GENETIC TOOLS TO ENABLE BIOFUEL PRODUCTION IN THERMOPHILIC CELLULOSE-DEGRADING BACTERIA

by

JOSEPH DOUGLAS GROOM, JR.

(Under the Direction of Janet Westpheling)

ABSTRACT

The plant cell wall provides a renewable reservoir of energy-rich molecules that can be converted by microbes into useful products. Converting lignocellulosic biomass to fuels would improve environmental impacts of transportation fuels and provide an additional, renewable source of energy. Cellulolytic thermophiles express and secrete an arsenal of enzymes that degrade complex plant cell wall polysaccharides, releasing sugars the bacteria can metabolize. To be able to harness the power of these organisms, tools must be developed for introduction of DNA, gene expression and gene deletion. This work describes the development of genetic tools for Gram-positive, anaerobic, cellulolytic thermophiles *Caldicellulosiruptor bescii, Caldicellulosiruptor hydrothermalis* and *Clostridium thermocellum* to enable studies of plant cell wall degradation and to facilitate metabolic engineering. A thermophilic plasmid replicon from *C. bescii* was introduced into the two other species to test its maintenance and use for gene expression**.** Gene deletions in *C. hydrothermalis* and *C. thermocellum* were made with the aim of improving genetic manipulation.

Deleting a restriction enzyme gene, *chyI*, in *C. hydrothermalis* increased plasmid transformation efficiency. In an attempt to overcome undesired DNA recombination in *C. thermocellum*, a *recA* deletion showed that this gene is required for replication of the native *C. bescii* thermophilic plasmid. To expand the abilities for gene expression in cellulolytic thermophiles, a thermophilic β-glucosidase was developed as a reporter gene and used to characterize three promoters for gene expression, one of which was regulated by maltose. These tools are important advances in the development of genetic systems in non-model cellulolytic thermophiles. They will impact not only our understanding and engineering of these particular organisms, but also for developing genetic systems in new non-model organisms that produce novel enzymes, biochemical reactions and pathways.

INDEX WORDS: genetics, plasmid, thermophile, bacteria, gene expression, promoter, gene regulation, reporter gene, gene deletion, DNA recombination, DNA replication, transformation, biofuel, lignocellulosic biomass, bioethanol, cellulase

GENETIC TOOLS TO ENABLE BIOFUEL PRODUCTION IN THERMOPHILIC CELLULOSE-DEGRADING BACTERIA

by

JOSEPH DOUGLAS GROOM, JR.

BA, Vanderbilt University, 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

GENETIC TOOLS TO ENABLE BIOFUEL PRODUCTION IN THERMOPHILIC CELLULOSE-DEGRADING BACTERIA

by

JOSEPH DOUGLAS GROOM, JR.

Major Professor:	Janet Westpheling
Committee:	Jorge Escalante
	David Garfinkel
	Debra Mohnen
	Ellen Neidle

Electronic Version Approved:

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

DEDICATION

This dissertation is dedicated to Dr. Jan Westpheling, who, as of the twenty-second day of June, 2017, had only one thing to say about it:

“It was beyond terrible, in fact hopeless which is why I didn't provide any comments. Consistent with your continued disappointment as a graduate student!”

ACKNOWLEDGEMENTS

There are several people I need to acknowledge.

My mother Mitzi Groom, for being by biggest supporter and accepting me completely, no matter what. My father Joe Groom, for being my biggest fan, and for finally learning to pronounce the word “genetics.” I’d like to say to both of you that I’m sorry for not going to medical school.

Elise Snyder for providing me a continuous supply of quality web content and buying me that orange pipette rack. Dylan Kerlin for driving me to the lab when I couldn’t wait for the bus with the peasants. Kyle Sander for making me salads and giving me the most memorable summer of my life.

Daehwan Chung for teaching me to perform an experiment, the hard way. Jenna Young primarily for putting up with me but also for teaching me everything I know about anaerobic technique. Sun-ki Kim for being my role model, and for somehow managing to always keep a smile on his face. Adam Guss for the innumerable, detailed suggestions that saved several of my projects. Yannick Bomble and Mike Himmel for the opportunity to work in a mile-high lab. Kelly Dyer and Caitlin Conn for assistance with the phylogenetic analysis.

My graduate committee members Ellen Neidle, Deb Mohnen, Jorge Escalante and David Garfinkel for punctuality and a breadth of expertise that I am still trying to wrap my head around. I want to acknowledge David Garfinkel

especially, for stepping in at the last minute. And Jan Westpheling, for being so doggedly convinced that I could, especially when I was convinced that I couldn’t.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

I. The plant biomass alternative to petroleum for renewable fuels and chemicals

The potential for a renewable energy source rests at the interface of microbial molecular biology and plant cell wall biochemistry. Plants store a variety of complex carbohydrates and sugars ranging from the starch granules in the amyloplast [1] to the lignocellulose in the plant cell wall [2]. These macromolecules can be used to produce fuel such as ethanol, butanol, and hydrogen. Microbial conversion is one of several options to produce these fuels from plant carbohydrates. Alternative fuels are becoming an increasingly popular and important option because of the documented effects of greenhouse gases such as carbon dioxide (CO_2), which is produced by the combustion of fossil fuels [3]. The strongest argument for converting plant biomass to a fuel source is the timescale of CO_2 incorporation: while it takes millions of years for fossil fuels to be created from atmospheric CO_2, the time required for CO_2 assimilation into plants is much shorter. It takes a few years for trees like poplar to mature, or as briefly as a season to grow crops like corn, sugarcane, and switchgrass.

As it is the only carbon-rich source on the global scale besides fossil fuels, biomass is considered to be one of, if not the only, renewable alternative for the production of transportation fuels and chemicals [4]. Although many advances

have been made in modes of transportation running on electricity and hydrogen, biofuels are singular in their ability to meet the demand for renewable dense liquid fuels in aviation, long-haul trucking, and ocean shipping [5]. Advantages of plant-based fuels include opportunities for rural development [6], environmental sustainability [7], and atmospheric carbon mitigation. In the specific case of bioethanol, using plant biomass for fuel instead of gasoline could reduce green house gas (GHG) emissions by 34% to 108% [8]. Furthermore, plants dedicated to biomass production are associated with greater GHG reduction than annual crops such as corn.

Though biomass from other plants has this advantage, corn-starch is the most widely used first-generation biofuel feedstock—95% of commercial bioethanol in the United States was produced from corn-starch in 2016 [9]. Alternative crops would be more suitable for different climates and landscapes. As such, the United States Department of Agriculture established multiple regional Biomass Research Centers focusing on specific crops including sorghum, barley, peas, and “sugar crops” like sugar cane, sugar beet and almond hulls [10]. Several facilities in the United States and worldwide also use waste cellulosic material such as corn stover or sugarcane bagasse [11]. Still, to avoid competition with food crops, emphasis in recent research has focused on dedicated biofuel crops**,** preferably those that would thrive on land that is marginal for farming.

This emphasis resides on perennial plants rather than annual plants (e.g. corn), and its specific focus has been on lignocellulose rather than starch.

Lignocellulose is a term used to describe the complex web of interlinked polysaccharides and phenolic compounds found in the cell walls of all plants [2, 12]. It is an abundant carbon source, albeit much more difficult to process than corn-starch [13]. It can be isolated from non-food plants like switchgrass, poplar trees, and the grass miscanthus, avoiding food/fuel competition. An added practical bonus of lignocellulose is that the lignin can be precipitated and used to supply electricity to biofuel production facilities [14]. Moreover, the previously mentioned GHG emission reduction combined with increased carbon sequestration by perennials lessen the environmental impact of these plants compared to traditional crops. The GHG reduction estimates for bioethanol also take into account the effects of land use changes required for the establishment of bioenergy crops. These effects differ considerably between corn and perennial grasses like switchgrass (*Panicum virgatum*) and miscanthus *(Miscanthus* x *giganteus*) [15, 16]. Lower GHG emissions result because perennials require less fertilizer and thus result in less emitted N_2O, and also because perennials sequester carbon in their extensive root networks called rhizomes [17, 18]. In comparison to corn, perennial plants increase biodiversity of other plant, arthropod, bird, and methanotrophic bacterial species in planned multifunctional agricultural landscapes [19]. Such landscapes using perennials can also take up and filter run-off water from heavily fertilized neighboring crops, thereby improving water quality and preventing erosion [18]. The increasing global temperatures expected in the coming years may even be beneficial to

switchgrass cultivation, as crop yields are predicted to increase with higher temperatures [20].

While these environmental benefits make a strong argument for fuel from biomass, renewable energy is not the only possible outcome of this system. Petroleum-based chemicals are in high demand in both developed and developing countries [21]. These petrochemicals are used to make commodities including textiles, construction and packaging materials, tires, and additives for the pharmaceutical and food industries. Plant biomass can replace petroleum as the carbon source for such chemicals. Many compounds are considered "platform chemicals", meaning they are valuable precursors to a variety of products. One of these is succinic acid, a fermentative end-product of many anaerobic microorganisms, which is used in detergents, metal corrosion inhibitors and antimicrobials. It is also a key platform chemical for a variety of products including polybutylene succinate, a bioplastic produced by esterification of succinic acid and 1,4-butanediol [22]. Another platform chemical is levulinic acid, formed by the hydrolysis of six-carbon sugars found in plant biomass [4]. With both a carbonyl and a carboxylic acid functional group, it affords many different reactive outcomes that can lead to the creation of herbicides, polymers and transportation fuels [23]. Many products can also add value to the production of biofuels—a recent example is the concomitant production of ethanol and the zero calorie sweetener D-psicose from waste cruciferous vegetable residue [24].

Projected environmental benefits and renewable chemical production from plant biomass encourage the development of agro-ecosystems, in which plants re-assimilate the carbon dioxide released by the combustion of their predecessors' fuel products [7]. This is certainly a lofty aspiration, and would need to be economically feasible and profitable to become practical. One issue concerning contamination, drying and spoiling of plant material has caused cellulosic ethanol plants like Dupont to use corn stover grown within just a 30 mile radius from their biofuel production facility [25]. However, bioethanol from corn-starch has already proven itself as a viable industry. In 2016, the United States produced a record amount of bioethanol—15.25 billion gallons [9]. This was accompanied by approximately 42 million metric tons of high-protein animal feed as a byproduct. Bioethanol production facilities fit within an "economies of unit *number*" model, in which a large number of smaller facilities meet local demands for energy with local feedstocks. This contrasts with the "economies of unit *scale*" model that is consistent with oil refineries, in which a few large-scale facilities process feedstocks and distribute products [26]. Still, techno-economic analyses reveal that both starch and lignocellulose feedstocks result in higher prices than fossil fuels per gigajoule (GJ) of energy produced [14]. This price increase does not result from the cost of the feedstock, but rather from the cost of the processing. Specifically, the enzymes—amylases for starch breakdown and cellulases for lignocellulose breakdown—are the most expensive component of biomass conversion [27]. Chemical, physical, and thermal pretreatments of lignocellulosic biomass decrease the necessary loading of cellulolytic enzymes.

However, pretreatment is also very costly [28], degrades useable substrate, and results in losses during processing [11]. Basic knowledge of plant cell wall structure and synthesis is required to optimize the processes of deconstruction and conversion.

II. Plant cell wall recalcitrance and biofuel

Plant cell wall structure. Unlike petroleum, lignocellulosic biomass is a solid, complex web of carbon-rich molecules. The plant cell wall has been termed "recalcitrant", referring to the fact that cell wall components including lignin, cellulose, and hemicellulose are resistant to degradation into their respective monomers [13]. Cellulose consists of highly polymerized D-glucose chains (or glucans) with β-(1,4) linkages, grouped into fibrils held together by both hydrogen bonds and van der Waals stacking interactions [29]. It comes in highly ordered crystalline and less ordered amorphous varieties. The crystalline form generally is less accessible to solutes and enzymes under common conversion conditions, and is therefore more recalcitrant [30]. Cellulose makes up ~35-50% of the plant cell wall in various plants of interest for lignocellulosic fuels (reviewed in [31]).

Hemicellulose proportions vary in the cell walls of different plants. It binds to the surface of cellulose and forms a matrix between its fibrils [12, 31]. It is less polymerized than cellulose but more branched, and is more heterogeneous in its sugar backbone molecules and in the linkages between them [32]. These polymers include mannans, mixed-linkage β-(1,3) and β-(1,4) glucans, xyloglucans with various monosaccharide side chains, and xylans with arabinan and glucuronic acid side chains. The heterogeneous structure of hemicellulose

suggests that many enzymatic activities are required for its full saccharification [12]. In grasses, glucuronoarabinoxylan can be acetylated on its arabinose residues, and these acetyl groups can form ester bonds with ferulic acid, a component of lignin [33-35]. These covalent linkages are players in larger lignin-carbohydrate complexes (LCCs), which are thought to be impenetrable by enzymes and chemicals [36]. Because of this, thermal and physical pretreatments are often used to remove lignin for better access to cellulose in cell wall deconstruction procedures [37, 38].

Lignin itself is a mixture of phenolic structures connected by ether and alkyl linkages comprising ~10-30% of the plant cell wall, with woody plants having more lignin [39]. It impacts many aspects of agriculture and the environment including cell wall rigidity, plant stress, and soil organic carbon cycling [40]. It is built from three primary monolignols: p-coumaryl alcohol, coniferyl alcohol, and synapyl alcohol. These monolignols generate the *p*-hydroxyphenyl (H), syringyl (S), and guaiacyl (G) subunits, respectively [12, 37]. The ratios of these monolignols are different in different plants. For instance, hardwoods contain predominantly G and S monolignols, while grasses also have significant levels of H monolignols. Because of its recalcitrant linkages, lignin is most often combusted for heat and electricity or gasified to make energy-rich syngas. However, it is also recognized as a potentially valuable source of carbon fibers [41].

While lignin and hemicellulose are covalently linked monocots such as grasses, these ferulic ester linkages exist between lignin and pectin in dicots [34,

42, 43]. Pectin is the most complex polysaccharide in the plant cell wall and is found mostly in the primary cell wall [44]. It makes up 35% of the cell wall polysaccharides in dicots like *Arabidopsis thaliana*, and about 10% in grasses [45]. All pectins contain 1,4-linked α-D-galactosyluronic acid residues as a general repeating unit; the three basic pectin polymer types are homoglacturonan, rhamnogalacturonan-I and substituted galacturonans. Pectins may be crosslinked to hemicellulose and proteins in the secondary cell wall (reviewed in [46]), which is notoriously more recalcitrant to degradation. In fact, a certain cell wall proteoglycan APAP1 incorporates covalently bound pectin and hemicellulose domains [47]. In addition to this, studies of Arabidopsis mutants in the pectin biosynthesis genes *qua1*, *irx8/gaut12, and glz1/gatl1* have less xylan [48, 49]. This could indicate that existing pectin in plant cell walls affects xylan biosynthesis or deposition, or that pectin modification like demethylesterification affects xylan deposition. Additionally, there appear to be clear associations of rhamnogalacturonan I and homogalacturonan with cellulose microfibrils in intact plant cell walls [50]. These observations and others have inspired a model in which pectins act as major players in the cohesiveness of the plant cell wall, and contribute to recalcitrance by blocking enzymes from reaching other cell wall polymers.

Sugars released by plant cell wall deconstruction can be converted to ethanol and other fuels (often by microbial fermentation) but the bottleneck in processing is the first step—the conversion of raw biomass into simple sugars [28]. However, studies of the plant cell wall's structure and synthesis [2] in

addition to its native deconstruction by microbes [51] are directly applicable to this problem. Better understanding of the plant cell wall has allowed targeting of specific genes responsible for synthesis. Even with these successes, there are poorly understood connections between different types of cell wall macromolecules, many of which contribute to plant biomass recalcitrance [49, 50, 52, 53].

Engineering plants to overcome cell wall recalcitrance. The heterogeneous and interwoven nature of the components of the plant cell wall—cellulose, hemicellulose, pectin, lignin—results in recalcitrance to degradation by chemicals, enzymes and microbes (reviewed in [12, 54]). As lignin seems to be the most recalcitrant part of the plant cell wall that hinders the access of cellulolytic enzymes [55], the aim of much of the genetic engineering work has been to modify lignin [39]. Much of the effort in poplar has focused on using natural variation in the syringyl : guaiacyl (S:G) ratio [56] or crossing plants to discover variants [57], as G lignin is often more branched and interconnected with cell wall polysaccharides than S lignin, and therefore more difficult to deconstruct [39]. However, these studies concluded that there are definitely other factors besides the S:G ratio that affect sugar release. Down-regulation of two lignin biosynthesis enzymes in alfalfa, hydroxycinnamoyl transferase (HCT) and caffeic acid 3-*O*-methyltransferase (COMT), achieved lower levels of lignin and also enhanced sugar release [58]. Similarly, down-regulating COMT in switchgrass met many biofuel goals including decreasing lignin content, reducing the S:G ratio and increasing the downstream bioethanol yield by 38% [59].

COMT-down-regulated switchgrass exhibits similar results when grown over multiple years in the field, verifying that these lignin modifications could be beneficial to real agro-ecosystems [60]. One alternative lignin modification scheme to produce easy-to-deconstruct lignin in the model plant *Arabidopsis thaliana* has been to replace native lignin monomers with truncated side chains, which reduces the complexity of linkages [61]. Another scheme employed the expression of a monolignol ferulate transferase to introduce chemically labile ferulate bonds into lignin, making the poplar cell wall more disposed to mild alkaline depolymerization [62].

Engineering pathways and enzymes besides lignin biosynthetic components has also achieved reductions in cell wall recalcitrance. For instance, knocking out an *Arabidopsis thaliana* glucuronoxylan methyltransferase (GXMT) increased glucuronoxylan release during hydrothermal treatment of the cell wall and altered lignin monomer composition [63]. In another case, down-regulating a galacturonoysltransferase GAUT12 responsible for pectin or xylan synthesis in poplar decreased both pectin and xylan levels, and increased glucose release upon enzymatic treatment [49]. It is also possible to introduce microbial endoglucanases into plants to reduce recalcitrance (reviewed in [64]). In fact, a functional endoglucanase E1 from the thermophilic bacterium *Acidothermus cellulolyticus* could be expressed in maize. Corn stover from these plants exhibited greater digestibility because E1 was actively hydrolyzing the cell wall during growth [65]. Transgenically expressed E1 also retains activity after ammonia fiber explosion (AFEX) pretreatment [66]. As pretreatments like AFEX

are important in making the plant cell wall accessible, it would be most efficient to synergistically coordinate plant cell wall engineering, various cell wall pretreatments, and downstream processing.

III. Traditional techniques for lignocellulosic deconstruction and conversion

Thermochemical conversion. Conversion of biomass into useful products fall into two general schemes: thermochemical and biochemical. Thermochemical conversion technology has advantages including low energy consumption, high conversion rates, short residence times of biomass in the facility, and versatility of possible feedstocks [21, 37]. Gasification is the partial combustion of biomass in the presence of oxygen, in which a mixture of combustible gases is formed (called syngas or synthesis gas) [67]. Syngas includes various concentrations of carbon monoxide (CO), hydrogen (H_2), methane (CH_4), nitrogen (N_2), various alkanes, and other gases [68]. Pyrolysis, on the other hand, occurs in the absence of oxygen at high temperatures to produce bio-oil, charcoal, non-condensable gases, acetic acid, acetone, and methanol [68]. While thermochemical processes are efficient, they differ from biochemical processes primarily in that they are not selective in their generation of products [37]. However, subsequent gas conversion methods can form specific downstream products from syngas. These methods include hydrocarbon chain synthesis (called Fischer-Tropsch synthesis), methanol synthesis, mixed alcohol synthesis and syngas fermentation [21].

Biochemical conversion. Less severe versions of thermochemical conversion techniques can be used as pretreatments to prepare the primary substrate for biochemical conversion. The overall process for deconstructing and biochemically converting plant biomass into products involves six general steps [54]:

(1) Size reduction to maximize surface area.

(2) Chemical or physical pretreatment of the biomass to open up the cell wall and remove the most recalcitrant components and linkages.

(3) Production of enzymes.

(4) Enzymatic hydrolysis of solids into sugars.

(5) Sugar conversion (most often fermentation to ethanol, butanol, etc.).

(6) Recovery of the product.

Broadly, four major types of pretreatments exist: physical, chemical, biological, and solvent-fractionation [37]. Besides heat treatments, physical pretreatments include milling and grinding to make particle sizes smaller, increase surface area, and reduce the degree of polymerization of cellulose. Because this requires high energy input, it is generally believed that milling is not economically feasible for lignocellulosic biomass processing [69]. A wide variety of chemical pretreatments have been developed to remove specific components of lignocellulose, including alkaline hydrolysis, acid hydrolysis, and ozonolysis (reviewed in [37, 54, 70, 71]). Most pretreatments are physicochemical, combining heat and pressure with various chemical additives to explode or oxidize biomass. A particularly interesting example is the previously mentioned

ammonia fiber explosion (AFEX) technique, which redistributes lignin and makes pores in the cell wall to render cellulose more accessible to enzymes [72]. Biological pretreatments mostly rely on white-rot fungi that degrade lignin and hemicellulose with minimal cellulose degradation, but this process is very time-consuming (reviewed in [71]). Prominent solvent fractionation techniques include fractionation with γ-valerolactone (GVL) [73] and co-solvent enhanced lignocellulosic fractionation (CELF) with tetrahydrofuran (THF) [74], both of which utilize sustainable plant cell wall-derived solvents. The effect of solvent fractionation is to separate the cellulose, hemicellulose and lignin components so they can be used individually, effectively reducing the cost of enzyme loading [75].

The enzyme cocktails used commercially for plant cell wall digestion exhibit three main cellulase activities: cellobiohydrolase, endoglucanase, and β-glucosidase [76]. Cellobiohydrolases cleave the cellobiose, a β-linked disaccharide of glucose, from the end of a glucan chain, while endoglucanases cleave the glucan chain in the middle, and β-glucosidases cleave cellobiose into two glucose molecules. Hemicellulases are also included when more mild pretreatments are used (reviewed in [12]), since less hemicellulose is degraded in mild pretreatments. In fact, both the Cellic cocktail from Novozyme and the Accellerase cocktail from Dupont contain hemicellulases [76]. More recently, enzymes called lytic polysaccharide monooxygenases (LPMOs) have been added, as they improve the rates and yields of bioethanol production [77, 78]. Most of the industrial cellulases are produced from aerobic cellulolytic fungi,

particularly *Trichoderma reesei* (also known as *Hypocrea jecorina*), as these fungi can produce crude cellulase at over 100 g/L [79].

This enzyme cocktail is loaded onto the biomass, and saccharification of the polymers produces soluble sugars. There are a variety of schemes for breakdown and fermentation of the plant biomass to fuel molecules. In separate hydrolysis and fermentation (SHF), the enzyme saccharification and fermentation are performed sequentially. This is beneficial because of the different operating temperatures of cellulases and traditional fermentative microbes. However, the released sugars glucose and cellobiose can inhibit cellulases, so simultaneous saccharification and fermentation (SSF) can instead be used to quickly convert the sugars and alleviate this inhibition [54]. Yeast fermentations from solid lignocellulose loadings routinely yield >50 g/L ethanol titers, which is necessary to reduce the energy required for ethanol recovery and to achieve savings on operational costs [80, 81]. Most companies in Brazil, Italy, and the USA currently producing ethanol from lignocellulosic biomass use yeast for ethanol fermentation, but DuPont uses *Zymomonas mobilis* [11], an ethanologenic bacterium that has many industrial benefits. *Z. mobilis* boasts a wide pH range of growth to deal with acidic or basic pretreatments, a high alcohol tolerance, the ATP-efficient Entner-Douderoff pathway for glycolysis, and a set of strain improvements that allow simultaneous utilization of the hexose glucose and the pentoses xylose and arabinose (reviewed in [82]). In sum, the input feedstock of raw plant material is converted to end products that include the desired chemical

(e.g. ethanol, butanol, succinic acid, levulinic acid) and often a by-product (e.g. heat, electricity).

The overall goals of plant biomass pretreatment fall into three categories: LCC cleavage and hemicellulose removal, lignin modification and redistribution, and cellulose decrystallization [37]. The disadvantages of pretreatments are loss of substrate, formation of inhibitors to downstream biological processing, and cost effectiveness [70, 83]. Alkaline treatment can remove acetyl groups and thereby overcome the ferulate linkages of lignin and hemicellulose. Acid pretreatment hydrolyzes hemicellulose, but it leads to the production of furfural, which can inhibit both metabolic enzymes and microbes [84, 85]. Xylooligomers [86] and phenols [87] released during many pretreatments inhibit cellulase activity. Also, the acids, bases and solvents must be recovered for these processes to be economically feasible on an industrial scale. Recently developed co-solvents such as tetrahydrofuran (THF) have a very low boiling point and are easily recovered [74], but not all pretreatment chemicals have this feature. Alternative enzymes to the *T. reesei* cellulases have also been characterized. One such enzyme is CelA [88], a multifunctional cellulase from the thermophilic bacterium *Caldicellulosiruptor bescii*. CelA can release double the amount of sugars from cellulose as the fungal cellulase Cel7A in the same amount of time, and its enzymatic activity on a molar basis is seven times that of the common processive fungal cellulase mixture [89]. The organism encoding CelA can actually grow on unpretreated biomass and in doing so, avoid the negative aspects of pretreatment [90].

Taken together, pretreatments pose problems for enzymatic hydrolysis of the cell wall and to downstream conversion of sugars to bioproducts, in addition to their high cost. In particular, the enzyme cocktails themselves are arguably the most expensive part of the process [27], and might not even be the best enzymes for the job. An exciting alternative to physicochemical pretreatments and purified enzyme cocktails is to harness the native ability of microorganisms that utilize plant biomass. This idea is consistent with a consolidated bioprocessing (CBP) scheme, relying on a single organism to both break down the plant cell wall and convert the released sugars to fuels and chemicals.

IV. Consolidated bioprocessing (CBP): simultaneous microbial deconstruction and conversion of the plant cell wall

Characteristics of potential CBP organisms. An ideal organism for CBP would be able to deconstruct cellulose and convert the released sugars efficiently into desired products. The primary advantage is that the organism itself would continually produce the cellulase cocktail as it grew on plant biomass, avoiding the expense of added enzymes. If the organism is a native cellulose degrader, pretreatments can be mild or even non-existent, which would decrease those costs as well. It would be extremely advantageous if the CBP organism could use both five-carbon (pentose) and six-carbon (hexose) sugars to fully metabolize the available plant cell wall sugars. Many traditional industrial organisms have been engineered with cellulases, with some success at breaking down cellulose and pretreated biomass (reviewed in [91]). Endoglucanases, cellobiohydrolases and β-glucosidases have been expressed in bacteria including *E. coli*, *Z. mobilis,*

Cornynebacterium glutamicum, Lactobacilus plantarum and yeasts including *S. cerevisiae* and *Kluyveromyces marxianus*. Expressing and secreting functional plant cell wall deconstruction enzymes in non-native hosts is difficult, and the cellulolytic activity and cellulosic ethanol production from these hosts is lower than native cellulose-degrading ethanologens [91-94]. Because of their ability to meet many of the criteria for plant biomass deconstruction and sugar utilization, cellulolytic thermophilic bacteria have been recent targets for CBP.

Thermophilic microorganisms are isolated from hotsprings and hydrothermal vents, and those that are cellulolytic can be enriched by growing environmental samples on different types of cellulose or plant material [95]. High temperatures increase the solubility and digestibility of plant substrates [51, 96]. They also eliminate contamination by mesophiles [97], and reduce the energy required for cooling after pretreatments of plant substrates in an industrial setting [98]. The fact that many target organisms for CBP are thermophiles overcomes the issue in simultaneous saccharification and fermention of low cellulase activity at the temperatures required for yeast fermentation [99]. Additionally, these thermophiles grow near or above the boiling point for several products such as ethanol (~78°C), which would allow for distillation and removal of valuable (and often inhibitory) end-products. While anaerobic organisms grow to lower cell densities when compared with aerobes [79], they are beneficial because more of the substrate can be directed toward end products rather than cell mass, and that oxygen does not need to be supplied for growth. Addition of oxygen is one of the most costly steps in an industrial microbial growth scheme, as exemplified by the

production of cellulase by the aerobe *T. reesei* [98]. Thermophilic anaerobes able to deconstruct cellulose have generated interest as industrial CBP organisms for these reasons.

The Gram-positive, thermophilic anaerobes in the genus *Caldicellulosiruptor* can deconstruct high loadings of plant biomass into simple sugars without the conventional pretreatments described above [90, 100, 101]. Found in various hotsprings worldwide, they are the most thermophilic cellulose-degrading bacteria that have been isolated and characterized, growing optimally at 68-80°C [102]. *Caldicellulosiruptor* species secrete free carbohydrate-active enzymes (CAZys) [103] with carbohydrate-binding modules to bind and degrade cell wall polysaccharides [104]. They can simultaneously utilize the wide range of hexoses, pentoses, and polysaccharides released from the plant cell wall with no evidence of carbon catabolite repression [105, 106]. These qualities make them well suited for CBP, although they naturally produce very little or no ethanol [90, 107, 108].

Although the cellulolytic thermophile *Clostridium thermocellum* (T_{opt}=60°C [109]), cannot use pentose sugars [110], it is still a leading candidate for CBP because it grows to relatively high cell densities for an anaerobe, exhibits the best biomass solubilization of any single microbe that has been discovered [111], and naturally produces ethanol [112]. *C. thermocellum* degrades the cell wall with a structure called the cellulosome, which is a cell-anchored multi-enzyme complex that attaches to biomass [113, 114]. The cellulosome is modular in that different types of saccharolytic enzymes can attach in different combinations to

the scaffold protein CipA (called the “primary scaffoldin”). *C. thermocellum* also secretes “cell-free” cellulosomes that can degrade plant cell wall polysaccharides remotely from the cell [115]. Recently a process called cotreatment has been tested, in which *C. thermocellum* is used to hydrolyze and ferment plant biomass while it is being milled [11, 111].

Other thermophiles are solventogenic but not cellulolytic—similar to *Z. mobilis*, they naturally produce end products such as ethanol. *Thermoanaero-bacterium saccharolyticum* is a xylan-consuming (hemicellulolytic but not cellulolytic) Gram-positive anaerobe that utilizess both hexoses and pentoses, and naturally produces ethanol [116]. Since it grows around the optimal temperature of *C. thermocellum*, it is easy to imagine a co-culture scenario in which *C. thermocellum* degrades plant biomass and *T. saccharolyticum* uses the released sugars to produce valuable end products. In fact, production of 38 g/L ethanol from plant biomass has been achieved with a co-culture of engineered strains of these two organisms [112].

There are certainly challenges and disadvantages to using cellulolytic thermophiles for CBP. Likely the greatest obstacle is the less efficient fermentative fuel production compared to yeast. Microorganisms remain susceptible to cell wall-derived microbial inhibitors like furfural and to fermentative end products like ethanol [117, 118]. High loadings of solid biomass can still be problematic, as this can inhibit cellulases [119]. Establishing and using genetic methods in cellulolytic thermophiles will enable their development as CBP organisms. Constructing deletions of genes and expressing heterologous

genes allows for the elimination of certain end products, formation of desirable products, resistance to inhibitors, and improved cellulose hydrolysis.

Engineering in cellulolytic thermophiles for CBP. Although it cannot directly break down cellulose, *T. saccharolyticum* has shown promise as a thermophilic, metabolically versatile alternative to yeast. Deletion of the phosphotransacetylase (*pta*), acetate kinase (*ack*) and lactate dehydrogenase (*ldh*) genes responsible for competing carbon flux pathways resulted in the production of 37 g/L ethanol from mixed sugars and ~20 g/L using a simultaneous saccharification and fermentation (SSF) scheme with crystalline cellulose [120]. More recently, by deleting exopolysaccharide synthesis genes and the oxidative stress response regulator *perR,* a strain of *T. saccharolyticum* was engineered that yields 0.46 g ethanol / g fermented sugar (comparable to hexose fermentions by yeast) and titers up to 70 g/L ethanol from cellobiose and maltodextrins [121].

Work in *C. thermocellum* has shown that more competing pathways must be eliminated for efficient ethanol production than were required for *T. saccharolyticum*. Single deletions of the phosphotransacetylase (*pta*) or the pyruvate:formate lyase (*pfl*) eliminated acetate or formate production, but had small effects on ethanol production [122, 123]. However, eliminating lactate dehydrogenase (*ldh*) [124], *ldh* and *pta* together [112, 125], or two different hydrogenase functions *(hydG* and *echA-F* deletions) [126] did improve the production of ethanol. Moreover, stacking four mutations in one strain (*ΔhydG Δldh Δpfl Δpta-ack*) considerably increased the ethanol yield [127]. When this

strain was subject to directed evolution, it eventually produced 22 g/L ethanol from crystalline cellulose at 0.39 g ethanol / g glucose [128]. As these modifications were made, it became clear that a significant amount of the carbon was being converted to amino acids. As a proof of concept, a glutamine synthase gene (*glnA*) was deleted, and ethanol yields increased by 53% over the wild-type strain, indicating that this is a promising way forward for new strain development. While ethanol production has been the main focus of work in *C. thermocellum*, improved isobutanol production from cellulose has also been demonstrated by expressing three different enzymes with empirically selected promoters [129].

