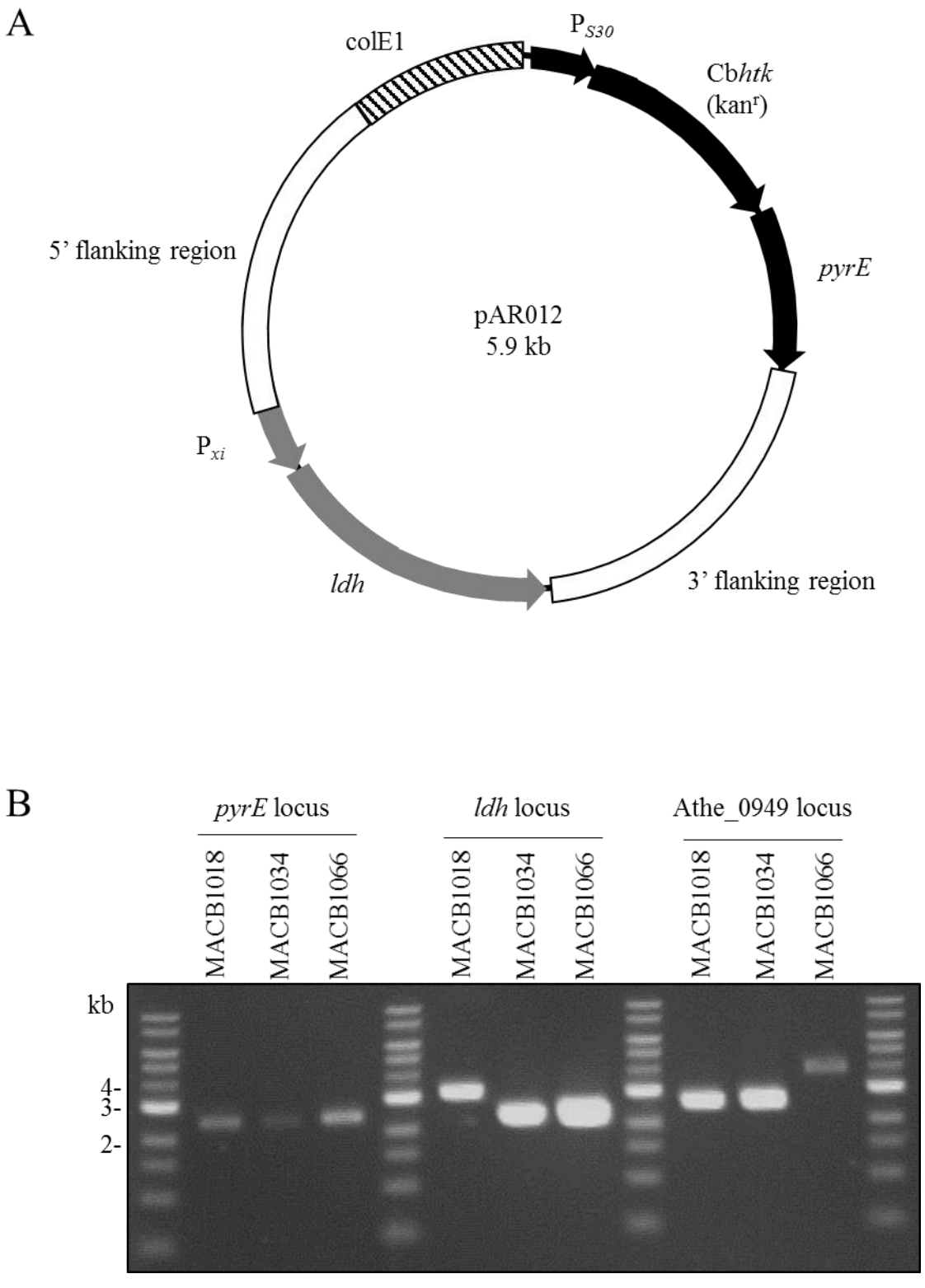
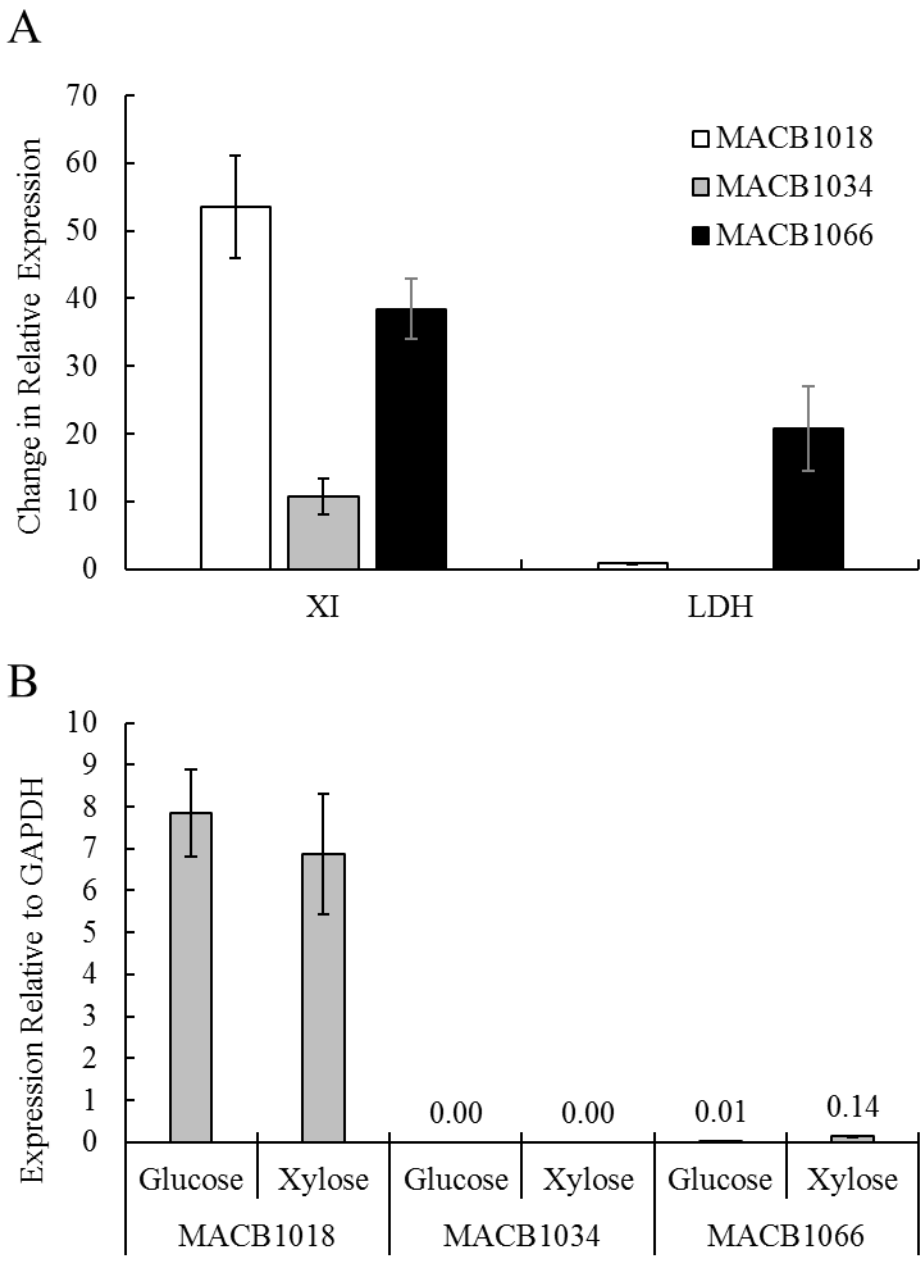

A
colE1
P_{S30}
Cb*htk*
(kan^r)
5' flanking region
pyrE
pAR012
5.9 kb
P_{xi}
ldh
3' flanking region
B
pyrE locus
ldh locus
Athe_0949 locus
MACB1018
MACB1034
MACB1066
MACB1018
MACB1034
MACB1066
MACB1018
MACB1034
MACB1066
kb
4-
3-
2-

Figure 4.3 Xylose dependence of *ldh* expression.

A) Change in relative expression of the genes encoding xylose isomerase (XI) and lactate dehydrogenase (LDH) in strains MACB1018 (white), MACB1034 (grey) and MACB1066 (black) when grown on 33.3 mM xylose as a carbon source compared to 27.8 mM glucose. **B)** Relative expression of LDH compared to GAPDH for strains MACB1018, MACB1034 and MACB1066 when grown on 27.8 mM glucose or 33.3 mM xylose.

Figure 4.3


A
Change in Relative Expression
70
60
50
40
30
20
10
0
MACB1018
MACB1034
MACB1066
XI
LDH
B
Expression Relative to GAPDH
10
9
8
7
6
5
4
3
2
1
0
0.00
0.00
0.01
0.14
Glucose
Xylose
Glucose
Xylose
Glucose
Xylose
MACB1018
MACB1034
MACB1066

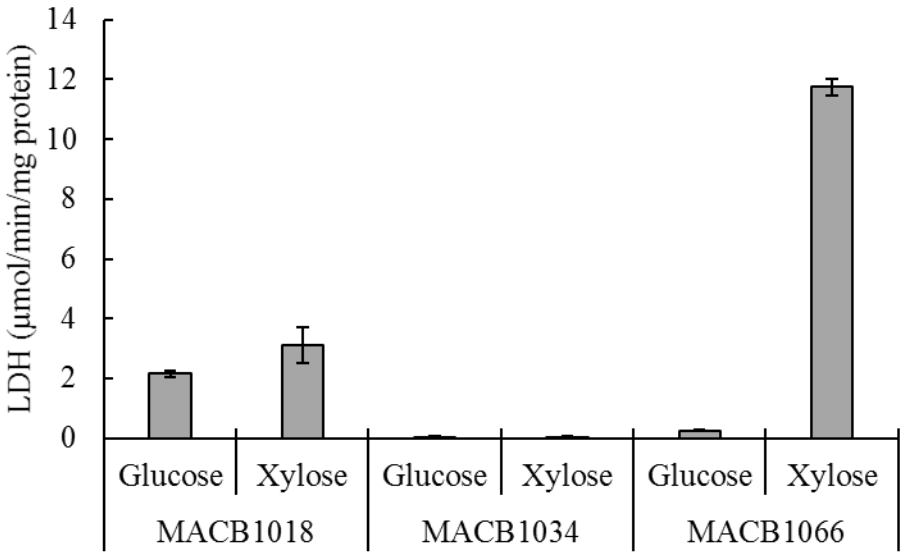


Figure 4.4 Specific activity of lactate dehydrogenase (LDH) of strains MACB1018, MACB1034 and MACB1066 grown on both 27.8 mM glucose and 33.3 mM xylose.

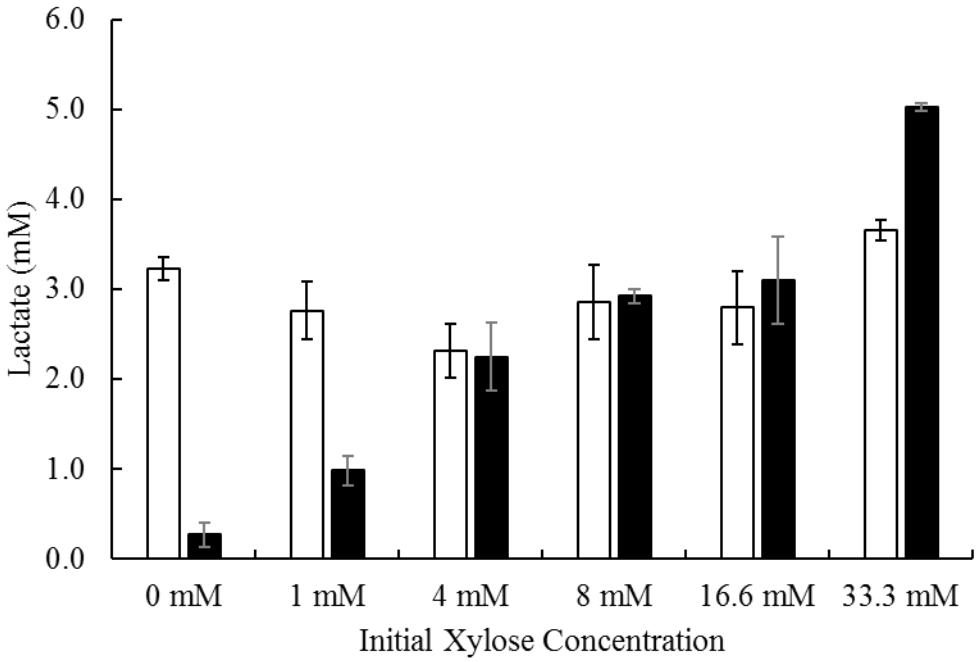


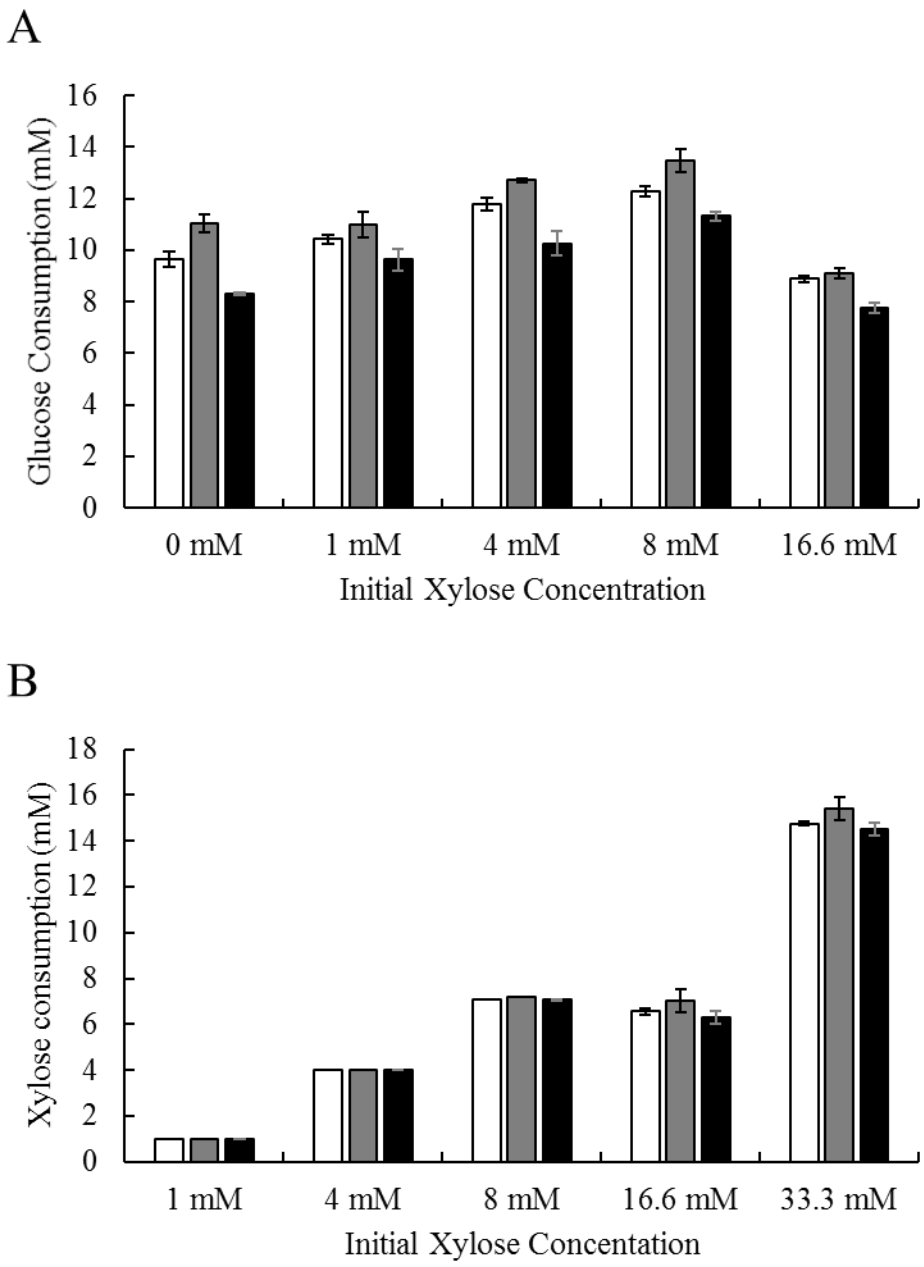
Figure 4.5 Lactate production for strains containing the natively controlled *ldh* (MACB1018, white) and the P_{xi} *ldh* (MACB1066, black).

Cultures were grown for 24 hours and samples were collected for lactate determination. Glucose was used as a supplemental carbon source for cultures with an initial xylose concentration below 33.3 mM as follows: 16.6 mM xylose cultures had an additional 13.9 mM glucose and 0-8 mM xylose cultures had 27.8 mM glucose added.

Figure 4.6 Glucose (A) and xylose (B) consumption after 24 hours by MACB1018 (white), MACB1034 (grey) and MACB1066 (black).

Glucose was used as a supplemental carbon source for cultures with an initial xylose concentration below 33.3 mM as follows: 16.6 mM xylose cultures had an additional 13.9 mM glucose and 0-8 mM xylose cultures had 27.8 mM glucose added. Sugar consumption was calculated as the difference between intial and final concentrations in the medium measured by HPLC for cultures in triplicate.

Figure 4.6


A
Glucose Consumption (mM)
16
14
12
10
8
6
4
2
0
0 mM
1 mM
4 mM
8 mM
16.6 mM
Initial Xylose Concentration
B
Xylose consumption (mM)
18
16
14
12
10
8
6
4
2
0
1 mM
4 mM
8 mM
16.6 mM
33.3 mM
Initial Xylose Concentation

CHAPTER 5

ENGINEERING REDOX-BALANCED ETHANOL PRODUCTION IN THE CELLULOLYTIC AND THERMOPHILIC BACTERIUM, *CALDICELLULOSIRUPTOR BESCII*[1]

[1]Williams-Rhaesa, A.M., Rubinstein, G.M., Scott, I.M., Lipscomb, G.L., Poole, F.L., Kelly, R.M., and Adams, M.W.W. Submitted to *Metabolic Engineering Communications*, 12/01/2017.

Abstract

Caldicellulosiruptor bescii is an extremely thermophilic cellulolytic bacterium with great potential for consolidated bioprocessing of renewable plant biomass. Since it does not natively produce ethanol, metabolic engineering is required to create strains with this capability. Previous efforts involved the heterologous expression of the gene encoding a bifunctional alcohol dehydrogenase, AdhE, which uses NADH as the electron donor to reduce acetyl-CoA to ethanol. Acetyl-CoA produced from sugar oxidation also generates reduced ferredoxin, but there is no known pathway for the transfer of electrons from reduced ferredoxin to NAD in *C. bescii*. Herein, we engineered a strain of *C. bescii* using a more stable genetic background than previously reported and heterologously-expressed *adhE* from *Clostridium thermocellum* (which grows optimally (T_{opt}) at 60°C) with and without co-expression of the membrane-bound Rnf complex from *Thermoanaerobacter* sp. X514 (T_{opt} 60°C). Rnf is an energy-conserving, reduced ferredoxin NAD oxidoreductase encoded by six genes (*rnfCDGEAB*). It was produced in a catalytically active form in *C. bescii*, the largest DNA construct to be expressed in this organism. The new genetic lineage containing AdhE resulted in increased ethanol production compared to previous reports. Ethanol production was further enhanced by the presence of Rnf, which also resulted in decreased production of pyruvate, acetoin and an uncharacterized compound as unwanted side-products. Using crystalline cellulose as the growth substrate for the Rnf-containing strain, 75 mM (3.5 g/L) ethanol was produced at 60°C, which is 5-fold higher than that reported previously. This underlines the importance of redox balancing and paves the way for achieving even higher ethanol titers in *C. bescii*.

Introduction

The need for sustainable liquid fuel production continues to grow as environmental impacts of fossil fuel use become more apparent (2). In order to overcome these consequences, use of renewable resources must increase. Lignocellulose is an attractive feedstock for biofuel production as it is highly abundant, does not compete with food supplies, and can be generated in sustainable ways (69). However, there are a number of challenges to using lignocellulose for biofuel production, most significantly the recalcitrance of plant biomass to deconstruction and the costs associated with pretreatment and the enzymatic hydrolysis of the plant wall polysaccharides (40). One potential way to minimize these costs is the use of consolidated bioprocessing (CBP) in which enzyme production, polysaccharide hydrolysis and sugar fermentation are all performed by the same microorganism in the same reaction vessel (42, 44).

The extremely thermophilic bacterium, *Caldicellulosiruptor bescii*, is of great interest for use as a CBP microbe as it has the highest growth temperature of any known cellulolytic bacterium (168). Moreover, it is capable of robust growth on a variety of carbon sources, including mono- and polysaccharides of both hexose and pentose sugars, as well as crystalline cellulose and plant biomass without pretreatment even at high solids loadings (105, 168, 179). *C. bescii* ferments these substrates to acetate, lactate, carbon dioxide and hydrogen (168). *C. bescii* is also amenable to metabolic engineering as a genetic system is available and has been used to construct strains with multiple gene deletions (183) and for heterologous expression of genes both from replicating plasmids (182, 190) and chromosomal integrations (85, 127, 185, 192). These studies have resulted in an increased understanding of the primary metabolism of *C. bescii* and its mechanisms of plant biomass utilization (187, 194). While the original genetic system has been useful, genome instability of the resulting recombinant strains was observed due to an active transposon, IS*Cbe4*

(192). A genetically-tractable strain generated through the use of a kanamycin resistance gene was therefore constructed and has been shown to be much more stable (127, 192).

In a previous study, the gene (*adhE*) encoding the bifunctional alcohol dehydrogenase from the bacterium *Clostridium thermocellum*, which grows optimally (T_{opt}) at 60°C (138), was heterogously-expressed in *C. bescii* using the original unstable genetic background (185). This enzyme reduces acetyl-CoA to ethanol (**Fig 1**). The resulting recombinant strain generated 14.0 mM and 12.8 mM ethanol, respectively, from crystalline cellulose and switchgrass at 60°C (185). The *adhE* and *adhB* (NADPH-dependent bifunctional alcohol dehydrogenase) from *Thermoanaerobacter pseudethanolicus* 39E (T_{opt} 65°C) were also expressed in the unstable lineage of *C. bescii* but this resulted in the production of only 2.3 mM and 1.4 mM, respectively, from crystalline cellulose at 75°C (186). These titers are well below levels required for the industrial scale use of *C. bescii*, which is 900 mM (40 g/L) (8).

The AdhE enzyme requires NADH as the reductant for ethanol production and while the glycolytic pathway of *C. bescii* does generate NADH, the conversion of pyruvate to acetyl-CoA, the substrate for AdhE, produces reduced ferredoxin (176) (**Fig 5.1**). One potential strategy to improve ethanol production in *C. bescii* is to enable the transfer of electrons from reduced ferredoxin to NADH. An enzyme with this capability commonly found in ethanol-generating thermophiles is the energy-conserving reduced ferredoxin NAD oxidoreductase (Rnf). This enzyme couples electron transfer from reduced ferredoxin to NAD to the generation of an ion gradient across the plasma membrane. However, *C. bescii* does not contain the six genes (*rnfCDGEAB)* that encode the Rnf complex (165, 176, 177, 243).

In this study, we compared ethanol production by *C. bescii* heterologously-expressing *C. thermocellum adhE* in the original strain and in the more genetically-stable background strain. In

addition, we attempted to further modify these strains by the heterologous expression of the six genes encoding the Rnf complex from *Thermoanaerobacter* X514 (T_{opt} 60°C: (169) to examine if redox balancing affected ethanol production. It should be noted that only single genes encoding cytoplasmic proteins have been heterologously-expressed previously in this organism and so successful expression of the six genes encoding the membrane-bound Rnf was a significant accomplishment. The results here revealed that the genetic background, the presence of Rnf and growth conditions all influenced carbon and electron flow to ethanol in *C. bescii* and, in combination, a more than five-fold improvement in ethanol production from crystalline cellulose was achieved compared to the previous report (185). As such, this paves the way for more sophisticated metabolic engineering strategies with this unique microorganism and illustrates its significant potential for CBP of renewable biomass.

Materials and Methods

Culture conditions

C. bescii was grown under anaerobic conditions in modified DSM 516 medium (C516) as described previously (127), with 5 g/L growth substrate at 60°C unless otherwise noted. For competent cell preparation, cultures were grown at 75°C in low osmolarity defined medium with amino acids (LOD-AA) under argon, as described (204, 214). Recovery and selection were performed using the C516 medium under a headspace of 80% (v/v) nitrogen and 20% (v/v) carbon dioxide (127). Uracil auxotrophic strains were grown in the presence of 20 µM uracil. Kanamycin and 5-fluoroorotic acid (5-FOA) were used at concentrations of 50 µg/mL and 4 mM, respectively (127). A solid growth medium was used during strain purification that was generated from C516 medium, as described (127) for growth at 70°C over 36-48 h. Glucose, xylose, beechwood xylan,

and Avicel PH-101 were obtained from Sigma. Switchgrass was sieved 20/80-mesh fraction (provided by Dr. Brian Davison, Oak Ridge National Laboratory, Oak Ridge, TN).

Fermentation cultures were grown at 500 mL scale using the DASGIP parallel bioreactor system with custom heating jackets to maintain growth temperatures between 59-61°C. Modified DSM 516 medium was also used with 20 g/L cellulose (Avicel) as the growth substrate. The culture was sparged at 2 L/h with 80% (v/v) nitrogen and 20% (v/v) carbon dioxide and stirred at 150 rpm. pH was maintained at 7.0 (25°C) through the addition of 10% sodium bicarbonate.

Strain and plasmid construction

Table 5.1 contains a list of the strains generated in this study. Plasmids were constructed from PCR products using NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs). PCR was performed using Takara PrimeSTAR Max DNA polymerase (Takara R045A). Genomic DNA, plasmid DNA, and PCR products were purified with kits from Zymo Research and Stratagene. The templates for PCR were either previously sequenced plasmids or genomic DNA from *C. bescii*, *C. thermocellum*, or *T.* sp. X514. All primers used in this study can be found in **Table 5.2** and plasmids constructed for this study can be found in **Figure 5.2**. Plasmids were cloned into NEB 10-beta competent *E. coli* (New England BioLabs C3019I), followed by Sanger sequencing through GeneWiz. Plasmids for insertion into *C. bescii* strains, still expressing the *cbeI* restriction enzymes, were methylated prior to transformation, as described (204). Methylation was verified by protection from digestion with *HaeIII* restriction enzyme (New England BioLabs).

Competent *C. bescii* cells were generated, as described (127). These were transformed by electroporation with Bio-Rad Gene Pulser with a voltage of 1.8-2.2 kV a resistance of 400-200 Ω and capacitance of 25 µF. Cells were transferred to 70°C recovery medium immediately after electroporation and 5mL samples transferred to selective medium, containing kanamycin, at

intervals between 0-2 h of recovery and incubated at 75°C until growth was observed, usually between 24-72 h. Transformants were screened and colony purified from solid medium at least once prior to counter selection for plasmid loss. Counter selection was performed on solid medium containing 5-FOA and 20 µM uracil. Final strains were confirmed by sequencing of the Athe_0949 locus (GeneWiz).

Quantitative reverse transcriptase polymerase chain reaction (qPCR)

RNA was harvested from 50 mL cultures of *C. bescii* strains grown to mid-exponential phase by phenol:chloroform extraction, followed by precipitation with isopropanol. Turbo DNA-*free* Kit (Ambion AM1907) was used to remove genomic DNA and RNA was subjected to a second round of phenol:chloroform extraction followed by isopropanol precipitation. cDNA was synthesized using the Agilent Affinity Script QPCR cDNA synthesis kit (Agilent #600559). RNA was quantified using a Thermoscientific NANODROP 2000c spectrophotometer. Quantitative reverse transcription PCR, (qPCR) was performed using the Brilliant III Ultra-Fast SYBR® Green QRT-PCR Master Mix (Agilent #600882) run on an Agilent MX3000p qPCR instrument and analysis was performed using the MxPro software (Agilent). Controls without reverse transcriptase were verified to have no amplification prior to use of cDNA for experimental quantification. Primers can be found in **Table 5.2**.

Enzymatic activity assays

Whole cell extracts were prepared from mid-exponential cultures of *C. bescii* strains grown on C516 media, with glucose as the carbon source either in 50 mL bottle cultures or in a 20 L fermentor, as described above. Cells were harvested by centrifugation at 6,000 g for 10 minutes or by continuous centrifugation, and immediately frozen in liquid nitrogen and stored at 80°C. Extracts were prepared anaerobically in a Coy chamber. Whole cell extracts were prepared from

50 mL cultures by suspending cells in 100 mM phosphate buffer pH 7.0 with 5 µM ferrous sulfate and 2 mM dithionite, as described (244). Cells were disrupted by sonication for 3 x 10 sec at a maximum of 40W with 1 min between each round of sonication. Membrane fractions were generated by suspension of 5 g of cells from exponentially growing 20 L fermentors in 50 mM MOPS buffer pH 7.0 with 2 mM dithionite. Cells were lysed using 1 mg/mL lysozyme and sonication. Membrane and cytoplasmic fractions were separated by ultra-centrifugation at 100,000 g for 1 h, followed by two washes with an additional volume of buffer and an additional 1 h of ultra-centrifugation. Membranes were suspended in buffer by sonication before and after each wash.

Both steps of the AdhE reaction were measured with aldehyde dehydrogenase (ALDH) representing the acetyl-CoA dependent reaction and alcohol dehydrogenase (ADH) representing the acetaldehyde dependent reaction. These were measured at 60°C in anaerobic cuvettes containing the following: 2 mL 50 mM MOPS buffer pH 7.0, 5 µM ferrous sulfate, and 0.5 mM NADH or NADPH. Reactions were started by the addition of either 0.35 mM acetyl-CoA (ALDH activity) or 10 mM acetaldehyde (ADH activity); an extinction coefficient of 6.22 $mM^{-1}cm^{-1}$ was used for NADH at 340 nm. Rnf activity was measured in membrane fractions of wild-type, MACB1052 and MACB1062 strains. This activity was measured at 60°C in anaerobic cuvettes containing, 2 mL of 50 mM MOPS buffer pH 7.0, and 1 mM benzyl viologen (BV). Reactions were started by the addition of 1 mM NADH or NADPH; an extinction coefficient of 7.4 $mM^{-1}cm^{-1}$ was used for reduced BV at 600 nm. Hydrogen uptake activity was measured for all membrane fractions to confirm similar activities in all strains. This was measured through the reduction of BV at 75°C in anaerobic cuvettes, which were flushed for 10 min with 100% hydrogen. Reactions were started by the addition of membrane fractions and absorbance at 600 nm was measured.

Cytoplasmic contamination was determined through pyruvate oxidoreductase (POR) activity, measured by pyruvate-dependent reduction of methyl viologen (MV). POR activity was measured anaerobically at 75°C in cuvettes containing: 2 mL EPPS pH 8.4, 0.2 mM CoASH, 0.4 mM thiamine pyrophosphate, 2.0 mM $MgCl_2$, 1.0 mM MV. The reaction was started by the addition of 10 mM pyruvate, and an extinction coefficient of 13.0 $mM^{-1}cm^{-1}$ was used for reduced MV at 600 nm. For all assays, anaerobic conditions were maintained by the addition of trace amounts (< 2 mM) of sodium dithionite. In all cases, activity was calculated as the difference in rate after the addition of substrate and specific activity is presented at µmol/min/mg of protein.

Fermentation analysis

Acetate, glucose, cellobiose, and pyruvate were measured in culture supernatant by high performance liquid chromatography (HPLC) on a 2690 separations module (Waters), equipped with using a Bio-Rad fermentation monitoring column (Bio-Rad #125-0115), a photodiode array detector (model 996; Waters) and a refractive index detector (model 410; Waters). A mobile phase of 5 mM sulfuric acid at a flow rate of 0.5 mL per min was used and concentrations were determined by comparison to standards. Culture supernatant was acidified to a final concentration of 0.1 M sulfuric acid. Ethanol and acetoin were quantified using an Agilent 7890A Gas-Chromatography instrument, equipped with a Carbowax/20 m column and an FID detector. Culture supernatant was acidified prior to analysis to a final concentration of 100 mM maleic acid prior to injection. Residual Avicel from 500 mL fermentations was measured as described (168). Cell densities were monitored using a Petroff-Hauser counting chamber.

Carbon balances were determined as described (105). Briefly, carbon consumed was determined by the glucose equivalents utilized given the measured residual Avicel, carbon recovered was the sum of the measured end products, acetate, ethanol, and pyruvate, as well as

glucose remaining in the medium. CO_2 production was calculated based on one molecule released for every molecule of acetate or ethanol produced. Biomass was calculated based on protein estimation by the Bradford method compared to a BSA standard (Bio-Rad).

Results

Construction and biochemical properties of the ethanologenic strains

The first goal was to determine if the two genetic lineages of *C. bescii* that are available for metabolic engineering showed any differences in their ability to produce ethanol by the heterologous expression of *adhE*. The second goal was to determine if in either lineage ethanol production could be enhanced by the co-expression of Rnf (ferredoxin NAD oxidoreductase). Rnf generates NADH from the reduced ferredoxin produced during sugar fermentation for the NADH-dependent production of ethanol. The two available genetic lineages of *C. bescii* are the original strains derived from JWCB005 (127, 204) and the new strain MACB1018 (127, 204). The latter lineage has been shown to be much more stable than the former in terms of overall genome arrangement and insertion elements, particularly involving the active IS*Cbe4* element (192).

Strains were generated in both lineages to investigate the effects of genome stability and co-expression of Rnf on ethanol production (**Table 5.1**). Strain JWCB032 has been characterized in a previous study (185) and it heterologously expresses the *adhE* from *C. thermocellum* (Ct*adhE*), which is the major enzyme responsible for ethanol production in this organism (245). JWCB032 was also used as the parent strain for the insertion of the *rnfCDGEAB* genes encoding Rnf from *Thermoanerobacter* sp. X514 (Tx*rnf*), to give MACB1052*, where the asterisk indicates that it is derived from a JWCB strain. In the MACB1018 background, the *Δldh* strain, MACB1034, was selected to act as the parent strain and was modified in two steps: first the addition of the Ct*adhE* generating strain MACB1058, followed by the insertion of the Tx*rnf* genes at the same

locus to give MACB1062 (**Fig. 5.3**). Note that all of the strains under study lack lactate dehydrogenase (*ldh*; **Table 5.1**) and therefore do not produce lactate as an end product.

In both genetic backgrounds, the Ct*adhE* gene was under the control of the promoter of the high level constitutively-expressed gene encoding the S-layer protein promoter. The Tx*rnf* genes were expressed by the promoter element upstream of the Athe_1295, which is responsible for the expression of the bifurcating hydrogenase (Athe_1295-1299). This enzyme is responsible for most of the H_2 production during fermentative growth of *C. bescii* (**Figs. 5.3**) (184). Expression of the Ct*adhE* and Tx*rnfE* genes was confirmed by qPCR relative to the highly-expressed constitutive *gapdh*, which encodes the glycolytic enzyme glyceraldehyde-3-phosphate dehydrogenase. Surprisingly, the expression of the Ct*adhE* gene was dramatically different in the two background lineages, with expression in the JWCB032 lineage strains approximately 4% of the expression of *gapdh,* while in the strains in the MACB1058 lineage expression was approximately 3-fold greater than that of *gapdh* (**Fig. 5.4A**). It should be noted that the genome location of the Ct*adhE* in these strains is not the same (see **Discussion**) and this may play a role in the observed differences in expression, despite the fact that both genes were under control of the same promoter. In the two new strains that also encoded the Tx*rnf* operon, Tx*rnfE* expression was similar to *gapdh* expression in MACB1052*, with expression in MACB1058 approximately half of that of *gapdh* (**Fig. 5.4B**).

Figure 5.5 shows the specific enzyme activities of bidirectional AdhE and Rnf for strains JWCB032, MACB1052*, MACB1058 and MACB1062. AdhE activities were measured in the physiologically-relevant forward direction using either acetyl-CoA or acetaldehyde reduction by NADH. These are referred to as ALDH and ADH activities, respectively. For all strains, the measured ALDH and ADH activities were comparable (**Fig. 5.5A**) and they were also specific to NADH as no activity was observed when NADPH replaced NADH (data not shown). In accord

with the measured expression levels of Ct*adhE* in JWCB032 and MACB1052*, both had much lower specific activities in cell-free extracts (0.46-0.83 µmol/min/mg) when compared to strains from the more recent lineage, MACB1058 (3.16-3.65 µmol/min/mg) and MACB1062 (1.69-2.01 µmol/min/mg) (**Fig. 5.5A**).

Rnf is a membrane-bound enzyme and its activity was measured by NADH-dependent reduction of the dye benzyl viologen (BV) in washed membrane fractions. This eliminated contaminating BV-associated activities in the cytoplasm, e.g., from the bifurcating hydrogenase (**Fig. 5.1**). As controls, the activities of the cytoplasmic enzyme pyruvate ferredoxin oxidoreductase (POR) and of the membrane-bound hydrogenase were measured, and these confirmed that cytoplasmic contamination of the membrane fraction was below 10% (**Table 5.3**). The *C. bescii* genome does not contain homologs of the Rnf genes and accordingly membranes prepared from wild type cells had no detectable Rnf activity (< 0.1 µmol/min/mg; **Table 5.3**). However, Rnf activity was readily detected in the membranes of both MACB1052* (7.18 µmol/min/mg) and MACB1062 (1.14 µmol/min/mg) (**Fig. 5.5B**). Once more, these activity results are in general agreement with transcriptional data where the expression of *rnfE* in MACB1052* was more than twice that measured in MACB1062.

Ethanol production by the new C. bescii strains

Having confirmed that both AdhE and Rnf are active in both genetic backgrounds of *C. bescii*, the ethanol production of all four strains was investigated. These strains contained AdhE in the original genetic background strain with (MACB1052*) and without (JWCB032) co-expression of Rnf, and in the new genetic background with (MACB1062) and without (MACB1058) co-expression of Rnf. The strains were initially grown in closed bottles on glucose at 60°C for 48 hr (**Fig. 5.6**). As expected, under these conditions, no ethanol was detected from either parent strain,

JWCB018 (old lineage) and MACB1034 (new lineage, data not shown), while ethanol was detected in all of other strains. As shown in **Fig. 5.6,** strains in the MACB1034 background produced almost twice as much ethanol as strains in the JWCB018 background (**Fig. 5.6**). It is likely that this increase is due to the higher expression level of Ct*adhE* and the corresponding increased AdhE activity in the MACB1034 background strains. However, under these growth conditions, co-expression of Rnf did not significantly affect ethanol production. The two strains based on the new lineage, MACB1058 (with AdhE) and MACB1062 (with AdhE and Rnf), which had superior ethanol production, were further characterized in closed cultures.

The two new MACB strains were grown between 55-75°C at 5°C increments to determine the optimal temperature for ethanol production. The highest ethanol concentration was produced was between 55 and 65°C for both strains (**Fig. 5.7A**). No ethanol production was observed for the parent strain, MACB1034, as expected, and none of these strains produced ethanol at 75°C (data not shown). This suggests that AdhE is inactive at this temperature, consistent with the much lower growth temperature optima of *C. thermocellum* (60°C) as well as a previous characterization of the Ct*adhE* expressed in *C. bescii* (original lineage: (185)). Acetate production for strains MACB1058 and MACB1062 declined proportionally with increasing ethanol production, while slightly more acetate was produced by the parent strain MACB1034 with increasing growth temperature (**Fig. 5.7B**) presumably as a result of increased growth (the optimum is 78°C).

Surprisingly, the C_4 compound acetoin was produced by all three strains at the lower temperature, being more pronounced in the ethanol production strains (~ 3 mM was produced) compared to the parent strain (~0.5 mM; **Fig. 5.7C**). Acetoin production was slightly lower in strain MACB1062 expressing the Tx*rnf* compared to its parent strain, MACB1058, lacking Rnf (**Fig. 5.7C**). Pyruvate (C3) was also detected in the growth medium of all three strains, although

at lower concentrations than acetoin (**Fig. 5.8**). There was a slight increase in pyruvate produced at temperatures below 65°C, from less than 0.5 mM at 70-75°C to between 0.5-1 mM at the lower temperatures. Differences between strains were not statistically significant. As expected, due to the overall increase in the amount of products generated by the two ethanol producing strains, they consumed more glucose (20-25 mM) compared to the parent (~10 mM; **Fig. 5.7D**). At 75°C, where no ethanol was produced, all strains consumed the same amount of glucose and produced approximately the same concentrations of end products (**Fig. 5.7A-D**).

The effect of carbon source on ethanol production by the two strains containing AdhE (± Rnf) was investigated at 60°C using glucose, xylose, xylan, crystalline cellulose (Avicel), and unpretreated switchgrass. The two strains generated similar amounts of ethanol on all of substrates (ranging from 18 – 32 mM), with the notable exception of switchgrass, where ~ 6 mM ethanol was generated, presumably because of the recalcitrance of the feedstock (**Fig. 5.9**). This was in spite of the fact that switchgrass yielded a comparable amount of acetate to that generated on glucose, yet the amount of ethanol was about 4-fold less than on glucose (**Fig. 5.9**). In contrast, growth on cellulose and xylan, the two major carbohydrate components of switchgrass, yielded about twice the amount of acetate and from 4-6 times the amount of ethanol. Interestingly, growth on switchgrass yielded no detectable acetoin or pyruvate in the growth media (data not shown), while on cellulose and xylan both of these end products were generated in the 0.5 – 2.0 mM range (**Fig. 5.9**).

Effect of fermentation conditions on ethanol production

All studies of ethanol production by *C. bescii* described to date were carried out in shaken, closed bottles with no pH control. As described above, the new ethanol-producing strains under these conditions produced more ethanol (**Fig. 5.6**) compared to previous studies, presumably due

to improved Ct*adhE* expression (**Fig. 5.4A**). Previously we showed that the use of pH-controlled, sparged bioreactors significantly improved substrate utilization and organic acid production by wild-type *C. bescii* at 78°C, its optimum growth temperature (105). Ethanol production by the two strains was therefore evaluated under these controlled conditions. Strains MACB1058 (with AdhE) and MACB1062 (with AdhE and Rnf), along with the parent strain, MACB1034, were grown on 20 g/L cellulose (Avicel) in bioreactors (0.5 L working volume) at 60°C. With all three strains, stationary phase (~5 x 10^8 cells/ml) was reached after about 40 hours and there was minimal cell lysis at least up to 200 hr (**Fig. 5.10**). As anticipated, there was an increase in both ethanol and acetate production when compared to growth in closed cultures (**Fig. 5.11**). Strain MAC1058, which lacked Rnf, produced a maximum of 60 mM ethanol after 120 hr, although the ethanol concentration began to decline thereafter (**Fig 5.11A**). In contrast to the closed cultures, the presence of Rnf had a dramatic impact as strain MACB1062, containing Rnf, generated 75 mM ethanol, a 25% increase. Moreover, strain MACB1062 showed little if any ethanol uptake even after more than 200 hr at 60°C (**Fig. 5.11A**).

In contrast to the decrease in acetate production by the ethanol production strains in closed bottles, acetate production was essentially the same in the parent strain as in the two AdhE-containing strains under pH-controlled, gas-sparging conditions (**Fig. 5.11B**). Acetate production was lower than expected (~60 mM) when compared to that produced by wild type *C. bescii* (127 mM acetate plus 22 mM lactate to give ~150 mM total organic acid; (105)). However, if total amount of end products for the two ethanol production strains are considered, they performed similarly to the wild type, generating 123 mM and 139 mM in total products (ethanol plus acetate) by MACB1058 and MACB1062, respectively (**Fig. 5.11**). In addition to ethanol and acetate, significant amounts of pyruvate were also detected in the growth media of all three strains,

approaching 12 mM for MACB1058 (**Fig 5.11C**). Once more, Rnf made a difference since pyruvate production was significantly lower throughout the growth phase and was comparable to that seen in the parent strain. On the other hand, in contrast to what was observed in closed bottles, acetoin was not detected for any of the three strains under bioreactor growth conditions (data not shown). Glucose was also detected in the media once cells reached stationary phase and increased rapidly once ethanol production ceased in the ethanol producing strains, reaching about 8 mM in MACB1058 and about half of that in the Rnf-containing MACB1062 and in the parent strains (**Fig. 5.10B**).

Both MACB1058 and MACB1062 reached stationary phase after 40 hr (**Fig. 5.10**), but continued to produce ethanol for over 150 hr (**Fig. 5.11**). The rate of ethanol production was higher during growth (0.7 - 0.9 mM/hr) and declined by more than half (0.2 - 0.4 mM/hr) once stationary phase was reached (**Table 5.4**). However, both strains produced ethanol during growth at about twice the rate that acetate was produced by the parent strain (0.24 mM/hr), even though the growth kinetics of the three strains were similar (**Fig. 5.10**). The parent strain continued to produce acetate during stationary phase at a rate also lower than that of ethanol production by the other two strains (0.16 mM/hr; **Table 5.4**). With all three strains, pyruvate began to accumulate at mid-exponential phase and continued to increase during stationary phase (**Fig. 5.11C**). For the parent strain, pyruvate production rate remained fairly constant at 0.1 mM/h (**Table 5.4**), but in both ethanol strains it ceased approximately 24 h before the maximum ethanol concentration was reached (**Fig. 5.11C**).

The ethanol producing strains consumed about 50% more cellulose (11.3 - 11.8 g/L) than the parent strain (7.6 g/L), as might be expected since they produce comparable amounts of acetate (**Fig. 5.11B**) but the parent does not produce ethanol (**Fig. 5.11A**). The carbon balance (cellulose

utilized versus products detected) was closed for the Rnf strain (MACB1062, 102.8% closed) but product recovery was somewhat lower (93.9% closed) for the strain lacking Rnf (MACB1058) and much lower (80.5% closed) for the parent strain (**Fig. 5.11D**). The lower carbon balance for the parent strain was surprising as analysis of wild-type *C. bescii* resulted in approximately 100% closure when grown under bioreactor conditions at 75°C, near its optimal growth temperature (105). This suggests that the parent strain generates significant amounts of an unidentified product (or products) when grown at 60°C but not at 75°C, a product that is not significantly generated in the ethanol strains, as these produce ethanol instead. In fact, this unknown product may be produced by AdhE-containing MACB1058 and it may account for the low carbon balance, while the presence of Rnf enables more ethanol to be produced by MACB1062 and little if any of the unknown product is generated. As shown in **Fig. 5.1**, the strains that produce ethanol generate less ATP than the parent that produces acetate. Consequently, the carbon balances reveal that in the ethanol production strains, a lower percentage of the cellulose carbon is used for cell biomass production (**Fig. 5.11D**).

Discussion

The goal of this work was to increase ethanol production in *C. bescii* by heterologous expression of the membrane-bound Rnf complex, which reduces NAD using reduced ferredoxin and should therefore yield a more redox balanced ethanol-producing pathway (**Fig. 1**). The presence of Rnf in the more stable genetic lineage resulted in the production of 75.5 mM ethanol under bioreactor conditions, which is 5-fold higher than the highest ethanol production previously reported with in this organism (14.8 mM; (185)). These results are very promising for the use of *C. bescii* for CBP as this ethanol titer is very similar to that reached by the cellulolytic *C. thermocellum* strain AG553, 73.4 mM, when the pathways for all other traditional fermentation

products were genetically-deleted (159). This strain was also subjected to adaptive evolution to improve the growth rate, resulting in the highest ethanol production by *C. thermocellum* to date of 580 mM ethanol (26.7 g/L; (160)). However, *C. thermocellum* has an important disadvantage compared to *C. bescii* as it is unable to utilize xylan for ethanol production (138). This is very significant because xylan represents a major fraction of the carbohydrate in plant biomass (~31% in switchgrass). In contrast, the AdhE/Rnf strain of *C. bescii* uses both xylan and xylose for ethanol production; indeed, more ethanol was produced from xylan than from cellulose (**Fig. 5.9**).