T. saccharolyticum and *C. thermocellum* are native ethanologens, but the more thermophilic C*. bescii* produces the desirable biofuel hydrogen near the Thauer metabolic limit of 4 mol H_2 / mol glucose [106]. It also produces acetate and lactate as major fermentation end products, and can use non-pretreated plant biomass as its sole carbon source [90]. The *ldh* gene was the first target for deletion, which resulted in increased production of hydrogen [130]. *C. bescii* was also engineered to produce ethanol directly from switchgrass by expressing alcohol dehydrogenases from other thermophiles [117, 131]. These experiments were the first examples bioethanol production from plant biomass without pretreatment.

Plant cell wall deconstruction is a characteristic of many cellulolytic thermophiles, but this process is susceptible to the same issues as enzyme cocktails. As such, improving cellulose hydrolysis and overcoming cellulase inhibition by breakdown products has been a goal of strain development. For

example, additional xylanases from *Acidothermus cellulolyticus* have been introduced into *Caldicellulosiruptor bescii* to reduce xylooligomer inhibition. Expression of these xylanases improved both cell growth and enzymatic activity on xylan substrates [132]. Similar results were obtained by the expression of surface-layer associated multidomain xylanases originating from *Caldicellulosiruptor kronotskyensis* [133]. Since furfural is a potent inhibitor of microbes and enzymes, a butanol dehydrogenase has been expressed in both *C. bescii* [134] and *Clostridium thermocellum* [135] to increase furfural and hydroxymethylfurfural tolerance. These types of genetic engineering indicate the path forward for making cellulolytic thermophiles a reality for CBP.

Gene deletions have also contributed to the understanding of both plant-microbe interactions and the basic structure of the plant cell wall. Recently, the gene encoding the multi-domain cellulase CelA was deleted in *C. bescii*, which revealed this enzyme's important *in vivo* role in plant biomass deconstruction [136]. Similarly, deleting a cluster of pectinase genes revealed these enzymes' crucial roles in growth on plant biomass [137]. Moreover, this work demonstrated that pectin was an important obstacle to overcome in order to deconstruct intact plant biomass, providing further evidence for the role of pectin in the structure and recalcitrance of the plant cell wall. Gene deletions of cellulosome components in *C. thermocellum* have revealed that the primary scaffoldin not only acts as an anchor for cellulolytic enzymes, but is also involved in global transcriptional regulation of genes involved in cellulose degradation, sugar transport, and metabolism [115]. These studies demonstrate the application of

cellulolytic organisms as a tool to study both plant cell wall structure and recalcitrance.

All achievements in engineering of cellulolytic and solventogenic thermophiles were made possible because of the development of genetic tools. They demonstrate the value of the ability to transform and manipulate potentially useful microorganisms. Genetic tool development is largely empirical for new strains of interest.

V. Genetic tools in thermophiles to facilitate CBP

Aspects of a genetic toolkit. In general, an ideal set of genetic tools and methods for an organism allows for the investigation of native gene functions and the introduction of new gene functions. Different types of plasmid vectors are necessary [138]. Integrating vectors enable either insertion of the entire circular plasmid into the genome or replacement of a particular segment of DNA with a marker. Plasmids harboring functional replication origins and replication proteins allow for the stable extrachromosomal expression of genes. Selectable markers are vital, and include both determinants for antibiotic resistance and nutritional markers. Various antibiotics will often target bacterial ribosomes, cell wall synthesis, or replication machinery [139]. The ability to create markerless manipulations is also important. For instance, a positive-negative selection cassette may be attached to a particular DNA sequence, so that the uptake of the sequence can be selected for with one marker, and the removal of all sequences except the sequence of interest can be counter-selected by growth of the organism in a different condition. Examples of positive selectable markers

would be antibiotic resistance determinants or various nutritional markers that repair various auxotrophies. Nutritional markers include *trpA* for tryptophan biosynthesis [140] and *pyrF* for uracil biosynthesis [141]. Negative selectable markers include genes that result in cell damage when grown in the presence of a certain compound, like *hpt* [142, 143], *tdk* [144], *pyrF* [141], or *sacB* [145]. The *hpt and tdk* gene products incorporate toxic nucleotide analogs into DNA; *pyrF* encodes the orotidine-monophosphate decarboxylase, which converts 5-fluoroorotic acid into a toxic compound; *sacB* leads to lethality when the cells are exposed to high levels of sucrose. As an alternative to negative selection, removal of certain sequences can be achieved in some organisms by the expression of site-specific recombinases like the Cre/*lox* and the Flp/*FRT* systems [146], which catalize site-specific recombination between homologous sites flanking the region to be deleted.

Developing DNA sequences and proteins that control gene expression are required to introduce new gene functions. A variety of transcriptional promoters is necessary—constitutive promoters for static gene expression and regulated promoters for dynamic expression [147]. Regulated promoters require specific transcriptional regulatory proteins, the likes of which are beginning to be studied in cellulytic thermophiles [148, 149]. There are even more sophisticated expression tools in some organisms such as altered translation initiation regions to control protein levels [150, 151].

Thermophilic genetic tools. Thermophilic archaea have been the subject of basic and applied research, as many of them have unique and useful central

carbon metabolisms and thermostable enzymes [152]. Genetic tools for *Pyrococcus furiosus*, a marine hydrogen-producing archaeon that grows at 100°C, include a strain naturally competent for linear DNA uptake [153], a replicating shuttle vector based on the chromosomal replication origin [154], and methods for rapid gene deletion [155]. There are also methods for linear DNA introduction in thermophilic *Sulfolobus* species [156, 157] and *Thermococcus kodakaraensis* [158], and replicating shuttle vectors based on native thermophilic plasmids for each [159-161].

Thermophilic genetic tools are not easily adapted from tools available in the genetic workhorse organisms *E. coli* and *B. subtilis.* This is not because of physical or chemical differences in the actual plasmid DNA between mesophiles and thermophiles [162], but rather because proteins required for DNA replication of plasmids are nonfunctional at extremely high temperatures. The plasmids for extrachromosomal replication must be derived from native, preferably small, plasmids in thermophilic hosts.

Genetic methods mostly rely on nutritional selection for DNA uptake and counter-selection, as antibiotics themselves are usually non-functional at temperatures ≥ 72°C, except kanamycin and neomycin (reviewed in [163]). The antibiotic resistance markers are also usually non-functional at very high temperatures, but mesophilic antibiotic resistance markers have been applied to more moderate thermophiles like *Thermoanaerobacterium saccharolyticum* [164] and *Clostridium thermocellum* [122, 165]. To overcome issues with more extreme thermophiles, heat-stable kanamycin markers from *Staphylococcus aureus* were

screened, mutated and evolved for greater thermostability, and applied to *Thermus thermophilus* (T_{opt} 70-75°C) in the 1980s and 1990s [166-169]. More recently, this same marker, called *htk* (for high temperature kanamycin) was applied to *Thermoanaerobacter ethanolicus* (T_{opt}=69°C). A codon-optimized version of this kanamycin resistance marker was developed for Gram-positive thermophilic bacterium *Caldicellulosiruptor bescii* (T_{opt}=78°C) [170].

With the recent interest in CBP, more focus has been placed on genetic tool development in cellulolytic and solventogenic thermophiles. Interestingly, various *Thermoanaerobacter* and *Thermoanaerobacterium* strains are naturally competent for DNA upake [171]. This greatly simplified the process for genetic manipulation, although replicating vector pIKM1 is available for *T. saccharolyticum* [164] and pTE16 for *T. ethanolicus* [172]. Recently, natural competence was used to create a negative selection system based on the *tdk* marker in *T. ethanolicus* [144].

Clostridium thermocellum and *Caldicellulosiruptor bescii* have not shown signs of natural competence, so genetic tools must rely on circular plasmids. While non-replicating vectors function in *C. bescii*, *C. thermocellum* seems to require a replication origin for transformation to occur. The pNW33N replication origin is most commonly used for these transformations, although it does not replicate at the optimum growth temperature of *C. thermocellum* (60°C). The chloramphenicol acetyltransferase gene (*cat*) is most commonly used as an antibiotic resistance determinant to thiamphenicol, a thermostable derivative of chloramphenicol [173]. Gene deletions have been performed with positive-

negative selection cassettes in *C. thermocellum*, both *pyrF*-based [122, 174] and *hpt*-based [112, 123, 124, 126, 127, 175]. The *tdk* marker is often used to select against the replication backbone of plasmids used for genetic manipulation, so that only the positive-negative *cat-hpt* cassette integrates into the genome. A set of endogenous promoters have been characterized using a thermostable *LacZ* reporter gene [176]. One of these promoters has been developed into a regulated promoter system based on the inducing sugar laminaribiose [149].

The *htk* kanamycin resistance marker was only recently developed for *C. bescii*, so most genetic work to date has relied on *pyrF* for positive and negative selection. A replicating shuttle vector [177] has been created from the native *C. bescii* plasmid pBAS2 [178]. This vector allows rapid expression of homologous and heterologous genes at the optimal growth temperature of *C. bescii*, both for improving cellulose hydrolysis and metabolic engineering [179, 180]. Non-replicating knockout vectors have been used to create a variety of gene deletions using the positive and negative selection capabilities of *pyrF* [130, 136, 181]. Chromosomal insertion sites have been chosen and utilized for heterologous gene expression. Two strong promoters have driven expression of selectable markers and heterologous genes: a ribosomal protein promoter and the S-layer protein promoter [117, 132, 182, 183].

Transforming a newly isolated organism with foreign DNA is the first step in the process of developing genetic tools, and work led by Daehwan Chung showed that this is anything but trivial. For *C. bescii*, achieving transformation required understanding and overcoming restriction-modification systems [184],

that is, site-specific endonucleases and their cognate DNA methyltransferases. Restriction systems, particularly those that act at short DNA sequences such as GGCC, can often act as absolute barriers to transformation, as was the case with *C. bescii* [185, 186] and the archaeon *Sulfolobus* [157, 187]. Restriction systems have not been prohibitive to transformation in *C. thermocellum* or *T. saccharolyticum*, but deletion of a particular restriction enzyme in *C. thermocellum* did increase transformation efficiency (Riley, Papanek, and Guss, unpublished data). Understanding restriction systems, transforming organisms with DNA and then developing and applying genetic tools in thermophiles will enable scientists to engineer new “bugs from the mud” and exploit the wide range of metabolisms and activities found in nature.

VI. Structure of the Dissertation and Major Issues Addressed

This work is presented in manuscript-style. Chapters 2 and 3 are first author published manuscripts, and Chapters 4 and 5 are to be submitted as first-author published manuscripts. Appendices A, B, and C are abstracts for middle-author publications. The following are the major issues that were addressed in each manuscript or planned manuscript, listed with the chapter of the dissertation in which they are addressed:

1. The need for genetic methods for other members of the *Caldicellulosiruptor* genus. Before this work, *C. bescii* was the only tractable genetic organism in this genus. Other species of this genus are needed for comparative analysis of biomass deconstruction ability. *C. hydrothermalis* is the least cellulolytic species, so it will provide a relatively naïve expression host to

analyze the effects of plant biomass deconstruction enzymes from more cellulolytic species.

In Chapter 2: Heterologous complementation of a *pyrF* deletion *in Caldicellulosiruptor hydrothermalis* generates a new host for the analysis of biomass deconstruction.

2. The need for plasmid replicons. There are very few plasmid replicons that work in thermophiles in general, and there were no extreme thermophilic autonomously replicating plasmids for *Clostridium thermocellum*, the most promising thermoplile for consolidated bioprocessing.

In Chapter 3: Promiscuous plasmid replication in thermophiles: Use of a novel hyperthermophilic replicon for genetic manipulation of *Clostridium thermocellum* at its optimum growth temperature.

3. An understanding of plasmid replication. Very little is known about how these plasmids replicate, what is required for stable replication or how to maintain plasmids in high copy.

In Chapter 4: Deletion of the *Clostridium thermocellum recA* gene reveals that it is required for thermophilic plasmid replication.

4. The need for a reporter gene. Reporter genes that allow rapid detection of gene expression do not exist in cellulolytic hyperthermophiles, nor are there characterized regulated promoters:

In Chapter 5: A new reporter gene for anaerobic cellulolytic thermophiles: Use for the analysis of maltose-regulated promoters

5. An understanding of the enzymes that degrade plant biomass and how they function.

In Appendix A: 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(3): e0119508.

6. Technology to identify DNA methylation relevant for transformation of non-model organisms.

In Appendix B: Yu M, Ji L, Neumann DA, Chung D, Groom J, Westpheling J, He C, Schmitz RJ. (2015) Base-resolution detection of N^4-methylcytosine in genomic DNA using 4mC-Tet-assisted-bisulfite sequencing. *Nucleic Acids Research* 43(21): e148

7. Understanding the effect of derived toxins on bacterial growth and product synthesis and establishing mechanisms of resistance.

In Appendix C: Kim S, 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. *Biotechnology for Biofuels* 10: 66.

CHAPTER 2

HETEROLOGOUS COMPLEMENTATION OF A *PYRF* DELETION IN *CALDICELLULOSIRUPTOR HYDROTHERMALIS* GENERATES A NEW HOST FOR THE ANALYSIS OF BIOMASS DECONSTRUCTION.[1]

[1] Groom J, Chung D, Young J and Westpheling J. 2014. *Biotechnology for Biofuels.* 7:132.

I. Abstract

Background: Members of the thermophilic, anaerobic Gram-positive bacterial genus *Caldicellulosiruptor* grow optimally at 65 to 78°C and degrade lignocellulosic biomass without conventional pretreatment. Decomposition of complex cell wall polysaccharides is a major bottleneck in the conversion of plant biomass to biofuels and chemicals and conventional biomass pretreatment includes exposure to high temperatures, acids, or bases as well as enzymatic digestion. Members of this genus contain a variety of glycosyl hydrolases, pectinases, and xylanases, but the contribution of these individual enzymes to biomass deconstruction is largely unknown. *C. hydrothermalis* is of special interest because it is the least cellulolytic of all the *Caldicellulosiruptor* species so far characterized, making it an ideal naïve system to study key cellulolytic enzymes from these bacteria.

Results: To develop methods for genetic manipulation of *C. hydrothermalis*, we selected a spontaneous deletion of *pyrF,* a gene in the pyrimidine biosynthetic pathway, resulting in a strain that was a uracil auxotroph resistant to 5-fluoroorotic acid (5-FOA). This strain allowed the selection of prototrophic transformants with either replicating or non-replicating plasmids containing the wild-type *pyrF* gene. Counter-selection of the *pyrF* wild-type allele on non-replicating vectors allowed the construction of chromosomal deletions. To eliminate integration of the non-replicating plasmid at the *pyrF* locus in the *C. hydrothermalis* chromosome, we used the non-homologous *Clostridium thermocellum* wild-type *pyrF* allele to complement the *C. hydrothermalis pyrF* deletion. The autonomously replicating shuttle vector was maintained at 25 to 115 copies per chromosome. Deletion of the ChyI restriction

enzyme in *C. hydrothermalis* increased the transformation efficiency by an order of magnitude and demonstrated the ability to construct deletions and insertions in the genome of this new host.

Conclusions: The use of *C. hydrothermalis* as a host for homologous and heterologous expression of enzymes important for biomass deconstruction will enable the identification of enzymes that contribute to the special ability of these bacteria to degrade complex lignocellulosic substrates as well as facilitate the construction of strains to improve and extend their substrate utilization capabilities.

II. Introduction

Plant biomass recalcitrance is one of the most important barriers to its use as a substrate for the production of fuels and chemicals using micro-organisms as catalysts. The plant cell wall consists of a complex web of polysaccharides and phenolics that function in plant structure and development [2]. Perennial plants like switchgrass could be incorporated into so-called agro-ecosystems, which would increase carbon storage and biofuel production, decrease carbon dioxide emissions, and improve water quality through wetland denitrification [7]. While the natural recalcitrance of plant biomass is a major barrier [13], several methods including direct microbial breakdown of cell wall structures can be used to liberate energy-rich sugars for conversion to useful biofuels and bioproducts.

Chemical and thermal pretreatments are often used to break down the raw substrate, but they are expensive and destructive to the sugars in the biomass [188], and they produce hydrolysates inhibitory to both cellulose degradation and microbial growth [189]. In contrast, thermophilic anaerobes in the genus *Caldicellulosiruptor*

can deconstruct high loadings of plant biomass into simple sugars without conventional pretreatment [90, 100, 101] and have recently been engineered to produce ethanol directly from switchgrass [117]. *Caldicellulosiruptor* species can simultaneously utilize the wide range of hexoses, pentoses, oligosaccharides, and polysaccharides released from the plant cell wall, and there is no evidence of carbon catabolite repression [105, 106]. These qualities make them well suited for consolidated bioprocessing (CBP), in which one microorganism is used for both biomass deconstruction and end-product formation.

Members of the *Caldicellulosiruptor* genus are anaerobic Gram-positive bacteria, and they are the most thermophilic cellulose-degrading organisms known [102]. They secrete free carbohydrate-active enzymes (CAZys) [103] with carbohydrate-binding modules that are well suited for binding and degrading cell wall polysaccharides [104]. *C. bescii* is one of the strongest crystalline cellulose degraders in the genus, whereas the closely related *C. hydrothermalis* is one of the weakest [100]. Interestingly, *C. hydrothermalis* lacks the multidomain CAZys found in more cellulolytic members of the genus [114] as well as a cluster of pectinases that affect *C. bescii* growth on biomass [137]. In addition to lacking multidomain enzymes, *C. hydrothermalis* completely lacks the glycosyl hydrolase (GH) domains GH9 and GH48 [102], the domains that comprise the most highly secreted cellulase CelA in *C. bescii* [190]. *C. hydrothermalis* thus provides a “blank slate” with which to study thermophilic enzymes important for biomass degradation *in vivo*, and is a promising system for heterologous expression of plant biomass deconstruction enzymes.

To establish methods for genetic manipulation of *C. hydrothermalis*, we took an approach similar to the one previously used for *C. bescii* that relied on the selection of a *pyrF* deletion mutant, which allows for nutritional selection of transformants [186]. Interestingly, *C. hydrothermalis* contains fewer mobile genetic elements than other members of the genus [191], so this species may have other advantages for genetic manipulation, including fewer issues with genome stability that could result from genetic selections and counter-selections. We transformed the *pyrF* deletion mutant with the pJGW07 shuttle vector that is based on a native plasmid, pBAS2, from *C. bescii* [192]. Both the *C. bescii* and the heterologous *Clostridium thermocellum* wild-type *pyrF* allele were shown to complement this deletion, restoring the mutant to uracil prototrophy. Deletion of the ChyI restriction enzyme in *C. hydrothermalis*, a homolog of a HaeIII-like restriction enzyme known to be a barrier to transformation in *C. bescii* [185, 193], increased the transformation efficiency by about an order of magnitude. The new strain *C. hydrothermalis* JWCH008 should facilitate the assessment of plant biomass deconstruction by the *Caldicellulosiruptor* genus and the molecular engineering of deconstruction enzymes.

III. Methods

Selection of a *pyrF* deletion. Wild-type *C. hydrothermalis* DSM 18901 was grown from a 0.5% inoculum in 50 mL of a low osmoloarity defined growth medium (LOD) supplemented with 40 μM uracil. Cultures were grown at 55°C, 60°C, 68°C, and 75°C. Cells in the late exponential phase were cooled to room temperature, harvested by centrifugation at 6,100 × g, and resuspended in 1X *C. bescii* partial base salts [90]. 100 μL of resuspended cells were mixed with a 1.5% agar overlay

and plated onto LOD media with 40 μM uracil and 8 mM 5-FOA. The plates were degassed in anaerobic chambers and incubated for 4 days at 68°C. The colonies were picked directly into 20 mL LOD media with 40 μM uracil, which was immediately degassed and incubated at 68°C. When the media was turbid, chromosomal DNA was extracted with a Zymo Research gDNA Extraction kit (Irvine, CA). The *pyrF* gene region was amplified from the wild type and deletion mutant with primers DC163 and DC164 (Figure 2.1A), and the PCR products were analyzed for the presence of deletions in a 1.5% w/v agarose gel by electrophoresis (Figure 2.1B). The sequences of all plasmids were verified by automatic sequencing (GENEWIZ).

Construction of a shuttle vector containing the *pyrF* gene from *Clostridium thermocellum* ATCC 27405. High-Fidelity Q5 polymerase (New England Biolabs (NEB), Ipswich, MA) restriction enzymes (NEB) and Fast-link™ DNA Ligase (Epicentre Technologies, Madison, WI) were used according to manufacturer instructions. pJGW07 (Figure 2.2A) was constructed by replacing the *C. bescii pyrF* gene (Cbes1377) on the replicating shuttle vector pDCW89 [177] with the *Clostridium thermocellum* homolog for *pyrF* (Cthe0951). Cthe0951 was amplified by PCR using primers JG021 and JG022, engineered to contain XbaI and NcoI restriction sites, respectively. This 945-bp PCR product was ligated directionally to a 6.75-kb pDCW89 PCR product amplified using primers JG023 and JG024, which also contain XbaI and NcoI sites. Correct clones were purified from *E. coli* with a Miniprep kit (Qiagen), and confirmed by restriction fragment analysis and DNA sequencing (GENEWIZ).

Transformation of *C. hydrothermalis* JWCH006. To prepare cells for transformation, 15 mL of a freshly grown JWCH006 (*ΔpyrF*) culture were inoculated into four 500-mL bottles of fresh LOD supplemented with 40 μM uracil and 19 amino acids, and incubated at 65°C to the early-exponential phase (OD_{680} approximately 0.04 to 0.05). The cultures were cooled to room temperature for 1 h, harvested by centrifugation (5,000 × g, 10 min) at 4°C and washed three times with 50 mL pre-chilled 10% sucrose. After the final wash, the cell pellets were resuspended in 250 µL pre-chilled 10% sucrose. 60-µL aliquots of competent JWCH006 were added to plasmid DNA (0.5 µg), either methylated with M.CbeI methyltransferase, as previously described [186], or unmethylated, gently mixed and incubated in 10% sucrose for 15 minutes at room temperature. Electrotransformation of the cell/DNA mixture was performed via a single exponentially decaying electric pulse (1.8 kV, 350Ω, and 25 microF) in a pre-chilled 1-mm cuvette using a Bio-Rad Gene Pulser (Bio-Rad, Hercules, CA). After pulsing, the cells were incubated in 10 mL low osmolarity complex growth medium (LOC) [194] at 68°C. After 4 h, the cultures growing in LOC medium were cooled to room temperature, harvested by centrifugation (6,100 x g, 10 min), and washed three times at room temperature with 1X *C. bescii* partial base salts [90] to remove the rich media. The cells were finally resuspended in 800 µL 1X base salts. For each plate, 100 µL of resuspended cells were mixed with 2 mL of a 1% agar overlay and plated onto LOD media lacking uracil to select for transformation. The plates were degassed in anaerobic chambers and incubated for 4 days at 65°C. The colonies were picked directly into 20 mL LOD media without uracil, which was immediately degassed and incubated at 65°C. Uracil-

prototrophic transformants were confirmed by PCR amplification of *C. hydrothermalis* DNA using primers JG021 and JF199, which are specific for pJGW07 (Figure 2.3, Figure 2.4).

Analysis of plasmid structure and stability in *C. hydrothermalis*. A single transformant colony was picked directly into 20 mL LOD media lacking uracil, which was immediately degassed and incubated at 65°C. This strain maintaining the pJGW07 shuttle vector was named JWCH009. JWCH009 was grown to the late exponential phase (OD_{680} approximately 0.15) in 50 mL LOD. Direct extraction of plasmid DNA from JWCH009 was performed as previously described [193, 195]. The plasmid DNA was digested with enzymes HaeIII, EcoRI, HhaI, and MboI (NEB).

To determine structural stability of the plasmid, DNA was extracted from *C. hydrothermalis* with a gDNA extraction kit (Zymo Research), and 2 μL of DNA was electro-transformed into *E. coli* DH5α via single electric pulse (2.5 kV, 200Ω, and 25 microF) in a pre-chilled 2 mm cuvette using a Bio-Rad gene Pulser. Cells were placed into 1 mL SOC media for 1 h with shaking at 37°C, and then plated onto LB agar supplemented with 50 μg/mL apramycin. Colonies were picked into 10 mL LB with 50 μg/mL apramycin, and plasmid DNA was extracted using a Miniprep kit (Qiagen) and screened with restriction enzymes EcoRI and ApaLI (NEB) (Figure 2.4B).

Plasmid copy number determination by quantitative PCR. To determine the pJGW07 copy number and maintenance over time, JWCH009 was grown to OD_{680} approximately 0.15 and serially subcultured for 5 days. The total DNA was isolated from the cultures and treated with RNase A. qPCR experiments were carried

out with a LightCycler 480 Real-Time PCR instrument (Roche, Indianapolis, IN) with LightCycler 480 SYBR Green I Master mix (Roche). Two independent sets of primers specific to either the pJGW07 plasmid (Q1/Q2 inside Cbes2777, Q3/Q4 inside Cbes2778) or the *C. hydrothermalis* chromosome (Q11/Q12 inside Calhy0897, Q13/Q14 inside Calhy1377) were used to compute the relative copy number of the plasmid to the chromosome. Three replicate reactions for each primer set were performed, and the average of the two primer sets on each sample was used to calculate the plasmid copy number in each serial subculture (Table 2.3) according to the method of Lee *et al* [196]. The amplification efficiency over 10^4-fold range of DNA concentration was 93.5%, within the ratio of 90 to 110% that is considered acceptable (Life Technologies).

Construction of a deletion of the *chyI* gene in *C. hydrothermalis*. A vector backbone was amplified from suicide vector pDCW88 [185] with primers DC081 and DC262. Flanking regions for the *chyI* gene (Calhy0408) were amplified using primers DC484 and DC485 (5′ flanking region) and DC486 and DC487 (3′ flanking region). The 5′ and 3′ flanking regions were combined into one fragment by overlap extension PCR and ligated into the pDCW88 vector backbone using restriction enzymes KpnI and ApaLI. The resulting plasmid pDCW151 was used to construct pJGW03 in which the *C. bescii pyrF* gene cassette (Cbes1377) was replaced with the *C. thermocellum pyrF* gene (Cthe0951) as described above. Competent *C. hydrothermalis* JWCH006 cells were prepared as described above, and 1.0 µg of non-replicating plasmid pJGW03 was added to 50-µL aliquots of competent JWCH006, gently mixed and incubated for 15 min at room temperature. Electrotransformation of the cell/DNA

mixture was performed via a single electric pulse (1.8 kV, 350Ω, and 25 microF) in a pre-chilled 1-mm cuvette using a Bio-Rad Gene Pulser. After pulsing, the cells were incubated in 10 mL low osmolarity complex growth medium (LOC) [194] at 68°C with shaking at 150 rpm. From this culture, 0.5% inocula were transferred into LOD medium lacking uracil at 65°C every hour for 4 h to select for integration into the *C. hydrothermalis* genome. When the cultures in the selective media were turbid, a 0.5% inoculum was transferred to 50 mL LOD supplemented with 40 μM uracil to allow a loopout of the wild-type *pyrF* allele to occur. This culture was then plated onto LOD plates supplemented with 40 μM uracil and 6 mM 5-FOA to select against the *pyrF* wild-type allele, and grown for 4 days at 68°C. Colonies were picked directly into 20 mL LOD medium with uracil, which was immediately degassed and incubated at 68°C. The resulting cultures were screened for a deletion using primers JG025 and JG026, and one deletion culture was purified twice on solid LOD media supplemented with 40 μM uracil.

IV. Results and Discussion

Selection for resistance to 5-FOA resulted in a spontaneous deletion of the *pyrF* gene in *C. hydrothermalis*. The *pyrF* gene encodes orotidine monophosphate decarboxylase, an enzyme in the pyrimidine biosynthesis pathway. Deletion of this gene results in uracil auxotrophy and resistance to 5-fluoroorotic acid (5-FOA), allowing prototrophic selection of transformants and counter-selection of the wild-type allele [141]. The optimal growth temperature for *C. hydrothermalis* is 65°C [197], and we had previously observed an increase in the spontaneous mutation rate in cells grown above and below 65°C. To obtain spontaneous deletions of *pyrF*, cells

were grown at non-optimal temperatures in the presence of uracil. The presence of uracil in the growth medium allowed for the maintenance of cells with spontaneous mutations in the *pyrF* gene. After growth at various temperatures, cells were plated onto a medium with 5-FOA selecting resistance and loss of *pyrF* function. One 5-FOA resistant mutant, JWCH006 (Table 2.1), that had been grown at 60°C contained a 99-bp deletion in *pyrF* (Figure 2.1A) and was confirmed to be a uracil auxotroph resistant to 5-FOA.

Uracil auxotrophy in *C. hydrothermalis* is complemented by heterologous expression of the *Clostridium thermocellum pyrF* gene. Electrocompetent *C. hydrothermalis* JWCH006 (*ΔpyrF*) cells were prepared based on the method for *C. bescii* [186] and transformed with a previously described shuttle vector, pDCW89 [177], containing the wild-type *C. bescii pyrF* allele. Transformation of *C. bescii* with plasmid DNA isolated from *E. coli* requires *in vitro* methylation by a methyltransferase, M.CbeI [186]. As *C. hydrothermalis* has a similar restriction profile to that of *C. bescii* [185], we anticipated that *in vitro* methylation by M.CbeI would protect plasmid DNA isolated from *E. coli* and allow transformation of *C. hydrothermalis*. pDCW89 DNA methylated with M.CbeI *in vitro* successfully transformed the *C. hydrothermalis ΔpyrF* mutant to prototrophy at an average frequency of 37 colony forming units (CFUs) per microgram of DNA (Table 2.2).

Since the deletion of *pyrF* in JWCH006 was only 99 bp (Figure 2.1A), and there is 95% DNA sequence homology between the *pyrF* genes in *C. bescii* and *C. hydrothermalis*, there was a large region of residual homology between the *pyrF* locus in the *C. hydrothermalis* chromosome and the *C. bescii pyrF* gene on the

plasmid. Although pDCW89 DNA methylated *in vitro* with M. CbeI transformed *C. hydrothermalis* JWCH006 to prototrophy, we found that the wild-type plasmid-encoded *pyrF* allele repaired the mutant *pyrF* locus (Figure 2.2) in almost every case, and the plasmid was not maintained autonomously over time. Since marker replacement in *C. hydrothermalis* relies on the integration of non-replicating plasmids at sites other than *pyrF*, the *C. bescii pyrF* cassette in pDCW89 was replaced by the *C. thermocellum* homolog (Cthe0951) to create pJGW07 (Figure 2.3A). While the two PyrF protein sequences are 45% identical, the *C. thermocellum pyrF* gene has very low DNA sequence homology with the *Caldicellulosiruptor pyrF* genes. Transformation of JWCH006 with methylated pJGW07 was successful (Figure 2.3B, Figure 2.4), and uracil prototrophic transformants were obtained at a frequency of 22 ± 10 CFU per microgram of DNA (Table 2.2).

Plasmid pJGW07 was purified directly from *C. hydrothermalis* based on a previously published method for other Gram-positive organisms [195]. Undigested plasmid isolated from *C. hydrothermalis* differs in its band pattern from plasmid isolated from *E. coli*, and we suggest that the difference is not in size but in methylation within the native host compared to *E. coli* (Figure 2.2B). Restruction endonuclease analysis using a panel of enzymes indicated that plasmid isolated from *C. hydrothermalis* is protected at GGCC sites by a HaeIII-like modification system as expected but not at GATC sites (Figure 2.5). This suggests that *C. hydrothermalis* lacks the DNA adenine methylase present in *E. coli*. EcoRI or HhaI recognition sites are not protected in either organism.

To confirm that the plasmid was replicating autonomously, DNA isolated from the *C. hydrothermalis* transformant JWCH009 (Table 2.1) was back-transformed into *E. coli* DH5α. Plasmid DNA recovered from 12 apramycin-resistant *E. coli* transformants was identical in its restriction patterns to pJGW07 transformed into *C. hydrothermalis,* suggesting that the plasmid was structurally stable during transformation, replication and back-transformation to *E. coli* (Figure 2.4). There was no evidence of plasmid integration (Figure 2.6) resulting from recombination between the *C. thermocellum pyrF* gene on the plasmid and the *C. hydrothermalis pyrF* gene in the chromosome.

These data show that, although *C. thermocellum* has an optimal growth temperature of 60°C, the *C. thermocellum* orotidine-5′-phosphate decarboxylase functions at temperatures up to at least 65°C and that the *C. thermocellum* gene is expressed at sufficient levels to complement the *C. hydrothermalis pyrF* deletion.

A shuttle vector derived from a native *C. bescii* plasmid is maintained at a high copy number in *C. hydrothermalis*. We recently reported the construction of a shuttle vector for *C. bescii* [177] based on the smaller of two native plasmids in that species [192]. The native plasmid pBAS2 is maintained in *C. bescii* at a copy number of approximately 75 [177]. Because the shuttle vector derived from this plasmid competed with the native plasmid it was derived from, the shuttle vector was maintained in low copy and was readily lost without selection. While an unstable plasmid is useful for some applications, stability and high copy number also have advantages. No native plasmid DNA was detected in *C. hydrothermalis* using conventional plasmid isolation methods, nor was one identified during the sequencing

of total DNA isolated from this strain [198]. We anticipated that the *C. bescii* shuttle vector would likely replicate in *C. hydrothermalis* and might be stably maintained at a high copy number. Quantitative polymerase chain reaction (qPCR) was performed as described in the Methods section to determine the copy number of the pJGW07 plasmid in *C. hydrothermalis,* and the results indicated that it was maintained between 25 and 115 copies per chromosome over five serial transfers, but was quickly lost without selection (Figure 2.7, Table 2.3).