The genetic lineage of the *C. bescii* strain was important since ethanol production was increased by about 75% in the more stable background, MACB1034, compared to what was reported with the original genetic lineage JWCB018 (**Fig 5.6;** (185)). This is probably due to the lower expression of the Ct*adhE* gene, which resulted in a 2-5 fold higher specific activity of AdhE in cell-free extracts of strains MACB1058 and MACB1062 compared to JWCB032 and MACB1052*. Since the Ct*adhE* is not in the same genome location in these two sets of strains, that could be a contributing factor, especially given the proximity of this gene to a transposase (Athe_0863) and a transcriptional regulator (Athe_0864) in the JWCB018 background (185). It is also possible that the decrease in *adhE* expression (**Fig. 5.4A**) could be due to the increased insertion element activity in the JWCB018 background, which has contributed to an increase in overall genome instability (192).

Although the effect of Rnf on ethanol production by *C. bescii* was not really evident in a closed culture system, it had very significant effects under bioreactor conditions with pH control and gas-sparging. Specifically, not only was there increased ethanol production by 30% when Rnf was co-expressed, there was delayed production of pyruvate and also apparently of an as yet uncharacterized compound as unwanted side-products (**Fig. 5.6, 5.7,** and **5.11**). These data suggest

that, as we anticipated, Rnf activity results in an increased supply of NADH and this enables a higher flux of carbon from pyruvate via acetyl-CoA to ethanol (**Fig. 5.1**). Hence, ethanol production in the AdhE-containing strain appears to be NADH-limited. In the absence of Rnf, pyruvate accumulates and is either excreted or is converted to acetoin or to an unknown side product (see below). However, the Rnf-containing strain still generates acetate as a major carbon end product and its role in generating NADH may become even more important in further metabolic engineering studies, such as the deletion of either the phosphotransacetylase and acetate thiokinase or of the bifurcating hydrogenase (**Fig. 5.1**).

The production of acetoin as a side-product of fermentation at sub-optimal growth temperatures, as observed here with *C. bescii*, has also been reported with other *Caldicellulosiruptor* species, *C. sacharolyticus* (T_{opt} 70°C: (246)), and also with the extremely thermophilic archaeon *Pyrococcus furiosus* (T_{opt} ~100°C; (247)). Acetoin is generated from pyruvate by the action of acetolactate synthase and the abiotic decarboxylation of acetolactate to acetoin at high temperatures (above 60°C). In *P. furiosus*, it was suggested that the increase in acetoin production at the lower temperature was likely due to a bottleneck in glycolysis at pyruvate due to a decrease in the rate of pyruvate oxidation by POR at lower growth temperatures (247, 248). The results here suggest that a similar bottleneck occurs in *C. bescii* at 60°C, resulting in pyruvate and acetoin in the growth medium, and that this can be minimized by the action of Rnf. The carbon balance analysis of the parent strain indicates that *C. bescii* also generates an as yet uncharacterized side product that presumably also stems from pyruvate, since it is minimized by ethanol production such that little if any is produced by the Rnf-containing strain (**Fig. 5.11B**).

Under pH controlled growth conditions, *C. bescii* excreted pyruvate instead of acetoin (**Fig. 5.11C**). Lactic acid producing bacteria produce acetoin as a way of maintaining pH homeostasis

in the cell, but high intracellular pyruvate concentrations are required to do this (249). A similar effect may be occurring with *C. bescii,* where at lower growth temperatures there is a metabolic bottleneck at pyruvate and this accumulates intracellularly and is used for pH homeostasis under non-pH controlled conditions. A possible metabolic engineering solution for the build-up of pyruvate in *C. bescii* could be the heterologous expression of pyruvate-oxidizing POR from a microorganism that has temperature optimum for growth near 60°C. Note that increased POR activity would generate increased reduced ferredoxin, so the activity of Rnf would be even more important. Accordingly, both acetoin and pyruvate were generated in higher amounts in the strain lacking Rnf.

Herein we engineered *C. bescii* to produce ethanol but it did so optimally near 60°C. While this is near the temperature optimum of *C thermocellum* AdhE, the ethanol producing enzyme, it is almost 20° below the optimum growth temperature of *C. bescii* (78°C). The ethanol-producing strains at the end of fermentation had used just over half of the cellulose (~11 g/L) that was originally present in the growth medium (20 g/L). In spite of the lower growth temperature, this is very similar to that (10.6 g/L) metabolized by wild-type *C. bescii* when grown on 20 g/L cellulose (105).This contrasts with the wild type strain at 78°C, as this solubilized 60% of cellulose 50 g/L (105). This is not surprising, however, given the lower activity of the cellulases at 60°C compared to 78°C. For example, CelA, the most abundant enzyme in the secretome of *C. bescii,* retains 37% of its activity at 60°C compared to 75°C (170). Clearly, cellulose degradation is not the limiting factor in ethanol production. Similarly, *C. bescii* generated comparable amounts of ethanol from simple (glucose, xylose) and complex sugars (cellulose, xylan) although significantly less was produced from the much more recalcitrant switchgrass (in closed cultures, **Fig. 5.9**). This

was expected, however, as *C. bescii* is very inefficient at biomass deconstruction at 60°C compared to the optimal temperature of 78°C (168).

At 78°C, the limiting factor for growth of the wild type strain was proposed to be accumulation of acetate (> 160 mM; (105)) and our hypothesis is that production of the less inhibitory ethanol instead of acetate would lead to increased substrate conversion. This was indeed the case at 60°C, where the ethanol-producing strains utilized about 50% more of the initial cellulose (20 g/L) than the non-ethanol producing wild type strain. However, to have the same effect at 78°C, the optimum growth temperature, heterologous expression in *C. bescii* of genes encoding AdhE and Rnf that are highly active at that temperature are required. Unfortunately, potential sources of such genes are microbes that naturally produce ethanol, but none are known that do so optimally above 70°C and only eight such bacteria are known (AdhE is not found in the archaea) that grow optimally in the 60 – 69°C range (see (250)). Hence, finding more thermostable examples of AdhE and Rnf that are functional near 80°C is key to maximizing ethanol production by *C. bescii* and particularly from recalcitrant plant biomass.

In conclusion, ethanol production by *C. bescii* was significantly improved through the use of the new, more stable background (higher expression of Ct*adhE*), co-expression of the six genes of Tx*rnf*, and use of pH-controlled fermentor growth conditions. Interestingly, growth of *C. bescii* at 60°C resulted in the production of acetoin and pyruvate, which was exacerbated by the expression of Ct*adhE*. Co-expression of the Tx*rnf* decreased acetoin and pyruvate production and also prevented ethanol uptake in stationary phase of controlled bioreactor cultures. Growth of *C. bescii* at these suboptimal temperatures is not ideal particularly when recalcitrant substrates are used. Thus it is critical to further optimize *C. bescii* for ethanol production at higher temperatures,

including the heterologous expression of more thermostable enzymes. Additionally, future work to further alter metabolism in *C. bescii* to eliminate acetate production will be necessary.

Acknowledgements

We would like to thank Gerrit Schut, Piyum Khatibi, Jonathan Conway, Chris Straub and Laura Lee for helpful discussions. We also thank Daehwan Chung and Janet Westpheling for supplying *C. bescii* strain JWCB032. This research was supported by a grant (DE-PS02-06ER64304) from the Bioenergy Science Center (BESC), Oak Ridge National Laboratory, a U.S. Department of Energy (DOE) Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science.

Tables and Figures

Table 5.1 Strains used in this study.

Strain Name	Genotype	Parent Strain	Reference
JWCB032	*ΔpyrFA, ldh::ISCbe4, Δcbe1,* $P_{S\text{-}layer}$ *Cthe-adhE*	JWCB018	(185)
MACB1034	Δ*pyrE*, Δ*ldh*	MACB1018	(127)
MACB1052*	*ΔpyrFA, ldh::ISCbe4, Δcbe1,* $P_{S\text{-}layer}$ *Cthe-adhE* $P_{BF\text{-}H2ase}$ *rnfCDGEAB*	JWCB032	This study
MACB1058	*ΔpyrE,* Δ*ldh,* $P_{S\text{-}layer}$ *Cthe-adhE*	MACB1034	This study
MACB1062	*ΔpyrE,* Δ*ldh,* $P_{S\text{-}layer}$ *Cthe-adhE* $P_{BF\text{-}H2ase}$ *rnfCDGEAB*	MACB1058	This study

Table 5.2 Primers used for plasmid constructions, qPCR and strain validation in this study.

Primer	Gene Target	Sequence (5' to 3')
Plasmid Construction		
GLCB062	Plasmid backbone	CGCATGTGAGCAAAAGGCGACAGTTTTCCCTTTGATATGTAACG
GLCB057	Plasmid backbone	TTATTTTTTCCTACTTCCAGGCTTC
GLCB018	pSBS4 backbone fragment	TCTGACGCTCAGTGGAACG
GLCB028	pSBS4 backbone fragment	CAACATCATATTGACCATCCTTTTAAAATGGA ATTCTTTTTGAAACATCAAC
GLCB029	pDCW89 backbone fragment	AAGGATGGTCAATATGATGTTGAACTTCTCTG ATAGGC
GLCB019	pDCW89 backbone fragment	CAGCCTAACTTCGATCATTGGAC
AR049	pAR007 and pAR009 5' flanking region	CCTGGAAGTAGGAAAAAATAATGCAAGTTTAATAGTAGTTTTAATACTGC
AR019	pAR007 and pAR009 5' flanking region	CTGCATAGCCTCTTTTAAATCCTGTTTATGTAAAATTCTTACTAAAAAGGAGG
AR018	pAR007 and pAR009 3' flanking region	TTTATTTTGTTTAATCATACCTCTTAGATAGCATAGATTACCAAAAAGGAGGC
AR052	pAR007 and pAR009 3' flanking region	TCGCCTTTTGCTCACATGCGTTTGAGAAAATCTAAGGTATATTGAGAGAG
AR007	SLP promoter region, pAR008 3' flanking region	ACAGGATTTAAAAGAGGCTATGCAG
AR005	SLP promoter region	AACTACTCACCAAACCTCCTTG
AR014	Terminator stem loop	TAATAAAGCTGAAATAAAAGAGGGTGAG
AR015	Terminator stem loop	ATCTAAGAGGTATGATTAAACAAAATAAAAGAGG
AR050	*C. thermocellum* AdhE	AGGAGGTTTGGTGAGTAGTTATGACGAAAATAGCGAATAAATACG
AR051	*C. thermocellum* AdhE	CTCACCCTCTTTTATTTCAGCTTTATTATCATTTCTTCGCACCTCCGT

AR035	pAR008 5' flanking region	GATGGGCATCTGAGGAATGGTTATGTAAAATTCTTACTAAAAAGGAGGC
AR053	pAR008 3' flanking region	GTCGCCTTTTGCTCACATGCGAGCACAAATCATACCGTTGTCG
AR036	*C. bescii* H2ase promoter	CCATTCCTCAGATGCCCATC
AR037	*C. bescii* H2ase promoter	AAATCTGCATCCCCCTTCTATTGTTAG
AR038	*T.* sp. X514 RnfCDGEAB	AAGGGGGATGCAGATTTATGAAGTTAGAAAATCTCACTTTTAAGG
AR039	*T.* sp. X514 RnfCDGEAB	CTGCATAGCCTCTTTTAAATCCTGTTTATTGTACTTTCTTCCTTACCTGTTC
qPCR		
Athe_1406qF	*C. bescii* GAPDH	GCTGCAGAAGGCGAATTAAAG
Athe_1406qR	*C. bescii* GAPDH	GGTCTGCAACTCTGTTGGAATA
AC04	*C. thermocellum* AdhE	GACCTTCTTTGATGTGTCTCCAG
AC05	*C. thermocellum* AdhE	CATACATCACCCACATGATTTTGG
AR058	*T.* sp. X514 RnfE	GGCACTAGGTATGTGTTCAG
AR059	*T.* sp. X514 RnfE	GCTGGAATACGAATTTTATCTGG
Strain Validation		
AR031	Athe_0949 genome region	CTTGCTCTGCTGTTTTGTCTC
AR032	Athe_0949 genome region	GAGTATCATTTGCTGCTAATTTATATCC

Table 5.3 Control activities for Rnf reactions.

Assay	Extract Type	Strain	Electron Donor	Electron Acceptor	Specific Activity (μmol/min/mg)
Hydrogen uptake	Membrane	WT	Hydrogen	BV	0.56
		MACB1052*	Hydrogen	BV	1.63
		MACB1062	Hydrogen	BV	1.17
Pyruvate oxidation	Membrane	WT	Pyruvate	MV	0.00
		MACB1052*	Pyruvate	MV	0.00
		MACB1062	Pyruvate	MV	0.02
	Cytoplasmic	WT	Pyruvate	MV	1.02
		MACB1052*	Pyruvate	MV	1.32
		MACB1062	Pyruvate	MV	0.26

Table 5.4 Rates of end product generation for different culture phases.

ND= not determined as product was not generated.

End Product	Strain	Exponential (mM/h)	Early Stationary (mM/h)	Late Stationary (mM/h)
Ethanol	MACB1034	ND	ND	
	MACB1058	0.70	0.43	
	MACB1062	0.92	0.18	
Acetate	MACB1034	0.24	0.16	NA
	MACB1058	0.29	0.17	0.48
	MACB1062	0.23	0.16	0.39
Pyruvate	MACB1034	0.07		
	MACB1058	0.13		
	MACB1062	0.10		

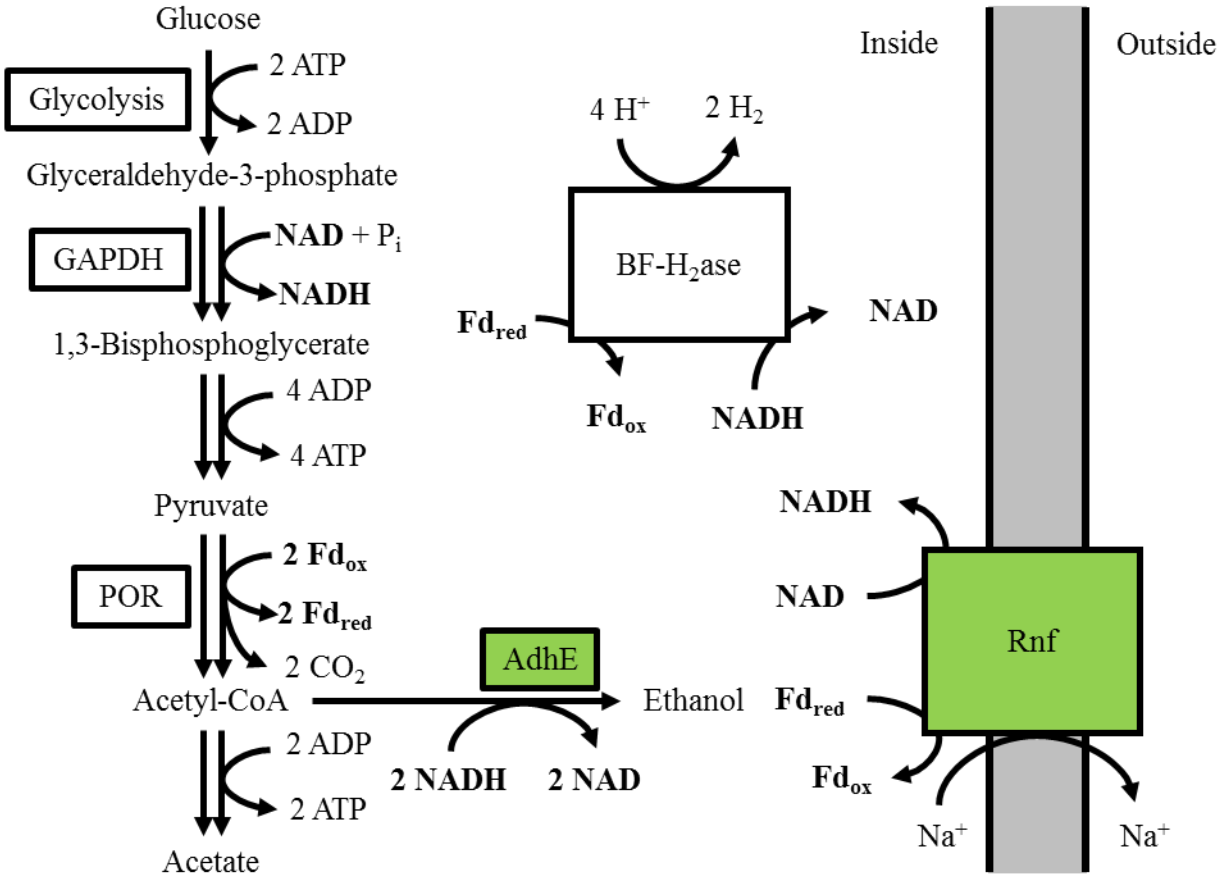


Figure 5.1 Pathway for redox balanced ethanol production in *C. bescii*.

The AdhE from *C. thermocellum* requires two NADH to generate ethanol from acetyl-CoA, but glycolysis in *C. bescii* generates one NADH and one reduced ferredoxin (Fd_{red}). The heterologous expression of Rnf allows for the transfer of electrons from reduced ferredoxin to NAD to generate the additional NADH needed for redox balanced ethanol production.

Figure 5.2 Plasmids for insertion of heterologous gene expression in *C. bescii* strains.

Elements required for plasmid cloning in *E. coli* are shown in grey. The selection cassette for *C. bescii* which allows for selection with kanamycin or uracil prototrophy and counter selection for plasmid loss with *pryE* are shown in black. 5' and 3' flanking regions are shown in white and heterologous genes and promoters for their expression are shown in green. pAR007 (A) for the insertion of the *adhE* from *C. thermocellum* (Ct*adhE*) at the Athe_0949 gene locus of a *ΔpyrE* background strain. pAR008 (B) for the insertion of the *rnfCDGEAB* from *T.* sp. X514 (Tx*rnf*) at the Athe_0949 gene locus into a strain already encoding the Ct*adhE* at this locus. pAR009 (C) for the insertion of Tx*rnf* at the Athe_0949 gene locus of a *ΔpyrF* strain without previous modification to this locus.

Figure 5.2

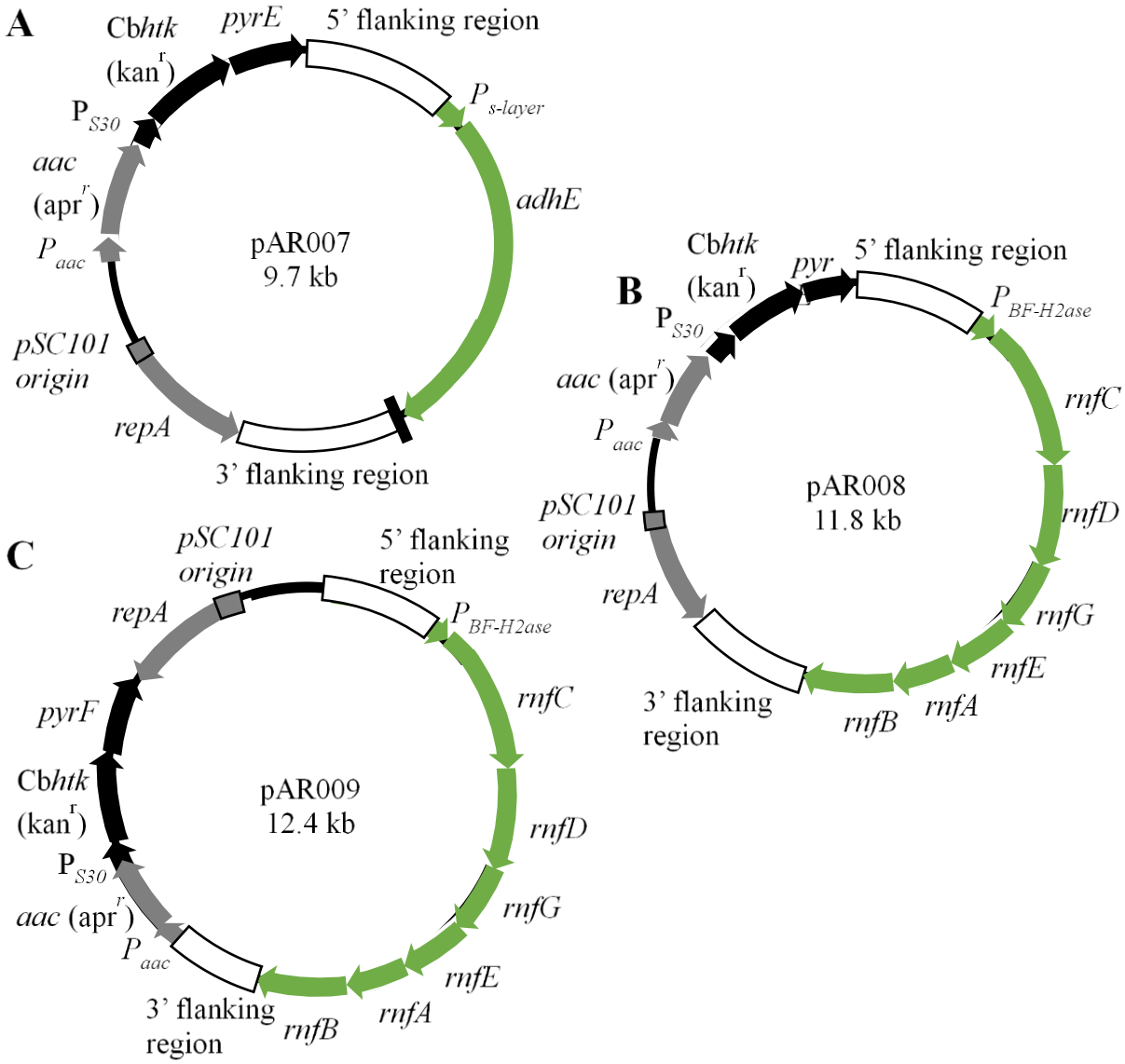
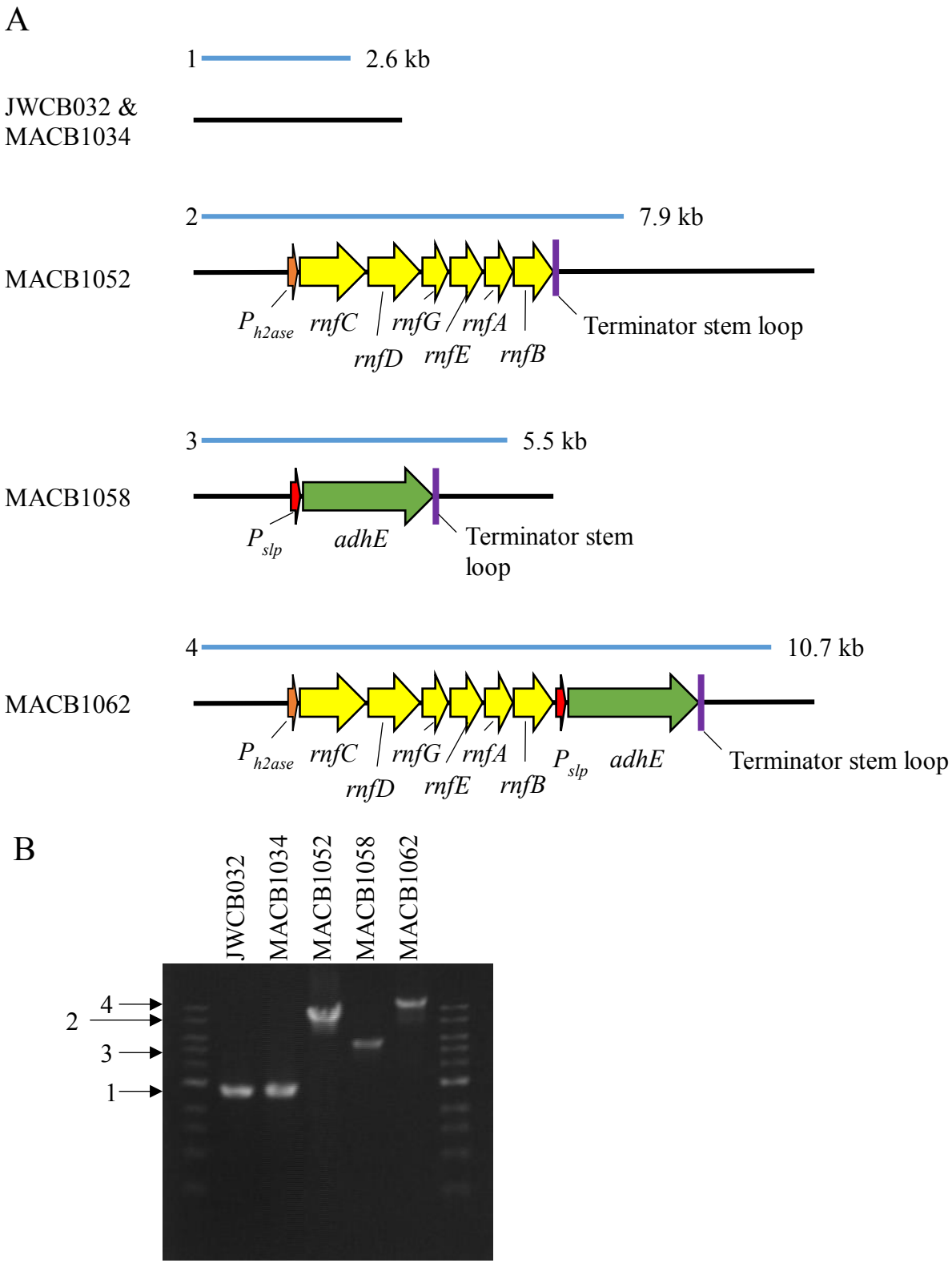

A
Cb*htk*
(kan^r)
pyrE
5' flanking region
$P_{s\text{-}layer}$
adhE
P_{S30}
aac
(apr^r)
P_{aac}
pAR007
9.7 kb
pSC101
origin
repA
3' flanking region
B
Cb*htk*
(kan^r)
P_{S30}
pyr
5' flanking region
$P_{BF\text{-}H2ase}$
rnfC
aac (apr^r)
P_{aac}
pAR008
11.8 kb
pSC101
origin
rnfD
repA
rnfG
rnfE
3' flanking
region
rnfB
rnfA
C
pSC101
origin
5' flanking
region
repA
$P_{BF\text{-}H2ase}$
rnfC
pyrF
Cb*htk*
(kan^r)
pAR009
12.4 kb
rnfD
P_{S30}
rnfG
aac (apr^r)
rnfE
P_{aac}
3' flanking
region
rnfB
rnfA

Figure 5.3 PCR verification of strains.