Deletion of the *C. hydrothermalis* ChyI restriction enzyme results in increased transformation efficiency. In developing methods for DNA transformation of *C. bescii,* we observed that restriction by a HaeIII isoschizomer, CbeI [193], was an absolute barrier to transformation of DNA from *E. coli*. We identified, cloned, expressed, and purified its cognate methyltransferase, M.CbeI, from *C. bescii* and showed that DNA methylated *in vitro* readily transformed *C. bescii* [186]. Deletion of *cbeI* in *C. bescii* relieved the requirement for *in vitro* methylation of plasmid DNA from *E. coli* by M.CbeI [185] and allowed efficient DNA transformation. The *C. hydrothermalis chyI* gene is an ortholog of *cbeI* with 96% DNA sequence identity and 100% sequence coverage. A deletion of *chyI* was constructed on a plasmid, pJGW03, that was transformed into *C. hydrothermalis* JWCH006 (Figure 2.8A). A deletion mutant was readily obtained in a screen of 50 colonies, and only two rounds of purification on low osmolarity defined (LOD) plating media containing uracil [194] were required to resolve the merodiploid, resulting in *C. hydrothermalis* JWCH008 *(ΔpyrF ΔchyI*) (Figure 2.8C). We note that the region of the *C. hydrothermalis* genome that contains ChyI is not identical to the region of the *C.*

bescii chromosome that contains CbeI. The *C. hydrothermalis* region contains an additional open reading frame (ORF) that apparently encodes a truncated form of the N-terminal portion of the ChyI protein. Sequence analysis revealed a premature stop codon likely resulting from a point mutation in the ChyI ORF. This altered gene structure had no obvious effect on enzyme function *in vivo*. The deletion we designed included both ORFs.

To assess whether the loss of the ChyI restriction enzyme resulted in an increase in transformation efficiency in *C. hydrothermalis*, we compared the transformation efficiencies of the two strains generated in this study, JWCH006 (*ΔpyrF*) and JWCH008 (*ΔpyrF ΔchyI*). As shown in Table 2.2, electrotransformation of the JWCH006 parent strain with either the pJGW07 shuttle vector containing the *C. thermocellum pyrF* gene, or the pDCW89 vector containing the *C. bescii pyrF* gene, was low but detectable, and methylation of plasmid DNA did not make a significant difference in transformability. The standard deviation in experiments with low numbers of transformants is substantial but not unexpected [199]. This extremely low transformation efficiency may be an underestimate of the actual efficiency, as the plating efficiency of *C. hydrothermalis* on a selective solid medium is less than 10^{-4} (plating 10^6 cells as determined by cell count resulted in fewer than 100 colonies). Transformation of JWCH008 containing the *chyI* deletion was an order of magnitude higher than that of JWCH006, and again, there is no significant difference between the transformation efficiency of methylated and unmethylated DNA, suggesting differences in the restriction/modification systems of *C. hydrothermalis* and *C. bescii*. The fact that methylation with M.CbeI made no difference in transformation efficiency

was somewhat surprising, especially considering the fact that deletion of ChyI increased the transformation frequency of unmethylated DNA. We interpret this to indicate that there are differences between *C. bescii* and *C. hydrothermalis* in their restriction/modification systems and perhaps additional enzymes in one or the other that account for the differences in transformation frequencies. It is also possible that the truncated form of the CbeI/ChyI orthologous proteins makes a difference in their activities (Figure 2.8B). We previously reported that *C. hydrothermalis* chromosomal DNA is resistant to digestion by BamHI and BspEI [185]. Both these enzymes have six base recognition sequences that are relatively rare compared to four base recognition sequences. The plasmids used in this study do not contain BamHI sites. While there are two BspEI restriction sites (TCCGGA), *E. coli* DH5α, the strain we used to make plasmid DNA, contains an adenine methyltransferase known to protect this site and may prevent cleavage by *C. hydrothermalis* during DNA transformation.

We emphasize that the observed transformation of the JWCH006 parent strain is not an indication that *C. hydrothermalis* is naturally competent for DNA uptake. Preparation of electrocompetent cells and electroporation was necessary to detect transformation. We have invested some effort to induce natural competence in members of this genus, but those efforts have not been successful. There are homologs to the natural competence genes *comEA, comEC, comGC,* and *dprA* [200], but in contrast to other thermophilic anaerobes [171], there is no evidence to date of natural competence in *Caldicellulosiruptor* species.

IV. Conclusions

Methods for genetic manipulation of *C. hydrothermalis,* based on those used for *C. bescii,* were successful and efficient. Restriction of DNA was not an absolute barrier to transformation, but deletion of the ChyI restriction enzyme in *C. hydrothermalis* increased the transformation efficiency by an order of magnitude. Heterologous expression of the *Clostridium thermocellum pyrF* gene was sufficient to complement the *C. hydrothermalis ΔpyrF* mutant, allowing both autonomous plasmid replication at relatively high copy (about 25 to 115 copies/cell) and marker replacement of the *chyI* gene in the *C. hydrothermalis* chromosome. The use of this new strain, *C. hydrothermalis* JWCH008, should allow for the expression of heterologous and homologous enzymes for both the identification and analysis of enzymes involved in biomass deconstruction of unpretreated plant biomass by the *Caldicellulosiruptor* genus. It will also enable the engineering of glycosyl hydrolases such as CelA and other important plant biomass deconstruction enzymes in a strain devoid of similar enzymes or activities.

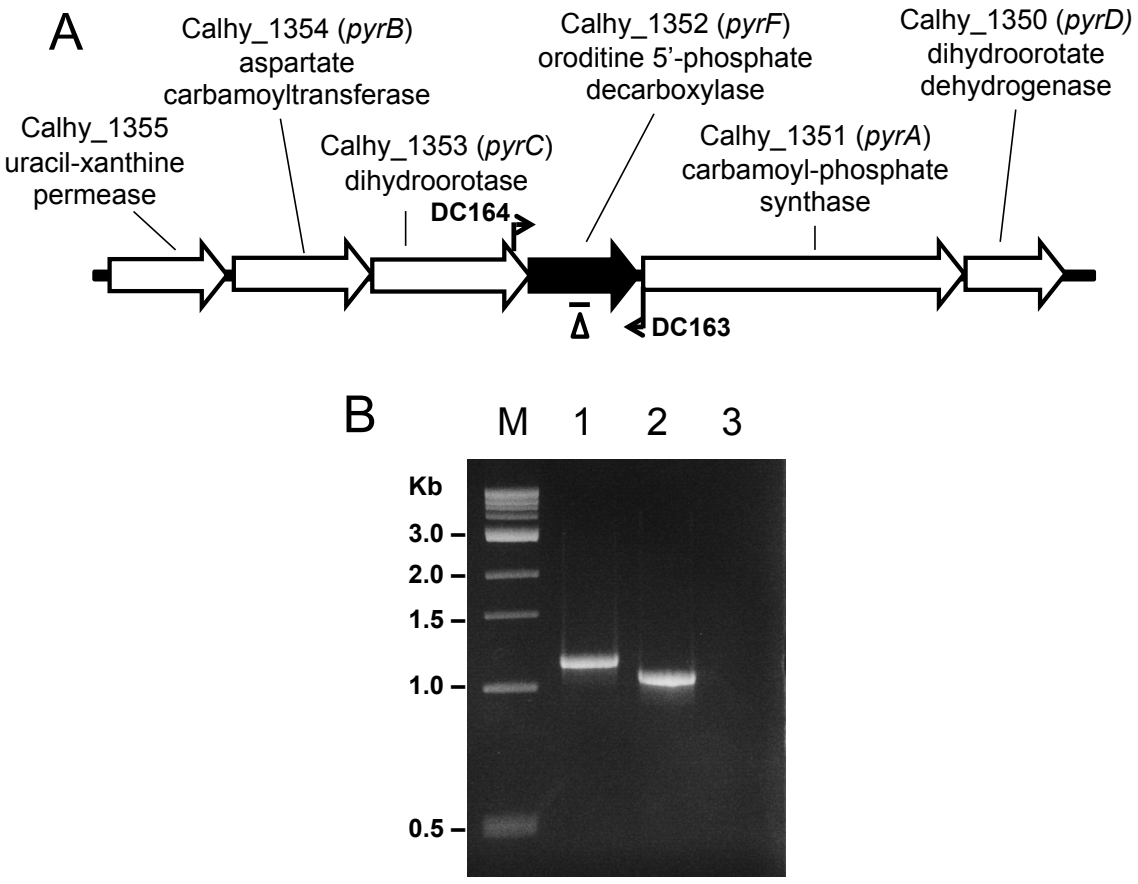


Figure 2.1. Isolation of a spontaneous pyrF mutant in C. hydrothermalis. (A) Chromosomal map of the uridine monophosphate (UMP) biosynthetic gene cluster in *C. hydrothermalis* JWCH006. The 99-bp spontaneous deletion in Δ*pyrF* is indicated by the line below the diagram. 462 bp lie upstream and 357 bp lie downstream of the deletion. Arrows depict primers used to verify the structure of the *pyrF* gene in the JWCH006 strain. **(B)** Gel depicting PCR products of the *pyrF* region in the wild-type strain (1.13 kb) compared to the JWCH006 strain (1.02 kb) amplified by the indicated primers (DC163 and DC164). M: 1 kb DNA ladder (NEB); 1: wild-type genomic DNA; 2: JWCH006 genomic DNA; 3: negative control.

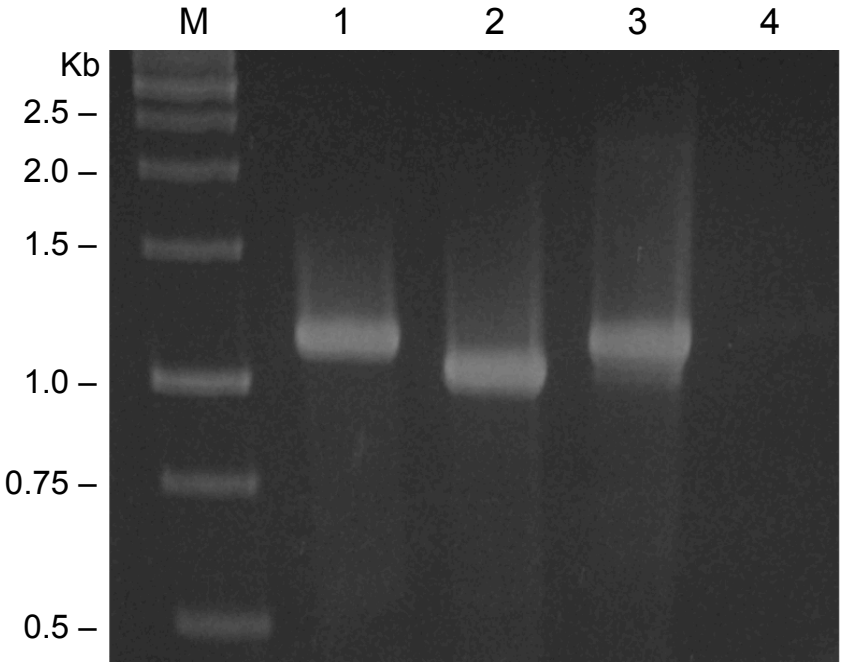


Figure 2.2. Repair of the *pyrF* gene in pDCW89 transformants. *C. hydrothermalis* JWCH006 (*ΔpyrF*) was transformed with M.CbeI-methylated pDCW89, and individual transformant colonies were picked. DNA was isolated from the strain, and PCR using *pyrF*-flanking primers DC163 and DC164 (diagrammed in Figure 2.1) was performed. M: molecular weight standards (NEB); 1: *C. hydrothermalis* wild-type genomic DNA; 2: *C. hydrothermalis* JWCH006 genomic DNA; 3: genomic DNA from *C. hydrothermalis* JWCH006 transformed with pDCW89; 4: negative control.

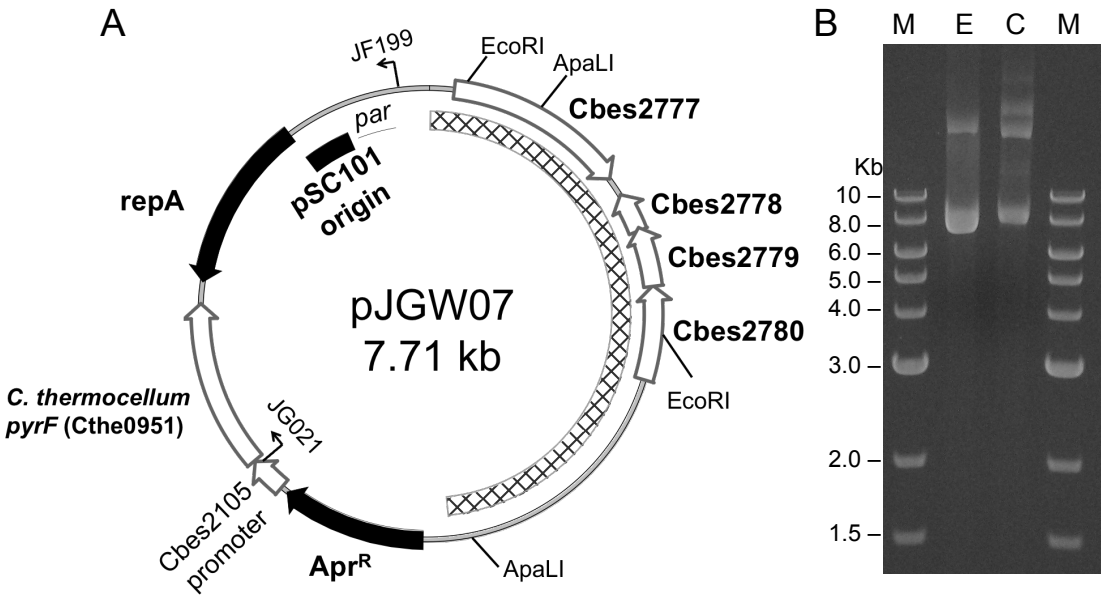
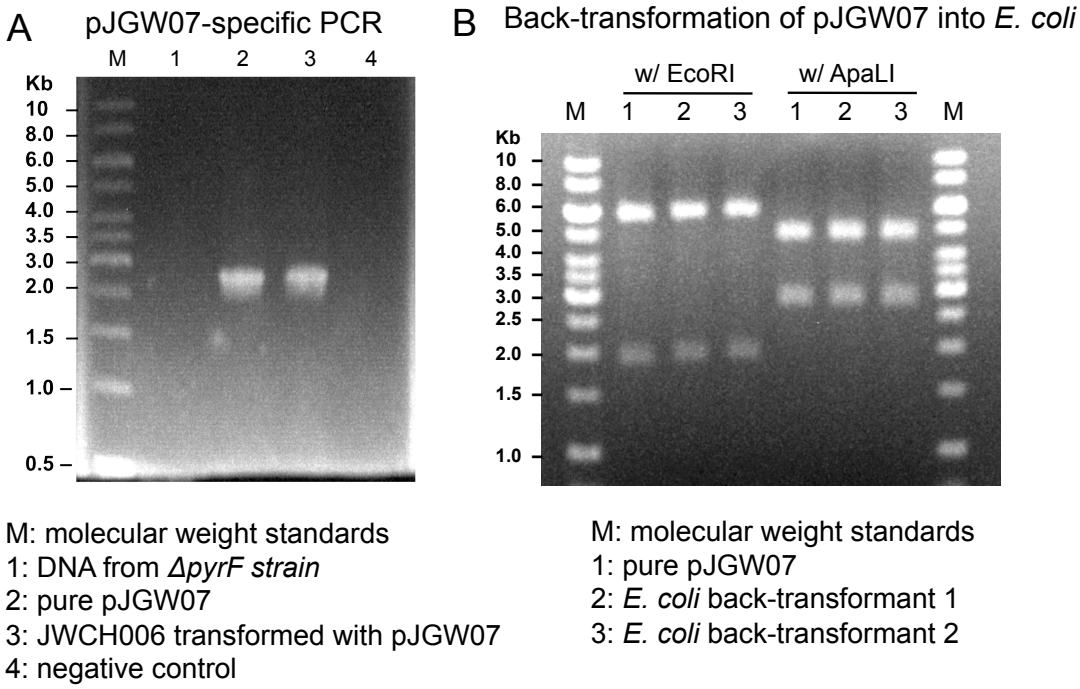


Figure 2.3. Plasmid pJGW07 isolated directly from *C. hydrothermalis* Δ*pyrF*. (A) pJGW07 constructed by replacing the *C. bescii pyrF* gene with the *C. thermocellum* ATCC24705 homolog Cthe0951. The hatched region was derived from *C. bescii* native plasmid pBAS2. Apr^R, apramycin resistance casette; *repA*, replication initiator for pSC101 replication origin; *par*, partitioning locus. **(B)** Agarose gel depicting pJGW07 plasmid DNA extracted from different sources. M, molecular weight standards; E, pJGW07 purified from *E. coli*; C, pJGW07 purified from *C. hydrothermalis.*



M: molecular weight standards
1: DNA from *ΔpyrF strain*
2: pure pJGW07
3: JWCH006 transformed with pJGW07
4: negative control

M: molecular weight standards
1: pure pJGW07
2: *E. coli* back-transformant 1
3: *E. coli* back-transformant 2

Figure 2.4. Evidence for transformation and stable replication of the *Caldicellulosiruptor/E. coli* shuttle vector pJGW07 in *C. hydrothermalis.*

A) pJGW07 is present in transformants**.** DNA was isolated from the indicated *C. hydrothermalis* strains in the legend. PCR was performed using two pJGW07-specific primers JG021 and JF199 (see Figure 2.3) with an expected amplicon of 2.6 kb.

B) pJGW07 is structurally stable in *C. hydrothermalis.* Back-transformation of pJGW07 isolated from *C. hydrothermalis* into *E. coli* DH5α. For *EcoRI* digests, expected bands: 5.8 kb and 1.9 kb. For *ApaLI* digests, expected bands: 5 kb and 2.7 kb (See Figure 2A). 12 *E. coli* colonies were analyzed, restriction digests of plasmid isolated from two colonies is shown.

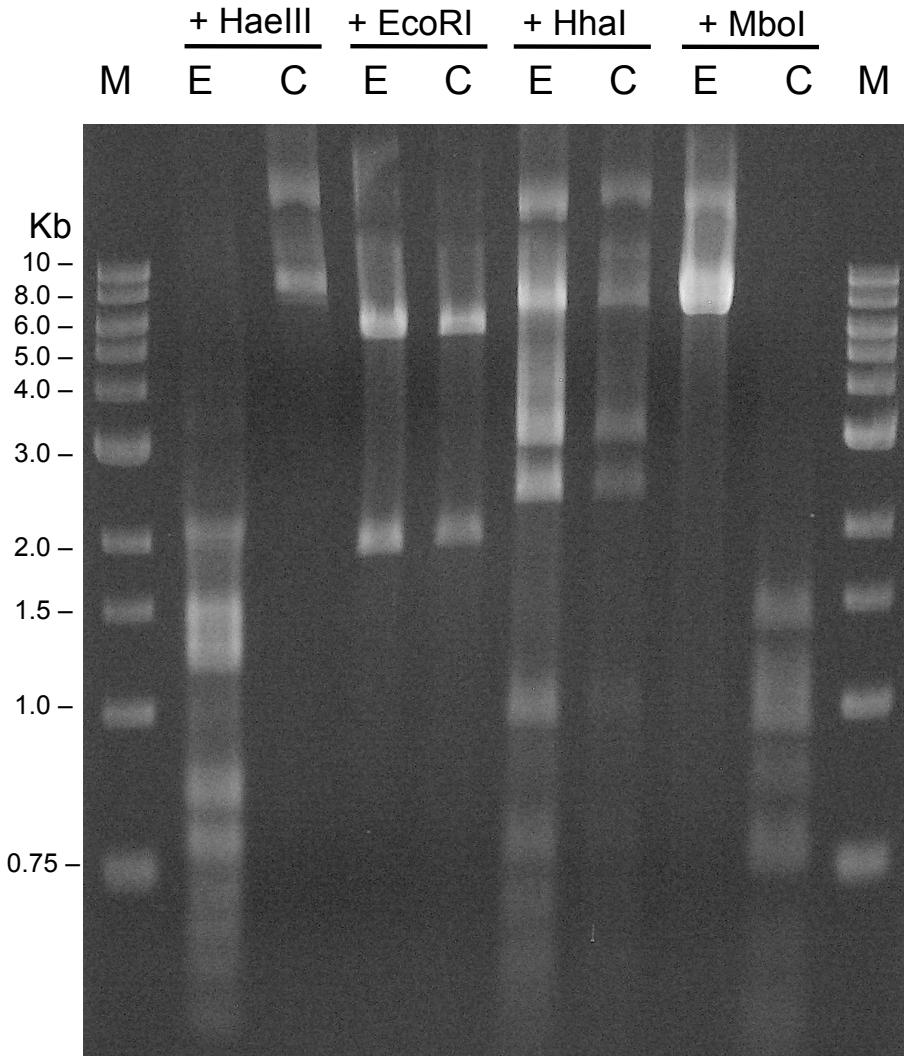


Figure 2.5. Restriction digest analysis of pJGW07. Plasmid pJGW07 purified from either *E. coli* (lanes labeled E) or *C. hydrothermalis* (lanes labeled C) was exposed to enzymes HaeIII, EcoRI, HhaI, or MboI in individual reactions, as indicated. Electrophoresis profiles of each digest reaction are shown. M, molecular weight standards (NEB).

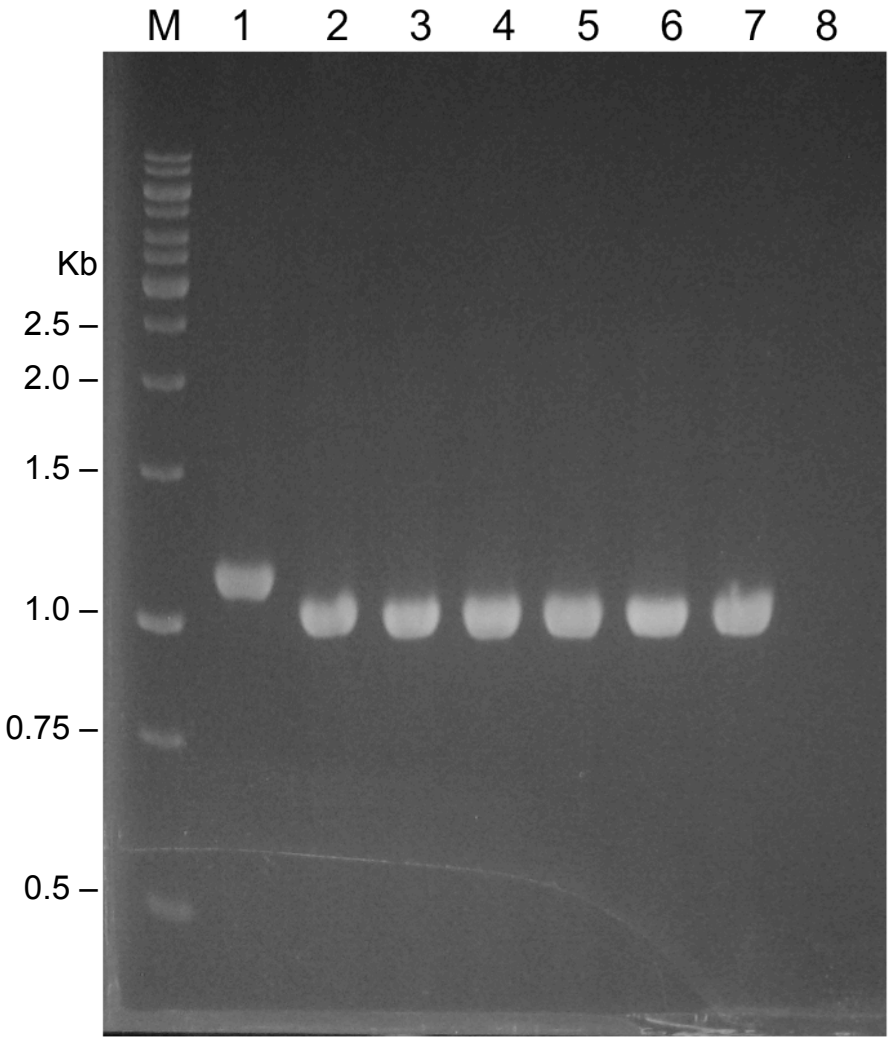


Figure 2.6. Maintenance of the mutated *pyrF* gene in pJGW07 transformants. *C. hydrothermalis* JWCH006 (*ΔpyrF*) was transformed with M.CbeI-methylated pJGW07, and individual transformant colonies were picked. DNA was isolated from the strain, and PCR using *pyrF*-flanking primers DC163 and DC164 (see Figure 2.1) was performed. M: molecular weight standards; 1: *C. hydrothermalis* wild-type genomic DNA; 2: *C. hydrothermalis* JWCH006 genomic DNA. 3 through 7: genomic DNA from five individual colonies of *C. hydrothermalis* JWCH006 transformed with pJGW07; 8: negative control.

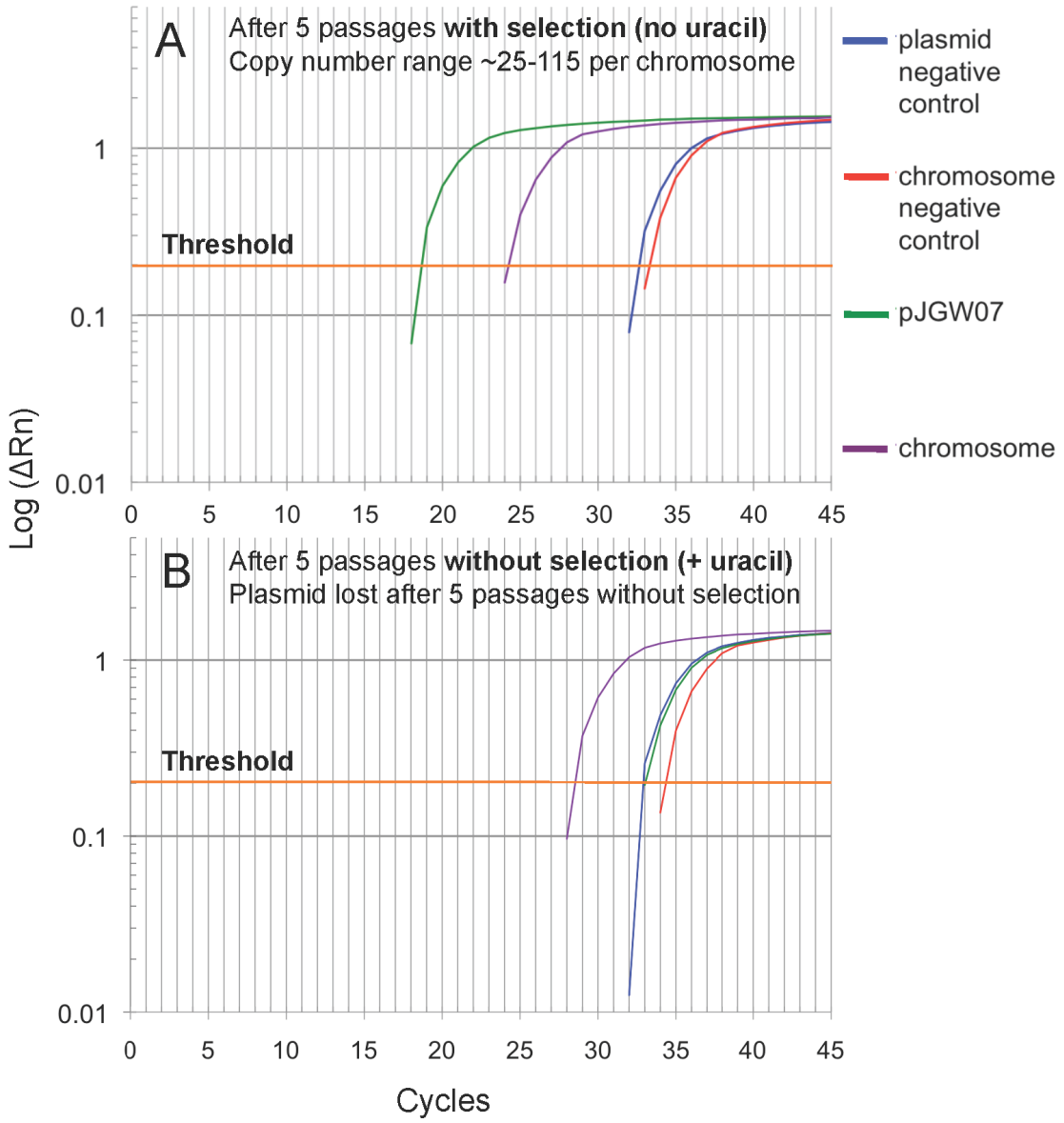


Figure 2.7. Determination of plasmid copy number. Quantitative PCR (qPCR) was used to detect the copy number of plasmid pJGW07 in relation to the chromosome. Shown are the results after the fifth passage through **(A)** selective and **(B)** non-selective media. The *x*-axis is the number of iterations of the polymerase chain reaction, and the *y*-axis displays the logarithm of ΔRn, which is the fluorescence of the SYBR green dye with the baseline fluorescence subtracted. The number of cycles required to cross a given threshold (cycles to threshold or Ct) is reflective of the plasmid copy number (PCN). The threshold is indicated by a horizontal line. PCN was calculated using the formula PCN = 2^| $Ct_{chromosome}$— $Ct_{plasmid}$ |. The copy number ranged from about 25 to 115 copies per chromosome in the cultures with selective media. The copy number was determined based on two independent biological samples with three technical replicates (See Table 2.3).

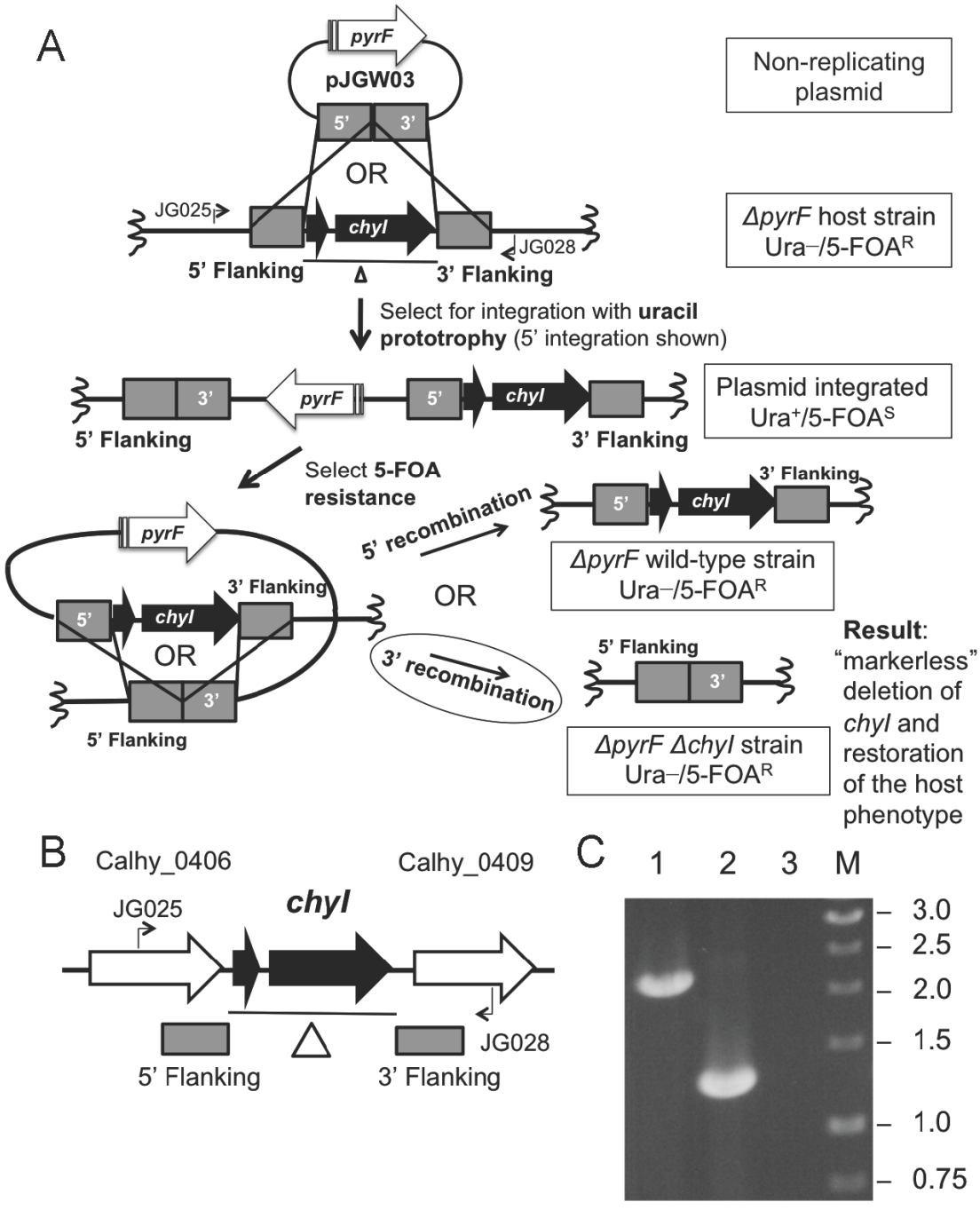

A
pyrF
pJGW03
5'
3'
OR
JG025
chyI
JG028
5' Flanking
Δ
3' Flanking
Non-replicating plasmid
ΔpyrF host strain Ura−/5-FOAR
Select for integration with uracil prototrophy (5' integration shown)
3'
pyrF
5'
chyI
5' Flanking
3' Flanking
Plasmid integrated Ura+/5-FOAS
Select 5-FOA resistance
5' recombination
3' Flanking
5'
chyI
ΔpyrF wild-type strain Ura−/5-FOAR
OR
pyrF
3' Flanking
5'
chyI
OR
3'
5' Flanking
3' recombination
5' Flanking
3'
ΔpyrF ΔchyI strain Ura−/5-FOAR
Result: "markerless" deletion of chyI and restoration of the host phenotype
B
Calhy_0406
Calhy_0409
JG025
chyI
JG028
5' Flanking
3' Flanking
C
1
2
3
M
3.0
2.5
2.0
1.5
1.0
0.75

Figure 2.8. Deletion of the gene encoding putative restriction enzyme ChyI. **(A)** Scheme for targeted gene deletion. The pJGW03 chyI knockout vector is transformed into JWCH006 (ΔpyrF), and uracil prototrophy selects for integration at one of the 500-bp flanking regions, denoted by the gray boxes. 5′ integration is shown. The uracil prototroph is then plated on 5-FOA to select for loop-out of the *pyrF* cassette via homologous recombination between flanking region sequences. The two possible results are a return to the wild-type sequence or a clean *chyI* deletion. **(B)** Chromosomal map of the locus containing the ORF for *chyI*. The deleted region is indicated by the line below the diagram. Bent arrows depict primers used to verify deletion of *chyI* in the JWCH008 strain. **(C)** Gel depicting PCR products of the *chyI* region in the JWCH006 Δ*pyrF* strain (2.3 kb) compared to the JWCH008 ΔpyrF ΔchyI strain (1.25 kb) amplified by the indicated primers (JG025 and JG028). 1: *C. hydrothermalis* JWCH006 genomic DNA; 2: *C. hydrothermalis* JWCH008 genomic DNA; 3: negative control; M: 1 kb DNA ladder (NEB).

Table 2.1. Strains and plasmids used in Chapter 2

Strain or plasmid	Genotype/phenotype
Caldicellulosiruptor	
JWCH001	*C. hydrothermalis* DSM 18901 wild-type (*ura*$^{+}$/5-FOA^{S})
JWCH006	*C. hydrothermalis ΔpyrF* (*ura*$^{-}$/ 5-FOA^{R})
JWCH008	*C. hydrothermalis ΔpyrF ΔchyI* (*ura*$^{-}$/5-FOA^{R})
JWCH009	*C. hydrothermalis ΔpyrF* harboring pJGW07 (*ura*$^{+}$/5-FOA^{S})
Escherichia coli	
JW401	DH5α containing pJGW03 (Apramycin^{R})
JW402	DH5α containing pJGW07 (Apramycin^{R})
Plasmids	
pDCW89	*E. coli/Caldicellulosiruptor* shuttle vector (*C. bescii pyrF*)
pDCW88	*Caldicellulosiruptor* non-replicating vector (*C. bescii pyrF*)
pDCW151	*Caldicellulosiruptor chyI* deletion vector (*C. bescii pyrF)*
pJGW03	*Caldicellulosiruptor chyI* deletion vector (*C. thermocellum pyrF)*
pJGW07	*E. coli/Caldicellulosiruptor* shuttle vector (*C. thermocellum pyrF*)

Table 2.2. Transformation efficiency of *C. hydrothermalis*.
Transformation efficiency results are reporter in colony forming units (CFU) / µg of transformed plasmid DNA. Results represent the average ± standard deviation of three biologically independent transformation experiments. The suffix M denotes plasmid methylated with *C. bescii* M.CbeI methyltransferase.

Strain	pJGW07	pJGW07M	pDCW89	pDCW89M
JWCH006 Δ*pyrF*	33 ± 24	22 ± 10	18 ± 20	37 ± 43
JWCH008 Δ*pyrF* Δ*chyI*	395 ± 301	241 ± 174	58 ± 48	100 ± 90

Table 2.3. Quantitative PCR data to determine plasmid copy number in *C. hydrothermalis*.

These data represent the amplification cycles required to cross a threshold based on SYBR green fluorescence. qPCR was performed on total DNA isolated from *C. hydrothermalis* over the course of five passages through selective media. Passages through non-selective media resulted in Ct values above 30, comparable with the negative control results, so these values were not interpreted as accurate [196]. PCN was calculated using the formula PCN = 2 ^ | $Ct_{chromosome}$— $Ct_{plasmid}$ |. Ct_P and Ct_C are the average of three replicate qPCR reactions. An amplication efficiency curve over four logs of DNA concentration revealed an efficiency of 93.5%, which is within the acceptable range of 90-110%.

Value	Day 1	Day 2	Day 3	Day 4	Day 5
Plasmid (Ct_P)	21.02	21.54	20.04	19.38	18.22
Chromosome (Ct_C)	26.74	26.38	26.08	26.24	23.99
$Ct_c - Ct_P$	5.72	4.84	6.04	6.87	5.78
Copy number (PCN)	52.77	28.54	65.72	116.84	54.76

Table 2.4. Primers used in Chapter 2.

Primer	Sequence (5′ → 3′)
DC081	AGAGAGGTACCACCAGCCTAACTTCGATCATGGGA
DC163	TCCTGAACCAATAACCAAAACCT
DC164	AGTGGGAAGTGAAAGAGGAAAAC
DC262	TGTGTGGTGCACTCTGACGCTCAGTGGAACGAA
DC484	AGACTCCGATCGATTCCCATGAGCCCACGAACAGT
DC485	ATGTGCGATTCCTTTTGCGGTTTGGTCCAT
DC486	ACCAAACCGCAAAAGGAATCGCACATCGAAAGTTGGGAGT
DC487	ACAACAGTGCACACTCCATGTAAAGCGATTTTCA
JF199	TCGCTAACGGATTCACCACT′
JG021	AGAATATCTAGAATGTTTATTGATACATTAATTGAAAAGATT AGAGAAAAGG
JG022	TGTAGTCCATGGTTACTTCCTGTCTCGCAACGC
JG023	TCTACTCCATGGTCATCTGTGATATGGACAGTTTTCC
JG024	AGATCATCTAGAGACCATCCTTTCTATGTAGAAA
JG025	CTGCCAAGTTAGAAAACAAGGAC
JG026	AGAACAAGGAATACCAAGCCA′
JG027	ACCTTGCTGTGATAGAAAACCT
JG028	CATACAATCGGGATTCAGCAGT
Q1	TGGGAAAGCCGTCCATAATC
Q2	TCTCCCGCTCTTCTCTCTTT
Q3	GTGCGTCTACAGGACCTTATTT
Q4	GGCAAGATTCTACAGGCAAGA
QH11	CACATCAGCAACAGCAAGTAAG
QH12	CCTCACAAGCAACTACTCTACC
QH13	GCTCGGTCGCTCTGAATATAAC
QH14	GAGTTGGAAAGCTCAGGTCATC

CHAPTER 3

PROMISCUOUS PLASMID REPLICATION IN THERMOPHILES: USE OF A NOVEL HYPERTHERMOPHILIC REPLICON FOR GENETIC MANIPULATION OF *CLOSTRIDIUM THERMOCELLUM* AT ITS OPTIMUM GROWTH TEMPERATURE.[2]

[2] Groom J, Chung D, Olson DG, Lynd LR, Guss AM, and Westpheling J. 2016. *Metabolic Engineering Communications.* 3: 30-38.

I. Abstract

Clostridium thermocellum is a leading candidate for the consolidated bioprocessing of lignocellulosic biomass for the production of fuels and chemicals. A limitation to the engineering of this strain is the availability of stable replicating plasmid vectors for homologous and heterologous expression of genes that provide improved and/or novel pathways for fuel production. Current vectors rely on replicons from mesophilic bacteria and are not stable at the optimum growth temperature of *C. thermocellum*. To develop more thermostable genetic tools for *C. thermocellum*, we constructed vectors based on the hyperthermophilic *Caldicellulosiruptor bescii* replicon pBAS2. Autonomously replicating shuttle vectors based on pBAS2 reproducibly transformed *C. thermocellum* at 60°C and were maintained in multiple copy. Promoters, selectable markers and plasmid replication proteins from *C. bescii* were functional in *C. thermocellum*. Phylogenetic analyses of the proteins contained on pBAS2 revealed that the replication initiation protein RepL is unique among thermophiles. These results suggest that pBAS2 may be a broadly useful replicon for other thermophilic Firmicutes.

II. Introduction

Clostridium thermocellum has been the subject of intense study because of its unique mechanism for solubilizing plant biomass. *C. thermocellum* secretes proteins that form an extracellular organelle, called a cellulosome [113, 114], for attachment to and digestion of complex plant biomass. Its ability to produce ethanol from cellulose has made it a leading candidate for consolidated

bioprocessing [98, 201]. Recent work to engineer *C. thermocellum* for increased ethanol production has eliminated competing fermentation pathways [112, 122, 202-204] and allowed synthesis of other fuel molecules like isobutanol [129].

Published genetic methods in *C. thermocellum* allow transformation of plasmid DNA [165, 205, 206], the generation of gene deletions [112] and some methods for gene expression [129, 149, 176, 202]. While gene deletions and gene expression have led to significant increases in ethanol production, the expression systems for robust high-level expression of homologous and heterologous proteins in *C. thermocellum* have limitations. The plasmid replicon from pNW33N that has been used extensively in published *C. thermocellum* transformation protocols [122] is not stable at elevated temperatures [207]. Therefore, current methods for DNA transformation require that selection of transformants be performed at 50-51°C [205], a suboptimal growth temperature for *C. thermocellum*, which grows optimally at 60°C [109]. Because this origin does not support stable replication at 60°C, it may contribute to both the observed plasmid instability [176] and chromosomal integration [129], even at 50-55°C. A plasmid that could faithfully replicate at 60°C would enable optimal *in vivo* expression and rapid screening of cell wall decomposition enzymes or metabolic enzymes for strain engineering.

Selectable markers that have been used in *C. thermocellum* [205] include *pyrF*, that confers uracil prototrophy in a *pyrF* deletion strain; chloramphenicol acetyl transferase (*cat)*, that confers thiamphenicol resistance; and neomycin acetyl transferase (*neo)*, that confers neomycin resistance. Counter-selectable

markers include *pyrF*, that confers sensitivity to 5-fluoroorotic acid (5-FOA); *tdk*, that confers sensitivity to 5-fluoro-2'-deoxyuridine (FUdR); and *hpt*, that confers sensitivity to toxic purine analogs such as 8-azahypoxanthine (8AZH) [142, 143]. Counter-selection with *tdk* may be performed in a wild type strain, while use of *pyrF* and *hpt* require the use of a strain deleted for the *C. thermocellum* chromosomal copy of the gene.

Many plasmids have been shown to be capable of intergeneric and even interphyletic replication, raising the possibility of finding additional origins of replication for use in *C. thermocellum*. Early work in mesophilic Gram-positive bacteria identified several such replicons. *Staphylococcus aureus* plasmid pC194, that replicates by a rolling circle mechanism [208], as well as other *Staphylococcus* plasmids successfully transformed *Bacillus subtilis* [209, 210]. *Enterococcus faecalis* plasmid pAMβ1 was used to transform *Lactobacillus acidophilius* [211] and *Bacillus subtilis* [212], and has since become a broad-range host vector for members of the Firmicutes. An example of interphyletic plasmid replication is pNG2, that was isolated from *Corynebacterium diphtheriae*, a member of the Actinobacteria phylum, but replicates in *Escherichia coli*, a member of the Proteobacteria phylum [213].

Unfortunately, many of the commonly used plasmid origins were isolated from mesophilic organisms, so functionality at thermophilic temperatures is a concern. For instance, most Staphylococcal plasmids can be cured by growth at 43°C [214]. The *Bacillus subtilis* plasmid pIM13 was used to transform *Staphylococcus aureus* [215], and more recently the replicon was used for the

construction of pIKM1 for use in the thermophile *Thermoanaerobacterium saccharolyticum* at 48°C*,* although its optimal growth temperature is 60°C [216]. The pIP404 replicon [217] from *Clostridium perfringens,* that grows optimally at 43°C, was used to transform *Thermoanaerobacter ethanolicus* JW200*,* albeit at a temperature much lower than the recipient's optimum growth temperature of 69°C [172]. Although the promiscuity of many plasmids has been demonstrated, there are limitations to mesophilic bacterial plasmids being used at the optimal growth temperatures of thermophiles. Issues of structural instability [176] chromosomal integration [129, 176], and decreased copy number [218] have been reported.

An alternative approach is to identify origins of replication from thermophilic organisms as a starting point for developing more thermostable shuttle vectors. One example is plasmid pBAS2 that is native to *Caldicellulosiruptor bescii* [192], a thermophilic Firmicute. *C. bescii* grows optimally at 75°C [90] and plasmids using the pBAS2 origin of replication have been shown to replicate at high copy at 65°C [219]. The replication functions of pBAS2 have been used to construct vectors for transformation of *C. bescii* [177, 179] and *Caldicellulosiruptor hydrothermalis* [219]. Bioinformatic analysis revealed significant sequence identity to other Gram-positive rolling circle replication (RCR) origins, but the mechanism of pBAS2 replication is not clear. The conserved nick site is not obvious, and single stranded DNA that typically accumulates as a consequence of rolling circle replication was not observed for pBAS2 [192]. The pBAS2 replicon encodes a replication protein with homology

to known RepL proteins and a Xer-like recombinase [177, 192], which is potentially responsible for the resolution of multimers of both chromosomes and plasmids that form during replication [220, 221]. The open reading frame encoding Cbes2777 (the Xer-like recombinase XerD) was shown to be necessary for autonomous replication in *C. bescii* [177]. While some replicons rely on host recombinases that work *in trans* to resolve multimers, pBAS2 encodes its own recombinase to resolve plasmid multimers and maintain sequence fidelity.

Based on the thermophilic source of pBAS2 and the presence of both a replication gene and a recombinase gene, we hypothesized that pBAS2-derived plasmids would be capable of independent replication in *C. thermocellum* at its optimal growth temperature of 60°C. We therefore built and tested plasmids carrying the pBAS2 origin of replication with different selectable markers to explore this possibility. Here we report the transformation and stable replication in *C. thermocellum* of plasmid vectors based on pBAS2. We show that the *pyrF* gene from *C. bescii* functions to complement a deletion of the *pyrF* gene in *C. thermocellum* and that promoters from *C. bescii* function to drive expression of genes at levels sufficient for selection of transformants. Transformation of *C. thermocellum* DSM 1313 was performed at 60°C, the optimal growth temperature of this strain, and plasmids derived from pBAS2 replicate stably in *C. thermocellum* at this temperature. Phylogenetic analysis of pBAS2 protein sequences suggests that the replication protein is novel and unique among known thermophilic proteins.

III. Materials and Methods

Bacterial media and growth conditions. *Clostridium thermocellum* DSM 1313 and its derivatives were grown anaerobically in modified CTFUD medium [205] at 60°C, under an atmosphere of 85% nitrogen, 10% CO_2, and 5% Hydrogen. Defined medium for transformation and selection was CTFUD-NY [205], which contains a vitamin mix of para-aminobenzoic acid, vitamin B12, biotin, and pyridoxamine in place of the yeast extract. CTFUD-NY was supplemented with 360 μM uracil when noted. Complex medium for recovery after transformation was similar to CTFUD but contained casein (0.2% w/v) and less yeast extract (0.1% w/v), which is referred to as CTFUD+C. *Escherichia coli* was grown in Luria-Bertani broth supplemented with 50 μg/mL apramycin when selecting for the presence of a plasmid.

Plasmid vector construction. All PCR reactions for cloning were performed with Q5 polymerase (New England Biolabs, Ipswich, MA) according to the manufacturer's instructions (98°C duplex denaturation, 60°C annealing temperature, 30 seconds per kb at 72°C for elongation). Plasmid pDCW196 was constructed by ligating *C. bescii* shuttle vector pDCW89 [177] to the *cat* gene from *C. thermocellum* vector pMU612 [122]. Primers X013 and X014 amplified a 7.69 kb fragment from pDCW89, and primers X015 and X016 amplified a 1.053 kb fragment from pMU612. These fragments were digested with *Bam*HI and *Nde*I (New England Biolabs) twice in succession for 45 minutes at 37°C and ligated with the FastLite ligation kit (Epicentre, Madison, WI) according to the manufacturer's instructions. Plasmid pJGW37 was constructed by amplifying a

7.467 kb fragment from pDCW196 using primers JG024 and JG099, digesting with *Xba*I (New England Biolabs) and ligating as described above to circularize the linear fragment. All plasmid sequences were verified by Sanger DNA sequencing (Genewiz). All primers are listed in Table 3.5.

Transformation of *C. thermocellum ΔpyrF.* A 20 mL starter culture of *C. thermocellum* DSM 1313 Δ*pyrF* was grown at 60°C to mid-exponential phase in defined CTFÜD medium [205] supplemented with uracil. A 15 mL aliquot of this culture was transferred to 150 mL of the same medium and grown to OD_{600} = 0.6. Cells were cooled to room temperature for 25 minutes, harvested aerobically at 7500 x g for 10 minutes, and washed twice with pre-chilled 10% sucrose. Competent cells were divided into 30 µL aliquots, and those not used immediately were frozen in a dry ice-ethanol slurry and stored at -80°C. Plasmid DNA pDCW89 (500 ng) was incubated with each aliquot for 15 minutes in pre-chilled 1 mm cuvettes. Cells were electrotransformed with a single exponential pulse with a Gene Pulser (BioRad, Hercules, CA) set at 1.8 kV, 25 µF, and 350 Ω, and placed immediately into 60°C CTFUD+C medium [205] supplemented with 360 µM uracil for recovery. A 0.25% inoculum was transferred to defined liquid CTFUD-NY medium lacking uracil every three hours over the period of recovery to select for transformants**.**

Transformation of *Clostridium thermocellum Δhpt.* A 10 mL starter culture of *C. thermocellum Δhpt* was grown at 60°C to mid-exponential phase in rich CTFUD medium [205]. This culture was transferred to 150 mL of the same medium and grown to OD_{600} = 0.9. Cells were cooled to room temperature for 25

minutes, harvested aerobically at 7500 x g for 8 minutes, and washed twice with pre-chilled 10% glycerol, 260 mM sucrose. Competent cells were divided into 30 µL aliquots, and those not used immediately were frozen in a dry ice-ethanol slurry and stored at -80°C. Plasmid DNA pJGW37 (350 ng) was incubated with each aliquot for 15 minutes in pre-chilled 1 mm cuvettes. Cells were electrotransformed with a single exponential pulse with a BioRad Gene Pulser (1.8 kV, 25 µF, 350 Ω), and placed immediately into warm CTFUD+C medium for 6 hours at 60°C, after which they were serially diluted onto CTFUD plates containing yeast extract (0.1% w/v) and 10 µg/mL thiamphenicol (Sigma, St. Louis, MO). Colonies appeared after 4-5 days of growth at 60°C under an atmosphere of 85% nitrogen, 10% CO_2, and 5% hydrogen. Transformation efficiency was calculated in CFU/µg of plasmid DNA, and experiments were performed in biological triplicate.

Total DNA isolation from *Clostridium thermocellum.* For plasmid copy number determination over the growth phase, a 0.1% (v/v) inoculum was grown without shaking at 60°C. Samples (2% (v/v)) were removed for qPCR analysis at the indicated time points (Figure 3.7). For plasmid maintenance experiments (Table 3.4), 0.25% inocula were grown in serial passages to exponential phase (OD_{600} ~ 0.4 for *C. thermocellum* Δ*pyrF*, ~0.7 for *C. thermocellum* Δ*hpt*) and 5 mL of culture was harvested at 4°C at 6100 x g. The pellet was resuspended in 200 µL 40 mM EDTA, 50 mM Tris-HCl, 25% sucrose (w/v) with 30 mg/mL lysozyme (Sigma, St. Louis, MO) and 1 µL/mL RNase A (Qiagen), and incubated for 30 minutes at 37°C. Cells were frozen and thawed five times using a dry ice-ethanol

slurry and a 37°C water bath. A 500 µL aliquot of 6M Guanidine-HCl pH 8.5 (Sigma, St. Louis, MO) was immediately added to the cell lysate and allowed to incubate at 75°C for 10 minutes. The lysate was washed twice with Phenol:chloroform:isoamyl alcohol and once with chloroform. To the aqueous layer, 3M sodium acetate pH 5.2 was added, and DNA was precipitated at -80°C for 3 hours, washed with cold 70% ethanol and resuspended in 10 mM Tris-HCl. Total DNA for *C. thermocellum* transformants containing pJGW37 was visualized on a 1% agarose gel (Sigma, St. Louis, MO) stained with ethidium bromide (Figure 3.5C).

Verification of plasmid transformation and structural stability. Taq polymerase (Sigma, St. Louis, MO) was used for PCR reactions to confirm the presence of the plasmid using total DNA purified from *C. thermocellum* transformants as template. Reactions were performed with primers DC091 and DC508 according to the manufacturer's instructions (94°C duplex denaturation, 56°C annealing temperature, 1 minute per kb at 72°C for elongation). PCR products were visualized on a 1% agarose gel with an NEB 1 kb Ladder for size verification (NEB, Ipswich, MA). To verify structural stability in *C. thermocellum,* total DNA was electrotransformed into *E. coli* BL21 with a BioRad Gene Pulser (Biorad, Hercules, CA) using an exponential pulse in a 2 mm cuvette (2.5 kV, 25 µF, 200 Ω). Selection in *E. coli* was performed with 50 µg/mL apramycin, and colonies were screened for the presence of the plasmid by performing restriction digests with EcoRI and ApaLI (NEB, Ipswich, MA).

Quantitiative polymerase chain reaction (qPCR). qPCR experiments were carried out with a LightCycler 480 Real-Time PCR instrument (Roche, Basel, Switzerland) with LightCycler 480 SYBR Green I master mix (Roche). Cycle thresholds resulting from amplification with two independent sets of primers specific to either the pJGW07 plasmid (Q1/Q2 inside Cbes2777, Q3/Q4 inside Cbes2778) or the *C. thermocellum* chromosome (CTQ1/CTQ2 inside Clo1313_0092, CTQ3/CTQ4 inside Clo1313_0090) were used to compute relative copy number of the plasmid to the chromosome. The formula for this calculation was PCN = 2^| $Ct_{chromosome}$— $Ct_{plasmid}$ |. Four replicate reactions for each primer set were performed, and the average of the two primer sets on each sample was used to calculate the plasmid copy number in each serial subculture (Table 3.5) according to the method of Lee *et al* [196]. Amplification efficiency over 10^4-fold range of DNA concentration was 102%, within the ratio of 90-110% that is considered acceptable.

Bioinformatic analysis. The National Center for Biotechnology Information (NCBI) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) were used to search for homologous proteins to the ORFs on plasmid pBAS2. Amino acid similarity was also determined with the NCBI Basic Local Alignment Search Tool (BLAST). Clustal Omega was used for multiple sequence alignment with default settings [222, 223]. ProtTest 3.4 [224] was used to estimate models of evolution for the multiple sequence alignments. To determine posterior probabilities, MRBAYES version 3.2.5 x64 [225] was run for 1,000,000 generations with the WAG+I+G+F model of evolution [226] for Cbes2777 and the

LG+I+G+F model [227] for Cbes2780. Figtree version 1.4.2 (http://tree.bio.ed.ac.uk/ software/figtree/) was used to visualize the phylogenetic trees generated by MRBAYES. MEME was used for motif analysis [228], searching for 6 motifs for Cbes2777 and 3 motifs for Cbes2780. MAST was used for motif searches with default settings [228]. CodonO software [229] was used with default settings on genome files uploaded from NCBI.

IV. Results and Discussion

The XerD protein encoded by the *C. bescii* native plasmid pBAS2 has homology to known thermophilic proteins but the replication initiation protein, RepL, does not. As noted previously [192], the protein sequences encoded by the pBAS2 open reading frames Cbes2777 and Cbes2780 (Figure 3.1) are homologous to Xer-like recombinases and the RepL family of replication initiation proteins, respectively. Open reading frames Cbes2778 and Cbes2779 are short ORFs that would encode proteins of 109 aa and 73 aa respectively and are annotated as hypothetical proteins of unknown function (Table 3.1, Table 3.2).

Xer recombinases are known to resolve multimers of both chromosomes and plasmids that form during replication [220, 221]. To better understand the replication machinery of the pBAS2 origin, we performed phylogenetic analysis of Cbes2777 and Cbes2780 with more recently acquired sequence data [230-232]. That analysis revealed that Cbes2777 is homologous to XerC and XerD proteins in other Firmicutes, and it weakly clusters with a group of proteins from the thermophilic, anaerobic genera *Thermoanaerobacter* and

Thermoanaerobacterium (Figure 3.2A). This is not surprising, as *Caldicellulosiruptor* is known to be closely related to these genera [233]. Motif analysis revealed that all of these Xer-like recombinases contain a phage integrase-like domain and a DNA breaking-rejoining domain (Figure 3.2B). The members of the cluster containing Cbes2777 have strong motif signatures for each of these protein domains [228, 234], in particular the RHRY conserved catalytic active site [220] (Figure 3.2B).

Xer recombination sites for multimer resolution like *cer* on ColE1 and *psi* on pSC101 require extra sequences for the binding of accessory proteins [235, 236]. One such protein for *cer* is the arginine repressor ArgR [237], which binds to specific sites on the plasmid to orient the recombination sites during multimer resolution. We were unable to identify canonical XerC- and XerD- binding sites [238] in the pBAS2 sequence, but a 1.4 kb region of non-coding sequence on plasmid pBAS2 contains predicted ArgR binding sites (Figure 3.1) existing as an approximate direct repeat for regulatory protein binding. This site has striking sequence similarity to known ArgR binding sites, particularly those from the *roc* and *car* operons from *Bacillus subtilis* [239]. Thus, one possibility is that this sequence behaves like *cer* and *psi* for the resolution of plasmid multimers using the XerD recombinase encoded *in cis* by plasmid pBAS2. This would require that the host chromosome encodes an *argR* gene which is, in fact, present in both *C. bescii* and *C. hydrothermalis*, where pBAS2 has previously been shown to replicate. Further, it is present in *C. thermocellum* and other thermophiles including *Thermoanaerobacterium saccharolyticum and Thermoanaerobacter*

ethanolicus [234], suggesting that pBAS2 may have a broader host range than the *Caldicellulosiruptor* genus.

In contrast to Cbes2777, homology searches for Cbes2780 revealed a limited number of homologs with relatively low sequence identity (≤30%). There were no plasmid-encoded thermophilic homologs, and very few Firmicute homologs (Figure 3.3A). Cbes2780 clusters with RepL proteins from diverse bacterial phyla including Proteobacteria, Cyanobacteria, and Firmicutes (Figure 3.3A and Figure 3.4). Among all the replication protein homologs there is a single strong consensus sequence ΦNPX_5G in a helix-turn-helix DNA binding domain (Figure 3.3B), although the *C. bescii* protein has many more arginine residues at this site. It is possible that these residues play a role in thermostability, as arginine residues have been associated with protein thermostability, particularly by increasing hydrophilicity when they replace lysine residues [240, 241], and when they exist in clusters [242]. These findings, in light of the lack of thermophilic homologs and the fact that Cbes2780 resides on a plasmid, suggest that this protein might have been co-opted by *C. bescii* from a distantly-related mesophilic organism, or alternatively from a currently undiscovered thermophilic organism. Importantly, unique Rep proteins provide for unique incompatibility groups [243] making these plasmids potentially compatible with other known replicons.

It is noteworthy that *Cyanothece* sp. PCC 8802 MarR2/RepL and *Sodalis glossinidius* RepL are encoded on plasmids that encode Xer-like recombinases, reminiscent of the arrangement of RepL and XerD on pBAS2. In fact, the use of

portions of pBAS2 to construct other vectors revealed that the XerD recombinase is required for plasmid replication in *C. bescii* [177]. The organization of pBAS2 may contribute to plasmid promiscuity with both a replication initiation protein and a multimer-resolving recombinase on the same plasmid.

A thermostable replicon from the *C. bescii* plasmid pBAS2 transforms *C. thermocellum* at 60°C. Two pBAS2-derived plasmids were used for testing transformation of *C. thermocellum*. Plasmid pDCW89 (Figure 3.5A) was previously constructed from the native *C. bescii* plasmid pBAS2 (Figure 3.1) [192] for use as an *E. coli/Caldicellulosiruptor* shuttle vector [177]. It contains the pBAS2 origin of replication, the *E. coli* plasmid pSC101 origin of replication, an apramycin resistance cassette for selection in *E. coli,* and the *C. bescii pyrF* wild type allele used to select uracil prototrophy in strains containing a deletion of the *pyrF* gene. Previously, a *C. thermocellum pyrF* deletion was constructed [122], allowing transformation of pDCW89 to be tested (see below) using uracil prototrophy as the selection. Because thiamphenicol selection is also commonly used for *C. thermocellum* transformations, we constructed plasmid pJGW37 (Figure 3.5B). This plasmid was based on pDCW89 (Figure 3.5A) but with the *pyrF* gene replaced with a chloramphenicol acetyltransferase (*cat*) gene for selection of transformants in *C. thermocellum* strains that are wild type at the *pyrF* locus.

Using a method developed for *C. bescii* [186], pDCW89 was electrotransformed into a *C. thermocellum* strain containing a deletion of the *pyrF* gene [122]. As with the *C. bescii pyrF* deletion, this strain is a uracil auxotroph

resistant to 5-fluoroorotic acid (5-FOA) (Table 3.3), allowing selection and counter-selection of the *pyrF* wild type allele. Transformants of pDCW89 were successfully selected for uracil prototrophy in defined liquid medium at 60°C. Further, plasmid pJGW37 was successfully transformed into *C. thermocellum* containing a deletion of the hypoxanthine phosphoribosyl transferase *(hpt)* [112] selecting thiamphenicol resistance, again at 60°C. Hpt forms nucleotide monophosphates from purines, and can lead to the incorporation of toxic purine analogs such as 8-azahypoxanthine (AZH) into DNA and RNA [143]. The *Δhpt* strain is resistant to AZH allowing selection of transformants that are thiamphenicol resistant and subsequent counter selection for AZH resistance.

Total DNA isolated from *C. thermocellum* transformants containing pDCW89 or *C. thermocellum* transformants containing pJGW37 was used to back-transform *E. coli*. Two different restriction endonuclease digests performed on plasmid DNA purified from nine independent *E. coli* colonies (three shown) resulted in identical banding patterns relative to the original plasmid (Figure 3.6). This proves that the plasmids were successfully transformed into *C. thermocellum* and suggests that there was no major structural instability of the plasmid during transformation and replication in *C. thermocellum* and back-transformation into *E. coli.* Plasmid pJGW37 was also purified directly from *C. thermocellum* and could be visualized on an agarose gel of total DNA (Figure 3.5C).

Plasmid stability, copy number and transformation efficiency. To assess transformation efficiency of the pBAS2 replicon, cells were transformed

with pJGW37 and plated after a 6 hour recovery period in rich recovery medium onto plates with thiamphenicol at 60°C. Transformation efficiency of pJGW37 was determined to be 3242 ± 575 colony-forming units per µg plasmid DNA (CFU / µg), demonstrating that the pBAS2 origin of replication can efficiently transform *C. thermocellum* at 60°C. This is in contrast to plasmids containing the pNW33N replicon, where the temperature limit for transformation is ~ 51° C (Olson & Lynd, 2012a).

To assess plasmid stability and copy number, individual *C. thermocellum* transformants were passaged in both selective and non-selective media five times, and the copy number was measured using quantitative polymerase chain reaction (qPCR), as described by Lee *et al* [196]. The plasmid copy number (PCN) varied depending on the selection method and the growth phase but was highest in late exponential phase for pJGW37 at 10-20 copies per chromosome (Figure 3.7A) and the PCN for pDCW89 varied from 2-10 during both exponential and stationary phases (Figure 3.7B). The reason for the differences in copy number with growth phase and growth rate may reflect the fact that copy number is an average across the population. After one passage without selection, the copy number is below 1, suggesting that either 10% of the population lost the plasmid, or more likely, most contain the plasmid in multiple copies, and some have lost it. Without selection, both plasmids decreased in average copy number with successive serial passages and were lost after five passages (Table 3.4).

Heterologous proteins and promoters from *C. bescii* function in *C. thermocellum*. Plasmid pDCW89 contains the *C. bescii pyrF* wild type allele and

this gene functions to complement the *pyrF* deletion in *C. thermocellum*. The PyrF proteins have 45% amino acid sequence identity. The G+C content of the *C. bescii pyrF* gene is 35.7%, and that of the *C. thermocellum pyrF* gene is 39.9%. In fact, CodonO software [229] indicates that codon usage bias between the entire *C. bescii* and *C. thermocellum* genomes is not statistically significant with a p value of >0.33. Both pDCW89 and pJGW37 rely on a *C. bescii* ribosomal protein promoter [177] to direct transcription of the selectable marker gene (Figure 3.5), the *pyrF* gene in pDCW89 and the *cat* gene in pJGW37. While we have no direct evidence for the efficiency of these promoters at this time, they clearly function at a level sufficient to allow selection. No promoter from *C. bescii* has been characterized experimentally so the transcription start site and RNA polymerase binding site for this promoter is unknown, but the sequence does contain a prototypical RNA polymerase binding site sequence for growth phase dependent transcription in Gram-positive bacteria.