A) Athe_0949-Athe_0950 locus of the parent strain, MACB1058 expressing *C. thermocellum adhE* and MACB1062 expressing *T.* sp. X514 *rnfCDGEAB*. B) Gel electrophoresis results verifying the insertion of the gene constructs at the Athe_0949-Athe_0950 locus.

Figure 5.3


A
1
2.6 kb
JWCB032 &
MACB1034
2
7.9 kb
MACB1052
P h2ase
rnfC
rnfG
rnfA
rnfD
rnfE
rnfB
Terminator stem loop
3
5.5 kb
MACB1058
P slp
adhE
Terminator stem
loop
4
10.7 kb
MACB1062
P h2ase
rnfC
rnfG
rnfA
rnfD
rnfE
rnfB
P slp
adhE
Terminator stem loop
B
JWCB032
MACB1034
MACB1052
MACB1058
MACB1062
2
4
3
1

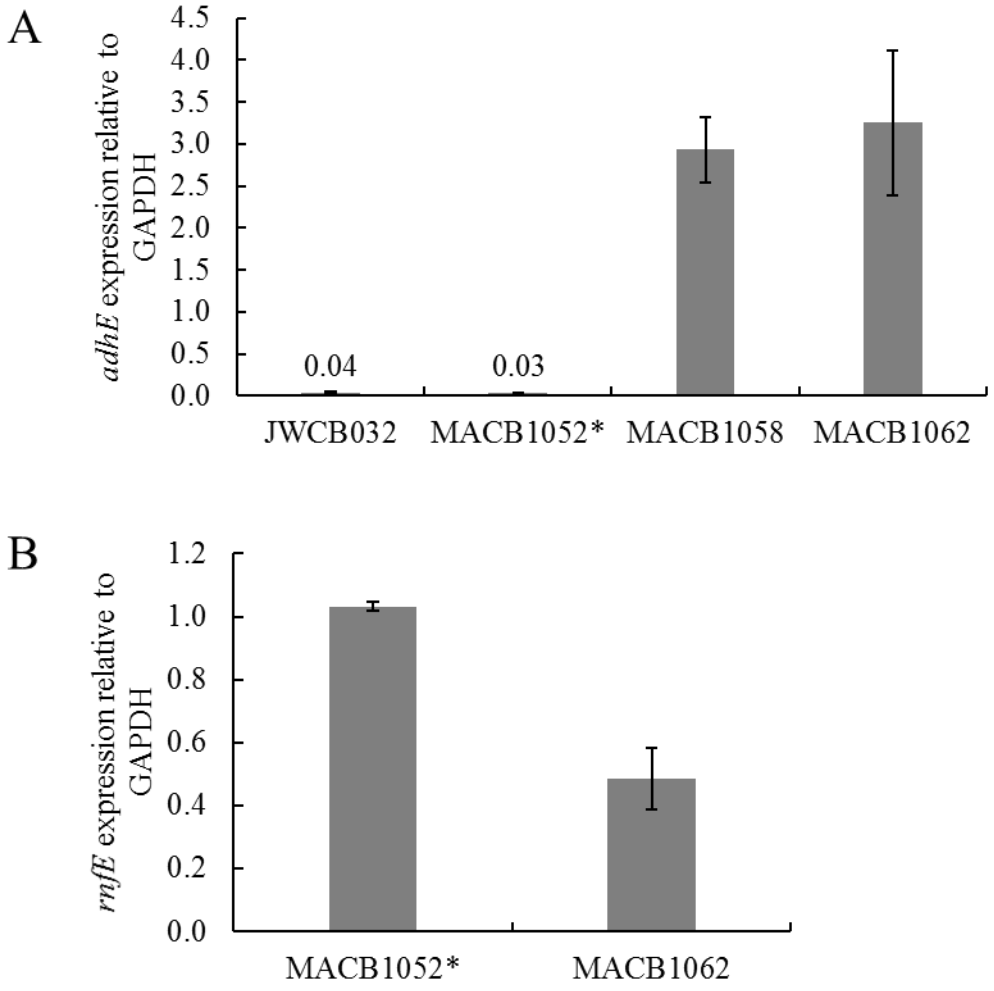
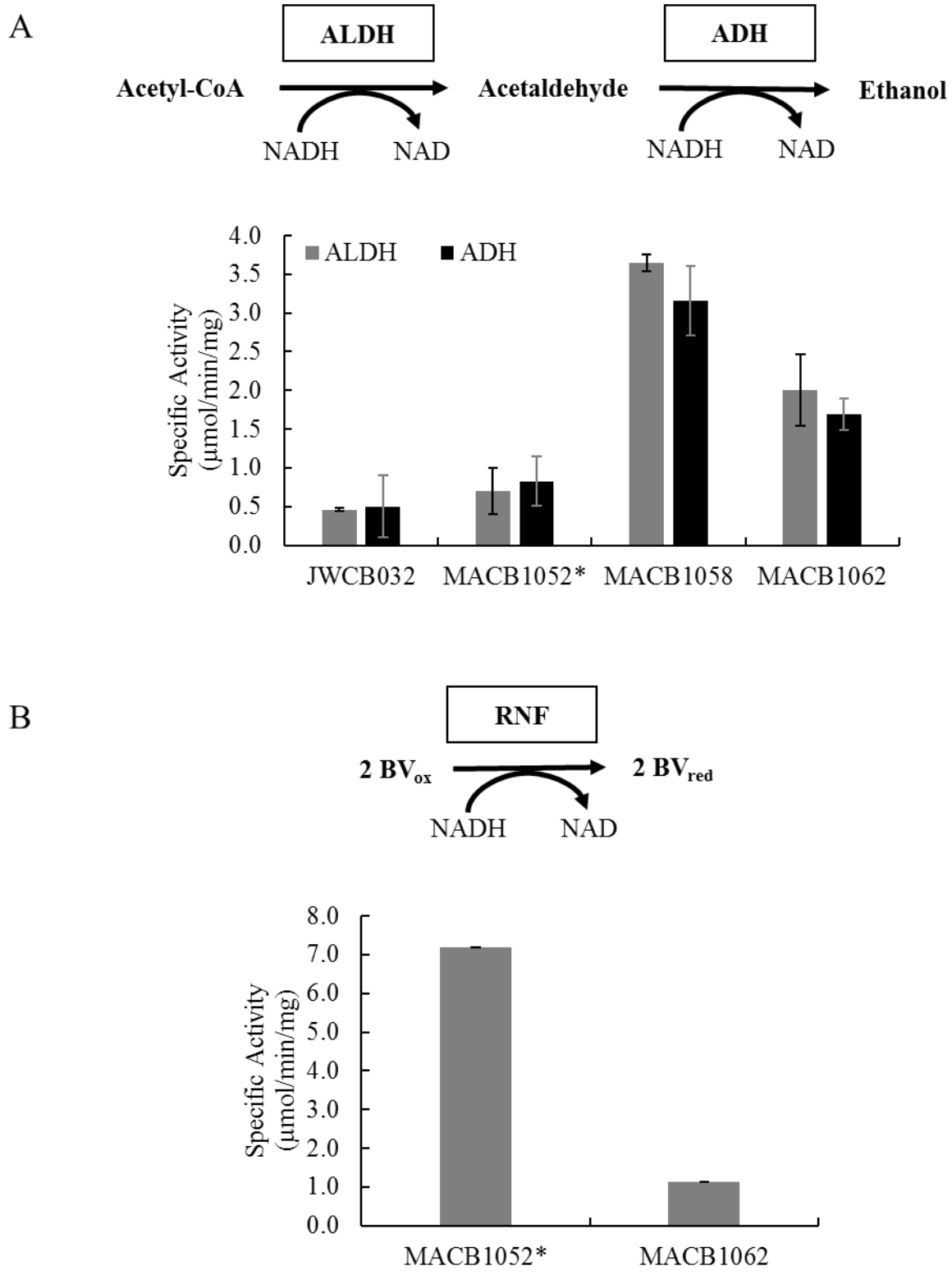


Figure 5.4 Expression of the *adhE* from *C. thermocellum* (A) and the *rnfE* from *T.* sp. X514 (B) relative to the native *gapdh* in *C. bescii* ethanol production strains.

Figure 5.5 Enzyme activities in cell-free extracts of *C. bescii* ethanol production strains.

Assays were conducted in the direction shown. Error bars represent the standard deviation (n=3) A) AdhE activity measured in whole cell extracts as separate reactions with Acetyl-CoA reduction shown as ALDH activity and acetaldehyde reduction shown as ADH activity. Both reactions were measured by following the oxidation of NADH to NAD. B) Rnf activity was measured in washed membranes for MACB1052* and MACB1062 by following NADH dependent the reduction of benzyl viologen (BV).

Figure 5.5


A
ALDH
ADH
Acetyl-CoA
Acetaldehyde
Ethanol
NADH
NAD
NADH
NAD
Specific Activity (µmol/min/mg)
ALDH
ADH
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
JWCB032
MACB1052*
MACB1058
MACB1062
B
RNF
2 BVox
2 BVred
NADH
NAD
Specific Activity (µmol/min/mg)
8.0
7.0
6.0
5.0
4.0
3.0
2.0
1.0
0.0
MACB1052*
MACB1062

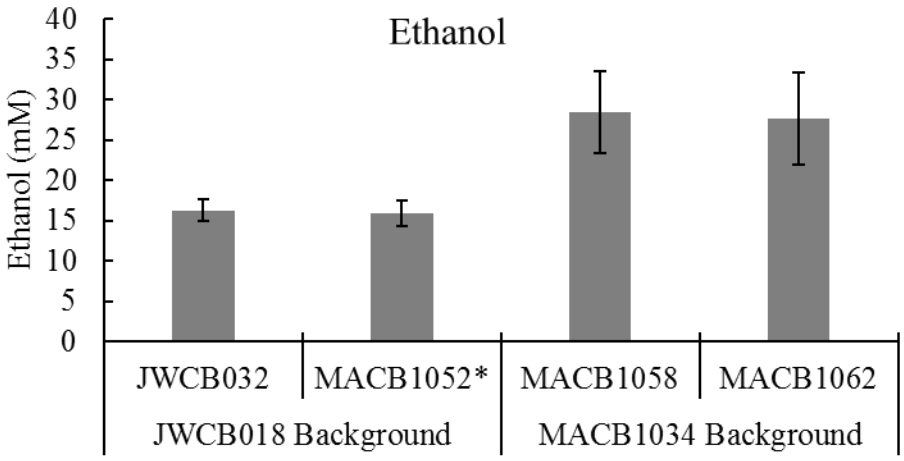


Figure 5.6 Ethanol production at 60°C from 50 mL scale cultures with glucose as the carbon source by strains containing the *C. thermocellum adhE* in the MACB0134 background (MACB1058, MACB1062) or the JWCB018 background (JWCB032, MACB1052*).

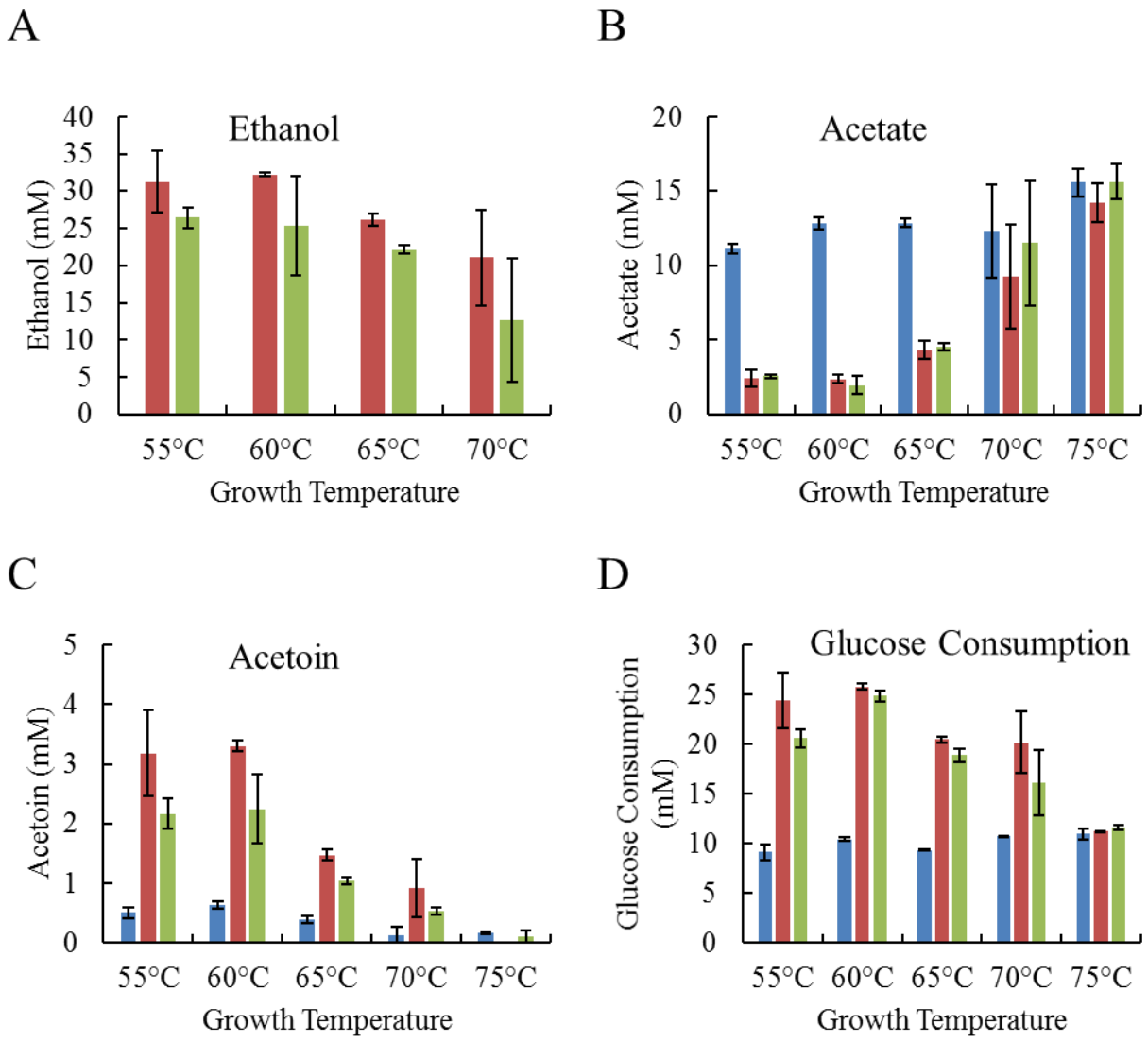


Figure 5.7 Ethanol (A), acetate (B), and acetoin (C) production and glucose consumption (D) for 50 mL cultures with glucose as the carbon source at temperatures between 55-75°C for strains MACB1034 (blue), MACB1058 (red), and MACB1062 (green).

No ethanol was detected for all strains at 75°C and for MACB1034 at any temperature (data not shown). Error bars represent the standard deviation (n=3).

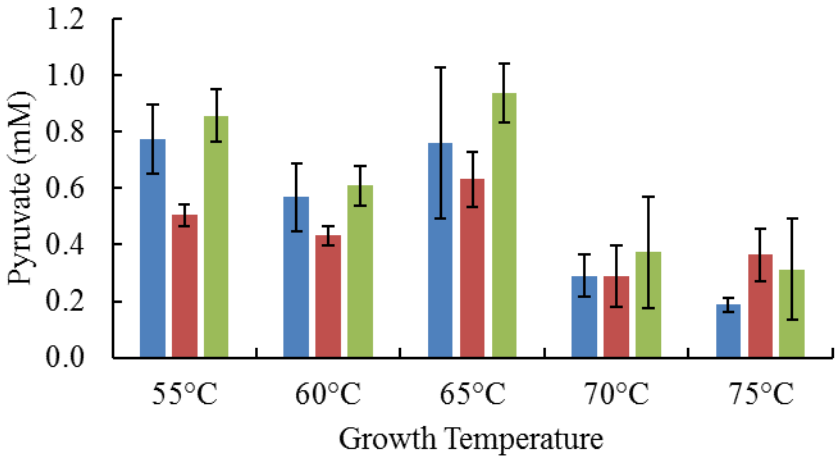
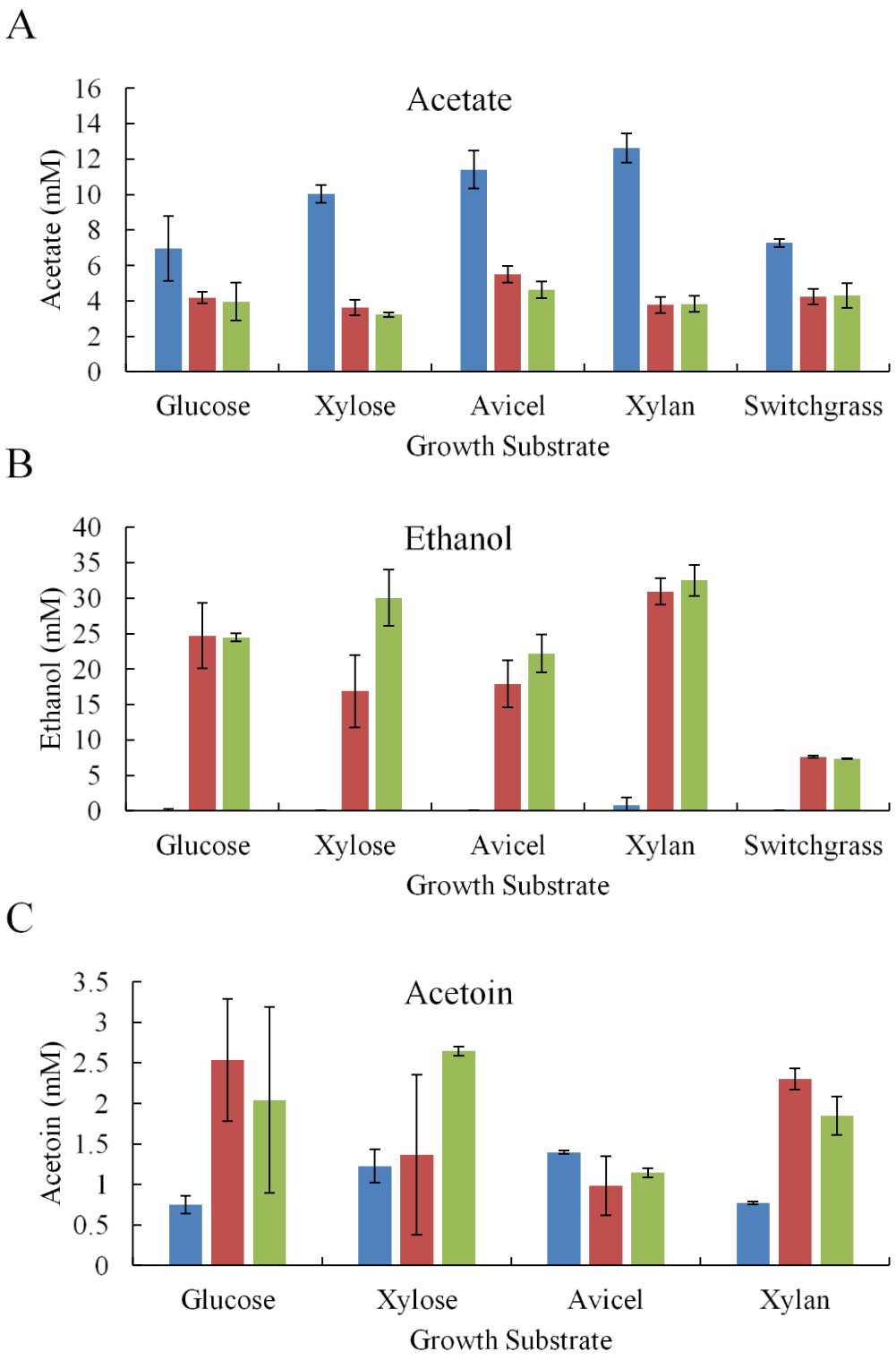


Figure 5.8 Pyruvate production by strains MACB1034 (blue), MACB1058 (red) and MACB1062 (green) during growth on glucose at temperatures between 55-75°C.

Figure 5.9 Acetate (A), ethanol (B), and acetoin (C) production for 50 mL cultures with glucose, xylose, Avicel, beechwood xylan, switchgrass and poplar as the carbon sources at 60°C for strains MACB1034 (blue), MACB1058 (red), and MACB1062 (green).

Error bars represent the standard deviation (n=3).

Figure 5.9


A
Acetate
Acetate (mM)
16
14
12
10
8
6
4
2
0
Glucose
Xylose
Avicel
Xylan
Switchgrass
Growth Substrate
B
Ethanol
Ethanol (mM)
40
35
30
25
20
15
10
5
0
Glucose
Xylose
Avicel
Xylan
Switchgrass
Growth Substrate
C
Acetoin
Acetoin (mM)
3.5
3
2.5
2
1.5
1
0.5
0
Glucose
Xylose
Avicel
Xylan
Growth Substrate

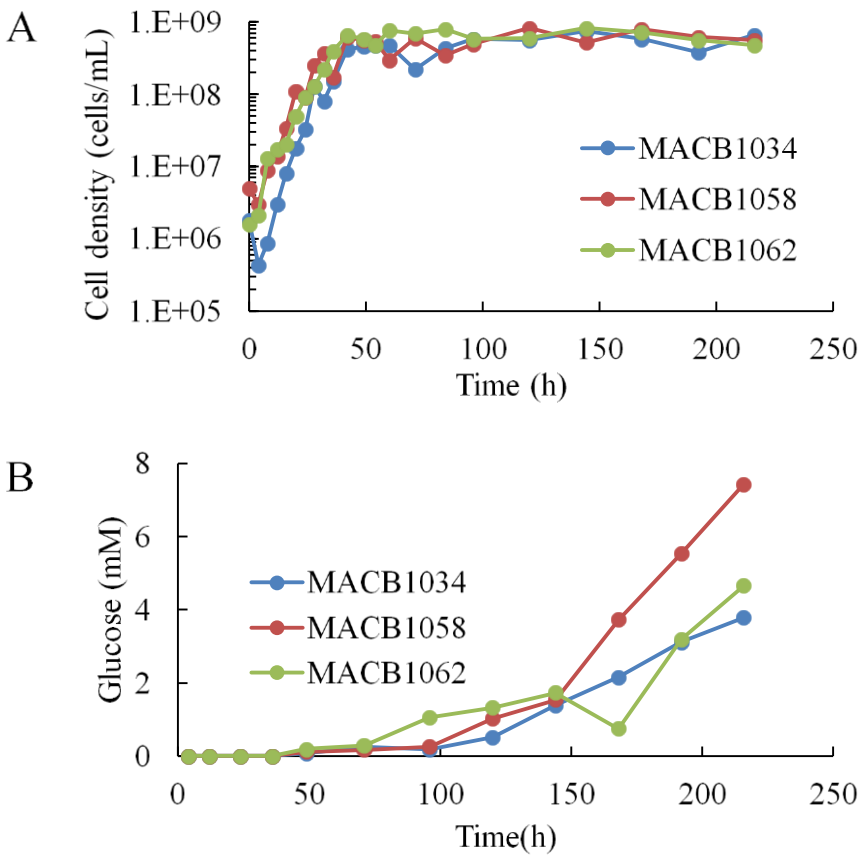


Figure 5.10 Growth and glucose production for bioreactor growths.

A) Growth of strains MACB1034, MACB1058 and MACB1062 under pH-controlled, gas-flushed bioreactor conditions on 20 g/L cellulose. B) Glucose accumulation in the medium during bioreactor growths.

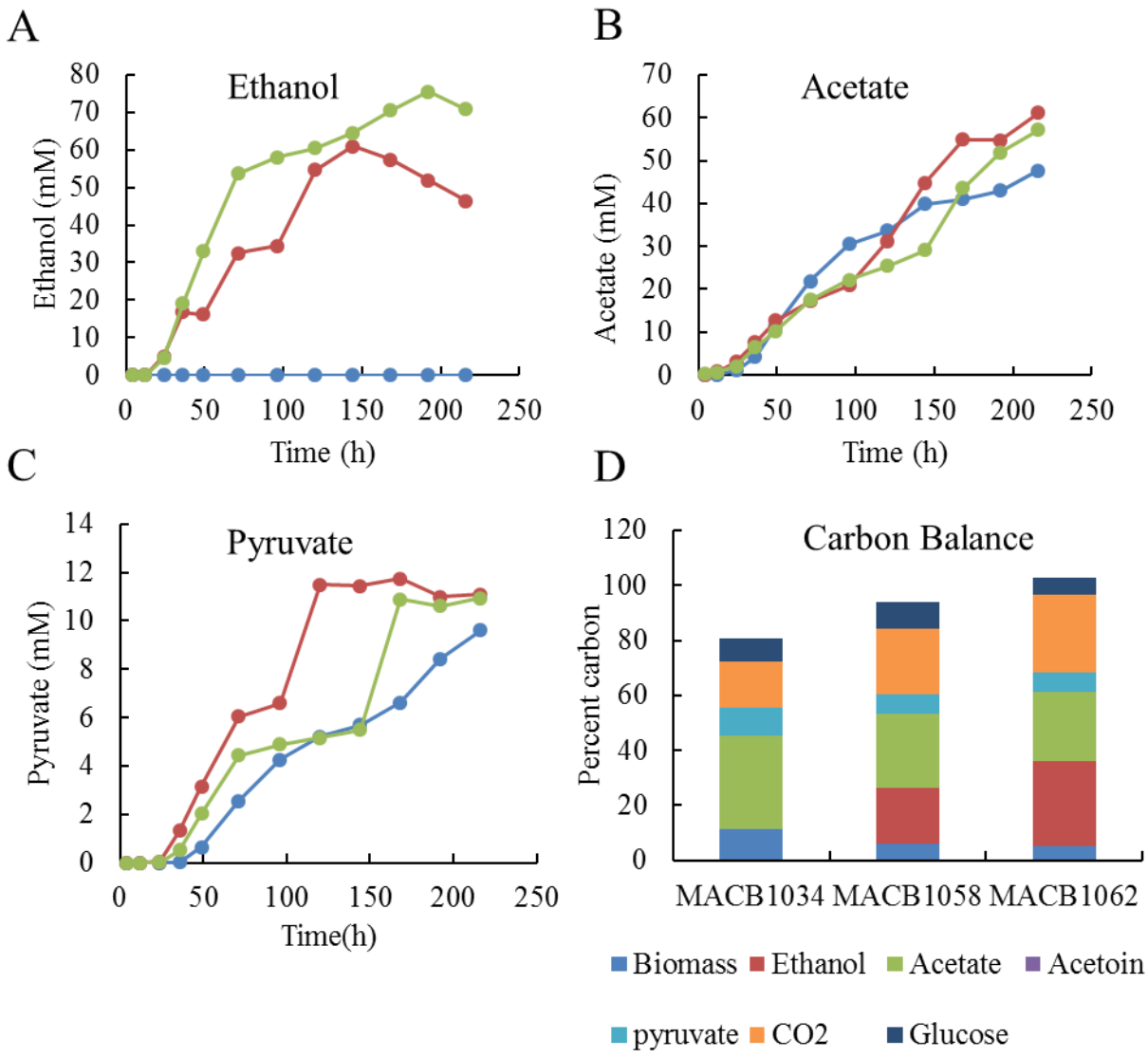


Figure 5.11 Acetate (A), ethanol (B), and pyruvate(C) production and carbon balances (D) for cultures grown in 1 L bioreactors with crystalline cellulose as the carbon sources at 60°C for strains MACB1034 (blue), MACB1058 (red), and MACB1062 (green).

Carbon balances were calculated based on measured glucose, pyruvate, acetoin, acetate, ethanol, cellular protein (biomass) and residual Avicel.

CHAPTER 6

DISCUSSION

In recent years, our understanding of the thermophilic cellulolytic *C. bescii* has increased dramatically leading to early proof of concept metabolic engineering of this organism for renewable biofuel production. The emergence of a genetic system for this organism has been critical to these advancements. The work here describes the limitations of wild-type *C. bescii* on high substrate loadings of crystalline cellulose and unpretreated switchgrass, advances in the genetics toolkit for use in *C. bescii*, and improvements to ethanol production by this organism. These studies have increased our understanding of the physiology and metabolism of *C. bescii* and have made large steps toward the use of this organism for consolidated bioprocessing (CBP).

Understanding and overcoming limitations of *C. bescii* fermentation

As outlined in Chapter 1, the generation of at least 40 g/L ethanol is critical for the economical production of this potential biofuel (8). This requirement must be considered then when testing organisms of interest for CBP as, clearly, high substrate loadings will be required to generate the necessary ethanol titers. Since theoretically 2.00 and 1.67 molecules of ethanol can be generated from glucose and xylose respectively, complete utilization of 72.9 g/L cellulose or 108 g/L switchgrass (0.45 g/g C_6 and 0.31 g/g C_5) would need to be achieved to generate 40 g/L ethanol. In Chapter 2, *C. bescii* was shown to grow on up to 200 g/L either crystalline cellulose or unpretreated switchgrass in closed serum bottle cultures (192). Additionally, 50 g/L cellulose and switchgrass were used to evaluate performance in pH-controlled and gas-flushed conditions.

In closed serum bottles, the pH decreased below 5.0 for cultures grown on crystalline cellulose (Avicel) and between 2.1-2.6 g/L was solubilized independent of initial concentration (**Table 2.1**). This effect was alleviated by growth under pH-controlled and gas-flushed conditions where a maximum of 30 g/L Avicel was solubilized, though not completely metabolized. The final factor found to be limiting to crystalline cellulose fermentation the generation of greater than 150 mM organic acids (192). Understanding the exact cause of this limitation is critical to future metabolic pathway engineering. Three possible stressors were identified, they are; the membrane decoupling effect of weak organic acids (216), salt tolerance, and osmotic pressure (214). The role of each of these factors in metabolism of soluble sugars was assessed. To test these, *C. bescii* was grown in increasing concentrations of sodium acetate (**Fig. 2.10**), sodium chloride and glucose (**Fig. 6.1**). 150 mM sodium acetate and 150 mM sodium chloride had similar effects with very long lag phases and slower growth rate to reach a similar final density as control cultures. There was a much higher tolerance for a neutral stressor represented by glucose where 500 mM was required to get a similar growth phenotype. Based on these results, it was hypothesized that the generation of a neutral end product would help to overcome the salt tolerance of *C. bescii* and would possibly lead to increased substrate utilization and fewer residual sugars in the fermentation broth (105).

Since ethanol is a neutral product, the strains described in Chapter 5 provide the opportunity to test this hypothesis. Unfortunately, the optimal temperature of the ethanol production pathway used was 60°C which results in slower growth of *C. bescii* and less overall substrate solubilization. However, when compared to the parent strain, this ethanol production strain utilized more of the available substrate under both closed serum bottle conditions (**Fig 5.7**) as well as pH-controlled bioreactor conditions, which supports the hypothesis generated based on

the wild-type physiology. One obvious next step to improve ethanol production under these conditions is to apply directed evolution to these strains on minimal medium with high cellulose loadings at 60°C to select for faster growing isolates. This has been performed for *C. thermocellum* by serial transfer in defined medium improving the growth rate of a modified strain with a poor growth phenotype from 0.06/h to 0.22/h, comparable to wild-type as well as improving substrate utilization (160).

For the real-world substrate, switchgrass, it was the recalcitrance of this substrate which seemed to limit solubilization and therefore fermentation proceeded in a carbon limited like state (105). Evidence for this limitation included utilization of the same percentage, 30% (**Table 2.3**), of substrate independent of initial substrate concentration and the need to supplement spent media with additional carbon to facilitate growth (**Fig. 2.12**). Additional work by another group has found that recalcitrance of plant material to digestion by *C. bescii* can be overcome through some genetic manipulations of the plants or through the identification of natural variants with lower recalcitrance (179). Also, since the development of the genetics system in this organism, some work has been undertaken to improve plant biomass utilization through expression of additional extracellular carbohydrate active enzymes (108).

Work performed in the wild-type strain also showed that growth of *C. bescii* was strongly inhibited by the presence of greater than 10 g/L dilute acid-pretreated switchgrass even after extensive washing to neutralize the material and remove simple sugars and water soluble inhibitors generated during pretreatment (105). It is concerning that *C. bescii* demonstrates such sensitivity as pretreatment, though costly, is likely critical for utilization of plant biomass at the industrial scale to generate a versatile, feedstock agnostic process (40, 64). Testing *C. bescii* compatibility with different pretreatments, including those shown to generate fewer inhibitory sugar degradation

products remains an under explored area. In addition to finding the most compatible pretreatment for use by *C. bescii*, engineering *C. bescii* for increased inhibitor tolerance is an important area of study (98). This also appears to be important as there is additional evidence that an inhibitory compound may be released during growth of *C. bescii* on unpretreated plant biomass (**Fig. 2.12**). Purification and characterization of the unknown inhibitory compound and its mechanism of action on *C. bescii* will be critical to determine how to overcome this limitation. As growth of another thermophile, *T. mathranii*, was not inhibited by this unknown compound (**Fig. 2.12**) it may be possible to determine a gene for heterologous expression in *C. bescii* to overcome this inhibitory effect.

Improvements to the genetic toolkit for *C. bescii*

The development of a system for genetic manipulation of *C. bescii* has been critical to the study of this organism and its metabolic engineering for ethanol production. However, genetics in non-model microorganisms are often not developed much beyond transformation. For *C. bescii* to be a useful platform organism for biotechnology, a robust and stable genetics system is critical and requires continued expansion of the genetic tool-kit available. Recent years have seen such expansion with the addition of an antibiotic resistance marker for use in *C. bescii* (127) and the characterization of a native xylose-inducible promoter (Chapter 5). Continued development is needed, but these are certainly large steps toward use of *C. bescii* as a biotechnological platform.

As described in Chapter 3, there are two genetic lineages of *C. bescii* both based on uracil auxotrophic strains generated in different ways (**Fig. 3.1**). The first was based on the spontaneous deletion of the *pyrFA* genes selected for by plating wild-type on solid medium containing both uracil and the toxic 5-fluorootoric acid (5-FOA)(182). The second made use of a codon optimized, thermostable kanamycin resistance gene for use in *C. bescii* which was used to generate a targeted

pyrE deletion (127). Insertion element movements were observed in strains generated in the random-deletion lineage prompting the use of Southern Blot analysis and PacBio whole genome sequencing to examine stability of these two lineages (192). We found (Chapter 3) that the random-deletion lineage had a dramatic increase in IS*Cbe*4, which was active in both lineages, more single nucleotide polymorphisms, insertions and deletions (**Table 3.3**) than the targeted deletion background and at least one instance of large-scale genome rearrangements (**Fig. 3.9**).

Interestingly, these resequencing efforts also provided some potential insights into the mode of action of the IS*Cbe4* insertion element. The increase in copy number without loss of the transposon at the original location in the parent strain strongly suggests a replicative mechanism. Additionally, the clean loss of the IS*Cbe4* element from a few locations, from the *ldh* gene of MACB1013 and one instance between wild-type and JWCB005, suggests that loss is a rare event which is not an active part of the transposition method of this element. IS*Cbe4* is a member of the IS*Lre2* family which is not fully understood, however circular forms have been reported (239) giving another bit of evidence to suggest the mechanism of IS*Cbe4*. Taken together, we propose the following model of IS*Cbe4* transposition displayed in **Figure 6.2**. IS*Cbe4* transposition begins with a single strand break at the 3' end of the insertion element. Strand transfer occurs with the free 3' hydroxyl group generating a transposon circle. Replication resolves this intermediate and the original transposon is either regenerated through bacterial host DNA repair enzymes or in rare events, recombination occurs at the direct repeat regions resulting in loss of the transposon. Further validation of this model will be needed including characterization of the transposase encoded by IS*Cbe4* as well as continued study of future transposition events that can be monitored by Southern blot and sequencing analysis. In addition, the development of a transposition assay and a way to examine extrachromosomal IS*Cbe4* would greatly facilitate research in this area.

It is important to note that as IS*Cbe*4 is active in both backgrounds of *C. bescii*, it is possible that in the future, the more a strain has been modified from the wild-type through several generations, the copy number of this element is likely to increase leading to instability in these strains as well. Given the potential importance of stability for industrially utilized strains, regular sequencing of these strains will be necessary. In Chapter 3, a method utilizing Southern Blot analysis probing for IS*Cbe*4 was proposed to monitor overall mutations with approximately 8-9 mutations predicted for each additional band on a Southern blot (192). Additionally, more individual isolates of each strain may need to be generated and selected based on IS*Cbe*4 copy number. As strain generation in *C. bescii* is not high-throughput, the task of deleting all copies of IS*Cbe*4 will be challenging, though it may prove to be a worthwhile endeavor if a strain of great industrial importance is developed.

An unexpected, but important finding from the PacBio resequencing of all strains in this study was a mis-annotation in the original genome sequence of *C. bescii* at the highly important glucan degradation locus (GDL) (**Fig. 3.11**). Due to the highly repetitive nature of this region with upwards of 2.5 kb regions with 100% identity at the sequence level, the shorter sequence of the original Sanger/454 sequencing made proper assembly of the GDL unlikely. This observation does not change the sequence of any particular gene, it merely results in a reordering of this locus, however, strains with alterations at this locus need to be re-evaluated to confirm the desired mutation was in fact generated. As an example, an earlier study reported broad effects of a c*elA* deletion in *C. bescii* for growth on a wide variety of substrates (191). However, a more recent study generated target deletions of each of the genes in the GDL, confirmed by PacBio sequencing of the locus in each strain, showed phenotypic evidence suggesting that the *celA* deletion strain may have a larger portion of the GDL deleted (191, 194).

Chapter 4 describes the characterization of a xylose-inducible promoter for use in *C. bescii* making use of a lactate dehydrogenase (*ldh*) reporter in an *ldh* deletion strain in the more stable parent background. It was shown that *ldh* expression levels and lactate dehydrogenase enzymatic activity were responsive to growth of this strain on glucose or xylose as a carbon source (**Fig. 4.3** and **Fig. 4.4**). Additionally, lactate production was tightly linked to initial xylose concentration in the medium (**Fig. 4.5**). Further bioinformatic and biochemical characterization of the regulatory mechanisms that control expression by this promoter will be needed. This promoter promises to be a useful addition to the *C. bescii* molecular biology tool kit allowing for condition expression or knockdown of genes of interest. One could imagine use of this promoter to titrate expression levels of certain pathway genes to determine optimal expression, which is often not constitutively high expression commonly used in proof of concept strain development. Additionally, conditional gene knockdown is possible with this system which would be ideal for increasing our understanding of essential genes in this organism. It is important to note that this promoter, while highly useful for lab scale studies, will not be useful for conditional deletion on plant biomass which is xylose rich. Ideally, a suite of promoters at a variety of expression levels, both inducible and constitutive, will be characterized for use in *C. bescii* in the coming years.

Metabolic engineering of *C. bescii* for ethanol production

The production of ethanol by *C. bescii* represents a first step to use of this organism as a platform for biofuel and bioproduct generation. *C. bescii* does not natively produce ethanol as a fermentation end product, though ethanol production was previously engineered into *C. bescii* through the expression of a bifunctional alcohol dehydrogenase (AdhE) (185, 186). In these two studies the *adhE* genes were from *Clostridium thermocellum* (185) and *Thermoanaerobacter pseudethanolicus* 39E (186) for ethanol production at 65°C and 75°C respectively. The strain with

the *adhE* from *C. thermocellum* (Ct*adhE*), JWCB032, produced 14.8 mM (0.7 g/L) ethanol from cellobiose (185) which was the highest ethanol production by *C. bescii* prior to the work presented in Chapter 5. This ethanol production is well below industrial metrics and thus strategies for improvement were needed and a first logical step was to consider the redox metabolism of *C. bescii*. During glycolysis, *C. bescii* generates both NADH and reduced ferredoxin which both must be recycled for metabolism to continue (**Fig. 1.4**). AdhE only makes use of NADH as a redox co-factor and *C. bescii* is not known to produce any enzyme which allows for the transfer of electrons from reduced ferredoxin to NAD (177).

Building off this previous study, strains for redox balanced ethanol production were developed in both the JWCB018 background, and the new more stable background. These strains were engineered for the expression of the energy-conserving reduced ferredoxin NAD oxidoreductase (Rnf) that facillitates the transfer of electrons from reduced ferredoxin to NAD to generate an ion gradient across the cell membrane (165, 243). Interestingly, strains in the new parent background generated more ethanol (**Fig. 5.6**) which was attributed to higher expression of the Ct*adhE* gene in this background (**Fig. 5.4B**). It was not clear if this difference in expression was due to the genomic location of the Ct*adhE* gene, a mutation or overall instability of the JWCB018 background (Chapter 5). The strain expressing both the Ct*adhE* and the *rnfCDGEAB* from *Thermoanaerobacter* sp. X514 generated the highest ethanol reported to date in *C. bescii*, 75 mM (**Fig. 5.11**). Three factors contributed to this increase: the new parent background (improved expression of the Ct*adhE*), pH-controlled, gas-flushed bioreactor growth conditions, and Rnf.

Importantly, due to the growth temperature optima (60°C) of the organisms from which *adhE* and *rnfCDGEAB* were taken (138, 169), this work was all performed 18°C below the optimum growth temperature for *C. bescii*. Little is known about what changes to *C. bescii*

metabolism can be expected at these lower temperatures. In fact, acetoin and pyruvate were excreted into the growth medium by ethanol production strains and the parent strain when grown at 60°C (**Fig. 5.7, 5.8, 5.9** and **5.11**). That these metabolites are produced in the growth media by *C. bescii* had not been previously reported. These data suggest a metabolic bottle neck at the pyruvate node of metabolism which can be explained by either difficulty in recycling ferredoxin by a decrease in the pyruvate oxidoreductase (POR) activity at this lower growth temperature. The data presented in Chapter 5 suggest that likely both explanations play a role. Since these products are seen in the parent strain grown at lower temperatures, it suggests a decrease in POR activity at these lower temperatures as this strain should not be exhibiting redox stress due to differential recycling of NADH and ferredoxin. Strains containing AdhE exacerbate acetoin and pyruvate excretion which could be due to a redox imbalance as NADH is consumed for ethanol production with more limited ferredoxin recycling. Rnf partially alleviates this phenotype, though not back to parent levels suggesting titration of AdhE and Rnf levels may be necessary to get this pathway to operate optimally in *C. bescii* (**Fig. 5.7** and **5.11**).

The xylose-inducible promoter described in Chapter 4 will present the opportunity for precise control of genes for metabolic engineering of *C. bescii*. A key target will be the bifurcating hydrogenase, which is the key NADH and ferredoxin recycling enzyme in *C. bescii* (**Fig. 1.4**). This is a major enzyme in *C. bescii* (~5% of the cytoplasmic protein) and efforts to delete this gene from both genetic backgrounds even in strains with ethanol production and Rnf have been unsuccessful (unpublished data). In fact, when attempting to replace the *hydABCD* genes, encoding the bifurcating hydrogenase, with a kanamycin resistance marker, counter selection to generate the knockout resulted in colonies with mutations in the *pyrE* gene allowing these strains to maintain both the kanamycin resistance gene and the bifurcating hydrogenase genes. Future efforts will

instead work to replace the native hydrogenase promoter with the xylose-inducible promoter. This strategy is promising for developing conditional knockdowns of many potentially important genes such as the phosphotransacetylase which may lead to further improvements to ethanol production.

Conclusions and outlook

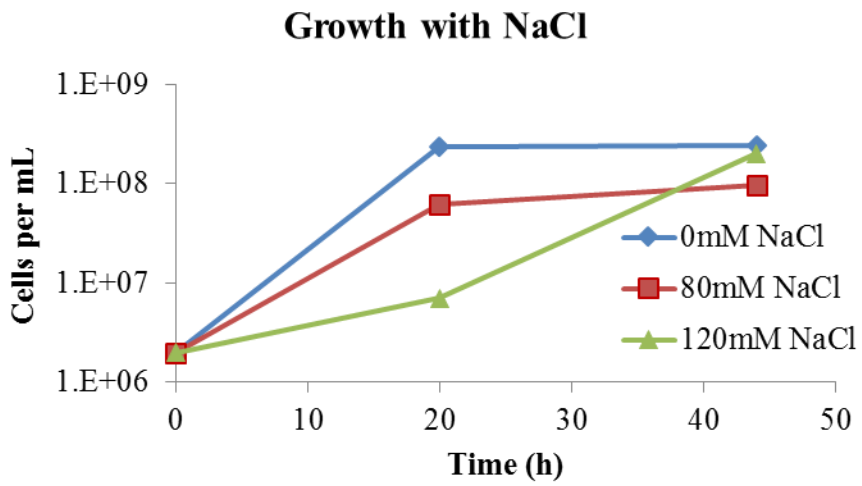
Recent years have seen great advances in our understanding of *C. bescii*. Work has included understanding the limits of fermentation by this organism (105), the development and advancement of the genetics system (Chapters 3 and 4; (127, 192, 204)), and metabolic engineering of *C. bescii* for ethanol production (Chapter 5, (185, 186)). Other work has involved more in-depth characterization of the native metabolism of *C. bescii*, in particular the discovery of a major tungsten-containing oxidoreductase that may well change our view of the primary metabolism of this organism (190). Additionally, biomass deconstruction has been examined for *C. bescii* and other members of this genus (85, 107, 136, 167, 168, 194). All of this work has led to major strides toward the use of *C. bescii* as platform organism for renewable biofuel production.