V. Conclusions

While many plasmid origins of replication are derived from mesophilic sources, genetic tools for thermophiles must function at elevated temperatures. The use of thermophilic sources offers a solution to this issue, and we have demonstrated that plasmid origin pBAS2 from thermophile *C. bescii* can replicate in *C. thermocellum* at its optimal growth temperature of 60°C. This expansion of the genetic tools available for *C. thermocellum* will facilitate more rapid genetic engineering toward the goal of developing an organism for the efficient conversion of lignocellulosic biomass to fuels and chemicals. Further, the

demonstration of stable, autonomous replication of the *C. bescii* pBAS2 replicon in *C. thermocellum* suggests that this replicon might serve as a new tool for plasmid-based expression in other thermophiles, as well.

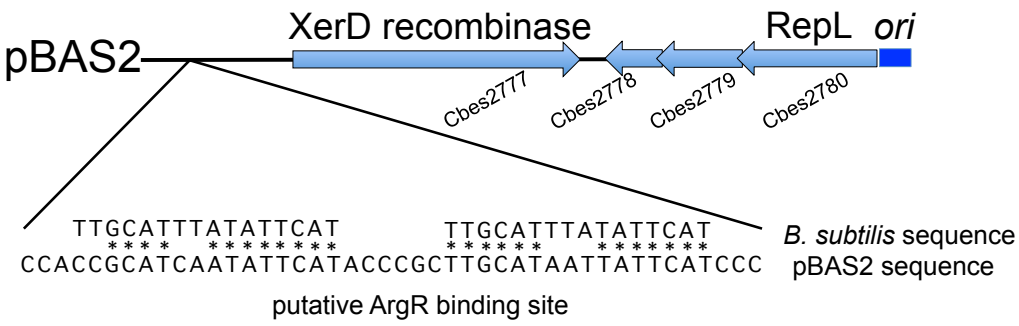


Figure 3.1. The annotated sequence of *Caldicellulosiruptor bescii* native plasmid pBAS2. Open reading frame numbers are shown below their respective genes, with predicated annotations above the genes. ***ori,*** conserved putative plasmid replication origin. **XerD recombinase**, site-specific tyrosine recombinase XerD. **RepL**, replication initiation protein. Magnified is the sequence identified by a MAST [228] motif search for the ArgR binding site. The ArgR binding site from *B. subtilis* [239] was used as a query motif. The putative site exists as an approximate direct repeat.

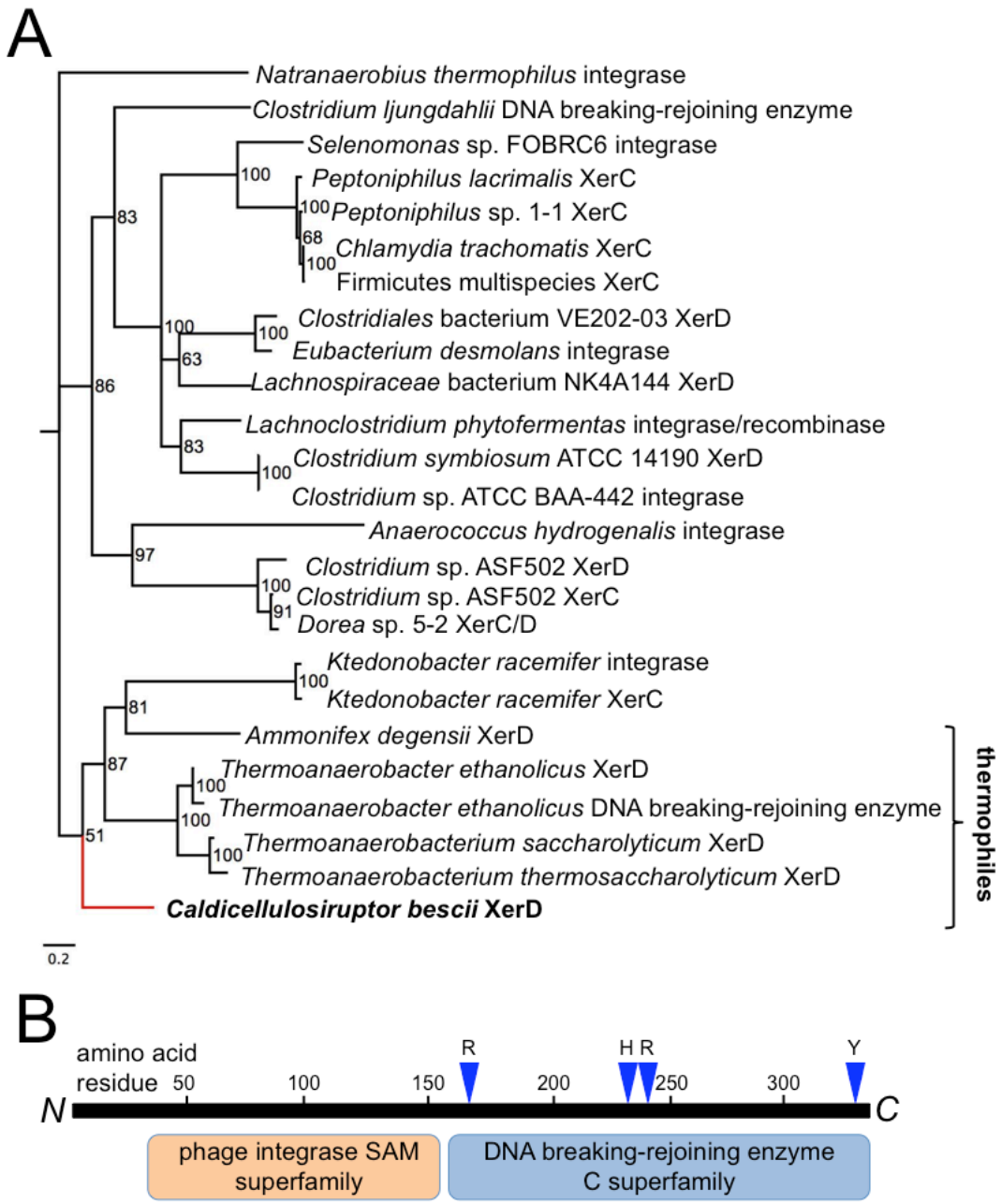
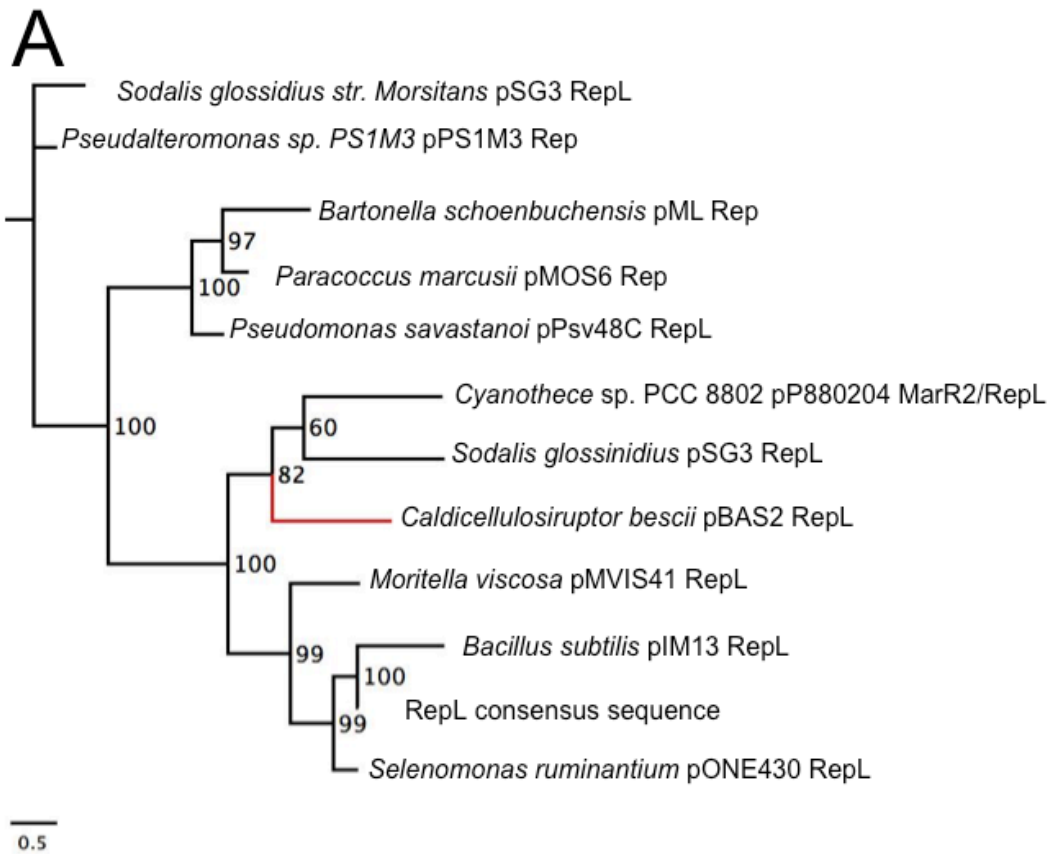
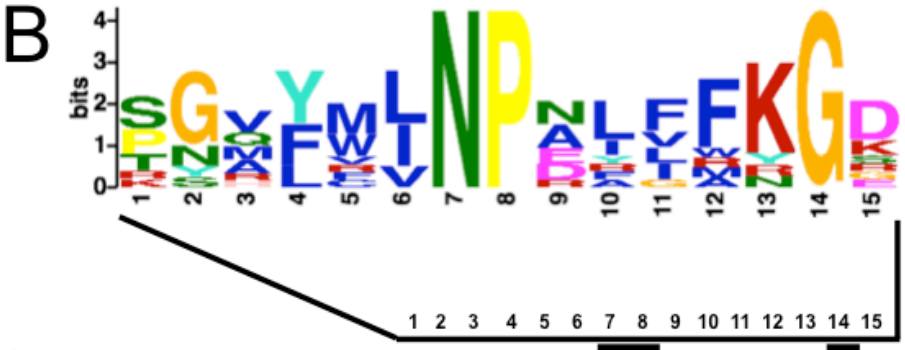

A
Natranaerobius thermophilus integrase
Clostridium ljungdahlii DNA breaking-rejoining enzyme
Selenomonas sp. FOBRC6 integrase
Peptoniphilus lacrimalis XerC
Peptoniphilus sp. 1-1 XerC
Chlamydia trachomatis XerC
Firmicutes multispecies XerC
Clostridiales bacterium VE202-03 XerD
Eubacterium desmolans integrase
Lachnospiraceae bacterium NK4A144 XerD
Lachnoclostridium phytofermentas integrase/recombinase
Clostridium symbiosum ATCC 14190 XerD
Clostridium sp. ATCC BAA-442 integrase
Anaerococcus hydrogenalis integrase
Clostridium sp. ASF502 XerD
Clostridium sp. ASF502 XerC
Dorea sp. 5-2 XerC/D
Ktedonobacter racemifer integrase
Ktedonobacter racemifer XerC
Ammonifex degensii XerD
Thermoanaerobacter ethanolicus XerD
Thermoanaerobacter ethanolicus DNA breaking-rejoining enzyme
Thermoanaerobacterium saccharolyticum XerD
Thermoanaerobacterium thermosaccharolyticum XerD
Caldicellulosiruptor bescii XerD
thermophiles
0.2
B
amino acid residue
50
100
150
200
250
300
R
H R
Y
N
C
phage integrase SAM superfamily
DNA breaking-rejoining enzyme C superfamily

Figure 3.2. Cbes2777 has thermophilic homologs and a conserved Xer-like catalytic domain. (A) Maximum likelihood tree of Xer-like homologs of Cbes2777. Cbes2777 resides in a cluster with predominantly thermophilic organisms ($T_{opt} \geq 60°C$), indicated by the bracket. Posterior probabilities determined by MRBAYES are shown at the nodes of the tree. The XerD homolog from *Caldicellulosiruptor bescii* is indicated by a red branch. The scale bar indicates the distance for 0.2 amino acid substitutions per site. **(B)** Protein domains of the Cbes2777 XerD recombinase. R164, H243, R246, and Y328 appear to constitute the RHRY conserved catalytic residues [220], indicated above the C-terminal catalytic domain.





	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Sodalis glossinidius str. morsitans	K	G	H	F	F	I	N	P	N	F	V	F	N	G	D
Pseudoalteromonas sp. PS1M3	K	G	T	Y	F	I	N	P	N	F	V	F	N	G	N
Bartonella schoenbuchensis	P	N	Q	Y	W	I	N	P	A	I	F	F	K	G	E
Pseudomonas savastanoi	P	S	M	Y	W	I	N	P	A	L	F	F	K	G	D
Paracoccus marcusii	P	N	Q	F	W	V	N	P	A	L	F	F	K	G	D
Moritella viscosa	S	G	V	L	M	L	N	P	E	I	I	F	K	G	K
Bacillus subtilis	T	G	V	L	M	L	N	P	E	L	L	M	R	G	D
Selenomonas ruminantium	S	G	A	Y	C	V	N	P	D	A	V	F	K	G	G
RepL consensus	S	G	V	Y	M	I	N	P	D	L	L	F	K	G	D
Cyanothece sp. PCC 8802	S	Y	A	F	R	L	N	P	N	Y	G	W	K	G	K
Sodalis glossinidius	T	G	M	Y	V	L	N	P	N	L	F	A	K	G	S
Caldicellulosiruptor bescii	R	G	R	F	M	L	N	P	R	R	I	R	Y	G	R

Figure 3.3. Cbes2780 RepL protein is unique, but exhibits a conserved motif. (A) Maximum likelihood tree of plasmid-encoded RepL-like homologs to Cbes2780. Plasmid replication proteins are listed with the plasmids that encode them. The RepL consensus sequence is from Sprincova *et al* [244]. The scale bar indicates the distance for 0.5 amino acid substitutions per site. **(B)** The consensus motif of the conserved RepL helix-turn-helix domain generated by MEME [228]. In the multiple sequence alignment of 11 plasmid-encoded rep proteins, the conserved N,P and G residues are indicated in black with white font. Sites where the *Caldicellulosiruptor bescii* sequence contains arginine residues that are absent in all other sequences are in red in the multiple sequence alignment.

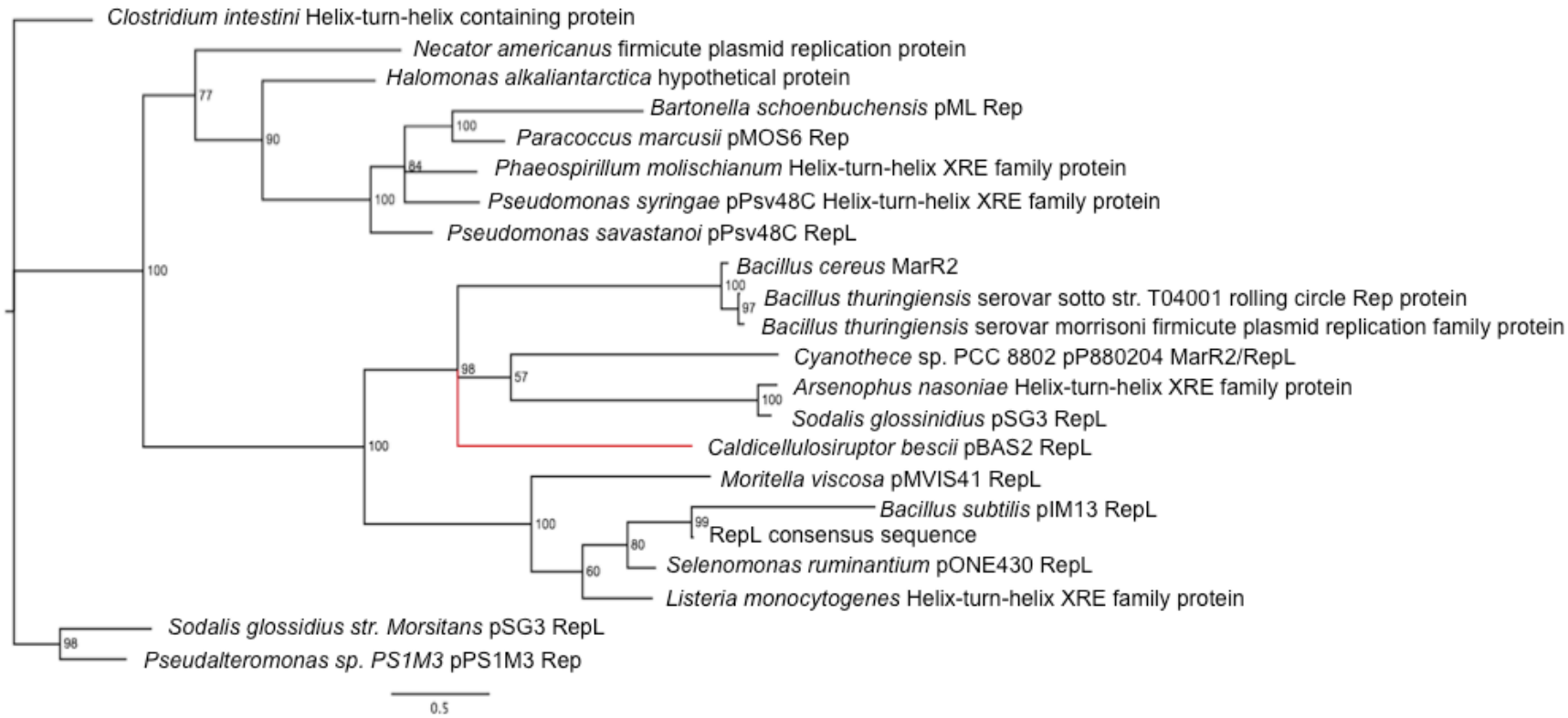

Clostridium intestini Helix-turn-helix containing protein
Necator americanus firmicute plasmid replication protein
Halomonas alkaliantarctica hypothetical protein
Bartonella schoenbuchensis pML Rep
Paracoccus marcusii pMOS6 Rep
Phaeospirillum molischianum Helix-turn-helix XRE family protein
Pseudomonas syringae pPsv48C Helix-turn-helix XRE family protein
Pseudomonas savastanoi pPsv48C RepL
Bacillus cereus MarR2
Bacillus thuringiensis serovar sotto str. T04001 rolling circle Rep protein
Bacillus thuringiensis serovar morrisoni firmicute plasmid replication family protein
Cyanothece sp. PCC 8802 pP880204 MarR2/RepL
Arsenophus nasoniae Helix-turn-helix XRE family protein
Sodalis glossinidius pSG3 RepL
Caldicellulosiruptor bescii pBAS2 RepL
Moritella viscosa pMVIS41 RepL
Bacillus subtilis pIM13 RepL
RepL consensus sequence
Selenomonas ruminantium pONE430 RepL
Listeria monocytogenes Helix-turn-helix XRE family protein
Sodalis glossidius str. Morsitans pSG3 RepL
Pseudalteromonas sp. PS1M3 pPS1M3 Rep
77
90
100
84
100
100
100
97
98
57
100
100
100
99
80
60
98
0.5

Figure 3.4. Maximum likelihood tree of full list of RepL-like homologs to Cbes2780. The tree contains both plasmid-encoded and chromosome-encoded homologs to Cbes2780. Plasmid replication proteins are listed with the plasmids that encode them. The RepL consensus sequence is from Sprincova *et al* [244]. The scale bar indicates the distance for 0.5 amino acid substitutions per site.

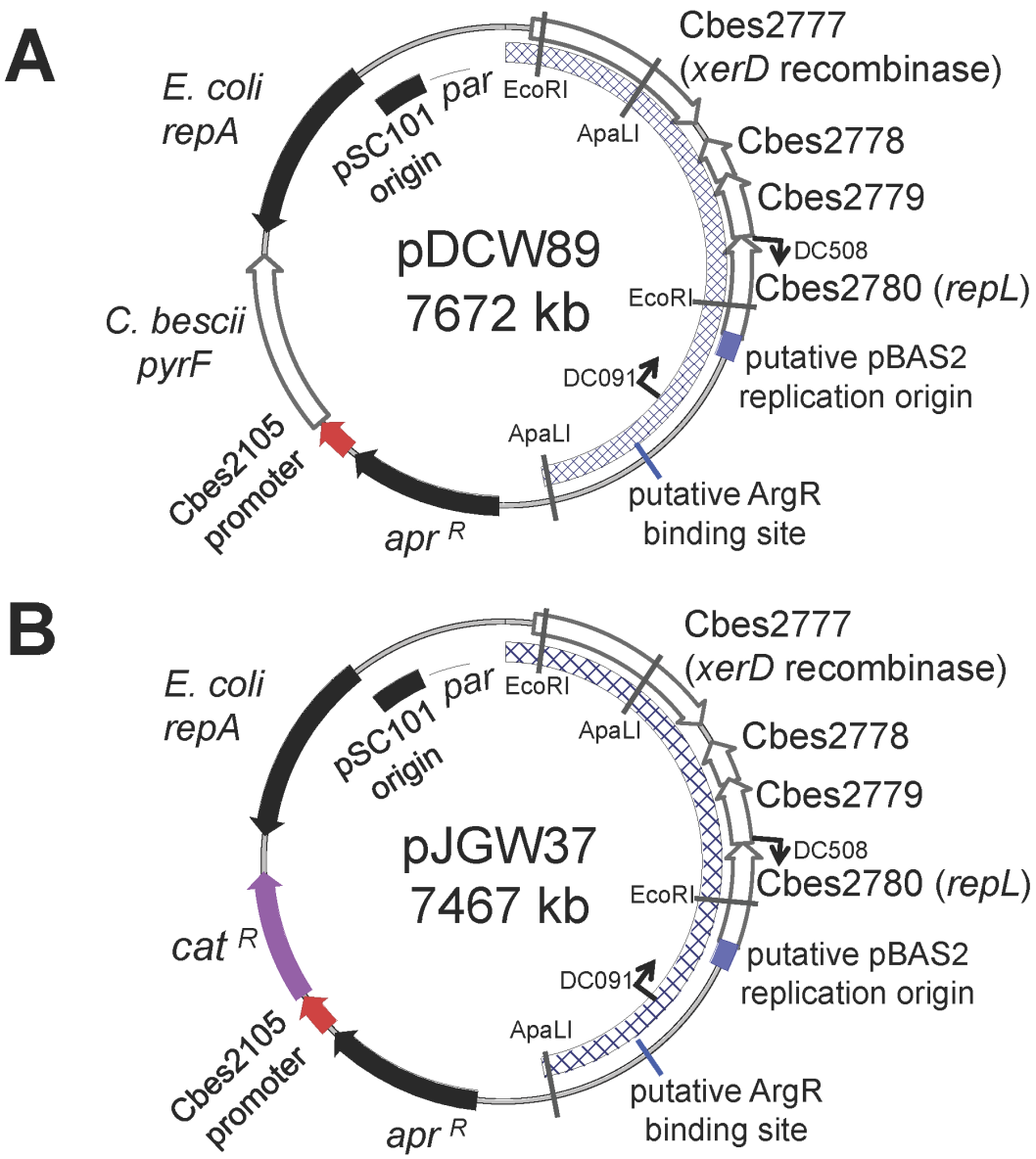

A
E. coli repA
pSC101 origin
par
EcoRI
ApaLI
Cbes2777 (xerD recombinase)
Cbes2778
Cbes2779
DC508
Cbes2780 (repL)
EcoRI
putative pBAS2 replication origin
DC091
ApaLI
putative ArgR binding site
apr R
Cbes2105 promoter
C. bescii pyrF
pDCW89
7672 kb
B
E. coli repA
pSC101 origin
par
EcoRI
ApaLI
Cbes2777 (xerD recombinase)
Cbes2778
Cbes2779
DC508
Cbes2780 (repL)
EcoRI
putative pBAS2 replication origin
DC091
ApaLI
putative ArgR binding site
apr R
Cbes2105 promoter
cat R
pJGW37
7467 kb
C
M
1
2
3
8.0
6.0
5.0
4.0
3.0
2.0
1.5
1.0

Figure 3.5. **Maps of vectors transformed into *Clostridium thermocellum*.** **(A)** pDCW89 constructed with the *C. bescii pyrF* gene driven by the *C. bescii* Cbes2105 ribosomal protein S30A promoter. The hatched region was derived from *C. bescii* native plasmid pBAS2. **Apr^R**, apramycin resistance casette; **repA**, replication initiator for *E. coli* pSC101 replication origin; **par**, partitioning locus for *E. coli*. Primers for PCR verification of transformation and restriction sites for structural verification are shown on the plasmid map**.** **(B)** pJGW37 is identical to pDCW89, but with the chloramphenicol acetyltransferase gene (*cat*) as the selectable marker. **(C)** Plasmid DNA can be visualized in total DNA directly purified from *C. thermocellum*. **1**: 500 ng pJGW37 purified from *E. coli*. **2**: 1.3 μg total DNA purified from *C. thermocellum Δhpt*. **3**: 1.3 μg total DNA purified from *C. thermocellum Δhpt* containing pJGW37**.**

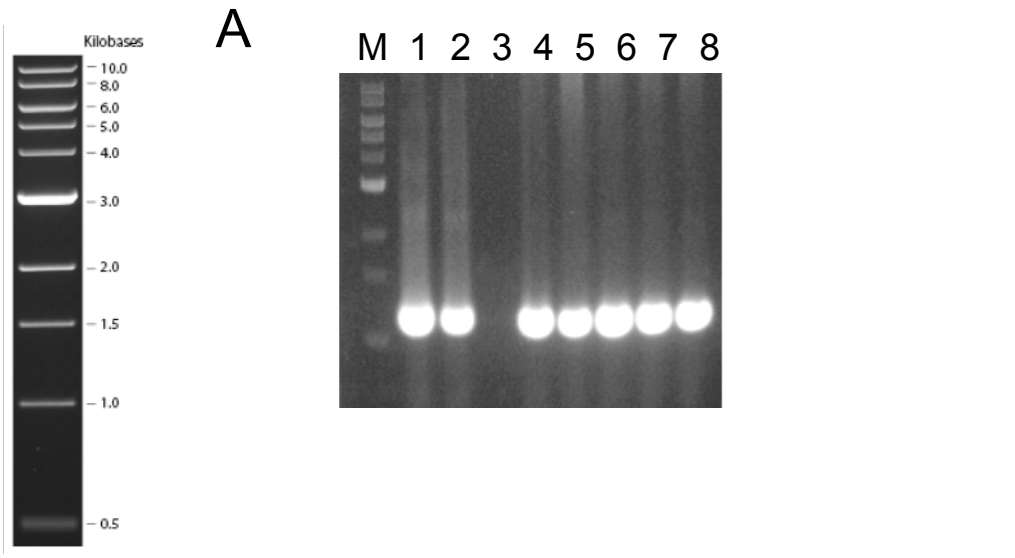
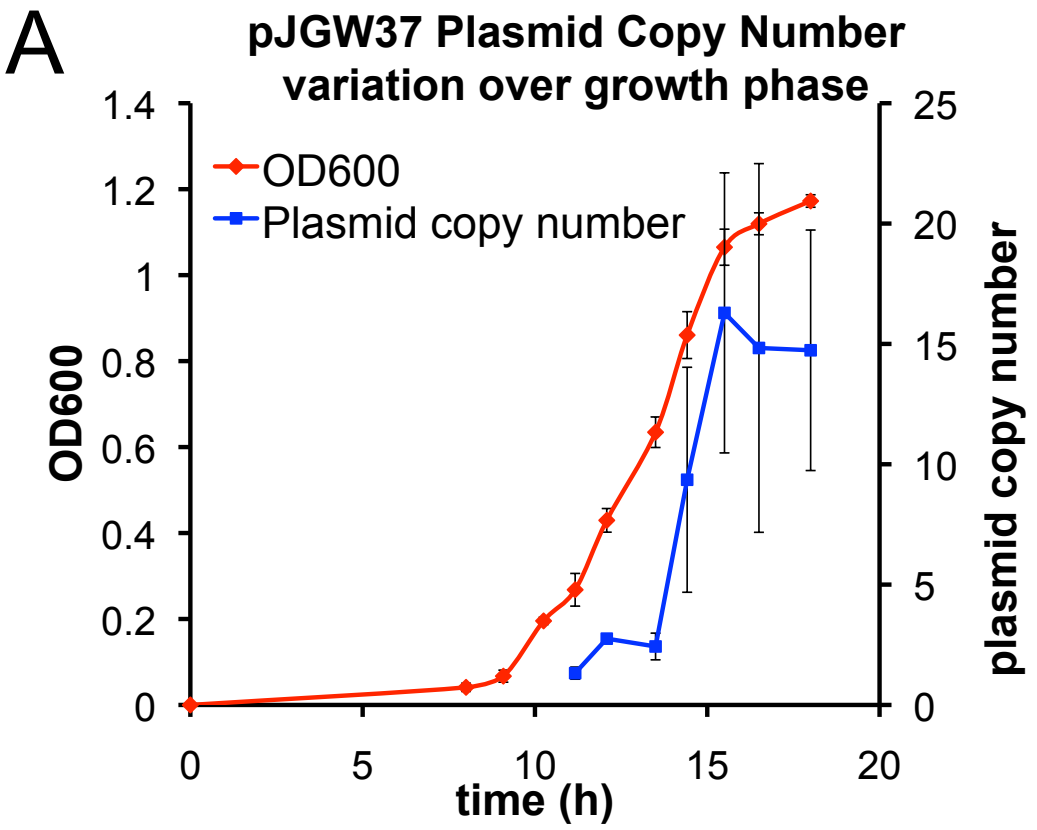

A
Kilobases
10.0
8.0
6.0
5.0
4.0
3.0
2.0
1.5
1.0
0.5
M 1 2 3 4 5 6 7 8
B
M 1 2 3 4 5 6 7 8 M
C
M 1 2 3 4 5 6 7 8 M

Figure 3.6. Plasmids with the pBAS2 replication origin are structurally stable in *C. thermocellum* DSM1313. (A) PCR verification of the pDCW89 plasmid in *C. thermocellum* Δ*pyrF* transformants using plasmid-specific primers DC091/DC508 (See Figure 3.5). **1:** plasmid pDCW89; **2:** plasmid pJGW37; **3:** negative control *C. thermocellum* Δ*pyrF* DNA; **4-6:** total DNA from uracil prototrophic pDCW89 transformants; **7-8**: total DNA from thiamphenicol resistant pJGW37 transformants. **(B)** Back-transformation of pJGW37 isolated from *C. thermocellum* into *E. coli*, followed by restriction digest. **1,5:** pJGW37 purified from *E. coli*; **2-4, 6-8:** plasmid isolated from individual *E. coli* colonies back-transformed with *C. thermocellum* DNA; **1-4:** Cut with EcoRI; **5-8:** Cut with ApaLI. (**C)** Back-transformation of pDCW89 isolated from *C. thermocellum* into *E. coli*, followed by restriction digest. **1,5:** pDCW89 purified from *E. coli*; **2-4, 6-8:** plasmid isolated from individual *E.coli* colonies back-transformed with *C. thermocellum* DNA; **1-4:** Cut with EcoRI; **5-8:** Cut with ApaLI. DNA ladder is the 1 kb Ladder from NEB.


A
pJGW37 Plasmid Copy Number
variation over growth phase
OD600
Plasmid copy number
OD600
plasmid copy number
time (h)
0
0.2
0.4
0.6
0.8
1
1.2
1.4
0
5
10
15
20
25

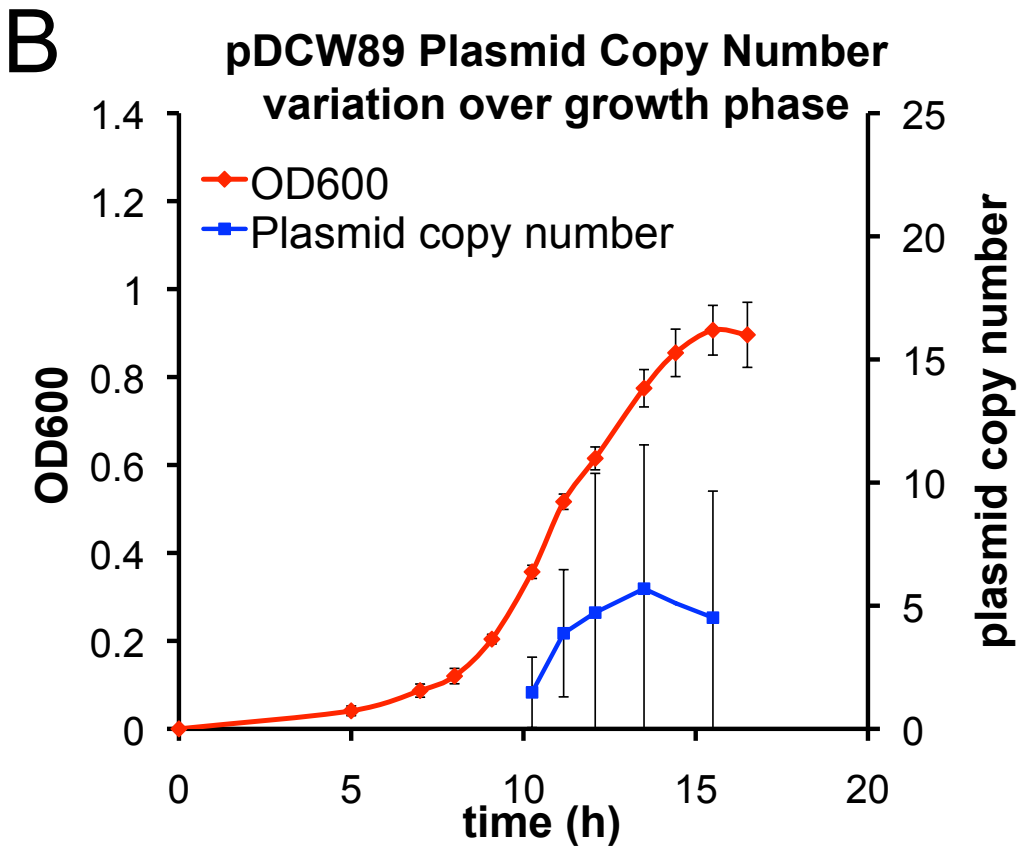

B
pDCW89 Plasmid Copy Number
variation over growth phase
OD600
Plasmid copy number
OD600
plasmid copy number
time (h)
0
0.2
0.4
0.6
0.8
1
1.2
1.4
0
5
10
15
20
25

Figure 3.7. Plasmid copy number is dependent on growth phase. Plasmid pJGW37 **(A)** was maintained with thiamphenicol resistance, while pDCW89 **(B)** was maintained with uracil auxotrophy. Growth of triplicate cultures at 60°C determined by OD_{600} is portrayed in red. Samples (2% (v/v)) for plasmid copy number analysis were taken from the standing cultures. Plasmid copy number (PCN), indicated by blue squares, represents the copies of plasmid per chromosome as measured by qPCR.