Generating 75 mM (3.5 g/L) ethanol is a major step toward demonstrating the industrial utility of *C. bescii* for CBP. In order to gain more perspective on this achievement, it is useful to compare the work in *C. thermocellum* which has been the subject of more metabolic engineering efforts toward CBP (156-161). Of particular note is the strain AG553, where all traditional fermentation products except ethanol were eliminated, allowing for the production of 73.4 mM ethanol from 20 g/L crystalline cellulose (Avicel) after 144 h (159). This performance is remarkably similar to the performance the *C. bescii* strain, MACB1062, co-expressing Rnf and AdhE in the new parent background. This strain of *C. bescii* generated 75 mM ethanol from 20 g/L Avicel after 200 h (**Fig. 5.11**). A subsequent study in *C. thermocellum* involved directed evolution of strain AG553 to improve the growth rate and substrate utilization resulting in the

production of 456 mM (22.4 g/L) ethanol from 60 g/L cellulose at 75% of theoretical yield (160). Despite this success, *C. thermocellum* still has a major disadvantage over *C. bescii* as it cannot utilize C_5 sugars which make up a large portion of plant biomass (~31% of switchgrass) and the former bacterium is in fact inhibited by xylose accumulation in growth media (138, 162). Adaptive evolution has yet to be applied to *C. bescii*, although given the results in *C. thermocellum* (160), use of this technique will likely be of great utility in future work.

Figures

A



B

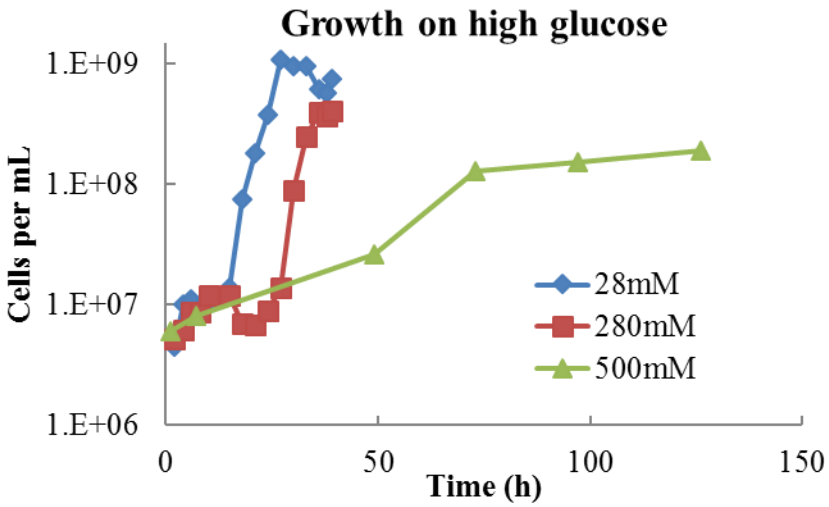


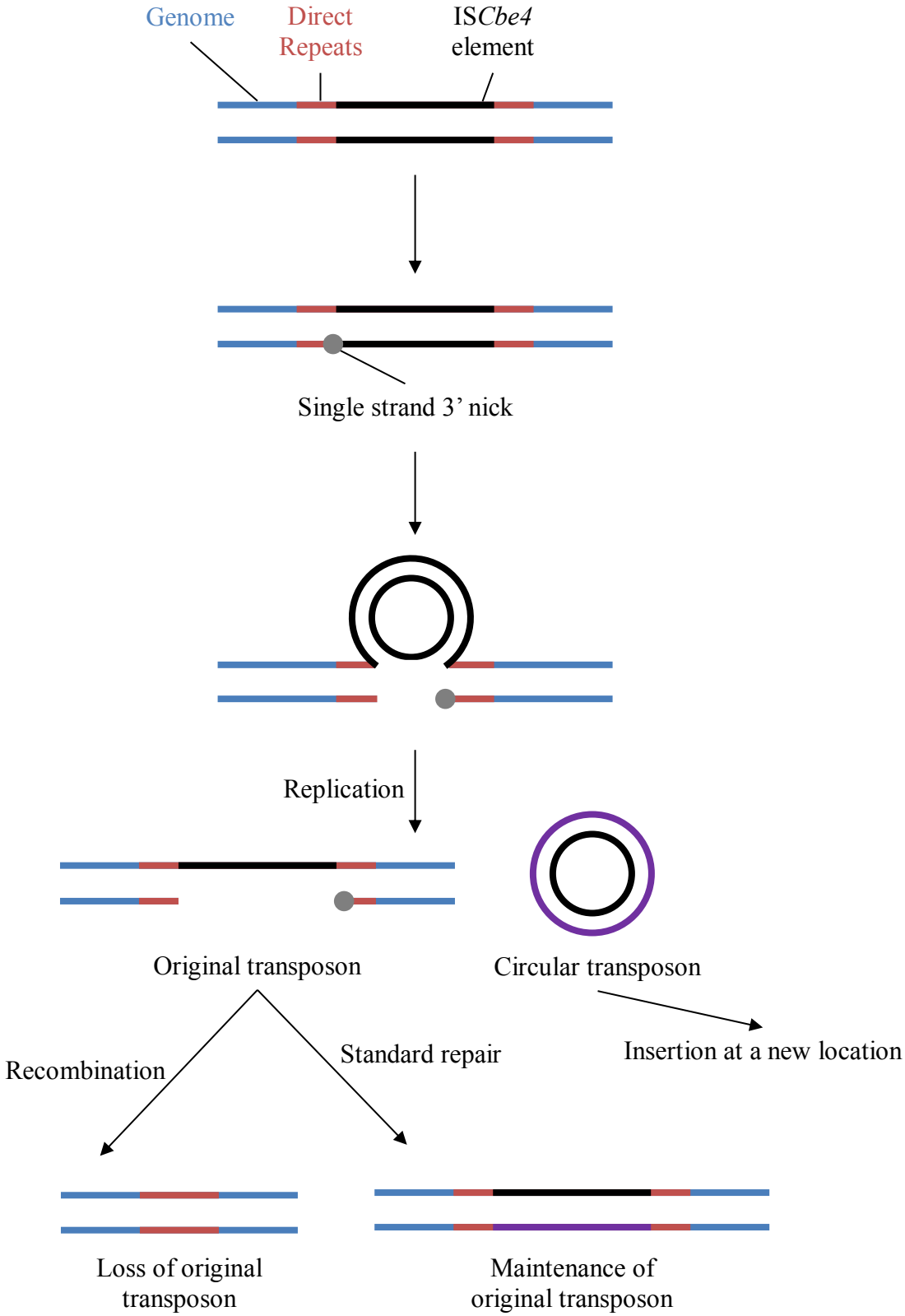
Figure 6.1 Growth of *C. bescii* in high salt and high glucose.

C. bescii was grown as described in Chapter 2 on standard growth medium containing 0-120 mM sodium chloride or 28-500 mM glucose.

Figure 6.2 Proposed mode of action for IS*Cbe4* insertion element.

Double stranded IS*Cbe4* element shown with the genome in blue, direct repeats in red and the insertion element in black. The grey circle represents a free 3' hydroxyl group. The purple denotes the new strand of a replicated transposon. This is a 'copy-out, paste-in' mechanism with the figure modified from (237).

Figure 6.2


Genome
Direct Repeats
IS*Cbe4* element
Single strand 3' nick
Replication
Original transposon
Circular transposon
Recombination
Standard repair
Insertion at a new location
Loss of original transposon
Maintenance of original transposon

REFERENCES

1. Sato M. 1990. Thermochemistry of the formation of fossil fuels. The Geochemical Society Special Publication U.S. Geological Survey, 2:271-283.
2. Demain AL. 2009. Biosolutions to the energy problem. J Ind Microbiol Biotechnol 36:319-332.
3. Demirbas A. 2009. Political, economic and environmental impacts of biofuels: A review. Appl Energy 86:S108-S117.
4. Monthly Energy Review. https://www.eia.gov/. Accessed 9/14/2017.
5. Dien BS, Cotta MA, Jeffries TW. 2003. Bacteria engineered for fuel ethanol production: current status. Appl Microbiol Biotechnol 63:258-266.
6. Zaldivar J, Nielsen J, Olsson L. 2001. Fuel ethanol production from lignocellulose: a challenge for metabolic engineering and process integration. Appl Microbiol Biotechnol 56:17-34.
7. Bothast RJ, Nichols NN, Dien BS. 1999. Fermentations with new recombinant organisms. Biotechnol Prog 15:867-875.
8. Huang W, Zhang Y-HP. 2011. Analysis of biofuels production from sugar based on three criteria: Thermodynamics, bioenergetics, and product separation, vol 4.
9. OECD. 2015. Biofuels. OECD-FAO Agricultural Outlook 2015 OECD Publishing.
10. Swinton SM, Babcock BA, James LK, Bandaru V. 2011. Higher US crop prices trigger little area expansion so marginal land for biofuel crops is limited. Energy Policy 39:5254-5258.

11. Rosegrant MW. 2008. Biofuels and grain prices: impacts and policy responses. International Food Policy Research Institute Washington, DC.

12. Hill J, Nelson E, Tilman D, Polasky S, Tiffany D. 2006. Environmental, economic, and energetic costs and benefits of biodiesel and ethanol biofuels. Proc Natl Acad Sci 103:11206-11210.

13. Robertson GP, Dale VH, Doering OC, Hamburg SP, Melillo JM, Wander MM, Parton WJ, Adler PR, Barney JN, Cruse RM, Duke CS, Fearnside PM, Follett RF, Gibbs HK, Goldemberg J, Mladenoff DJ, Ojima D, Palmer MW, Sharpley A, Wallace L, Weathers KC, Wiens JA, Wilhelm WW. 2008. Sustainable Biofuels Redux. Science 322:49-50.

14. Food Prices. https://ourworldindata.org/food-prices/. Accessed 9/21/2017.

15. Demirbas A. 2008. Biofuels sources, biofuel policy, biofuel economy and global biofuel projections. Energy Convers Manag 49:2106-2116.

16. Carroll A, Somerville C. 2009. Cellulosic Biofuels. Annu Rev Plant Biol 60:165-182.

17. Favaro L, Cagnin L, Basaglia M, Pizzocchero V, van Zyl WH, Casella S. 2017. Production of bioethanol from multiple waste streams of rice milling. Bioresour Technol 244:151-159.

18. Mattila H, Kuuskeri J, Lundell T. 2017. Single-step, single-organism bioethanol production and bioconversion of lignocellulose waste materials by phlebioid fungal species. Bioresour Technol 225:254-261.

19. Edwards MC, Williams T, Pattathil S, Hahn MG, Doran-Peterson J. 2014. Replacing a suite of commercial pectinases with a single enzyme, pectate lyase B, in *Saccharomyces cerevisiae* fermentations of cull peaches. J Ind Microbiol Biotechnol 41:679-86.

20. Wang D, Lu Y, Huang H, Xu J. 2012. Establishing oleaginous microalgae research models for consolidated bioprocessing of solar energy. Adv Biochem Eng Biotechnol 128:69-84.

21. Hansen TL, Jansen JlC, Davidsson Å, Christensen TH. 2007. Effects of pre-treatment technologies on quantity and quality of source-sorted municipal organic waste for biogas recovery. Waste Manag 27:398-405.

22. Council NR. 1996. Use of reclaimed water and sludge in food crop production. National Academies Press.

23. Woertz I, Feffer A, Lundquist T, Nelson Y. 2009. Algae grown on dairy and municipal wastewater for simultaneous nutrient removal and lipid production for biofuel feedstock. J Environ Eng 135:1115-1122.

24. Brunecky R, Hobdey SE, Taylor LE, 2nd, Tao L, Tucker MP, Himmel ME, Decker SR. 2014. High temperature pre-digestion of corn stover biomass for improved product yields. Biotechnol Biofuels 7:170.

25. Jin M, Balan V, Gunawan C, Dale BE. 2011. Consolidated bioprocessing (CBP) performance of *Clostridium phytofermentans* on AFEX-treated corn stover for ethanol production. Biotechnol Bioeng 108:1290-7.

26. Jin M, Gunawan C, Balan V, Dale BE. 2012. Consolidated bioprocessing (CBP) of AFEX-pretreated corn stover for ethanol production using *Clostridium phytofermentans* at a high solids loading. Biotechnol Bioeng 109:1929-36.

27. Papa G, Rodriguez S, George A, Schievano A, Orzi V, Sale KL, Singh S, Adani F, Simmons BA. 2015. Comparison of different pretreatments for the production of bioethanol and biomethane from corn stover and switchgrass. Bioresour Technol 183:101-110.

28. Speers AM, Reguera G. 2012. Consolidated bioprocessing of AFEX-pretreated corn stover to ethanol and hydrogen in a microbial electrolysis cell. Environ Sci Technol 46:7875-81.

29. Cruse RM, Herndl CG. 2009. Balancing corn stover harvest for biofuels with soil and water conservation. J Soil Water Conserv 64:286-291.

30. Campbell JE, Lobell DB, Genova RC, Field CB. 2008. The Global Potential of Bioenergy on Abandoned Agriculture Lands. Environ Sci Technol 42:5791-5794.

31. Sannigrahi P, Ragauskas AJ, Tuskan GA. 2010. Poplar as a feedstock for biofuels: a review of compositional characteristics. Biofuels Bioprod Biorefin 4:209-226.

32. Mitchell RB, Schmer MR, Anderson WF, Jin V, Balkcom KS, Kiniry J, Coffin A, White P. 2016. Dedicated energy crops and crop residues for bioenergy feedstocks in the central and eastern USA. BioEnergy Res 9:384-398.

33. Schmer MR, Liebig MA, Vogel KP, Mitchell RB. 2011. Field-scale soil property changes under switchgrass managed for bioenergy. Glob Change Biol Bioenergy 3:439-448.

34. Follett RF, Vogel KP, Varvel GE, Mitchell RB, Kimble J. 2012. Soil carbon sequestration by switchgrass and no-till maize grown for bioenergy. BioEnergy Res 5:866-875.

35. Mitchell R, Vogel KP, Sarath G. 2008. Managing and enhancing switchgrass as a bioenergy feedstock. Biofuels, Bioprod Biorefin 2:530-539.

36. Klemm D, Heublein B, Fink H-P, Bohn A. 2005. Cellulose: fascinating biopolymer and sustainable raw material. Angew Chem Int Ed 44:3358-3393.

37. Zhao X, Zhang L, Liu D. 2012. Biomass recalcitrance. Part I: the chemical compositions and physical structures affecting the enzymatic hydrolysis of lignocellulose. Biofuels Bioprod Biorefin 6:465-482.

38. Davison BH, Parks J, Davis MF, Donohoe BS. 2013. Plant cell walls: Basics of structure, chemistry, accessibility and the influence on conversion. Aqueous Pretreatment of Plant

Biomass for Biological and Chemical Conversion to Fuels and Chemicals John Wiley & Sons, Ltd:23-38.

39. Arora R, Behera S, Kumar S. 2015. Bioprospecting thermophilic/thermotolerant microbes for production of lignocellulosic ethanol: A future perspective. Renew Sustain Energy Rev 51:699-717.

40. Brethauer S, Studer MH. 2015. Biochemical conversion processes of lignocellulosic biomass to fuels and chemicals - a review. Chimia (Aarau) 69:572-81.

41. Lynd LR. 1996. Overview and evaluation of fuel ethanol from cellulosic biomass: technology, economics, the environment, and policy. Annu Rev Energy Environ 21:403-465.

42. Lynd LR, van Zyl WH, McBride JE, Laser M. 2005. Consolidated bioprocessing of cellulosic biomass: an update. Curr Opin Biotechnol 16:577-83.

43. Lynd LR, Weimer PJ, van Zyl WH, Pretorius IS. 2002. Microbial cellulose utilization: Fundamentals and biotechnology. Microbiol Mol Biol Rev 66:506-577.

44. Olson DG, McBride JE, Shaw AJ, Lynd LR. 2012. Recent progress in consolidated bioprocessing. Curr Opin Biotechnol 23:396-405.

45. Poovaiah CR, Nageswara-Rao M, Soneji JR, Baxter HL, Stewart CN. 2014. Altered lignin biosynthesis using biotechnology to improve lignocellulosic biofuel feedstocks. Plant Biotechnol J 12:1163-1173.

46. den Haan R, van Rensburg E, Rose SH, Gorgens JF, van Zyl WH. 2015. Progress and challenges in the engineering of non-cellulolytic microorganisms for consolidated bioprocessing. Curr Opin Biotechnol 33:32-8.

47. Jones L, Ennos AR, Turner SR. 2001. Cloning and characterization of irregular xylem4 (irx4): a severely lignin-deficient mutant of Arabidopsis. Plant J 26:205-216.

48. Li M, Pu Y, Yoo CG, Gjersing E, Decker SR, Doeppke C, Shollenberger T, Tschaplinski TJ, Engle NL, Sykes RW, Davis MF, Baxter HL, Mazarei M, Fu C, Dixon RA, Wang ZY, Neal Stewart C, Jr., Ragauskas AJ. 2017. Study of traits and recalcitrance reduction of field-grown COMT down-regulated switchgrass. Biotechnol Biofuels 10:12.

49. Loqué D, Scheller HV, Pauly M. 2015. Engineering of plant cell walls for enhanced biofuel production. Curr Opin Plant Biol 25:151-161.

50. Fu C, Mielenz JR, Xiao X, Ge Y, Hamilton CY, Rodriguez M, Chen F, Foston M, Ragauskas A, Bouton J, Dixon RA, Wang Z-Y. 2011. Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. Proc Natl Acad Sci 108:3803-3808.

51. Adams MWW, Biswal AK, Gelineo-Albersheim I, Hao Z, Hunt KD, Kataeva I, Mohnen DA. 2012. Plants with altered cell wall biosynthesis and methods of use. Google Patents.

52. Fan Z, Yuan L. 2010. Production of multifunctional chimaeric enzymes in plants: a promising approach for degrading plant cell wall from within. Plant Biotechnol J 8:308-315.

53. Gray BN, Ahner BA, Hanson MR. 2009. High-level bacterial cellulase accumulation in chloroplast-transformed tobacco mediated by downstream box fusions. Biotechnol Bioeng 102:1045-54.

54. Alizadeh H, Teymouri F, Gilbert TI, Dale BE. 2005. Pretreatment of switchgrass by ammonia fiber explosion (AFEX). Appl Biochem Biotechnol 121-124:1133-41.

55. Dougherty MJ, Tran HM, Stavila V, Knierim B, George A, Auer M, Adams PD, Hadi MZ. 2014. Cellulosic biomass pretreatment and sugar yields as a function of biomass particle size. PLoS One 9:e100836.

56. Nguyen TY, Cai CM, Kumar R, Wyman CE. 2015. Co-solvent pretreatment reduces costly enzyme requirements for high sugar and ethanol yields from lignocellulosic biomass. ChemSusChem 8:1716-25.

57. Paudel SR, Banjara SP, Choi OK, Park KY, Kim YM, Lee JW. 2017. Pretreatment of agricultural biomass for anaerobic digestion: Current state and challenges. Bioresour Technol doi:https://doi.org/10.1016/j.biortech.2017.08.182.

58. Haghighi Mood S, Hossein Golfeshan A, Tabatabaei M, Salehi Jouzani G, Najafi GH, Gholami M, Ardjmand M. 2013. Lignocellulosic biomass to bioethanol, a comprehensive review with a focus on pretreatment. Renew Sustain Energy Rev 27:77-93.

59. Wyman CE. 1994. Ethanol from lignocellulosic biomass: technology, economics, and opportunities. Bioresour Technol 50:3-15.

60. D. Wright J, Wyman C, Grohmann K. 1988. Simultaneous saccharification and fermentation of lignocellulose - Process evaluation, vol 18.

61. Favaro L, Jooste T, Basaglia M, Rose SH, Saayman M, Gorgens JF, Casella S, van Zyl WH. 2013. Designing industrial yeasts for the consolidated bioprocessing of starchy biomass to ethanol. Bioengineered 4:97-102.

62. Lee J, Sim SJ, Bott M, Um Y, Oh MK, Woo HM. 2014. Succinate production from CO(2)-grown microalgal biomass as carbon source using engineered *Corynebacterium glutamicum* through consolidated bioprocessing. Sci Rep 4:5819.

63. Luo Z, Zhang Y, Bao J. 2014. Extracellular secretion of beta-glucosidase in ethanologenic *E. coli* enhances ethanol fermentation of cellobiose. Appl Biochem Biotechnol 174:772-83.

64. Lynd LR, Wyman CE, Gerngross TU. 1999. Biocommodity engineering. Biotechnol Prog 15:777-793.

65. VanderHoeven D. 2016. Cellulosic ethanol: feedstock costs predominant, sugar cane cheapest. Bio Based Press, https://www.biobasedpress.eu/2016/02/cellulosic-ethanol-feedstock-costs-predominant-sugar-cane-cheapest/.

66. LuxResearch. 2016. Lux: Cellulosic ethanol price hinges on feedstock cost, http://biomassmagazine.com/articles/12958/lux-cellulosic-ethanol-price-hinges-on-feedstock-cost.

67. Weekly Retail Gasoline and Diesel Prices. https://www.eia.gov/. Accessed 8/29/2017.

68. Hasunuma T, Okazaki F, Okai N, Hara KY, Ishii J, Kondo A. 2013. A review of enzymes and microbes for lignocellulosic biorefinery and the possibility of their application to consolidated bioprocessing technology. Bioresour Technol 135:513-22.

69. Lynd LR, Liang X, Biddy MJ, Allee A, Cai H, Foust T, Himmel ME, Laser MS, Wang M, Wyman CE. 2017. Cellulosic ethanol: status and innovation. Curr Opin Biotechnol 45:202-211.

70. Qiu Z, Jiang R. 2017. Improving *Saccharomyces cerevisiae* ethanol production and tolerance via RNA polymerase II subunit Rpb7. Biotechnol Biofuels 10:125.

71. Lambertz C, Garvey M, Klinger J, Heesel D, Klose H, Fischer R, Commandeur U. 2014. Challenges and advances in the heterologous expression of cellulolytic enzymes: a review. Biotechnol Biofuels 7:135.

72. la Grange DC, den Haan R, van Zyl WH. 2010. Engineering cellulolytic ability into bioprocessing organisms. Appl Microbiol Biotechnol 87:1195-208.

73. Mazzoli R, Lamberti C, Pessione E. 2012. Engineering new metabolic capabilities in bacteria: lessons from recombinant cellulolytic strategies. Trends Biotechnol 30:111-9.

74. Zeldes BM, Keller MW, Loder AJ, Straub CT, Adams MW, Kelly RM. 2015. Extremely thermophilic microorganisms as metabolic engineering platforms for production of fuels and industrial chemicals. Front Microbiol 6:1209.

75. Antonov E, Schlembach I, Regestein L, Rosenbaum MA, Buchs J. 2017. Process relevant screening of cellulolytic organisms for consolidated bioprocessing. Biotechnol Biofuels 10:106.

76. Lee CR, Sung BH, Lim KM, Kim MJ, Sohn MJ, Bae JH, Sohn JH. 2017. Co-fermentation using recombinant *Saccharomyces cerevisiae* yeast strains hyper-secreting different cellulases for the production of cellulosic bioethanol. Sci Rep 7:4428.