Table 3.1. Homologous proteins to Cbes2778. BLAST analysis identified five hypothetical protein potential homologs to Cbes2778 in the *Caldicellulosiruptor* genus with relatively high query coverage and sequence identity. Multiple additional similar proteins were identified, mostly from eukaryotes.

Organism	Annotation	Query cover	% identity
Caldicellulosiruptor saccharolyticus	Hypothetical protein	94%	43%
Caldicellulosiruptor sp. Rt8.B8	Hypothetical protein	75%	42%
Caldicellulosiruptor kronotskyensis	Hypothetical protein	90%	35%
Spiroplasma apis	FeS assembly protein	76%	32%
Lumbricus terrestris	extracellular hemoglobin linker chain (N-terminal)	69%	33%
Lumbricus terrestris	Hemoglobin complex, chain 2	72%	34%
Lumbricus terrestris	Extracellular hemoglobin linker L3 subunit precursor	72%	34%
Lumbricus terrestris	Chain O, Lumbricus Erythrocruorin	72%	34%
Acidithiobacillus thiooxidans	hypothetical protein	47%	43%
Caldicellulosiruptor saccharolyticus	hypothetical protein	42%	45%
Acidithiobacillus thiooxidans	hypothetical protein	47%	43%
Kluyvera ascorbata	hypothetical protein	61%	27%

Table 3.2. Homologous proteins to Cbes2779. Cbes2779 has no homologs in *Caldicellulosiruptor* and shows only weak sequence similarity to any other bacterial proteins.

Organism	Annotation	Query cover	% identity
Eubacterium cellulolosolves	Transketolase	48%	33%
Buchnera aphidocola	GTPase EngA	84%	32%
Prolixibacter bellariivorans	Hypothetical protein	54%	31%
Stanieria cyanosphaera	Ferritin Dps family protein	41%	33%

Table 3.3. Strains and plasmid used in Chapter 3.

Strain or plasmid	Genotype/phenotype	Source
C. thermocellum		
LL1005	*C. thermocellum* DSM 1313 Δ*pyrF* (*ura*$^{-}$/ -FOA^{R})	Tripathi et al., 2010
M1354	*C. thermocellum* DSM 1313 Δ*hpt*	Argyros et al., 2011
JWCT02	*C. thermocellum Δpyrf* + pDCW89 (*ura*$^{+}$/5-FOA^{S})	This work
JWCT03	*C. thermocellum Δpyrf* + pDCW196 (*ura*$^{+}$/5-FOA^{S}/ Tm^{R})	This work
JWCT04	*C. thermocellum Δhpt* + pDCW196 (*ura*$^{+}$/5-FOA^{S}/ Tm^{R})	This work
JWCT05	*C. thermocellum Δhpt* + pJGW37 (Tm^{R})	This work
E. coli		
JW421	BL21 containing pJGW37 (Apramycin^{R})	This work
JW422	BL21 containing pDCW89 (Apramycin^{R})	This work
Plasmids		
pDCW89	*E. coli/C. bescii* shuttle vector ($P_{Cbes2105}$ - *pyrF*)	Chung et al., 2012
pDCW196	*E. coli/C. bescii* shuttle vector ($P_{Cbes2105}$ - *pyrF* / P_{gapDH} - *cat*)	This work
pJGW37	*E. coli/C. bescii* shuttle vector ($P_{Cbes2105}$ - *cat*)	This work

Table 3.4. Plasmid copy number in *Clostridium thermocellum.* Plasmid copy number was determined at mid-log phase after the indicated number of passages. Maintenance experiments were performed in biological duplicate. Uracil prototrophy was used to select for pDCW89. Resistance to thiamphenicol was used to select for pJGW37.

	Passages with selection		Passages without selection		
Plasmid	**1**	**5**	**1**	**2**	**5**
pDCW89	3.0 ± 0.4	9.2 ± 0.9	1.7 ± 1.1	0.1 ± 0.04	0
pJGW37	6.7 ± 0.6	7.5 ± 1.7	0.9 ± 0.1	0.1 ± 0.05	0

Table 3.5. Primers used in Chapter 3.

Primer	Sequence (5′ → 3′)
DC091	ATCTTCATGATTTTCCCAGGA
DC508	AGAGGTACCAGTTCCTGCTTTGTTAACATTCCTTG
X013	AGAGGATCCTTACTTCCTGTCTCGCAACGC
X014	AGACATATGGACAGTTTTCCCTTTGATATGT
X015	AGAGGATCCTGCAATAATTACTGTATCTCTCTGGCA
X016	AGACATATGCTTCAAACTTCCCAAAGGCGAGCCCT
JG024	AGATCATCTAGA GACCATCCTTTCTATGTAGAAA
JG099	AAGATCTAGAATGAACTTTAATAAAATTGATTTAGACAATTGGAA
Q1	TGGGAAAGCCGTCCATAATC
Q2	TCTCCCGCTCTTCTCTCTTT
Q3	GTGCGTCTACAGGACCTTATTT
Q4	GGCAAGATTCTACAGGCAAGA
CTQ1	CCAAACCTCCTTCCCGATATAC
CTQ2	CTCTCAGCTCCTCATCCTCTAT
CTQ3	GGAACCGGAGTGAATGTCATAG
CTQ4	CTGGGAATTGTAGCCCGAATAA

CHAPTER 4

DELETION OF THE *CLOSTRIDIUM THERMOCELLUM RECA* GENE REVEALS THAT IT IS REQUIRED FOR THERMOPHILIC PLASMID REPLICATION.[3]

[3] Groom J, Chung D, Kim S, Guss AM and Westpheling J. To be submitted to *Applied and Environmental Microbiology.*

I. Abstract

Efficient, facile genetic tools are essential for making interesting and important microbes useful. A limitation to the engineering of cellulolytic thermophiles is the availability of functional, thermostable (≥ 60°C) replicating plasmid vectors for rapid expression of genes that provide improved or novel fuel molecule production pathways. A series of plasmid vectors for genetic manipulation of the cellulolytic thermophile *Caldicellulosiruptor bescii* has previously been extended to *Clostridium thermocellum,* another cellulolytic thermophile that very efficiently solubilizes plant biomass and produces ethanol. While this replicon is thermostable, the use of homologous promoters, signal sequences and genes results in undesired recombination into the bacterial chromosome, a result that is shared with less thermostable replicating vectors. In an attempt to overcome undesired DNA recombination in *C. thermocellum*, a *recA* deletion showed that this gene is required for replication of the *C. bescii* thermophilic plasmid pBAS2. *C. thermocellum recA* mutants also showed impaired growth in chemically defined medium and an increased susceptibility to UV damage. Understanding thermophilic plasmid replication is not only important for engineering of these particular cellulytic thermophiles, but also for developing genetic systems in new potentially useful non-model organisms.

II. Introduction

The interactions between lignocellulose-degrading microbes and the plant cell wall have spurred advances towards a renewable energy source. Complex cell wall polysaccharides are recalcitrant to degradation, and this degradation is

the major bottleneck in the conversion of plant biomass to biofuels and chemicals [13]. A variety of schemes have been proposed to break down the plant cell wall and convert the released sugars into fuel molecules and bioproducts. One such scheme, Consolidated Bioprocessing (CBP), consists of one microbe performing both tasks [245]. The cellulolytic thermophile *Clostridium thermocellum* has been the subject of study for many years because of its cellulosome, an interesting extracellular protein supercomplex for degrading the plant cell wall [113, 114]. Its natural ethanologenic capabilities suggest its application as a biofuel-producing microbe, and much work has recently been done to engineer this strain to produce more ethanol from plant biomass [112, 128, 175, 202-204].

A reliable thermostable plasmid-based expression system is required for rapid screening of enzymes to improve carbon metabolism and plant cell wall degradation. We recently developed a series of replicating plasmid vectors for the cellulolytic thermophile *Caldicellulosiruptor bescii* and used them for the expression of homologous and heterologous genes [132, 134, 177, 179]. These vectors have proven useful to extend genetic methods developed for *C. bescii* to another member of this genus, *C. hydrothermalis*, a naïve host for the study of cellulolytic enzymes [219]. Surprisingly, they have also proven useful for *Clostridium thermocellum.* Vectors derived from pBAS2 transform C*. thermocellum* DSM 1313 at 60°C, with a transformation efficiency greater than 3000 colony forming units per µg of transformed DNA and a copy number ~10 per bacterial chromosome [246]. pBAS2 has recently been used to express a thermostable butanol dehydrogenase gene in *C. thermocellum* that provides

resistance to 5-hydroxymethylfurfural, a plant biomass derived microbial inhibitor [135]

Despite these advances, the use of homologous sequences as small as a 300-bp promoter region results in plasmid integration using both pBAS2 [135] and the less thermostable replicon pNW33N [129]. Overcoming this integration is necessary to preserve the integrity of the background strain while still using the best promoters [176], signal peptides and enzymes to meet strain development goals. One way to avoid recombination is to use heterologous sequences. In fact, several promoters from *Caldicellulosiruptor bescii* have been shown to function in *C. thermocellum* [135, 246]. However, because many of useful genetic tools will be encoded by endogenous (that is, homologous) sequences, a deletion of the *recA* gene would be useful. Although the RecA protein has been shown to have multiple functions in model organisms like *E. coli* [247], viable *recA* mutants deficient in homologous recombination have been selected for many decades [248].

The canonical RecA protein in the model organisms *E. coli* and *Bacillus subtilis* is a single-stranded DNA (ssDNA)-binding protein that initiates homologous recombination [247, 249]. RecA monomers join together to form a helical filament on ssDNA, protecting and preparing the DNA for strand invasion into duplex DNA. Once assembled, the RecA filament forms a second DNA binding site that allows strand invasion. This complex scans the invaded duplex and freezes when sequence homology is detected so that replication, subsequent branch migration and recombination can occur. Besides the RecA

protein's function in homologous recombination, it recognizes DNA damage and triggers the SOS response. It does this by promoting autocleavage of the global transcriptional repressor LexA, thereby alleviating repression on a variety of genes involved in DNA repair. Also, RecA and other ssDNA-binding proteins are required for mutagenic translesion synthesis by DNA polymerase V in cases of DNA damage.

Unsurprisingly, there are a variety of phenotypes observed with *recA* mutants in different bacteria. *recA* mutants in *Thermus thermophilus* [250]*, E. coli* [251]*,* and *B. subtilis* [252] all exhibit decreased cell viability. The thermophile *T. thermophilus* has extremely low viability (only 1 colony-forming unit in 10^3-10^5 cells) while more moderate decreases are observed in the other organisms (50-90% decrease in colony formation). The *T. thermophilus* strain also has a hypermutable phenotype, which could be attributed to a higher mutation rate at high temperatures [250]. In *Streptomyces lividans,* a Gram-positive actinomycete, *recA* mutants are inviable unless suppressor mutations are selected [253]. Nucleoid morphologies often change as well, with ~10% of *E. coli* mutants being anucleate due to defective chromosome partitioning [254]. Many organisms exhibit increased susceptibility to DNA damage like UV rays or mitomycin C, since the RecA protein is not available to initiate recombination-mediated repair at DNA lesions [255]. However, *recA* mutants in some organisms like *Thiobacillus ferrooxidans* are less sensitive to radiation than mutants in *E. coli* and *B. subtilis* [256].

We have deleted the *recA* gene in *Clostridium thermocellum* to show the phenotype of the strain with regard to homologous recombination. The primary goal was not only to stabilize a plasmid-based gene expression system by eliminating undesired recombination, but also to learn about the basic biology of RecA in Gram-positive anaerobic thermophiles. We show that growth and DNA repair ability is compromised by removing *recA* in *C. thermocellum*, and that transformation with a replicating thermophilic plasmid that functions in *C. thermocellum* requires the RecA protein.

III. Materials and Methods

Bacterial growth and media composition. *Clostridium thermocellum* DSM 1313 and its derivatives were grown anaerobically in modified CTFUD medium [205, 246] at 60°C, under an atmosphere of 85% nitrogen, 10% CO_2, and 5% hydrogen. Defined medium for selection was CTFUD-NY [205], which contains a vitamin mix of para-aminobenzoic acid, vitamin B12, biotin, and pyridoxamine in place of the yeast extract. Cells were grown without shaking for transformation and for growth curves. Defined medium CTFUD-NY was supplemented with 360 μM uracil for the comparative growth curve to minimize the effect of the *pyrF* auxotrophic strain. Complex medium for recovery after transformation of *C. thermocellum* LL1005 was similar to CTFUD but contained casein (0.2% w/v) and less yeast extract (0.1% w/v), which is referred to as CTFUD+C. *Escherichia coli* was grown in Luria-Bertani broth supplemented with 50 μg/mL apramycin when selecting for the presence of a plasmid.

Construction of plasmid vectors. Q5 Polymerase (NEB, Ipswich, MA) was used for the polymerase chain reaction (PCR), according the manufacturer's instructions. Fast-link Ligase (Epicentre, Madison, WI) was used for all ligation reactions. Multiple cloning steps were used to construct *pyrF* marker replacement vector pJGW61. Primers X118 (with a PstI site) and X119 (with a BamHI site) were used to amplify from genomic DNA from the region ~1100 bp upstream and downstream around the *Clostridium thermocellum recA* gene (Clo1313_1163). Primers X102 (with a BamHI site) and X103 (with a PstI site) amplified backbone vector pDCW140, and both fragments were subjected to a BamHI/PstI digest followed by ligation to create pCTH07. The *pyrF* gene on the backbone vector was removed by amplification with primers DC176 and DC230, and T4 polynucleotide kinase phosphorylation followed by blunt end ligation re-circularized the fragment, creating pJGW57. The *recA* coding sequence was removed in a similar fashion from pJGW57 using primers X120 (with a KpnI site) and X122 (with a SphI site) to create pJGW58. Next, primers JG137 (with a BglII site) and JG138 (with a BamHI site) were used to amplify the pNW33N replicon from plasmid pMU1162 [122]. Primers JG133 (with a BglII site) and X103 (with a BamHI site) were used to amplify the pJGW58 backbone, and a BglII/BamHI restriction enzyme digest of both fragments followed by ligation created pJGW59. To insert the *pyrF* gene between the *recA* flanking regions, the *pyrF* cassette from pJGW07 [219] was amplified using primers JG163 (with a KpnI site) and JG165 (with a SphI site). A PCR fragment was amplified from pJGW59 using X120 (with a KpnI site) and X122 (with a SphI site), and a KpnI/SphI restriction

digest and ligation with the *pyrF* cassette from pJGW07 was used to create pJGW59.5. To create the *tdk* cassette for negative selection, primers JG168 (with a BamHI site) and DC576 (with a PstI site) were used to amplify a fragment from backbone vector pDCW140 containing the Cbes2105 promoter [117], and this fragment was digested and ligated to a PCR fragment of the *Thermoanaerobacterium saccharolyticum tdk* gene (Tsac_0324) amplified by primers JG166.2 (with a PstI site) and JG167 (with a BamHI site). The resulting vector pJGW60, containing the *tdk* gene driven by the Cbes2105 promoter, was used as a template for PCR with primers JG169 (with a NheI site) and JG170 (with an AatII site). This *tdk* cassette was ligated to the PCR fragment amplified from pJGW59.5 by primers DC176 with a NheI site and DC230 with an AatII site to create the final knockout vector pJGW61.

Plasmid pJGW91 was the precursor to plasmids pJGW92 and pJGW93. pJGW91 was created by amplifying pSKW001 [135] with primers DC576 (with a PstI site) and DC700 (with an AvrII site) and amplifying the *recA* gene from *C. thermocellum* DSM1313 genomic DNA with primers JG248 (with a PstI site) and JG249 (with an AvrII site), digesting these products with PstI and AvrII, and ligating them together. To swap the selectable marker in pJGW92, JG024 (with an XbaI site) and X014 (with an NdeI site) amplified pJGW91, JG099 (with an XbaI site) and JG250 (with an NdeI site) amplified the *cat* marker from pDCW196 [246], and the fragments were digested with XbaI and NdeI and ligated. To swap the replication origin for pJGW93, JG251 (with a BglII site) and JG252 (with a BamHI site) amplified pJGW92, JG137 (with a BglII site) and JG138 (with a

BamHI site) amplified the pNW33N replicon from pMU1162 [122], and the fragments were digested with BglII and BamHI and ligated.

Selection for *Clostridium thermocellum recA* replacement with *pyrF* from *Caldicellulosiruptor bescii*. *C. thermocellum* LL1005 was grown to OD_{600} ~ 0.7 in 150 mL of defined CTFUD-NY medium [173, 246] supplemented with 360 µM uracil and 0.4 g/L glycine (Sigma, St. Louis, MO). Cells were harvested in three 50 mL aliquots at 7000 x g for 8 minutes, and washed twice with a solution of ice-cold 10% glycerol (w/v) and 10% sucrose (w/v). Each of the three pellets was resuspended in 350 µL of the wash solution, and pelleted in one Eppendorf tube with a refrigerated desktop centrifuge for 5 minutes at 13,000 RPM. The pellet was homogenized with a pipet, and divided into 30 µL aliquots. 1 µg of pJGW61 plasmid was mixed with an aliquot, and cells were electroporated in a 1-mm gap cuvette with an exponential pulse using a BioRad Gene Pulser (1.8 kV, 25 µF, 350 ohms). Cells were immediately recovered in 10 mL CTFUD-C at 50 °C to allow replication of the pNW33N replicon. Subcultures of 50 µL were transferred every 2 hours into 20 mL CTFUD-NY at 60°C to select for uracil prototrophy.

After 3-5 days, several subcultures became turbid, and these cultures were diluted again to ensure that cells were prototrophs. The fresh cultures were plated onto CTFUD medium with 4.5 g/L yeast extract and 0.8% (w/v) agar, supplemented with 10 µg/mL 5-fluoro-2′-deoxy-uridine (FUDR) to select against the plasmid backbone. Colonies formed after 2-3 days, and were picked into CTFUD-NY medium. Only 2 of 20 colonies grew in CTFUD-NY medium, both of

which were shown by PCR with internal primers to be merodiploids (Figure 4.3). After a subsequent round of plating on solid CTFUD-NY medium, 1 in 20 colonies sustained a clean deletion of the *recA* gene. This strain was named JWCT10.

UV sensitivity assay. Cultures of 25 mL of both LL1005 and JWCT10 were grown in CTFUD medium to early exponential phase (OD_{600} = 0.35-0.4). Cells were divided into 3 mL aliquots and chilled on ice for 20 minutes before being harvested at 4000 x g for 6 minutes. The cells were then suspended in 3 mL ice cold 0.1 M Mg_2SO_4 (JT Baker, Phillipsburg, NJ, USA) as recommended in the literature to eliminate absorption of UV light by the rich medium [257]. Resuspended cells were then exposed on plastic Petri dishes to varying intensities of UV light ranging from 0 to 1000 μJ/cm^2. Although no apparent photolyase homologs are present in the *C. thermocellum* genome that would photoreactivate DNA in response to UV damage [234], cells were manipulated and incubated in the dark after UV exposure. Cells were diluted and plated on CTFUD medium at 60°C for 3 days. Percent survival was calculated for each exposed sample by comparing colony forming units (CFU) to the CFU of cells not exposed to UV light.

Transformation of *C. thermocellum* Δ*recA::pyrF.* *C. thermocellum* LL1005 was grown to OD_{600} ~ 0.4 in 150 mL CTFUD medium with 1 g/L yeast extract [173, 246] supplemented with 40 μM uracil and 0.4 g/L glycine (Sigma, St. Louis, MO, USA). Cells were harvested in three 50 mL aliquots at 7000 x g for 8 minutes, and washed twice with a solution of ice-cold 10% glycerol (w/v) and

10% sucrose (w/v). Each of the three pellets was resuspended in 350 µL of the wash solution, and pelleted in one Eppendorf tube with a refrigerated desktop centrifuge for 5 minutes at 13,000 RPM. The pellet was homogenized with a pipet, and divided into 30 µL aliquots. Plasmid (700 ng) was mixed with each aliquot, and cells were electroporated in a 1-mm gap cuvette with an exponential pulse using a BioRad Gene Pulser (1.8 kV, 25 µF, 350 ohms). Cells were immediately recovered in 10 mL CTFUD-C at 53°C for plasmids with the pNW33N replicon or 60°C for plasmids with the pBAS2 replicon. Subcultures of 1 mL were transferred three times over the course of 24 h into 10 mL CTFUD with 6 µg/mL thiamphenicol at 60°C to select for resistance.

After 3-4 days, several subcultures became turbid, and these cultures were diluted again to ensure that cells were thiamphenicol resistant. The fresh cultures were plated onto CTFUD medium with 1 g/L yeast extract and 0.8% (w/v) agar, supplemented with 10 µg/mL thiamphenicol. Colonies formed after 2 days and were picked into the same liquid medium.

Analysis of transformed plasmid stability. *C. thermocellum* culture (5 mL) was harvested at 3,220 x g for 7 minutes, and the pellets were used for extraction of DNA with the Quick-gDNA Extraction kit (Zymo Research, Irving, CA, USA). PCR with Taq polymerase (Sigma, St. Louis, MO, USA) was performed with two sets of primers (JG162/JG144 and JG161/JG155) to check for internal bands, and JG161 and DC232 to check for the *pyrF* marker, according to the manufacturer's instructions (94°C duplex denaturation, 56°C annealing temperature, 1 minute per kb at 72°C for elongation). Approximately

180 ng of the mixture of genomic and plasmid DNA was electrotransformed into *E. coli* DH5α via single electric pulse (2.5 kV, 200 Ω, and 25 microF) in a pre-chilled 2-mm cuvette using a Bio-Rad Gene Pulser. The cells were placed into 1 mL SOC medium for 1 h with shaking at 37°C, and then plated onto LB agar supplemented with 50 µg/mL apramycin. The colonies were picked into 10 mL LB with 50 µg/mL apramycin, and the plasmid DNA was extracted using a Miniprep kit (Qiagen, Valencia, CA, USA) and screened with restriction enzymes EcoRI and AvaI for pJGW92 (NEB).

IV. Results

The *Caldicellulosiruptor bescii pyrF* gene functions efficiently for marker replacement of *recA* in *Clostridium thermocellum.* We used a *pyrF* gene from *Caldicellulosiruptor bescii* to transform a Δ*pyrF* strain of *Clostridium thermocellum* DSM 1313 to uracil prototrophy (Figure 4.1). By constructing a plasmid with the *pyrF* gene placed between 5′ and 3′ flanking regions of the *C. thermocellum recA* gene, we were able to select for a replacement of the *recA* gene in the genome. Employing the thymidine kinase *(tdk)* gene from *Thermoanaerobacterium saccharolyticum,* we subsequently counter-selected for the presence of the plasmid backbone sequence, in doing so isolating transformants that had acquired the *pyrF* gene but had not integrated the entire plasmid into the *C. thermocellum* genome.

A polymerase chain reaction (PCR) using genomic primers outside of the 5′ and 3′ flanking regions verified this gene replacement (Figure 4.2). Additionally, PCR with multiple sets of internal primers demonstrated the

absence of the *recA* gene in the genome of the new strain (Figure 4.3), named JWCT10 (*Clostridium thermocellum* DSM 1313 Δ*pyrF* Δ*recA::pyrF*, see Table 4.1).

Deletion of *recA* results in UV sensitivity. The Δ*recA* strain exhibited a long lag phase in comparison to its Δ*pyrF* parent strain and to the *C. thermocellum* DSM 1313 wild-type strain (Figure 4.4). It also exhibited a much slower growth rate and reached much lower cell densities in defined medium (Figure 4.4A, Table 4.2), but these effects were less apparent in rich medium (Figure 4.4B). In rich medium, the growth rate of *C. thermocellum* Δ*recA* was comparable or even slightly better than its parent strain, although the lag in growth was appreciable (Table 4.2).

In addition to growth phenotypes, Δ*recA* strains often exhibit sensitivity to UV light due to the deficiency in recombination-mediated double strand break repair [247, 248]. Fewer colonies of *C. thermocellum* Δ*recA* formed after exposure to UV light than formed after the same treatment of the parent strain *C. thermocellum* Δ*pyrF*, indeed showing that Δ*recA* cells are more UV sensitive (Figure 4.5). The effect was pronounced at UV intensities below 1000 μJ/cm^2. Results for the wild-type strain were comparable to previous studies of UV sensitivity in *C. thermocellum* [257].

The Δ*recA::pyrF* mutant has a much lower transformation efficiency than the wild-type strain and requires complementation for pBAS2 replication. Fifteen attempts to transform the Δ*recA* strain to thiamphenicol resistance with pJGW37 on solid medium failed. Attempts to complement the

Δ*recA* strain with a wild-type copy of *recA* (plasmid pJGW92, Figure 4.6A) were also unsuccessful on solid medium. However, when cells electroporated in the presence of pJGW92 were selected in liquid medium containing thiamphenicol, dense cultures grew after 3-4 days. Thiamphenicol-resistant cultures could be plated onto solid medium, and colonies were viable.

The *C. thermocellum* Δ*recA* transformants maintained plasmids that could be back-transformed into *E. coli* cells by DNA extracted from *C. thermocellum* (Figure 4.6B). Polymerase chain reaction (PCR) performed on the extracted DNA showed that the transformed strains still had the *C. bescii pyrF* gene in place of the *recA* gene (Figure 4.7). PCR using primers with homology to the wild-type *recA* sequence did not amplify DNA with 30 cycles. However, if the number of cycles was extended to 40, a faint band was observed in the transformed Δ*recA* strain that migrated around the same size as the predicted wild-type fragment. Control reactions containing purified plasmid pJGW92 mixed with purified Δ*recA* genomic DNA revealed that this was almost certainly a PCR artifact. It likely results from hybridization and amplification between the plasmids and genomic DNA that is only visible with extended cycles (Figure 4). This strongly suggests that these strains are not contaminants; it is expected that if a wild-type contaminant were grown under selection for thiamphenicol resistance and *recA* function in liquid medium, it would outcompete the mutant strain and this would be observable as stronger amplification of wild-type alleles by PCR.

V. Discussion

We report that plasmids based on thermophilic replicon pBAS2 cannot successfully transform a *recA* mutant of *Clostridium thermocellum.* This is not useful for overcoming the issues of recombination with homologous sequences on pBAS2-based plasmids, but these results do shed light on the replication mechanism of this plasmid.

The pBAS2 replicon has a conserved Firmicute origin of replication for rolling-circle plasmids, but the single stranded intermediates characteristic of rolling-circle origins were not observed [178]. While it is not an inherent feature of rolling-circle plasmids to require recombination, there is a Xer-like recombinase encoded on pBAS2 [177, 246]. The presence of this recombinase is consistent with the necessity of recombination for replication, and therefore the necessity of RecA. Because of the unsuccessful transformation attempts, it appears that pBAS2 might require recombination, and RecA function, for faithful replication. Similar results have been observed in *Streptococcus pneumoniae.* Establishment of multi-copy rolling-circle plasmid pMV158 was blocked but not completely abolished by a *recA* mutation [258]. However, this effect was not observed in plasmid establishment in *recA* mutants of *B. subtilis*, even though the established plasmid also had a rolling-circle replication mechanism [259].

Thiamphenicol resistance was only observed when selection was performed in liquid medium rather than solid medium. The absence of transformation on solid medium is reminiscent of results observed in *T. thermophilus*, in which colony formation was 100,000-fold worse in a Δ*recA* strain

[250]. The combination of a powerful antibiotic like thiamphenicol and the already poor colony formation of a Δ*recA* strain could feasibly result in the lack of transformed colonies that we observe. However, the liquid selection method yielded thiamphenicol-resistant colonies harboring intact plasmids, as long as a copy of *recA* was present on the plasmid.

Additional experiments will clarify the effect of the *recA* deletion. Further passages of the complemented Δ*recA* strain will determine whether plasmids eventually integrate at the small regions of homology to the chromosome on the plasmid (~20 bp of residual *recA* sequence). Further characterization of the spontaneous mutation rate would also be informative. It has been shown that the mutation rate of *C. thermocellum* to 5-FOA resistance is ~2×10^{-6} [257]. The Δ*recA* strain could be put to a similar test to see if the mutation rate is higher like *T. thermophilus* Δ*recA*, or if like *E. coli* Δ*recA*, it is defective in DNA damage-mediated mutagenesis (via translesion synthesis) and actually has a lower mutation rate [250]. Although this Δ*recA* strain will likely not be useful with plasmid replicon pBAS2, it may prove useful with other thermophilic replicons that are applied to the organism, like pNW33N.

VI. Future Directions

pNW33N is also believed to be a rolling-circle plasmid [260], but it is not known whether recombination is required for stable replication. Use of pNW33N as a replication origin in the *C. thermocellum* Δ*recA* strain could facilitate a recombination-free plasmid-based expression system. pNW33N and a pNW33N-based *recA* complementation plasmid will be transformed into *C. thermocellum*

Δ*recA* to see if a similar effect is observed as with pBAS2. If pNW33N can readily transform *C. thermocellum* Δ*recA* without complementation, plasmids will be transformed and tested for 1) functional expression and 2) absence of chromosomal integration of the plasmid, using homologous promoters for expression of the β-glucosidase reporter gene developed in Chapter 5 for *C. bescii.*

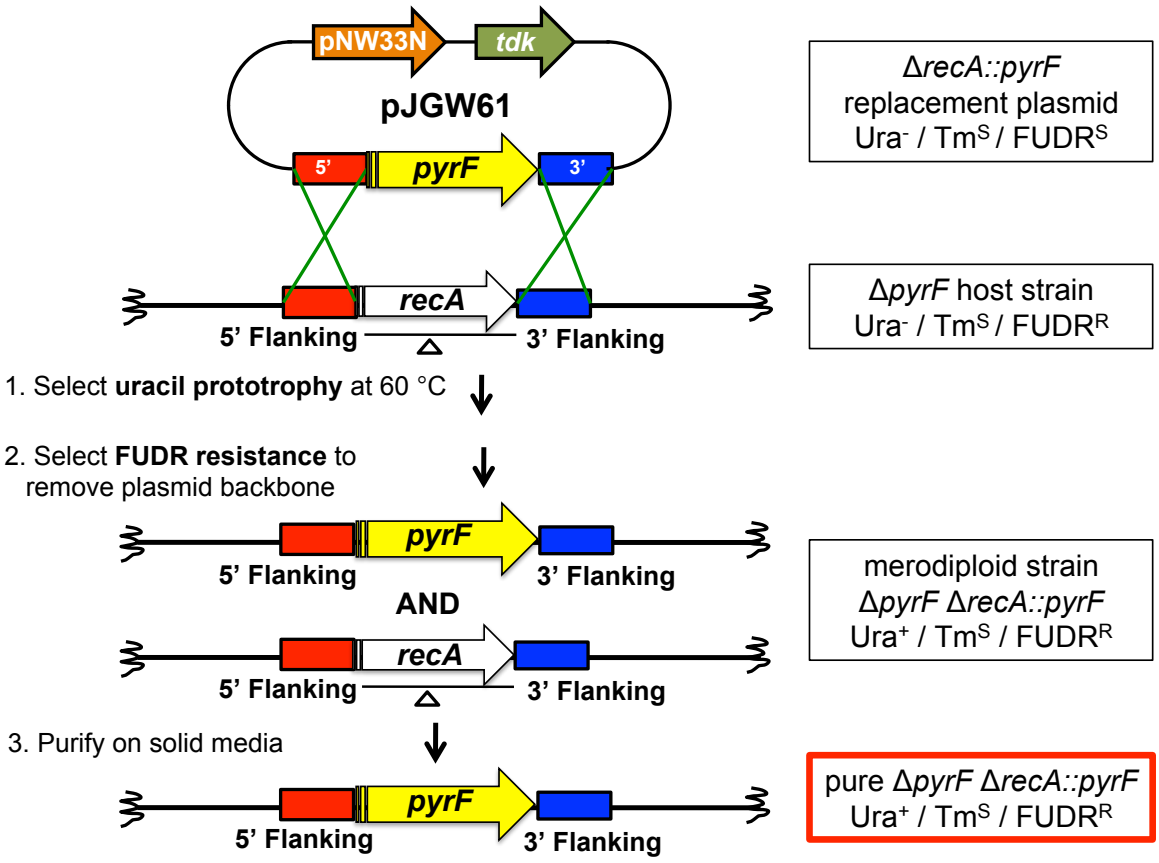


Figure 4.1. Scheme to delete the *recA* gene in *C. thermocellum*. Gene replacement plasmid pJGW61 transformed *C. thermocellum* ΔpyrF to uracil prototrophy. Then a combination of uracil protrophic selection and FUDR exposure selected for those cells which had acquired the pyrF gene but lost the tdk gene on the plasmid backbone. Finally the merodiploid was cured on solid media to generate a pure ΔrecA strain. **pyrF**: *C. bescii* orotidine 5'-phosphate decoarboxylase-encoding gene (Cbes1377). **recA:** *Clostridium thermocellum* DSM 1313 RecA-encoding gene (Clo1313_1163). **tdk:** *Thermoanaerobacterium saccharolyticum* thymidine kinase-encoding gene (Tsac_0324). **FUDR:** f-fluoro-2'-deoxy-uridine. **pNW33N**: replication origin from plasmid pNW33N.

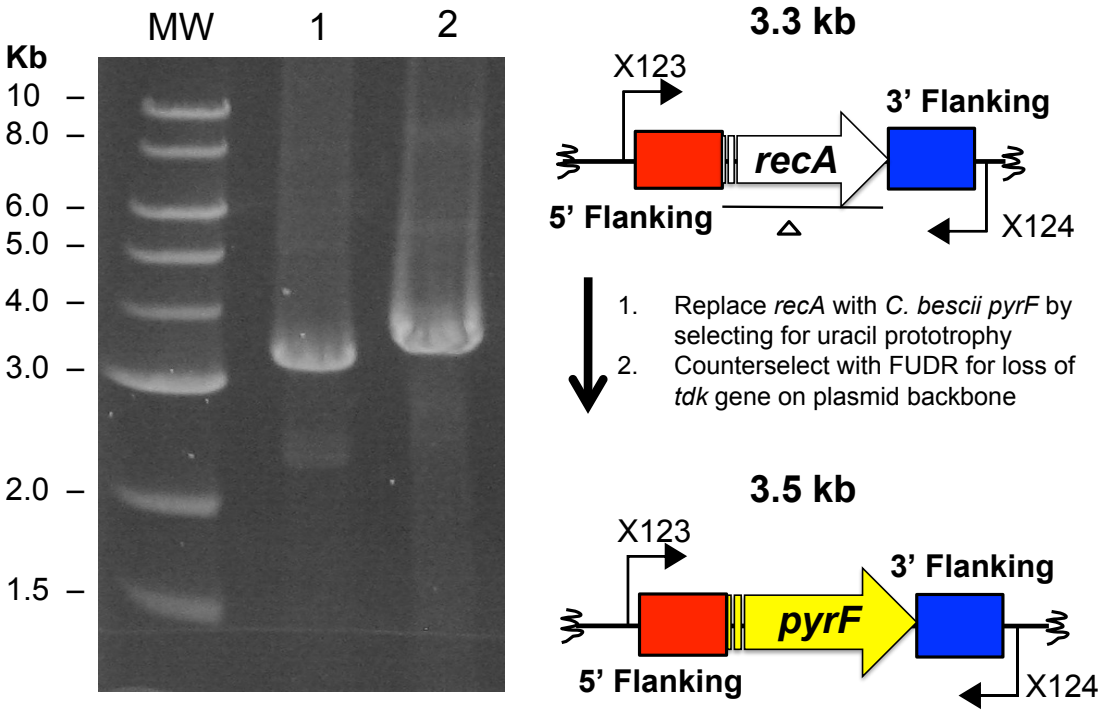


Figure 4.2. Verification of the *pyrF* replacement of *recA*. PCR using primers X123 and X124. Wild-type band, 3.3 kb. *pyrF* recplacement band, 3.5 kb. MW: molecular weight standards. Lane 1: wild-type *C. thermocellum* genomic DNA. Lane 2: *C. thermocellum* Δ*recA::pyrF* genomic DNA.

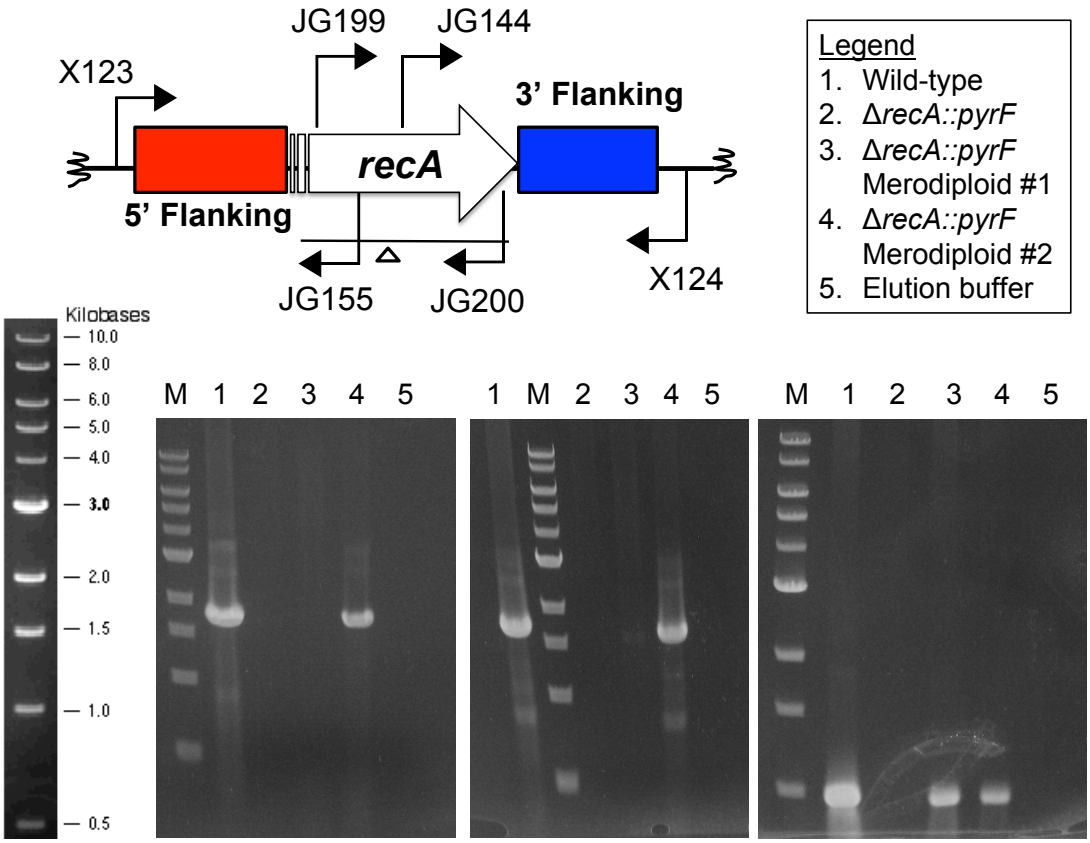


Figure 4.3. Internal PCRs to verify the deletion of *recA*. From left to right, Gel 1: PCR using primers X123 and JG155. Gel 2: PCR using primers JG144 and X124. Gel 3: PCR using primers JG199 and JG200.

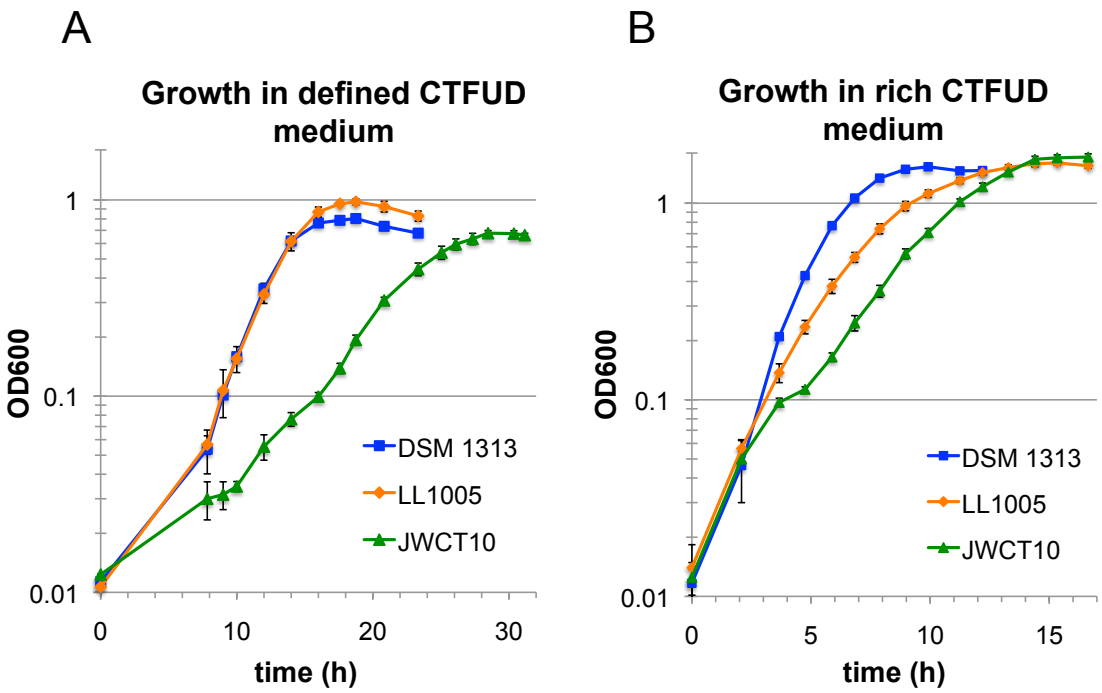


Figure 4.4. Growth of *C. thermocellum* Δ*recA*. **(A)** Growth in defined medium supplemented with uracil as measured by OD_{600}. **(B)** Growth in rich medium (4.5 g/L yeast extract) supplemented with uracil as measured by OD_{600}.

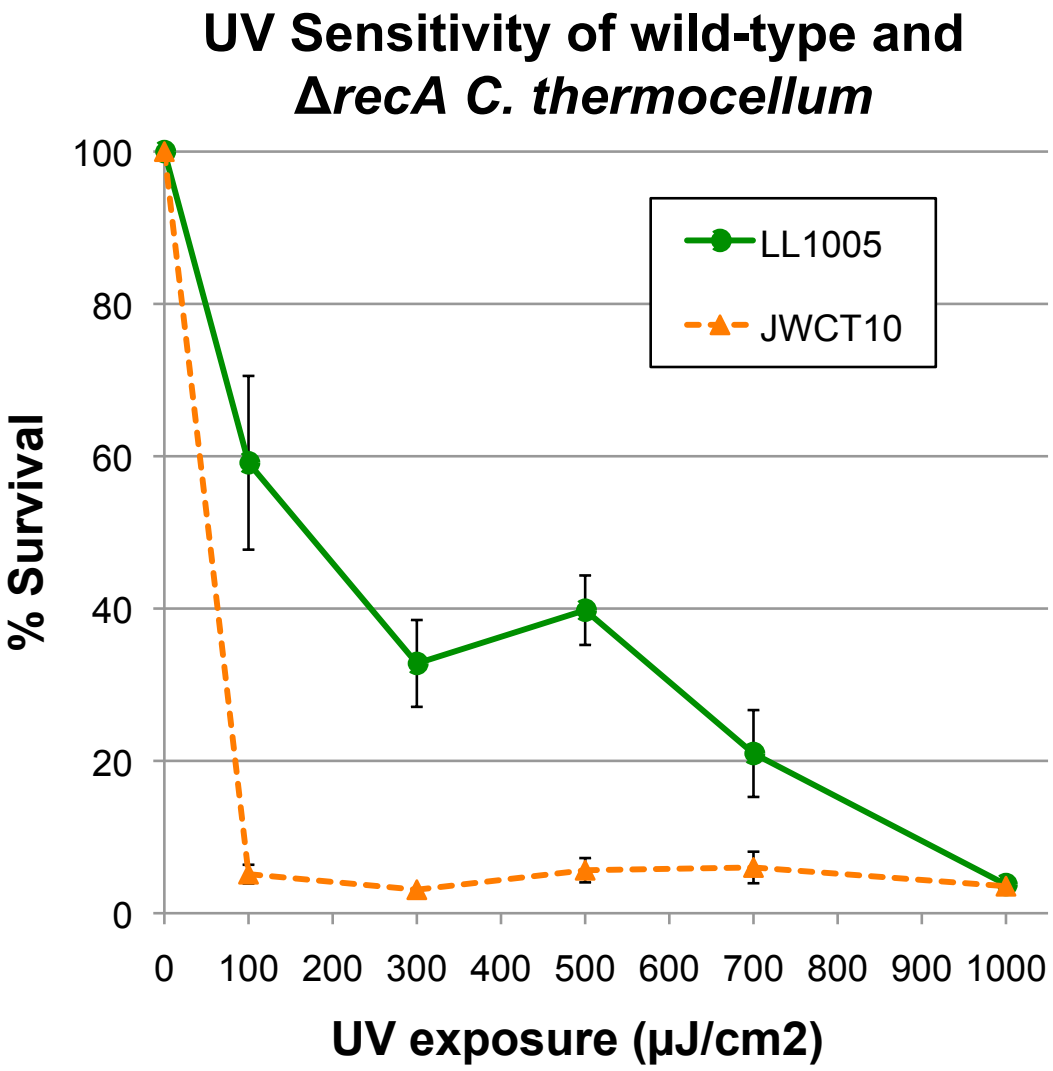


Figure 4.5. UV sensitivity of wild-type and Δ*recA C. thermocellum.* Percent survival is calculated by colony forming units from cells treated with increasing amounts of UV compared to untreated cells. Error bars show the standard deviation resulting from two biological replicates. **LL1005:** *C. thermocellum* Δ*pyrF*. **JWCT10:** *C. thermocellum* Δ*pyrF* Δ*recA::pyrF.*

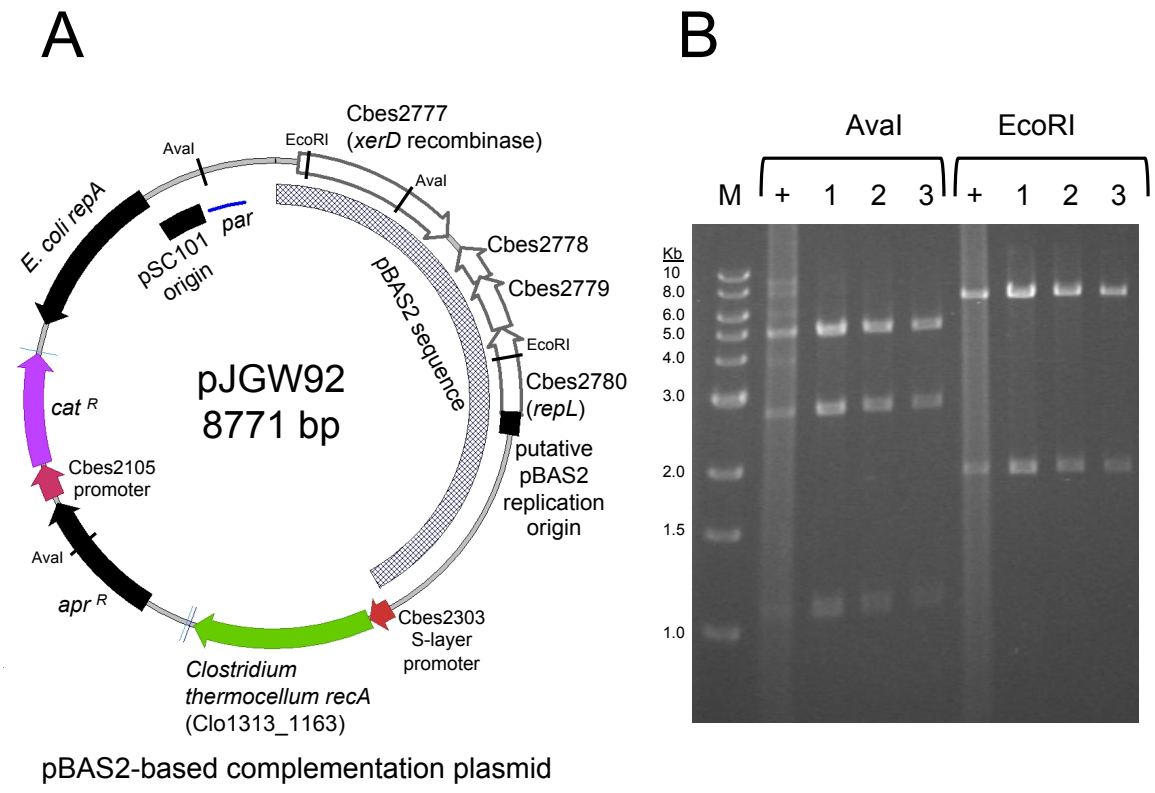


Figure 4.6. *C. thermocellum* Δ*recA* is transformable when it is complemented with a wild-type *recA* gene. A) *recA* complementation plasmid pJGW92. The hatched region was derived from *Caldicellulosiruptor bescii* native plasmid pBAS2. ***apr*^R^**, apramycin resistance casette; ***cat*^R^**, thiamphenicol resistance casette; ***repA***, replication initiaton protein for the E. coli pSC101 replication origin; ***par***, partitioning locus for E. coli. Restriction sites for structural verification are shown on the plasmid map. **B)** Restriction digests with AvaI and EcoRI. The expected bands from AvaI are 5076 bp, 2628 bp, and 1067 bp. The expected bands from EcoRI are 6908 bp and 1863 bp. **+**: purified pJGW92 from E. coli. **Lanes 1-3:** plasmids isolated from E. coli transformed with DNA isolated from JWCT26 (Δ*recA* + pJGW92).

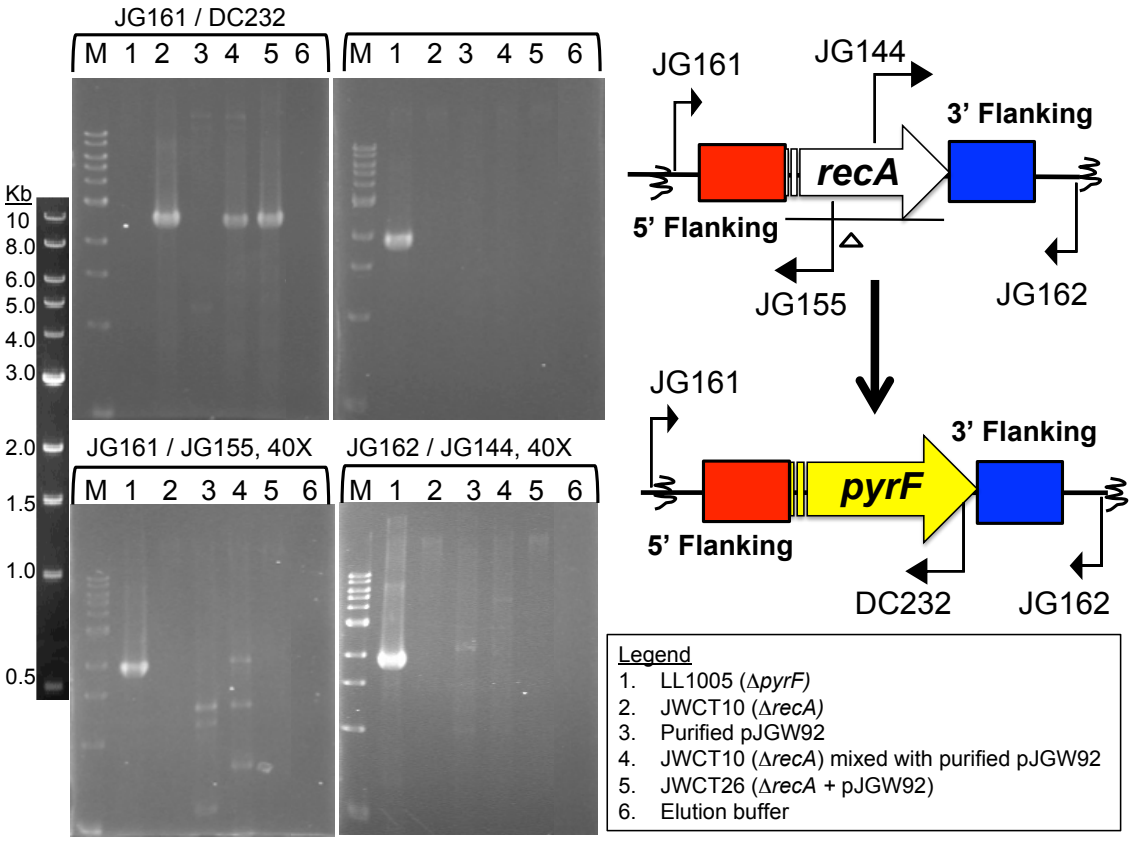


Figure 4.7. Δ*recA* strains transformed with complementation plasmids retain the chromosomal deletion of *recA*. Primers indicated in the gene diagram were used to amplify DNA extracted from *C. thermocellum* strains. JG161 and DC232 verified the replacement of *recA* by the *C. bescii pyrF* gene. Two different primer pairs (JG161/JG155 and JG162/JG144) were used to ensure that the transformed strain retained the *recA* deletion. PCR was performed for both 30 cycles (30X) and 40 cycles (40X) using primer pair JG162/JG144.

Table 4.1. Strains and plasmids used in Chapter 4

Strain or plasmid	Genotype / phenotype	Source
Clostridium thermocellum		
LL1005	*C. thermocellum* DSM 1313 Δ*pyrF* (*ura*$^-$/ 5-FOA^R)	[122]
JWCT02	*C. thermocellum* Δ*pyrF* + pDCW89 (*ura*$^+$/5-FOA^S)	[246]
JWCT10	*C. thermocellum* Δ*pyrF* Δ*recA*::$P_{Cbes2105}$-*pyrF* (*ura*$^+$/ 5-FOA^S)	This work
JWCT26	*C. thermocellum* Δ*pyrF* Δ*recA*::$P_{Cbes2105}$-*pyrF* + pJGW92 (*ura*$^+$/ 5-FOA^{S}Tm^R)	This work
E. coli		
JW421	BL21 containing pJGW37 (Apr^R)	[246]
JW442	DH5α containing pJGW57 (Apr^R)	This work
JW444	DH5α containing pJGW58 (Apr^R)	This work
JW445	DH5α containing pJGW59 (Apr^R)	This work
JW446	DH5α containing pJGW59.5 (Apr^R)	This work
JW447	DH5α containing pJGW60 (Apr^R)	This work
JW448	DH5α containing pJGW61 (Apr^R)	This work
JW449	BL21 containing pJGW61 (Apr^R)	This work
JW478	DH5α containing pJGW91 Apr^R)	This work
JW479	DH5α containing pJGW92 (Apr^R, Cm^R)	This work
JW480	BL21 containing pJGW92 (Apr^R, Cm^R)	This work
JW506	BL21 containing pCTH07	This work
Plasmids		
pDCW89	*E. coli/ C. bescii/C. thermocellum* shuttle vector with pBAS2 replicon ($P_{Cbes2105}$ – *pyrF*, Apr^R)	[177]
pDCW196	*E. coli/C. thermocellum* shuttle vector with pBAS2 replicon ($P_{Cbes2105}$ – *pyrF* / P_{gapDH} – *cat*, Apr^R, Cm^R)	[246]
pJGW37	*E. coli/C. thermocellum* shuttle vector with pBAS2 replicon ($P_{Cbes2105}$ –*cat,* Apr^R, Cm^R)	[246]
pJGW61	*recA* (Clo1313_1163) marker replacement vector with pNW33N replicon ($P_{Cbes2105}$ – *pyrF*, Apr^R)	This work
pJGW91	*E. coli/C. thermocellum* shuttle vector expressing *rec* with pBAS2 replicon ($P_{Cbes2105}$ – *pyrF,* Apr^R)	This work
pJGW92	*E. coli/C. thermocellum* shuttle vector expressing *rec* with pBAS2 replicon ($P_{Cbes2105}$ – *cat,* Apr^R, Cm^R)	This work

Table 4.2. Comparative doubling times of wild-type and Δ*recA* strains.

Medium	DSM 1313	LL1005 Δ*pyrF*	JWCT10 Δ*recA::pyrF*
CTFUD with 4.5 g/L yeast extract	1.9 ± 0.02 h	3.4 ± 0.2 h	3.1 ± 0.05 h
CTFUD without yeast extract	2.0 ± 0.04 h	2.0 ± 0.07 h	4.7 ± 0.2 h

Table 4.3. Primers used in Chapter 4.

Primer	Sequence (5'→3')
DC176	AGA GCT AGC TCC AAC GTC ATC TCG TTC TC
DC230	AGA GAC GTC TCA TCT GTG CAT ATG GAC AG
DC232	AGAGACGTCTTAAGAGATTGCTGCGTTGATA
DC576	AGA CTG CAG CTC ACC AAA CCT CCT TGT ATG AT
DC700	AGA CCT AGG CAT CAC CAT CAC CAT CAC TAA TAA T
JG024	AGA TCA TCT AGA GAC CAT CCT TTC TAT GTA GAA A
JG099	AAGA TCT AGA ATG AAC TTT AAT AAA ATT GAT TTA GAC AAT TGGAA
JG133	TCA TCA AGA TCT AGC CAA TTT AAT CCA AAT GCG TTA TCG T
JG137	AAC AAC AGA TCT TAT GGG AAA CAA AAT ATT GCG TAT GCGA
JG138	AGA AGA GGA TCC AAG TCA ATC CCG TTT GTT GAA
JG144	AAC CAA CTC AGA GAA AAG GTA GGA
JG155	CGT CGT CCT GGA TTT GTT GAT
JG161	CCA GGT GGT TTC GCT GATT
JG162	ACA CTT ATG CAG CAC AGG TA
JG163	TCT TGT GGT ACC TTC AAC AAC CAG AGA CAC TTG GGA
JG165	ACA TCT GCA TGC TTA AGA GAT TGC TGC GTT GAT A
JG166.2	TGT TGT CTG CAG TGT ATG GGC CTA AAG ACC A
JG167	TCT TCT GGA TCC TGC TGC ACC TCC TTA AAA ATC
JG168	TGT TGT GGA TCC TGG AAG AAC TTG AAA GCA GGC T
JG169	ACA ACA GCT AGC ACA GTT TGA TTA CAG TTT AGT CAG AGCT
JG170	AGA AGA GAC GTC TGC TGC ACC TCC TTA AAA ATC
JG248	AGA AGA CTGCAG ATG ATT GAG AAG AAA AAA GCT CTG GA
JG249	TGT TGT CCTAGG TTC TTC ATC TTC CAA ATC CAC TTC A
JG250	TCT TCT CAT ATG CTT CAA ACT TCC CAA AGG
JG251	AGA AGA AGA TCT ACC AGC CTA ACT TCG ATC ATT GGA
JG252	AGA AGA GGA TCC ACA GTT TGA TTA CAG TTT AGT CAG AGCT
X014	AGA CAT ATG GAC AGT TTT CCC TTT GAT ATG
X102	AGA CTG CAG ACC AGC CTA ACT TCG ATC ATT GGA
X103	AGA AGA GGATCC ACT CAC GTT AAG GGA TTT TGG TC
X118	AGA CTG CAG TGG AGT AAT GCT AAG CTC GCT G
X119	AGA AGA TCT AGC CAA TTT AAT CCA AAT GCG TTA TCGT
X120	AGA GGT ACC TCC AGA GCT TTT TTC TTC TCA ATC A
X122	AGA GCA TGC TGA AGT GGA TTT GGA AGA TGA AGA

CHAPTER 5

A NEW REPORTER GENE FOR ANAEROBIC CELLULOLYTIC THERMOPHILES: USE FOR THE ANALYSIS OF MALTOSE-REGULATED PROMOTERS[4]

[4] Groom J, Kim S, Yi X, and Westpheling J. To be submitted to *Biotechnology for Biofuels.*

I. Abstract

The availability of facile genetic tools is critical for the manipulation of all microbes but is especially important for the study of non-model systems. Reporter genes that allow visual inspection of colonies on plates are virtually nonexistent for hyperthermophilic anaerobes, limiting the ease of assaying for gene expression under a variety of genetic and environmental conditions. Here we demonstrate the use of a thermostable β-D-glucosidase from *Acidothermus cellulolyticus* in *Caldicellulosiruptor bescii* and its use as a reporter gene for the analysis of maltose regulated promoters. The *in vitro* zymogram assay demonstrates that one of the promoters responds to growth on maltose but not to cellobiose or glucose. These experiments establish precedent for further gene expression analyses in *C. bescii* and other cellulolytic thermophiles, with future applications ranging from proof of gene essentiality to controlled expression for biofuel engineering.

II. Introduction

The rapid analysis of gene expression in bacteria has relied on genes encoding enzymes that can convert substrates into colors detectable in colonies growing on agar surfaces. It would be difficult to overstate the contribution that β-galactosidase as a reporter gene has made to the study of gene regulation in *E. coli*, other bacteria, yeast, and even mammals [261]. While enzyme assays are not a direct measure of transcriptional activity, reporter genes provide a facile indication of regulated promoter activity and the regulatory factors involved.

Thermophilic β-galactosidases have been developed in the past few decades in aerobic thermophilic bacteria and archaea like *Thermus* and *Sulfolobus* spp. [262]. Fluorescent reporter genes exist that are able to fold and fluoresce properly above 70°C [263, 264], and anaerobic fluorescent reporter genes have also been engineered from blue-light photoreceptors [265]. For organisms requiring both conditions—high temperature and anaerobiosis— *in vivo* reporter genes are more difficult to find, since the enzymes they encode either cannot fold properly or cannot be detected visually in a colony without oxygen. Even if the oxygen requirement is omitted for an *in vitro* reporter gene assay, anaerobic cellulolytic thermophiles present a problem because they encode many carbohydrate-active enzymes. For instance, *Caldicellulosiruptor bescii* has 52 glycoside hydrolase enzymes that can degrade sugar-analog substrates typical of reporter genes [266], making results from *in vitro* assays with a wild-type strain uninformative because of the overwhelming background from the endogenous genes.

Still, advances in genetic tools have been made in Gram-positive, cellulolytic anaerobes. In *Caldicellulosiruptor bescii*, both replicating and integrative plasmids have been instrumental for the expression of homologous and heterologous genes from promoters identified by transcriptional profiling. These promoters have been used successfully for metabolic engineering of pathways for the production of ethanol directly from unpretreated plant biomass [117] and for analyzing and improving the cellulolytic capabilities of the organism [132, 179, 182]. Improvements to existing replicating vectors for another

cellulolytic thermophile, *Clostridium thermocellum*, have enabled enzyme expression [267] and the rapid assessment of endogenous promoters [176]. The application of the pBAS2 plasmid replication origin from *C. bescii* [177, 178] to *C. thermocellum* provides a more thermostable alternative to existing vectors [246] and has been used for the expression of a butanol dehydrogenase enzyme that achieves tolerance to common bacterial inhibitors found in plant biomass [135]. Control of gene expression is important for engineering goals, and while some exciting achievements have been made using a *glyR3*-regulated system in *C. thermocellum* [149], regulated promoters in *C. bescii* have not been identified. Of special importance in this context is the ability to demonstrate essentiality of genes involved in central metabolism using regulated expression of the genes of interest. Most of the information on pathways for energy metabolism, product synthesis and energy flux are predicted from bioinformatic analysis. An inability to generate a deletion often leads to the assumption that the gene in question is essential, but showing that viability is dependent on induction of the gene is proof of essentiality [268].

Promoters derived from *C. bescii* have been used for gene expression in their native host [117] and heterologously in *C. thermocellum* [135, 246]. More promoter options will help coordinate different gene expression programs during growth. Additionally, using heterologous promoters in *C. thermocellum* would be beneficial because it would prevent the homologous integration that has been observed at promoter sequences [129, 135]. While recent strides have been made in the application of synthetic and combinatorial promoters to bacteria

[269], a necessary first step for promoter development is to use naturally occurring promoters controlled by the presence of certain nutrients. For instance, *C. bescii* can utilize starch and its breakdown product maltose. Preliminary global gene expression data suggested that cells grown on maltose had high RNA levels corresponding to genes involved in maltodextrin transport, but low RNA levels on switchgrass, crystalline cellulose and glucose (BioEnergy Science Center, unpublished results). As maltose is an inexpensive substrate, we sought to investigate this phenomenon further in the hopes of developing a regulated promoter system using maltose, based on the native promoter regions of maltodextrin utilization genes.

Maltodextrins are storage glycans in both animals and plants. They are polymers of glucose residues linked by α-1,4 glycosidic bonds with α-1,6-linked glucose branches [270]. Animals contain the more extensively branched and compact macromolecule glycogen, while plants contain both linear and branched chains called starch [1]. Microorganisms ranging from human pathogens to plant biomass-degraders to deep-sea vent hyperthermophiles have systems to utilize the α-linked glucose polymers they find in their environments—both to degrade the polymers and to further metabolize the various breakdown products. The breakdown products include glucose, the disaccharide maltose as well as longer maltosaccharides such as maltotriose and maltotetraose (generally called maltodextrins), and the more complex pullulan that retains both α-1,4- and α-1,6-linked bonds. Some organisms are able to import maltodextrins with up to seven glucose units [271]. Diverse bacteria use similar enzymes to break down starch

(i.e. amylases, pullulanases, and cyclomaltodextrinases) [270]. They also express similar types of protein complexes to import the breakdown products. In some cases these are phosphoenolpyruvate-dependent phosphotransferase system (PTS) complexes [272], and in others they are ABC transporters [273, 274]). Although the molecular mechanisms for import are analogous, the genomic organization and gene regulation of the involved structural genes exhibit many differences. For instance, the Gram-negative enteric bacterium *E. coli* uses the same ABC transporter system to import both maltose and longer maltodextrins [274], whereas the Gram-positive pathogen *Streptococcus pneumoniae* has distinct ABC transporter systems for the import of maltose, maltotriose, and maltotetraose [275].