77. Zverlov VV, Kellermann J, Schwarz WH. 2005. Functional subgenomics of *Clostridium thermocellum* cellulosomal genes: Identification of the major catalytic components in the extracellular complex and detection of three new enzymes. J Proteomics 5:3646-3653.

78. Nakazawa H, Okada K, Kobayashi R, Kubota T, Onodera T, Ochiai N, Omata N, Ogasawara W, Okada H, Morikawa Y. 2008. Characterization of the catalytic domains of *Trichoderma reesei* endoglucanase I, II, and III, expressed in *Escherichia coli*. Appl Microbiol Biotechnol 81:681-9.

79. Chen X. 2017. Yeast cell surface display: An efficient strategy for improvement of bioethanol fermentation performance. Bioengineered 8:115-119.

80. Hong J, Wang Y, Kumagai H, Tamaki H. 2007. Construction of thermotolerant yeast expressing thermostable cellulase genes. J Biotechnol 130:114-23.

81. Liu Z, Inokuma K, Ho SH, den Haan R, van Zyl WH, Hasunuma T, Kondo A. 2017. Improvement of ethanol production from crystalline cellulose via optimizing cellulase ratios in cellulolytic *Saccharomyces cerevisiae*. Biotechnol Bioeng 114:1201-1207.

82. Matano Y, Hasunuma T, Kondo A. 2013. Cell recycle batch fermentation of high-solid lignocellulose using a recombinant cellulase-displaying yeast strain for high yield ethanol production in consolidated bioprocessing. Bioresour Technol 135:403-9.

83. Wei H, Wang W, Alahuhta M, Vander Wall T, Baker JO, Taylor LE, 2nd, Decker SR, Himmel ME, Zhang M. 2014. Engineering towards a complete heterologous cellulase secretome in *Yarrowia lipolytica* reveals its potential for consolidated bioprocessing. Biotechnol Biofuels 7:148.

84. Blumer-Schuette SE, Alahuhta M, Conway JM, Lee LL, Zurawski JV, Giannone RJ, Hettich RL, Lunin VV, Himmel ME, Kelly RM. 2015. Discrete and structurally unique proteins (tapirins) mediate attachment of extremely thermophilic *Caldicellulosiruptor* species to cellulose. J Biol Chem 290:10645-56.

85. Conway JM, Pierce WS, Le JH, Harper GW, Wright JH, Tucker AL, Zurawski JV, Lee LL, Blumer-Schuette SE, Kelly RM. 2016. Multi-domain, surface layer associated glycoside hydrolases contribute to plant polysaccharide degradation by *Caldicellulosiruptor* species. J Biol Chem 291:6732-6747.

86. Lu Y, Zhang Y-HP, Lynd LR. 2006. Enzyme–microbe synergy during cellulose hydrolysis by *Clostridium thermocellum*. Proc Natl Acad Sci 103:16165-16169.

87. Fan L-H, Zhang Z-J, Yu X-Y, Xue Y-X, Tan T-W. 2012. Self-surface assembly of cellulosomes with two miniscaffoldins on *Saccharomyces cerevisiae* for cellulosic ethanol production. Proc Natl Acad Sci 109:13260-13265.

88. Kotaka A, Bando H, Kaya M, Kato-Murai M, Kuroda K, Sahara H, Hata Y, Kondo A, Ueda M. 2008. Direct ethanol production from barley β-glucan by sake yeast displaying *Aspergillus oryzae* β-glucosidase and endoglucanase. J Biosci Bioeng 105:622-627.

89. Vinuselvi P, Lee SK. 2012. Engineered *Escherichia coli* capable of co-utilization of cellobiose and xylose. Enzyme Microb Technol 50:1-4.

90. Fan LH, Zhang ZJ, Mei S, Lu YY, Li M, Wang ZY, Yang JG, Yang ST, Tan TW. 2016. Engineering yeast with bifunctional minicellulosome and cellodextrin pathway for co-utilization of cellulose-mixed sugars. Biotechnol Biofuels 9:137.

91. Wei N, Oh EJ, Million G, Cate JH, Jin YS. 2015. Simultaneous utilization of cellobiose, xylose, and acetic acid from lignocellulosic biomass for biofuel production by an engineered yeast platform. ACS Synth Biol 4:707-13.

92. Caspeta L, Castillo T, Nielsen J. 2015. Modifying yeast tolerance to inhibitory conditions of ethanol production processes. Front Bioeng Biotechnol 3:184.

93. Kitagawa T, Tokuhiro K, Sugiyama H, Kohda K, Isono N, Hisamatsu M, Takahashi H, Imaeda T. 2010. Construction of a beta-glucosidase expression system using the multistress-tolerant yeast *Issatchenkia orientalis*. Appl Microbiol Biotechnol 87:1841-53.

94. Liu ZL, Moon J, Andersh BJ, Slininger PJ, Weber S. 2008. Multiple gene-mediated NAD(P)H-dependent aldehyde reduction is a mechanism of in situ detoxification of furfural and 5-hydroxymethylfurfural by *Saccharomyces cerevisiae*. Appl Microbiol Biotechnol 81:743-53.

95. Yona AH, Manor YS, Herbst RH, Romano GH, Mitchell A, Kupiec M, Pilpel Y, Dahan O. 2012. Chromosomal duplication is a transient evolutionary solution to stress. Proc Natl Acad Sci 109:21010-21015.

96. Linville JL, Rodriguez M, Jr., Land M, Syed MH, Engle NL, Tschaplinski TJ, Mielenz JR, Cox CD. 2013. Industrial robustness: understanding the mechanism of tolerance for the Populus hydrolysate-tolerant mutant strain of *Clostridium thermocellum*. PLoS One 8:e78829.

97. Clarkson SM, Hamilton-Brehm SD, Giannone RJ, Engle NL, Tschaplinski TJ, Hettich RL, Elkins JG. 2014. A comparative multidimensional LC-MS proteomic analysis reveals mechanisms for furan aldehyde detoxification in *Thermoanaerobacter pseudethanolicus* 39E. Biotechnol Biofuels 7:165.

98. Chung D, Verbeke TJ, Cross KL, Westpheling J, Elkins JG. 2015. Expression of a heat-stable NADPH-dependent alcohol dehydrogenase in *Caldicellulosiruptor bescii* results in furan aldehyde detoxification. Biotechnol Biofuels 8:102.

99. Kim SK, Groom J, Chung D, Elkins J, Westpheling J. 2017. Expression of a heat-stable NADPH-dependent alcohol dehydrogenase from *Thermoanaerobacter pseudethanolicus* 39E in *Clostridium thermocellum* 1313 results in increased hydroxymethylfurfural resistance. Biotechnol Biofuels 10:66.

100. Anasontzis GE, Christakopoulos P. 2014. Challenges in ethanol production with *Fusarium oxysporum* through consolidated bioprocessing. Bioengineered 5:393-5.

101. Anasontzis GE, Kourtoglou E, Villas-Boas SG, Hatzinikolaou DG, Christakopoulos P. 2016. Metabolic engineering of *Fusarium oxysporum* to improve its ethanol-producing capability. Front Microbiol 7:632.

102. Anasontzis GE, Zerva A, Stathopoulou PM, Haralampidis K, Diallinas G, Karagouni AD, Hatzinikolaou DG. 2011. Homologous overexpression of xylanase in *Fusarium oxysporum* increases ethanol productivity during consolidated bioprocessing (CBP) of lignocellulosics. J Biotechnol 152:16-23.

103. Carere CR, Rydzak T, Verbeke TJ, Cicek N, Levin DB, Sparling R. 2012. Linking genome content to biofuel production yields: a meta-analysis of major catabolic pathways among select H2and ethanol-producing bacteria. BMC Microbiol 12:295.

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

105. Rhaesa AM, Basen M, Kataeva I, Prybol CJ, Scott IM, Poole FL, Adams MW. 2014. Degradation of high loads of crystalline cellulose and of unpretreated plant biomass by the thermophilic bacterium *Caldicellulosiruptor bescii*. Bioresour Technol 152:384-92.

106. Kataeva I, Foston MB, Yang S-J, Pattathil S, Biswal AK, Poole Ii FL, Basen M, Rhaesa AM, Thomas TP, Azadi P, Olman V, Saffold TD, Mohler KE, Lewis DL, Doeppke C, Zeng Y, Tschaplinski TJ, York WS, Davis M, Mohnen D, Xu Y, Ragauskas AJ, Ding S-Y, Kelly RM, Hahn MG, Adams MWW. 2013. Carbohydrate and lignin are simultaneously solubilized from unpretreated switchgrass by microbial action at high temperature. Energy Environ Sci 6:2186-2195.

107. Blumer-Schuette SE, Lewis DL, Kelly RM. 2010. Phylogenetic, microbiological, and glycoside hydrolase diversities within the extremely thermophilic, plant biomass-degrading Genus *Caldicellulosiruptor*. Appl Environ Microbiol 76:8084-8092.

108. Chung D, Young J, Cha M, Brunecky R, Bomble YJ, Himmel ME, Westpheling J. 2015. Expression of the *Acidothermus cellulolyticus* E1 endoglucanase in *Caldicellulosiruptor bescii* enhances its ability to deconstruct crystalline cellulose. Biotechnol Biofuels 8:113.

109. Kim SK, Chung D, Himmel ME, Bomble YJ, Westpheling J. 2016. Engineering the N-terminal end of CelA results in improved performance and growth of *Caldicellulosiruptor bescii* on crystalline cellulose. Biotechnol Bioeng doi:10.1002/bit.26242.

110. Kim SK, Chung D, Himmel ME, Bomble YJ, Westpheling J. 2017. In vivo synergistic activity of a CAZyme cassette from *Acidothermus cellulolyticus* significantly improves the cellulolytic activity of the *C. bescii* exoproteome. Biotechnol Bioeng doi:10.1002/bit.26366.

111. Maki ML, Armstrong L, Leung KT, Qin W. 2013. Increased expression of beta-glucosidase A in *Clostridium thermocellum* 27405 significantly increases cellulase activity. Bioengineered 4:15-20.

112. Thomas L, Joseph A, Gottumukkala LD. 2014. Xylanase and cellulase systems of *Clostridium* sp.: an insight on molecular approaches for strain improvement. Bioresour Technol 158:343-50.

113. Ali SS, Nugent B, Mullins E, Doohan FM. 2013. Insights from the fungus *Fusarium oxysporum* point to high affinity glucose transporters as targets for enhancing ethanol production from lignocellulose. PLoS One 8:e54701.

114. Ali SS, Nugent B, Mullins E, Doohan FM. 2016. Fungal-mediated consolidated bioprocessing: the potential of *Fusarium oxysporum* for the lignocellulosic ethanol industry. AMB Express 6:13.

115. de Almeida MN, Guimaraes VM, Falkoski DL, Visser EM, Siqueira GA, Milagres AM, de Rezende ST. 2013. Direct ethanol production from glucose, xylose and sugarcane bagasse by the corn endophytic fungi *Fusarium verticillioides* and *Acremonium zeae*. J Biotechnol 168:71-7.

116. Puseenam A, Tanapongpipat S, Roongsawang N. 2015. Co-expression of endoxylanase and endoglucanase in *Scheffersomyces stipitis* and its application in ethanol production. Appl Biochem Biotechnol 177:1690-700.

117. Wu W, Davis RW, Tran-Gyamfi MB, Kuo A, LaButti K, Mihaltcheva S, Hundley H, Chovatia M, Lindquist E, Barry K, Grigoriev IV, Henrissat B, Gladden JM. 2017. Characterization of four endophytic fungi as potential consolidated bioprocessing hosts for conversion of lignocellulose into advanced biofuels. Appl Microbiol Biotechnol 101:2603-2618.

118. Cavka A, Alriksson B, Rose SH, van Zyl WH, Jönsson LJ. 2011. Biorefining of wood: combined production of ethanol and xylanase from waste fiber sludge. Journal of Industrial Microbiology & Biotechnology 38:891-899.

119. Blumer-Schuette SE, Brown SD, Sander KB, Bayer EA, Kataeva I, Zurawski JV, Conway JM, Adams MW, Kelly RM. 2014. Thermophilic lignocellulose deconstruction. FEMS Microbiol Rev 38:393-448.

120. Gomes E, de Souza AR, Orjuela GL, Da Silva R, de Oliveira TB, Rodrigues A. 2016. Applications and benefits of thermophilic microorganisms and their enzymes for industrial biotechnology. Gene Expression Systems in Fungi: Advancements and Applications Springer International Publishing:459-492.

121. Kristjansson JK. 1989. Thermophilic organisms as sources of thermostable enzymes. Trends Biotechnol 7:349-353.

122. Turner P, Mamo G, Karlsson EN. 2007. Potential and utilization of thermophiles and thermostable enzymes in biorefining. Microb Cell Fact 6:9.

123. Chien A, Edgar DB, Trela JM. 1976. Deoxyribonucleic acid polymerase from the extreme thermophile *Thermus aquaticus*. J Bacteriol 127:1550-1557.

124. Hoseki J, Yano T, Koyama Y, Kuramitsu S, Kagamiyama H. 1999. Directed evolution of thermostable kanamycin-resistance gene: a convenient selection marker for *Thermus thermophilus*. J Biochem 126:951-6.

125. Matsumi R, Manabe K, Fukui T, Atomi H, Imanaka T. 2007. Disruption of a sugar transporter gene cluster in a hyperthermophilic archaeon using a host-marker system based on antibiotic resistance. J Bacteriol 189:2683-2691.

126. Waege I, Schmid G, Thumann S, Thomm M, Hausner W. 2010. Shuttle vector-based transformation system for *Pyrococcus furiosus*. Appl Environ Microbiol 76:3308-3313.

127. Lipscomb GL, Conway JM, Blumer-Schuette SE, Kelly RM, Adams MW. 2016. A highly thermostable kanamycin resistance marker expands the tool kit for genetic manipulation of *Caldicellulosiruptor bescii*. Appl Environ Microbiol 82:4421-8.

128. Liao H, McKenzie T, Hageman R. 1986. Isolation of a thermostable enzyme variant by cloning and selection in a thermophile. Proc Natl Acad Sci USA 83:576-80.

129. Iwasaki Y, Kita A, Sakai S, Takaoka K, Yano S, Tajima T, Kato J, Nishio N, Murakami K, Nakashimada Y. 2013. Engineering of a functional thermostable kanamycin resistance marker for use in *Moorella thermoacetica* ATCC39073. FEMS Microbiol Lett 343:8-12.

130. Liu B, Wang C, Yang H, Tan H. 2012. Establishment of a genetic transformation system and its application in *Thermoanaerobacter tengcongensis*. J Genet Genomics 39:561-70.

131. Taylor MP, Esteban CD, Leak DJ. 2008. Development of a versatile shuttle vector for gene expression in *Geobacillus* spp. Plasmid 60:45-52.

132. Hashimoto Y, Yano T, Kuramitsu S, Kagamiyama H. 2001. Disruption of *Thermus thermophilus* genes by homologous recombination using a thermostable kanamycin-resistant marker. FEBS Lett 506:231-4.

133. Han D, Norris SM, Xu Z. 2012. Construction and transformation of a *Thermotoga-E. coli* shuttle vector. BMC Biotechnol 12:2-2.

134. Akinosho H, Yee K, Close D, Ragauskas A. 2014. The emergence of *Clostridium thermocellum* as a high utility candidate for consolidated bioprocessing applications. Front Chem 2:66.

135. Verbeke TJ, Zhang X, Henrissat B, Spicer V, Rydzak T, Krokhin OV, Fristensky B, Levin DB, Sparling R. 2013. Genomic evaluation of *Thermoanaerobacter* spp. for the construction of designer co-cultures to improve lignocellulosic biofuel production. PLoS One 8:e59362.

136. Blumer-Schuette SE, Giannone RJ, Zurawski JV, Ozdemir I, Ma Q, Yin YB, Xu Y, Kataeva I, Poole FL, Adams MWW, Hamilton-Brehm SD, Elkins JG, Larimer FW, Land ML, Hauser LJ, Cottingham RW, Hettich RL, Kelly RM. 2012. *Caldicellulosiruptor* core and pangenomes reveal determinants for noncellulosomal thermophilic deconstruction of plant biomass. J Bacteriol 194:4015-4028.

137. Thauer RK, Jungermann K, Decker K. 1977. Energy conservation in chemotrophic anaerobic bacteria. Bacteriol Rev 41:100-180.

138. Ng TK, Weimer PJ, Zeikus JG. 1977. Cellulolytic and physiological properties of *Clostridium thermocellum*. Arch Microbiol 114:1-7.

139. Bayer EA, Kenig R, Lamed R. 1983. Adherence of *Clostridium thermocellum* to cellulose. J Bacteriol 156:818-827.

140. Currie DH, Herring CD, Guss AM, Olson DG, Hogsett DA, Lynd LR. 2013. Functional heterologous expression of an engineered full length CipA from *Clostridium thermocellum* in *Thermoanaerobacterium saccharolyticum*. Biotechnol Biofuels 6:32.

141. Fierobe HP, Mingardon F, Chanal A. 2012. Engineering cellulase activity into *Clostridium acetobutylicum*. Methods Enzymol 510:301-16.

142. Hong W, Zhang J, Feng Y, Mohr G, Lambowitz AM, Cui GZ, Liu YJ, Cui Q. 2014. The contribution of cellulosomal scaffoldins to cellulose hydrolysis by *Clostridium thermocellum* analyzed by using thermotargetrons. Biotechnol Biofuels 7:80.

143. Hyeon JE, Yu KO, Suh DJ, Suh YW, Lee SE, Lee J, Han SO. 2010. Production of minicellulosomes from *Clostridium cellulovorans* for the fermentation of cellulosic ethanol using engineered recombinant *Saccharomyces cerevisiae*. FEMS Microbiol Lett 310:39-47.

144. Kovacs K, Willson BJ, Schwarz K, Heap JT, Jackson A, Bolam DN, Winzer K, Minton NP. 2013. Secretion and assembly of functional mini-cellulosomes from synthetic chromosomal operons in *Clostridium acetobutylicum* ATCC 824. Biotechnol Biofuels 6:117.

145. Liu JM, Xin XJ, Li CX, Xu JH, Bao J. 2012. Cloning of thermostable cellulase genes of *Clostridium thermocellum* and their secretive expression in *Bacillus subtilis*. Appl Biochem Biotechnol 166:652-62.

146. Willson BJ, Kovacs K, Wilding-Steele T, Markus R, Winzer K, Minton NP. 2016. Production of a functional cell wall-anchored minicellulosome by recombinant *Clostridium acetobutylicum* ATCC 824. Biotechnol Biofuels 9:109.

147. Morais S, Barak Y, Caspi J, Hadar Y, Lamed R, Shoham Y, Wilson DB, Bayer EA. 2010. Cellulase-xylanase synergy in designer cellulosomes for enhanced degradation of a complex cellulosic substrate. MBio 1:00285-10.

148. Vazana Y, Moraïs S, Barak Y, Lamed R, Bayer EA. 2012. Chapter twenty-three - Designer cellulosomes for enhanced hydrolysis of cellulosic substrates. Methods Enzymol Academic Press, 510:429-452.

149. Holwerda EK, Thorne PG, Olson DG, Amador-Noguez D, Engle NL, Tschaplinski TJ, van Dijken JP, Lynd LR. 2014. The exometabolome of Clostridium thermocellum reveals overflow metabolism at high cellulose loading. Biotechnology for Biofuels 7:155.

150. Taillefer M, Rydzak T, Levin DB, Oresnik IJ, Sparling R. 2015. Reassessment of the transhydrogenase/malate shunt pathway in *Clostridium thermocellum* ATCC 27405 through kinetic characterization of malic enzyme and malate dehydrogenase. Appl Environ Microbiol 81:2423-32.

151. Tyurin MV, Desai SG, Lynd LR. 2004. Electrotransformation of *Clostridium thermocellum*. Appl Environ Microbiol 70:883-890.

152. Brown SD, Guss AM, Karpinets TV, Parks JM, Smolin N, Yang S, Land ML, Klingeman DM, Bhandiwad A, Rodriguez M, Jr., Raman B, Shao X, Mielenz JR, Smith JC, Keller M, Lynd LR. 2011. Mutant alcohol dehydrogenase leads to improved ethanol tolerance in *Clostridium thermocellum*. Proc Natl Acad Sci USA 108:13752-7.

153. Shao X, Raman B, Zhu M, Mielenz JR, Brown SD, Guss AM, Lynd LR. 2011. Mutant selection and phenotypic and genetic characterization of ethanol-tolerant strains of *Clostridium thermocellum*. Appl Microbiol Biotechnol 92:641-52.

154. Timmons MD, Knutson BL, Nokes SE, Strobel HJ, Lynn BC. 2009. Analysis of composition and structure of *Clostridium thermocellum* membranes from wild-type and ethanol-adapted strains. Appl Microbiol Biotechnol 82:929-39.

155. Zhu X, Cui J, Feng Y, Fa Y, Zhang J, Cui Q. 2013. Metabolic adaption of ethanol-tolerant *Clostridium thermocellum*. PLoS One 8:e70631.

156. Biswas R, Prabhu S, Lynd LR, Guss AM. 2014. Increase in ethanol yield via elimination of lactate production in an ethanol-tolerant mutant of *Clostridium thermocellum*. PLoS One 9:e86389.

157. Biswas R, Zheng T, Olson DG, Lynd LR, Guss AM. 2015. Elimination of hydrogenase active site assembly blocks H2 production and increases ethanol yield in *Clostridium thermocellum*. Biotechnol Biofuels 8:20.

158. Rydzak T, Lynd LR, Guss AM. 2015. Elimination of formate production in *Clostridium thermocellum*. J Ind Microbiol Biotechnol 42:1263-72.

159. Papanek B, Biswas R, Rydzak T, Guss AM. 2015. Elimination of metabolic pathways to all traditional fermentation products increases ethanol yields in *Clostridium thermocellum*. Metab Eng 32:49-54.

160. Tian L, Papanek B, Olson DG, Rydzak T, Holwerda EK, Zheng T, Zhou J, Maloney M, Jiang N, Giannone RJ, Hettich RL, Guss AM, Lynd LR. 2016. Simultaneous achievement of high ethanol yield and titer in *Clostridium thermocellum*. Biotechnol Biofuels 9:116.

161. Hon S, Olson DG, Holwerda EK, Lanahan AA, Murphy SJL, Maloney MI, Zheng T, Papanek B, Guss AM, Lynd LR. 2017. The ethanol pathway from *Thermoanaerobacterium saccharolyticum* improves ethanol production in *Clostridium thermocellum*. Metab Eng 42:175-184.

162. Verbeke TJ, Giannone RJ, Klingeman DM, Engle NL, Rydzak T, Guss AM, Tschaplinski TJ, Brown SD, Hettich RL, Elkins JG. 2017. Pentose sugars inhibit metabolism and increase expression of an AgrD-type cyclic pentapeptide in *Clostridium thermocellum*. Sci Rep 7:43355.

163. Georgieva T, Mikkelsen M, Ahring B. 2007. High ethanol tolerance of the thermophilic anaerobic ethanol producer *Thermoanaerobacter* BG1L1, vol 2, p 364.

164. He Q, Hemme CL, Jiang H, He Z, Zhou J. 2011. Mechanisms of enhanced cellulosic bioethanol fermentation by co-cultivation of *Clostridium* and *Thermoanaerobacter* spp. Bioresour Technol 102:9586-92.

165. Hemme CL, Fields MW, He Q, Deng Y, Lin L, Tu Q, Mouttaki H, Zhou A, Feng X, Zuo Z, Ramsay BD, He Z, Wu L, Van Nostrand J, Xu J, Tang YJ, Wiegel J, Phelps TJ, Zhou J. 2011. Correlation of genomic and physiological traits of *Thermoanaerobacter* species with biofuel yields. Appl Environ Microbiol 77:7998-8008.

166. Hamilton-Brehm SD, Mosher JJ, Vishnivetskaya T, Podar M, Carroll S, Allman S, Phelps TJ, Keller M, Elkins JG. 2010. *Caldicellulosiruptor obsidiansis* sp. nov., an anaerobic, extremelythermophilic, cellulolytic bacterium isolated from Obsidian Pool, Yellowstone National Park. Appl Environ Microbiol 76:1014-1020.

167. Vanfossen AL, Verhaart MR, Kengen SM, Kelly RM. 2009. Carbohydrate utilization patterns for the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus* reveal broad growth substrate preferences. Appl Environ Microbiol 75:7718-24.

168. Yang SJ, Kataeva I, Hamilton-Brehm SD, Engle NL, Tschaplinski TJ, Doeppke C, Davis M, Westpheling J, Adams MWW. 2009. Efficient degradation of lignocellulosic plant biomass, without pretreatment, by the thermophilic anaerobe "*Anaerocellum thermophilum*" DSM 6725. Appl Environ Microbiol 75:4762-4769.

169. Roh Y, Liu SV, Li G, Huang H, Phelps TJ, Zhou J. 2002. Isolation and characterization of metal-reducing *Thermoanaerobacter* strains from deep subsurface environments of the piceance basin, Colorado. Appl Environ Microbiol 68:6013-6020.

170. Brunecky R, Alahuhta M, Xu Q, Donohoe BS, Crowley MF, Kataeva IA, Yang SJ, Resch MG, Adams MW, Lunin VV, Himmel ME, Bomble YJ. 2013. Revealing nature's cellulase diversity: the digestion mechanism of *Caldicellulosiruptor bescii* CelA. Science 342:1513-6.

171. Alahuhta M, Brunecky R, Chandrayan P, Kataeva I, Adams MW, Himmel ME, Lunin VV. 2013. The structure and mode of action of *Caldicellulosiruptor bescii* family 3 pectate lyase in biomass deconstruction. Acta Crystallogr D Biol Crystallogr 69:534-9.

172. Alahuhta M, Chandrayan P, Kataeva I, Adams MW, Himmel ME, Lunin VV. 2011. A 1.5 A resolution X-ray structure of the catalytic module of *Caldicellulosiruptor bescii* family 3 pectate lyase. Acta Crystallogr Sect F Struct Biol Cryst Commun 67:1498-500.

173. Alahuhta M, Taylor LE, 2nd, Brunecky R, Sammond DW, Michener W, Adams MW, Himmel ME, Bomble YJ, Lunin V. 2015. The catalytic mechanism and unique low pH

optimum of *Caldicellulosiruptor bescii* family 3 pectate lyase. Acta Crystallogr D Biol Crystallogr 71:1946-54.

174. Brunecky R, Donohoe BS, Yarbrough JM, Mittal A, Scott BR, Ding H, Taylor Ii LE, Russell JF, Chung D, Westpheling J, Teter SA, Himmel ME, Bomble YJ. 2017. The multi domain *Caldicellulosiruptor bescii* CelA cellulase excels at the hydrolysis of crystalline cellulose. Sci Rep 7:9622.

175. Groom J, Chung D, Young J, Westpheling J. 2014. Heterologous complementation of a pyrF deletion in *Caldicellulosiruptor hydrothermalis* generates a new host for the analysis of biomass deconstruction. Biotechnol Biofuels 7:132.

176. Kataeva IA, Yang SJ, Dam P, Poole FL, Yin Y, Zhou FF, Chou WC, Xu Y, Goodwin L, Sims DR, Detter JC, Hauser LJ, Westpheling J, Adams MWW. 2009. Genome sequence of the anaerobic, thermophilic, and cellulolytic bacterium "*Anaerocellum thermophilum*" DSM 6725. J Bacteriol 191:3760-3761.

177. Yang SJ, Kataeva I, Wiegel J, Yin YB, Dam P, Xu Y, Westpheling J, Adams MWW. 2010. Classification of '*Anaerocellum thermophilum*' strain DSM 6725 as *Caldicellulosiruptor bescii* sp. nov. Int J Syst Evol Microbiol 60:2011-2015.

178. Zurawski JV, Conway JM, Lee LL, Simpson HJ, Izquierdo JA, Blumer-Schuette S, Nookaew I, Adams MW, Kelly RM. 2015. Comparative analysis of extremely thermophilic *Caldicellulosiruptor* species reveals common and unique cellular strategies for plant biomass utilization. Appl Environ Microbiol 81:7159-70.

179. Zurawski JV, Khatibi PA, Akinosho HO, Straub CT, Compton SH, Conway JM, Lee LL, Ragauskas AJ, Davison BH, Adams MWW, Kelly RM. 2017. Bioavailability of

carbohydrate content in natural and transgenic switchgrasses for the extreme thermophile *Caldicellulosiruptor bescii*. Appl Environ Microbiol doi:10.1128/aem.00969-17.

180. Yee KL, Rodriguez M, Jr., Tschaplinski TJ, Engle NL, Martin MZ, Fu C, Wang ZY, Hamilton-Brehm SD, Mielenz JR. 2012. Evaluation of the bioconversion of genetically modified switchgrass using simultaneous saccharification and fermentation and a consolidated bioprocessing approach. Biotechnol Biofuels 5:81.

181. Chung D, Farkas J, Westpheling J. 2013. Overcoming restriction as a barrier to DNA transformation in *Caldicellulosiruptor* species results in efficient marker replacement. Biotechnol Biofuels 6:82.

182. Chung D, Cha M, Farkas J, Westpheling J. 2013. Construction of a stable replicating shuttle vector for *Caldicellulosiruptor* species: Use for extending genetic methodologies to other members of this Genus 2013. PLoS One 8.

183. Cha M, Chung D, Elkins JG, Guss AM, Westpheling J. 2013. Metabolic engineering of *Caldicellulosiruptor bescii* yields increased hydrogen production from lignocellulosic biomass. Biotechnol Biofuels 6:1-8.

184. Cha M, Chung D, Westpheling J. 2016. 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 100:1823-1831.

185. Chung D, Cha M, Guss AM, Westpheling J. 2014. Direct conversion of plant biomass to ethanol by engineered *Caldicellulosiruptor bescii*. Proc Natl Acad Sci USA 111:8931-6.

186. Chung D, Cha M, Snyder EN, Elkins JG, Guss AM, Westpheling J. 2015. Cellulosic ethanol production via consolidated bioprocessing at 75 °C by engineered *Caldicellulosiruptor bescii*. Biotechnol Biofuels 8:1-13.

187. Chung D, Pattathil S, Biswal AK, Hahn MG, Mohnen D, Westpheling J. 2014. Deletion of a gene cluster encoding pectin degrading enzymes in *Caldicellulosiruptor bescii* reveals an important role for pectin in plant biomass recalcitrance. Biotechnol Biofuels 7:1-12.

188. Chung D, Young J, Bomble YJ, Vander Wall TA, Groom J, Himmel ME, Westpheling J. 2015. Homologous expression of the *Caldicellulosiruptor bescii* CelA reveals that the extracellular protein is glycosylated. PLoS One 10:e0119508.

189. Kim SK, Chung D, Himmel ME, Bomble YJ, Westpheling J. 2016. Heterologous expression of family 10 xylanases from *Acidothermus cellulolyticus* enhances the exoproteome of *Caldicellulosiruptor bescii* and growth on xylan substrates. Biotechnol Biofuels 9:176.

190. Scott IM, Rubinstein GM, Lipscomb GL, Basen M, Schut GJ, Rhaesa AM, Lancaster WA, Poole FL, 2nd, Kelly RM, Adams MW. 2015. A new class of tungsten-containing oxidoreductase in the genus of the plant biomass-degrading, thermophilic bacteria *Caldicellulosiruptor*. Appl Environ Microbiol 81:7339-7347.

191. Young J, Chung D, Bomble YJ, Himmel ME, Westpheling J. 2014. Deletion of *Caldicellulosiruptor bescii* CelA reveals its crucial role in the deconstruction of lignocellulosic biomass. Biotechnol Biofuels 7:1-8.

192. Williams-Rhaesa AM, Poole FL, 2nd, Dinsmore JT, Lipscomb GL, Rubinstein GM, Scott IM, Conway JM, Lee LL, Khatibi PA, Kelly RM, Adams MWW. 2017. Genome stability in engineered strains of the extremely thermophilic lignocellulose-degrading bacterium *Caldicellulosiruptor bescii*. Appl Environ Microbiol 83.

193. Cha M, Wang H, Chung D, Bennetzen JL, Westpheling J. 2013. Isolation and bioinformatic analysis of a novel transposable element, IS*Cbe4*, from the hyperthermophilic bacterium, *Caldicellulosiruptor bescii*. J Ind Microbiol Biotechnol 40:1443-8.

194. Conway JM, McKinley BS, Seals NL, Hernandez D, Khatibi PA, Poudel S, Giannone RJ, Hettich RL, Williams-Rhaesa AM, Lipscomb GL, Adams MWW, Kelly RM. 2017. Functional analysis of the Glucan Degradation Locus (GDL) in *Caldicellulosiruptor bescii* reveals essential roles of component glycoside hydrolases in plant biomass deconstruction. Appl Environ Microbiol doi:10.1128/aem.01828-17.

195. Alvira P, Tomas-Pejo E, Ballesteros M, Negro MJ. 2010. Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: A review. Bioresour Technol 101:4851-4861.

196. Yang B, Wyman CE. 2008. Pretreatment: the key to unlocking low-cost cellulosic ethanol. Biofuels Bioprod Biorefin 2:26-40.

197. Ng TK, Benbassat A, Zeikus JG. 1981. Ethanol production by thermophilic bacteria - Fermentation of cellulosic substrates by cocultures of *Clostridium thermocellum* and *Clostridium thermohydrosulfuricum*. Appl Environ Microbiol 41:1337-1343.

198. Argyros DA, Tripathi SA, Barrett TF, Rogers SR, Feinberg LF, Olson DG, Foden JM, Miller BB, Lynd LR, Hogsett DA, Caiazza NC. 2011. High ethanol titers from cellulose by using metabolically engineered thermophilic, anaerobic microbes. Appl Environ Microbiol 77:8288-8294.

199. Demain AL, Newcomb M, Wu JHD. 2005. Cellulase, clostridia, and ethanol. Microbiol Mol Biol Rev 69:124-154.

200. van de Werken HJG, Verhaart MRA, VanFossen AL, Willquist K, Lewis DL, Nichols JD, Goorissen HP, Mongodin EF, Nelson KE, van Niel EWJ, Stams AJM, Ward DE, de Vos WM, van der Oost J, Kelly RM, Kengen SWM. 2008. Hydrogenomics of the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus*. Appl Environ Microbiol 74:6720-6729.

201. Taylor MP, Eley KL, Martin S, Tuffin MI, Burton SG, Cowan DA. 2009. Thermophilic ethanologenesis: future prospects for second-generation bioethanol production. Trends Biotechnol 27:398-405.

202. Van Fossen AL, Verhaart MRA, Kengen SMW, Kelly RM. 2009. Carbohydrate utilization patterns for the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus* reveal broad growth substrate preferences. Appl Environ Microbiol 75:7718-7724.

203. Svetlitchnyi VA, Kensch O, Falkenhan DA, Korseska SG, Lippert N, Prinz M, Sassi J, Schickor A, Curvers S. 2013. Single-step ethanol production from lignocellulose using novel extremely thermophilic bacteria. Biotechnol Biofuels 6:31.

204. Chung D, Farkas J, Huddleston JR, Olivar E, Westpheling J. 2012. Methylation by a unique alpha-class N4-cytosine methyltransferase is required for DNA transformation of *Caldicellulosiruptor bescii* DSM6725. PLoS One 7:e43844.

205. Kristensen JB, Felby C, Jorgensen H. 2009. Yield-determining factors in high-solids enzymatic hydrolysis of lignocellulose. Biotechnol Biofuels 2:11.

206. Jin MJ, Gunawan C, Balan V, Dale BE. 2012. Consolidated bioprocessing (CBP) of AFEX (TM)-pretreated corn stover for ethanol production using *Clostridium phytofermentans* at a high solids loading. Biotechnol Bioeng 109:1929-1936.

207. Desvaux M, Guedon E, Petitdemange H. 2001. Kinetics and metabolism of cellulose degradation at high substrate concentrations in steady-state continuous cultures of *Clostridium cellulolyticum* on a chemically defined medium. Appl Environ Microbiol 67:3837-3845.

208. Norland S, Fagerbakke KM, Heldal M. 1995. Light element analysis of individual bacteria by X-ray microanalysis. Appl Environ Microbiol 61:1357-1362.

209. Dharmagadda VSS, Nokes SE, Strobel HJ, Flythe MD. 2010. Investigation of the metabolic inhibition observed in solid-substrate cultivation of *Clostridium thermocellum* on cellulose. Bioresour Technol 101:6039-6044.

210. Ellis LD, Holwerda EK, Hogsett D, Rogers S, Shao XJ, Tschaplinski T, Thorne P, Lynd LR. 2012. Closing the carbon balance for fermentation by *Clostridium thermocellum* (ATCC 27405). Bioresour Technol 103:293-299.

211. Olofsson K, Bertilsson M, Liden G. 2008. A short review on SSF - an interesting process option for ethanol production from lignocellulosic feedstocks. Biotechnol Biofuels 1:7.

212. Jönsson LJ, Alriksson B, Nilvebrant NO. 2013. Bioconversion of lignocellulose: inhibitors and detoxification. Biotechnol Biofuels 6:16.

213. Kridelbaugh DM, Nelson J, Engle NL, Tschaplinski TJ, Graham DE. 2013. Nitrogen and sulfur requirements for *Clostridium thermocellum* and *Caldicellulosiruptor bescii* on cellulosic substrates in minimal nutrient media. Bioresour Technol 130:125-35.

214. Farkas J, Chung D, Cha M, Copeland J, Grayeski P, Westpheling J. 2013. Improved growth media and culture techniques for genetic analysis and assessment of biomass utilization by *Caldicellulosiruptor bescii*. J Ind Microbiol Biotechnol 40:41-9.

215. van Niel EWJ, Claassen PAM, Stams AJM. 2003. Substrate and product inhibition of hydrogen production by the extreme thermophile, *Caldicellulosiruptor saccharolyticus*. Biotechnol Bioeng 81:255-262.

216. Baronofsky JJ, Schreurs WJA, Kashket ER. 1984. Uncoupling by acetic acid limits growth of and acetogenesis by *Clostridium thermoaceticum*. Appl Environ Microbiol 48:1134-1139.

217. Johnson EA, Reese ET, Demain AL. 1982. Inhibition of *Clostridium thermocellum* cellulase by end products of cellulolysis. J Appl Biochem 4:64-71.

218. Jin YS, Jeffries TW. 2004. Stoichiometric network constraints on xylose metabolism by recombinant *Saccharomyces cerevisiae*. Metab Eng 6:229-238.

219. Attmannspacher U, Scharf BE, Harshey RM. 2008. FliL is essential for swarming: motor rotation in absence of FliL fractures the flagellar rod in swarmer cells of *Salmonella enterica*. Mol Microbiol 68:328-41.

220. Gibson DG. 2011. Enzymatic assembly of overlapping DNA fragments. Methods Enzymol 498:349-361.

221. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10:563-9.

222. Otto TD, Dillon GP, Degrave WS, Berriman M. 2011. RATT: Rapid Annotation Transfer Tool. Nucleic Acids Res 39:e57.

223. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119.

224. Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides NC. 2010. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. Nat Methods 7:455-7.

225. Lowe TM, Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res 25:955-64.

226. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glockner FO. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res 35:7188-96.

227. Nawrocki EP, Kolbe DL, Eddy SR. 2009. Infernal 1.0: inference of RNA alignments. Bioinformatics 25:1335-7.

228. Markowitz VM, Mavromatis K, Ivanova NN, Chen IM, Chu K, Kyrpides NC. 2009. IMG ER: a system for microbial genome annotation expert review and curation. Bioinformatics 25:2271-8.

229. Nordberg H, Cantor M, Dusheyko S, Hua S, Poliakov A, Shabalov I, Smirnova T, Grigoriev IV, Dubchak I. 2014. The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. Nucleic Acids Res 42:D26-31.

230. Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One 5:e11147.

231. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647-9.

232. Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdigian C, Fuellen G, Gilbert JG, Korf I, Lapp H, Lehvaslaiho H, Matsalla C, Mungall CJ, Osborne BI, Pocock MR, Schattner P, Senger M, Stein LD, Stupka E, Wilkinson MD, Birney E. 2002. The Bioperl toolkit: Perl modules for the life sciences. Genome Res 12:1611-8.

233. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res 34:D32-6.

234. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinformatics 10:421.

235. Basen M, Schut GJ, Nguyen DM, Lipscomb GL, Benn RA, Prybol CJ, Vaccaro BJ, Poole FL, Kelly RM, Adams MWW. 2014. Single gene insertion drives bioalcohol production by a thermophilic archaeon. Proc Natl Acad Sci USA 111:17618-17623.

236. Hennig S, Ziebuhr W. 2008. A transposase-independent mechanism gives rise to precise excision of IS256 from insertion sites in *Staphylococcus epidermidis*. J Bacteriol 190:1488-1490.

237. Curcio MJ, Derbyshire KM. 2003. The outs and ins of transposition: from Mu to Kangaroo. Nat Rev Mol Cell Biol 4:865-877.

238. Varani AM, Siguier P, Gourbeyre E, Charneau V, Chandler M. 2011. ISsaga is an ensemble of web-based methods for high throughput identification and semi-automatic annotation of insertion sequences in prokaryotic genomes. Genome Biol 12:R30.

239. Guérillot R, Siguier P, Gourbeyre E, Chandler M, Glaser P. 2014. The diversity of prokaryotic DDE transposases of the mutator superfamily, insertion specificity, and association with conjugation machineries. Genome Biol Evol 6:260-272.

240. Glass JI, Assad-Garcia N, Alperovich N, Yooseph S, Lewis MR, Maruf M, Hutchison CA, 3rd, Smith HO, Venter JC. 2006. Essential genes of a minimal bacterium. Proc Natl Acad Sci USA 103:425-30.

241. Jones RW, Garland PB. 1977. Sites and specificity of the reaction of bipyridylium compounds with anaerobic respiratory enzymes of *Escherichia coli*. Effects of permeability barriers imposed by the cytoplasmic membrane. Biochem J 164:199-211.

242. Basen M, Sun J, Adams MW. 2012. Engineering a hyperthermophilic archaeon for temperature-dependent product formation. MBio 3:e00053-12.

243. Biegel E, Schmidt S, Gonzalez JM, Muller V. 2011. Biochemistry, evolution and physiological function of the Rnf complex, a novel ion-motive electron transport complex in prokaryotes. Cell Mol Life Sci 68:613-34.

244. Zheng T, Olson DG, Tian L, Bomble YJ, Himmel ME, Lo J, Hon S, Shaw AJ, van Dijken JP, Lynd LR. 2015. Cofactor specificity of the bifunctional alcohol and aldehyde dehydrogenase (AdhE) in wild-type and mutant *Clostridium thermocellum* and *Thermoanaerobacterium saccharolyticum*. J Bacteriol 197:2610-9.

245. Lo J, Zheng T, Hon S, Olson DG, Lynd LR. 2015. The bifunctional alcohol and aldehyde dehydrogenase gene, *adhE*, is necessary for ethanol production in *Clostridium thermocellum* and *Thermoanaerobacterium saccharolyticum*. J Bacteriol 197:1386-1393.

246. Isern NG, Xue J, Rao JV, Cort JR, Ahring BK. 2013. Novel monosaccharide fermentation products in *Caldicellulosiruptor saccharolyticus* identified using NMR spectroscopy. Biotechnol Biofuels 6:47.

247. Nguyen DM, Lipscomb GL, Schut GJ, Vaccaro BJ, Basen M, Kelly RM, Adams MW. 2016. Temperature-dependent acetoin production by *Pyrococcus furiosus* is catalyzed by

a biosynthetic acetolactate synthase and its deletion improves ethanol production. Metab Eng 34:71-9.

248. Ma K, Hutchins A, Sung SJ, Adams MW. 1997. Pyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon, *Pyrococcus furiosus*, functions as a CoA-dependent pyruvate decarboxylase. Proc Natl Acad Sci U S A 94:9608-13.

249. Tsau J-L, Guffanti AA, Montville TJ. 1992. Conversion of pyruvate to acetoin helps to maintain pH homeostasis in *Lactobacillus plantarum*. Appl Environ Microbiol 58:891-894.

250. Keller MW, Lipscomb GL, Nguyen DM, Crowley AT, Schut GJ, Scott I, Kelly RM, Adams MWW. 2017. Ethanol production by the hyperthermophilic archaeon *Pyrococcus furiosus* by expression of bacterial bifunctional alcohol dehydrogenases. Microb Biotechnol 10:1535-1545.

APPENDIX A

LIST OF PUBLICATIONS

1. **Williams-Rhaesa, AM,** Awuku, NK, Lipscomb, GL, Poole, FL, Rubinstein, GM, Conway, JM, Kelly, RM, and Adams, MWW. 2017. Native xylose inducible promoter expands the genetic tools for the biomass-degrading extremely thermophilic bacterium *Caldicellulosiruptor bescii*. Submitted to Applied and Environmental Microbiology.
2. **Williams-Rhaesa, AM,** Rubinstein, GM, Scott, IM, Lipscomb, GL, Poole, FL, Kelly, RM, and Adams, MWW. 2017. Engineering redox-balanced ethanol production in the cellulolytic and thermophilic bacterium, *Caldicellulosiruptor bescii*. Submitted to Metabolic Engineering.
3. Conway, JM, McKinley, BS, Seals, NL, Hernandez, D, Khatibi, PA, Poudel, S, Giannone, RJ, Hettich, RL, **Williams-Rhaesa, AM**, Lipscomb, GL, Adams, MWW, Kelly, RM 2017. Functional analysis of the Glucan Degradation Locus (GDL) in *Caldicellulosiruptor bescii* reveals essential roles of compenent glycoside hydrolases in plant biomass deconstruction. Appl Environ Microbiol. doi:10.1128/aem.01828-17.
4. **Williams-Rhaesa, AM**, Poole, FL, Dinsmore, J, Lipscomb, GL, Rubinstein, GM, Scott, IM, Conway, JM, Lee, LL, Khatibi, PA, Kelly, RM, and Adams, MWW. 2017. Genome stability of engineered strains of the extremely thermophilic, lignocellulose-degrading bacterium *Caldicellulosiruptor bescii*. Appl Environ Microbiol. doi:10.1128/aem.00444–17.

5. Scott, IM, Rubinstein, GM, Lipscomb, GL, Basen, M, Schut, GJ, **Rhaesa**, **AM,** Lancaster, WA, Poole II, FL, Kelly, RM, and Adams, MWW. 2015. A new class of tungsten-containing oxidoreductase in the primary metabolism of the plant biomass-degrading, thermophilic bacterium *Caldicellulosiruptor bescii*. Appl Environ Microbiol. 81:7339–7347.
6. **Rhaesa, AM**, Basen, M, Kataeva, I, Prybol, CJ, Scott, IM, Poole, FL and Adams, MWW. 2014. Degradation of High Loads of Crystalline Cellulose and of Unpretreated Plant Biomass by the Thermophilic Bacterium *Caldicellulosiruptor bescii*. Bioresour Technol. 152 (2014) 384–392.
7. Kataeva, I, Foston, MB, Yang, SJ, Pattathil, S, Biswal, AK, Poole II, FL, Basen, M, **Rhaesa, AM**, Thomas, TP, Azadi, P, Olman, V, Saffold, TD, Mohler, KE, Lewis, DL, Doeppke, C, Zeng, Y, Tschaplinski, TJ, York, WS, Davis, M, Mohnen, D, Xu, Y, Ragauskas, AJ, Ding, SY, Kelly, RM, Hahn, MG, Adams, MWW. 2013. Carbohydrate and lignin are simultaneously solubilized from unpretreated switchgrass by microbial action at high temperature. Energy Environ Sci. 6:2186–2195.