ABC transporters consist of several protein subunits localized to the plasma membrane of both Gram-negative and Gram-positive bacteria [276]. Although gene names between species differ for historical reasons, complexes associated with maltose and maltodextrin have the general structure $MalEFGK_2$ (Fig. 1B and 1C). MalE is an extracellular solute binding protein. MalF and MalG are both integral membrane proteins—the permease components of the complex. The two MalK subunits are ATPases that perform the ATP hydrolysis necessary for substrate import. While the *malEFG* genes are usually located near one another in the genome or even in the same operon, the *malK* ATPase component is usually separate in the genome in Gram-positives [276]. In fact, in some organisms it has been shown that one ATPase functions for multiple systems [277], or that the ATPases are interchangeable [271]. The *S.*

pneumoniae malXCD gene cluster is analogous to *malEFG* in *E. coli,* but functions specifically for the uptake of maltodextrins with higher numbers of glucose units than the disaccharide maltose [275, 278].

While the basic components of these maltose/maltodextrin ABC transporters are similar with regard to amino acid sequence and protein complex architecture, many different regulatory mechanisms have been discovered. MalT is the activator of the *mal* regulon in *E. coli* that is induced by maltotriose [279]. MalI, a LacI homolog, is a repressor in *E. coli* that controls genes involved in metabolism of internal inducers of the maltose activator MalT [280]. These regulatory functions are not apparent in most Gram-positives. However, the MalR protein has considerable sequence similarity to *E. coli* MalI [281] and is the repressor of maltose import and utilization genes in *Streptococcus, Streptomyces* and others [273, 282]. In these systems maltose binds to MalR and relieves the repression of transcription from the maltose/maltodextrin ABC transporter genes. Interestingly, a MalR homolog has been shown to function as an activator in *Lactococcus lactis*—an unusual function of LacI-like proteins [283]. In general, members of the LacI/GalR family of proteins have high sequence similarity, particularly in their helix-turn-helix (HTH) DNA-binding domains [284]. These HTH domains recognize and bind to operator sites in DNA, portions of which are well conserved between Gram-negatives and Gram-positives [284, 285]. The other domains of LacI-like proteins confer further specificity with regard to activity: they are for binding particular inducers or co-repressors, or for multimerization.

Firmicutes, low G+C Gram-positive organisms, typically contain multiple LacI-like proteins [286]. *C. bescii* is no exception with eight LacI homologs, two of which are clear MalR homologs. One such MalR homolog resides in a cluster of genes with sequence homology to members of the maltodextrin uptake locus from *S. pneumoniae*. We analyzed the DNA sequence of this region and identified conserved MalR-binding sites upstream of the start codons of several genes. Since repressor proteins bind operator sites in other LacI-like systems to control transcription, we chose two regulatory regions to characterize with regard to their control of gene expression during growth on different carbon sources. The promoter region for an ABC transporter permease was expressed highly on maltose versus other sugars, while the promoter of an extracellular solute binding protein was constitutively expressed. Overexpression of the MalR protein, rather than making regulation tighter, seemed to reduce expression even in the presence of maltose. This is the first characterization of a regulated promoter in *C. bescii* by a reporter gene, and shows promise for application to gene expression studies in *C. bescii* and other industrially useful thermophiles.

III. Materials and Methods

Bioinformatic analysis. The National Center for Biotechnology Information (NCBI) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) were used to search for homologous proteins to those encoded by the genes in the *mal* cluster of *C. bescii*. BPROM was used for bacterial promoter sequence prediction [287]. The Erpin [288] and RNAmotif [289] algorithms were used for

terminator hairpin prediction on the ARNold web server (http://rna.igmors.u-psud.fr/toolbox/arnold/index.php).

MAST software (http://meme-suite.org/tools/mast) in the MEME suite was used for motif searches in *C. bescii* genomic DNA sequences with default settings [290]. The published *Streptococcus pneumoniae* D39 MalR binding site consensus sequence 5′—CGCAAACGTTTGCR—3′ [273] was used for an initial motif search in the *C. bescii* gene clusters of interest. Then, the motifs that were identified in *C. bescii* sequences were used to create a degenerate consensus motif specific to *C. bescii*, which was iteratively used to search relevant regions of *C. bescii* genomic DNA. This approach identified more motifs missed by searching with the *S. pneumoniae* MalR binding site consensus sequence.

Bacterial media and growth conditions. *Caldicellulosiruptor bescii* was grown anaerobically in low osmolarity defined medium (LOD) and low osmolarity complex medium (LOC) as previously described [194]. *Escherichia coli* was grown in Luria-Bertani broth supplemented with 50 µg/mL apramycin when selecting for the presence of a plasmid.

Plasmid vector construction. All PCR reactions for cloning were performed with Q5 polymerase (New England Biolabs, Ipswich, MA) according to the manufacturer's instructions (98°C duplex denaturation, 60°C annealing temperature, 30 seconds per kb at 72°C for elongation). All restriction enzymes were ordered from New England Biolabs. Ligations were performed with the Fastlink Ligation Kit (Epicentre, Madison, WI). All primers are listed in Table 5.2.

Plasmids for testing the *mal* promoters are derivations of the pBAS2-based β-glucosidase expression vector pSKW18 (Kim et al., 2017, submitted). This plasmid is a replicating *E. coli* / *C. bescii* shuttle vector derived from pJGW07 [219] that contains the 1.8 kb Acel_0133 β-glucosidase gene, the regulatory region of Cbes2303 (encoding the S-layer protein), a C-terminal 6X Histidine-tag, and a Rho-independent transcriptional terminator. For pJGW76 and pJGW80, primers SK044 and JG201 were used to amplify a 9.269 kb backbone fragment from pSKW18. To generate the *malF* regulatory region for pJGW76, a 282 bp fragment was amplified from the *C. bescii* genome with primers JG204 and JG206. For the *malE* regulatory region for pJGW80, primers JG209 and JG211 amplified a 597 bp region from the *C. bescii* genome. For both pJGW76 and pJGW80, the respective fragments were digested with SphI and BglII and ligated overnight at 17°C.

pJGW84 was generated in two steps. First, the *malR* gene (Cbes2575) was placed under the control of the S-layer promoter (P_{slp}) from *C. bescii* [117]. Primers JG168 and DC576 amplified a fragment from pDCW140 [117], and primers JG202 and JG203 amplified the Cbes2575 ORF from the *C. bescii* genome. These fragments were digested with BamHI and PstI, and ligated overnight to generate pJGW74. The P_{slp}–*malR* cassette was amplified with primers JG217 and JG220, and pJGW80 was used as a template to amplify a fragment with primers DC283 and JG221. These fragments were digested with KpnI and ClaI and ligated overnight to create pJGW84.

Transformation of *C. bescii.* Transformation of *C. bescii* was performed as previously described [185, 186]. Briefly, 50 mL starter cultures of strain JWCB018 were inoculated into 500 mL of LOD supplemented with 40 μM uracil and a mixture of 19 amino acids. Cells were grown to OD_{680} ~ 0.03, washed twice with 10% (w/v) sucrose, and divided into 50 μL aliquots. Plasmid DNA (500 ng) was incubated with each aliquot for 10 minutes in pre-chilled 1 mm cuvettes. Cells were electrotransformed with a single exponential pulse with a Gene Pulser (BioRad, Hercules, CA) set at 1.8 kV, 25 μF, and 350 Ω, and placed immediately into 75°C LOC medium supplemented with 40 μM uracil for recovery. A 0.5% inoculum was transferred to LOD medium lacking uracil at 65°C every hour for three hours. Prototrophic transformants were passaged on solid LOD medium once before making frozen stocks in 10% glycerol/10% DMSO solution. Structural stability of plasmids was verified by back-transformation into *E. coli* DH5α (Figure 5.5).

Protein preparation and zymogram analysis. Protein for zymogram analysis and reporter gene assays were prepared using a procedure modified from Kim et al. (2017, submitted). To collect cell free extracts (CFE), *C. bescii* cells were grown in 50 mL of LOD medium with 40 mM MOPS in closed bottles with shaking at 150 rpm at 65°C to an OD_{680} of 0.25-0.3, and then harvested by centrifugation at 6,000 × g at 4°C for 10 min. Cells were washed once with 50 mM Tris-HCl pH 8.0, and suspended in CelLytic B cell lysis reagent (Sigma, USA). Lysis was achieved by three freeze-thaw cycles followed by sonication (twice for 15 s at 40 amps with one min rests on ice). Samples were centrifuged

for 10 minutes at 15,000 × g in a desktop centrifuge at 4°C to separate protein lysate from cell debris, and the supernatant was used as CFE. Protein concentrations were determined using the Bio-Rad protein assay kit with bovine serum albumin (BSA) as the standard. CFE (60 µg) samples were electrophoresed in 4-20% gradient Mini-Protein TGX gels (Bio-Rad) and protein bands were visualized by staining with Coomassie Brilliant Blue R-250. For the zymogram analysis, the gel was soaked for 1 h in 2.5% (v/v) Triton X-100 solution to remove the SDS and washed in distilled water. After incubating the gel at 75°C for 20 min in reaction buffer containing 5 mM 4-methylumbelliferyl β-D-glucopyranoside (Sigma), 20 mM MES (pH 5.5), 1 mM dithiothreitol (DTT), 1 mM $CaCl_2$, and 1 mM $MgCl_2$, the presence of fluorescent reaction product was visualized under UV light. The quantification of band intensity was carried out using densitometry software (ImageJ). Biological triplicates (n=3) were used for zymogram expression quantification, and equal loading of protein was verified by Coomassie gel staining. Student's T-tests were performed on the quantification results, accounting for a two-tailed distribution and heteroscedastic variance in the samples.

IV. Results

***C. bescii* contains homologs to genes involved in maltodextrin transport, utilization and regulation.** The *C. bescii* genome contains separate gene clusters responsible for the metabolism of maltose and maltooligosaccharides, as does the genome of the closely related thermophile *Caldicellulosiruptor saccharolyticus* [291]. The presence of separate systems is

consistent with gram-positive pathogen *Streptococcus pneumoniae* [275], but is different from *E. coli*, which has one system for both maltose and maltodextrin uptake [274]. The distinct *C. bescii* gene cluster encoding maltose uptake proteins has a configuration similar to that of the *mal* cluster in the gram-positive actinomycete *Streptomyces coelicolor* [292]. The *S. pneumoniae* maltodextrin gene cluster consists of three operons: *malXCD* encoding the extracellular solute binding protein (MalX) and the two permease proteins of the ABC transporter (MalC and MalD); *malMP* encoding amylomaltase (MalM) and α-glucan phosphorylase (MalP), enzymes involved in maltooligosaccharide polymerization and depolymerization; and *malAR* encoding a maltose utilization enzyme (MalA) and the transcriptional regulator of the maltodextrin regulon (MalR) [278, 285]. In both *S. pneumoniae* and *C. bescii*, the gene coding for the ATPase subunit of the ABC transporter, MalK, is located elsewhere in the genome (Figure 5.1B). This has been noted to be true of many Gram-positive bacterial species [276].

Besides this common feature, the organization of the gene cluster in *C. bescii* is quite different from *S. pneumoniae*. However, it is relatively similar to a gene cluster in the gram-positive acidophilic thermophile *Alicyclobacillus acidocaldarius* (Figure 5.1A), although the proteins encoded by the *A. acidocaldarius* cluster are involved in the uptake of both maltose and maltooligosaccharides [293]. Unlike *S. pneumoniae,* the *A. acidocaldarius* cluster has the *malR* transcriptional regulator directly downstream of the ABC transporter genes, rather than in a completely different transcriptional unit. It also encodes more enzymes for maltodextrin and maltose utilization like

cyclomaltodextrinase (CdaA), amylopullulanase (AmyA), and alpha-glucosidase (GlcA). *C. bescii* contains *cdaA* and *malP* in its maltodextrin gene cluster, and has the *malR* gene directly downstream of the permease genes and *malP* (Figure 5.1A). There are no clear *malM* or *malA* homologs in the *C. bescii* genome. The gene encoding the extracellular solute binding protein, *malE*, is downstream of all the other genes, and contains its own predicted promoter (Figure 5.2A). Other promoters with strong sequence similarity to known -10 and -35 boxes were predicted upstream of *malP, malF, and cdaA* with BPROM [287]. A predicted transcriptional terminator exists downstream of the *malE* gene in *C. bescii* (ΔG= -15.9 kcal/mol), whereas a predicted terminator is found after *glcA* in *A. acidocaldarius* [293]. We emphasize that these analyses are solely based on sequence homology and bioinformatic prediction—to date no work has been performed to determine operon structure or transcriptional start sites, or to characterize promoters in detail in *C. bescii.*

MalR binding sites are located in the maltodextrin gene cluster.

Sequence analysis of the *C. bescii* maltodextrin gene cluster revealed several candidate promoter regions for developing a regulated gene expression system. Because operator sites in DNA bind repressor proteins that conditionally inhibit transcription, the nucleotide sequence of the entire gene cluster was queried for DNA motifs similar to the consensus MalR operator site for *S. pneumoniae* [273, 285]. Potential operator sites were enriched in this gene cluster upstream of or near the ATG of the *cdaA*, *malF,* and *malE* open reading frames (Figure 5.2A). These DNA motifs were used to create a working consensus *C. bescii* MalR

operator site, from here on referred to as *malO* (Figure 5.2B). The *malO* sequences are imperfect closely spaced inverted repeats, strikingly similar to those observed upstream of the *malM* and *malX* genes in *S. pneumoniae* [285].

Although it is clear that the *C. bescii* operator sequences are very similar to those described for *S. pneumoniae*, the number of predicted *malO* sites in the regulatory regions varies from those published for the model system. In *C. bescii* each of the sequences around the ATG codons of the *cdaA*, *malF*, and *malE* ORFs contains at least two predicted *malO* sites, three in the case of *malF* (Figure 5.2A). *cdaA* and *malE* each have one *malO* site directly 3′ of the predicted -10 box and another *malO* site within the coding sequence of the gene. The *malF* gene cassette contains two *malO* sites in similar relative locations as the other genes, and an additional site 5′ of the predicted -10 and -35 boxes. There are ~200-300 bp of spacer between the *malO* sites of a specific gene. As the goal of this study was to find promoters useful for regulated gene expression, nucleotide sequences upstream of the start codon for the *malE* and *malF* genes were chosen to test for regulation using a thermophilic β-glucosidase assay.

The *malF* regulatory region responds to maltose, but not to glucose or cellobiose. In a variety of bacterial systems, the MalR protein acts as a transcriptional repressor of genes required for maltodextrin uptake and utilization. Maltose act as the inducer, preventing the MalR repressor from binding its operator sequence and thereby allowing RNA polymerase to transcribe genes [285]. To examine whether gene expression driven by the regulatory regions of *malE* and *malF* could be controlled, these DNA sequences were placed

upstream of a β-glucosidase gene from *Acidothermus cellulolyticus* [294] and transformed into *C. bescii* on a replicating shuttle vector (Figure 5.2C) [177]. The P_{malE} and P_{malF} promoters were compared to the promoter of the Cbes2303 gene, encoding an S-layer protein highly expressed in *C. bescii* during growth on a variety of carbon sources [117, 179].

The expression of the β-glucosidase protein using these promoters could be observed and quantified using a zymogram analysis, in which the entire protein lysate from a *C. bescii* sample is separated by SDS-PAGE before treating the gel with a fluorescent substrate specific for β-glucosidase enzymes. Since *C. bescii* expresses multiple endogenous β-glucosidase enzymes, there is a very high background β-glucosidase activity. Particularly, there is a highly expressed β-glucosidase enzyme that migrates at ~115 kDa on the zymograms that is present even in strains harboring empty expression vectors (Figure 5.3). The unique molecular weight of the Acel_0133 β-glucosidase makes it possible to observe the specific band of enzymatic activity encoded by the heterologous gene once the proteins are separated by electrophoresis, overcoming the issue of background noise.

All three regulatory regions tested—P_{slp}, P_{malE}, and P_{malF}—drove expression of the reporter gene when *C. bescii* was grown on maltose (Figure 5.3A and Figure 5.3B). A striking decrease in expression from P_{malF} was observed when the cells were grown on either glucose (Figure 5.3A) or cellobiose (Figure 5.3C). Interestingly, expression from P_{malE} did not decrease when cells were grown on non-maltose sugars, but rather drove expression

similarly to P_{slp} (Figure 5.3A and Figure 5.3C). Because of this, P_{malF} was chosen for further study. Enzymatic activity of the heterologous β-glucosidase was quantified in triplicate, confirming that maltose increases expression from P_{malF} by a factor of about 7- and 10-fold when compared to glucose and cellobiose, respectively (Figure 5.4A).

Glucose or cellobiose cannot repress expression from P_{malF}, yet overexpression of the MalR protein decreases expression. As it has been observed that the presence of glucose can repress the expression of maltose utilization genes in *Streptomyces coelicolor* [282, 292], we also grew *C. bescii* on combinations of maltose and other sugars to test for carbon catabolite repression. The presence of either glucose (Figure 5.3B) or cellobiose (Figure 5.3D) in addition to maltose did not lead to repression of P_{malF}. This finding is particularly useful if P_{malF} were to be used as a maltose-inducible promoter, as the presence of either glucose or cellobiose in the growth medium would not repress gene expression.

The MalR protein is in the LacI-GalR transcriptional repressor family [281]. LacI-like repressors typically bind operator sequences to prevent transcription by RNA polymerase, but when their inducer is present they cannot bind [295]. Overexpression of the maltodextrin-associated MalR in *S. pneumoniae* allowed tightly controlled expression of a green fluorescent protein-encoding reporter gene [296]. In contrast, overexpressing MalR in *C. bescii* did not result in repression. Rather, overexpression of MalR resulted in low gene expression

even in the presence of maltose, in what could be interpreted as reduced induction by maltose (Figure 5.4B).

V. Discussion

The organization of the *C. bescii* gene cluster and sequence homology with the encoded proteins is more similar to *A. acidocaldarius* than *S. pneumoniae.* Biochemical studies of the ABC transporter complex have not been performed for the *C. bescii* proteins, so it is not known whether the complex is specialized for longer maltodextrin uptake or whether, like the *A. acidocaldarius* complex [293], it imports both substrates. Since *malK* is located elsewhere in the genome, it is possible that it provides ATPase function for more than one ABC transporter. This would not be unusual, as the *Streptomyces reticuli* homolog *msiK* provides ATPase function for the import of both maltose and cellobiose [277].

Our motif searches in *C. bescii* revealed multiple *malO* sites near various start codons in the maltodextrin gene cluster, but multiple operator sites for MalR-regulated promoters were not reported for *S. pneumoniae.* The *S. pneumoniae malO* sites were determined by DNA footprinting analysis rather than bioinformatic prediction, and those determined experimentally are likely *bona fide* MalR-binding sites [285]. There are a few possible explanations for our observations in *C. bescii.* **(1)** The specific site involved in the regulation requires the palindrome 5′ GCAA-N4-TTGC 3′ found in $malO_{F1}$ and $malO_{F2}$ for the repression to be detected by our reporter assay. The $malO_{E1}$ site lacks the 5′ G nucleotide, but the site within the ORF of *malE* that was not analyzed with the

repoter gene assay, *malO*$_{E2}$, does contain the palindrome (Figure 5.2B). **(2)** Multiple *malO* operator sites are required. Multiple *lacO* sites have been shown to be involved in full repression via DNA looping and binding by a tetramer of LacI repressor proteins in *E. coli* [297, 298]. It could be that there are actually multiple *malO* sites in the intergenic region between *S. pneumoniae malM* and *malX* (Figure 5.1A) but this was not recognized before. The other sites might be important *in vivo* but are not bound as strongly, and therefore could not be identified by DNA footprinting experiments reported by Nieto *et al* (1997). Consistent with this idea, the sequence near the divergent promoter P_X in *S. pneumoniae* has a clear regulatory influence [299]. This result might be the effect of a second *malO* site located there. Our own motif searches in the *S. pneumoniae* sequence reveal a potential *malO* site near the P_X promoter oriented towards the P_M promoter in addition to the P_X *malO* site previously determined by footprinting [285]. Additionally, motif searches of the *A. acidocaldrius* gene cluster reveal three clear MalR binding sites upstream of the *malE* gene that are each separated by 200-300 bp, similar to the spacing between predicted *malO* sites in the *C. bescii* gene cluster.

The results of the β-glucosidase reporter assays (Figure 5.3) are consistent with a model in which maltose induces the *malF* gene, and the absence of maltose (i.e. the presence of either glucose or cellobiose) results in repression by the MalR protein. This is the case in *S. pneumoniae*: electrophoretic mobility shift assays demonstrated that maltose is the inducer of the *malM* and *malX* genes, preventing the MalR protein from binding the operator

site [285]. Our results from cell growth on multiple sugars did not provide evidence for catabolite repression by glucose or cellobiose over maltose (Figure 5.3B and Figure 5.3D). Carbon catabolite repression (CCR) is observed in other organisms and is usually exerted by glucose over other sugars [300]. This result is consistent with the fact that the *C. bescii* genome contains no phosphenolpyruvate-dependent phosphotransferase system for glucose, and also consistent with the lack of observed CCR in another closely related member of the *Caldicellulosiruptor* genus [105]. Mearls *et al.* (2015) obtained a similar result for a LacI-regulated promoter in *C. thermocellum* when they demonstrated a lack of cellobiose repression on the laminaribiose-induced promoter for *celC.*

The regulatory region for the *malE* gene did not exhibit the same regulatory patterns as that of *malF* (Figure 5.3A and 5.3C). The fact that P_{malE} directed similar expression on cellobiose and maltose is supported by extracellular proteomic data showing that the MalE protein encoded by ORF Cbes2574 is highly expressed and secreted during logarithmic growth on crystalline cellulose [301]. This could indicate that the MalE protein is involved with uptake of other substrates including cellobiose or longer chain β-glucans, or that a basal level of MalE is always available for uptake of maltodextrins.

Rather than making regulation tighter, *malR* overexpression appeared to abolish induction by maltose from P_{malF} (Figure 5.4B). It has been shown that overexpression of the *S. pneumoniae malR* gene from a *tet* promoter on a plasmid achieved tight regulation of a reporter gene [296]. However, expressing *malR* from a multicopy plasmid did reduce the level of depression of the *malM*

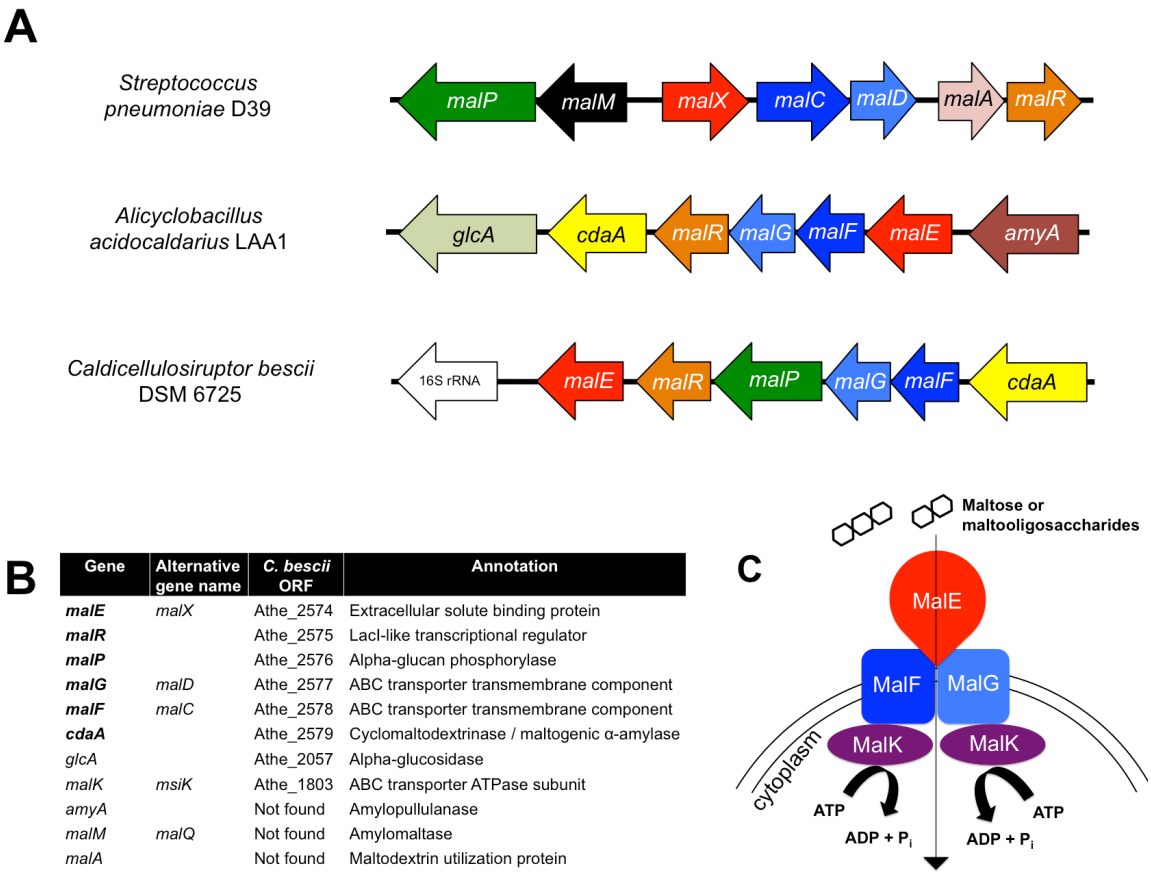
gene in the presence of maltose [281], which is more consistent with our results**.** Overexpression of the maltose-specific *malR* in *Streptomyces coelicolor* actually prevents the utilization of maltose entirely [282]. This was not expected to occur with the *C. bescii malR*, because the *malR* gene (ORF Cbes2575) is located in the presumed maltodextrin import/utilization cluster rather than the maltose-specific cluster. What we observe for *C. bescii malR* overexpression could result from the behavior of LacI regulatory proteins. The *E. coli* LacI inducer, allolactose, reduces the affinity of the LacI repressor for its specific operator sequence [302]. An inducer-bound repressor protein retains nearly identical affinity for non-operator genomic DNA [303]. In fact, LacI rarely dissociates from DNA *in vivo*, spending ~90% of its time bound to and diffusing along DNA strands non-specifically [304]. Thus, if MalR in *C. bescii* behaved similarly to *E. coli* LacI, expressing MalR with the powerful S-layer promoter would make more repressors available to bind operator sequences that could prevent transcription even in the presence of maltose. Alternatives for future use of this system would be to express MalR with a weaker promoter, the native sequence upstream of the *malR* gene that contains potential -10 and -35 boxes, or even the regulated P_{malF} sequence, similar to what was done with the *glyR3* regulator in *Clostridium thermocellum* [149].

VI. Future Directions

The MalR-regulated gene expression system will be tested in *C. thermocellum.* Another regulated promoter system would be very useful for this promising cellulolytic organism, especially a system that is induced by a

substrate like maltose—an inexpensive sugar that the organism cannot utilize as a sole carbon source.

This system will be developed further in *C. bescii* by deleting two annotated β-glucosidase genes that are likely responsible for the high molecular weight bands visibile on the zymogram. One is *bglA* (Cbes0458), encoding a family 1 glycosyl hydrolase, and the other is *bglX* (Cbes2354), encoding a family 3 glycosyl hydrolase. If successful, this would allow a high throughput *in vitro* assay rather than a zymogram.



Gene	Alternative gene name	*C. bescii* ORF	Annotation
malE	*malX*	Athe_2574	Extracellular solute binding protein
malR		Athe_2575	LacI-like transcriptional regulator
malP		Athe_2576	Alpha-glucan phosphorylase
malG	*malD*	Athe_2577	ABC transporter transmembrane component
malF	*malC*	Athe_2578	ABC transporter transmembrane component
cdaA		Athe_2579	Cyclomaltodextrinase / maltogenic α-amylase
glcA		Athe_2057	Alpha-glucosidase
malK	*msiK*	Athe_1803	ABC transporter ATPase subunit
amyA		Not found	Amylopullulanase
malM	*malQ*	Not found	Amylomaltase
malA		Not found	Maltodextrin utilization protein

Figure 5.1. A gene cluster in *Caldicellulosiruptor bescii* encodes an ABC transporter and enzymes for maltodextrin metabolism. (A) Genomic regions of maltodextrin ABC transporters in *Streptococcus pneumoniae*, *Alicyclobacillus acidocaldarius*, and *Caldicellulosiruptor bescii* with indicated open reading frames. **(B)** Gene table containing assigned gene names for *C. bescii* with published alternative gene names in other organisms where needed. Gene annotations come from Nieto *et al*. (1997), Hülsmann *et al.* (2000), and BLAST results for *C. bescii* proteins. **(C)** General model for the ABC transporter complex $MalEFGK_2$. Protein names correspond to the gene names and annotations in **(B).**

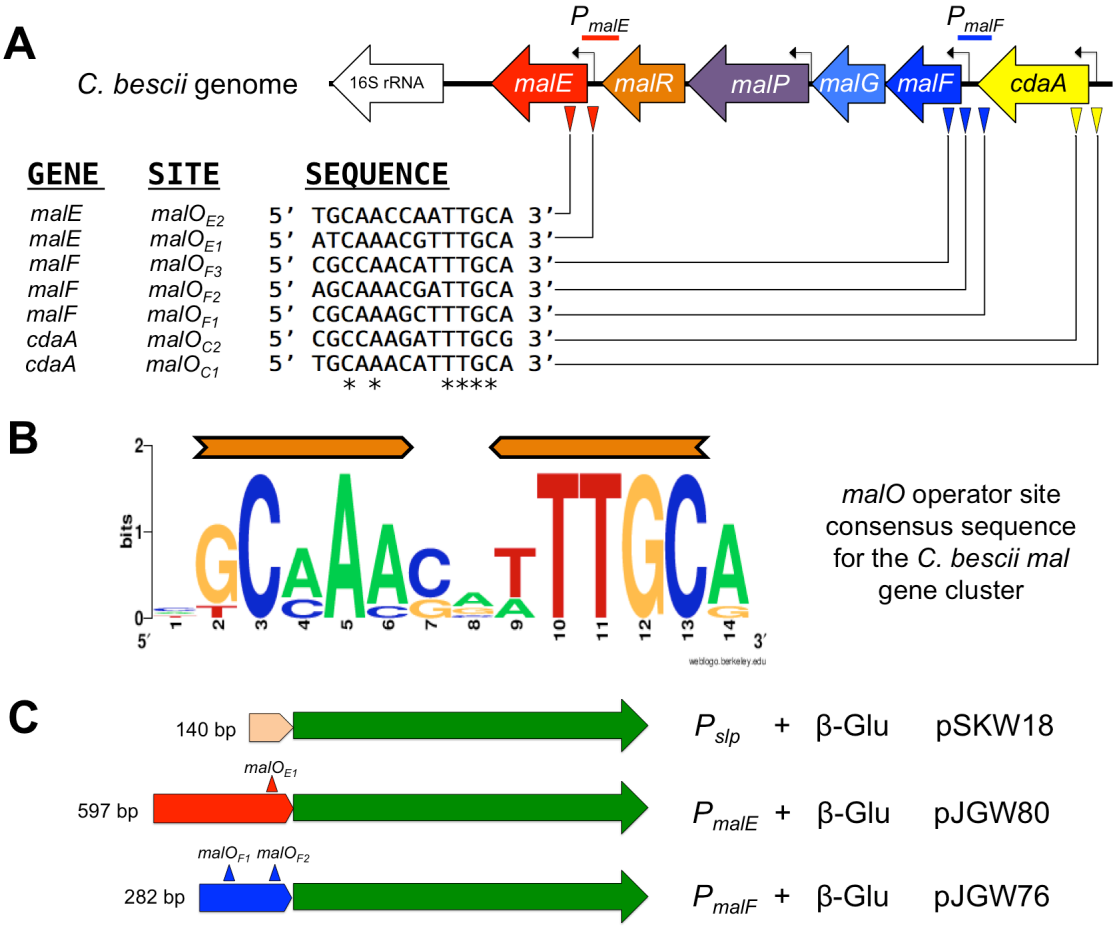


Figure 5.2. MalR-binding operator sites are found in the *C. bescii* maltodextrin utilization gene cluster. (A) *C. bescii* maltodextrin utilization gene cluster with predicted MalR-binding sites, labeled as *malO*. Distinct sites are named with an abbreviated gene name in the subscript as well a number representing the 5' to 3' orientation of the site. Conserved nucleotides are indicated by an asterick (*). Chosen promoter regions are indicated by colored lines above the gene diagram. **(B)** The *malO* consensus site determined for the *C. bescii* maltodextrin gene cluster. Imperfect inverted repeats are indicated with orange arrows. **(C)** Map of constructs used for reporter gene assays, with predicted *malO* operator sites indicated. The promoter names and plasmid names are listed next to their respective maps.

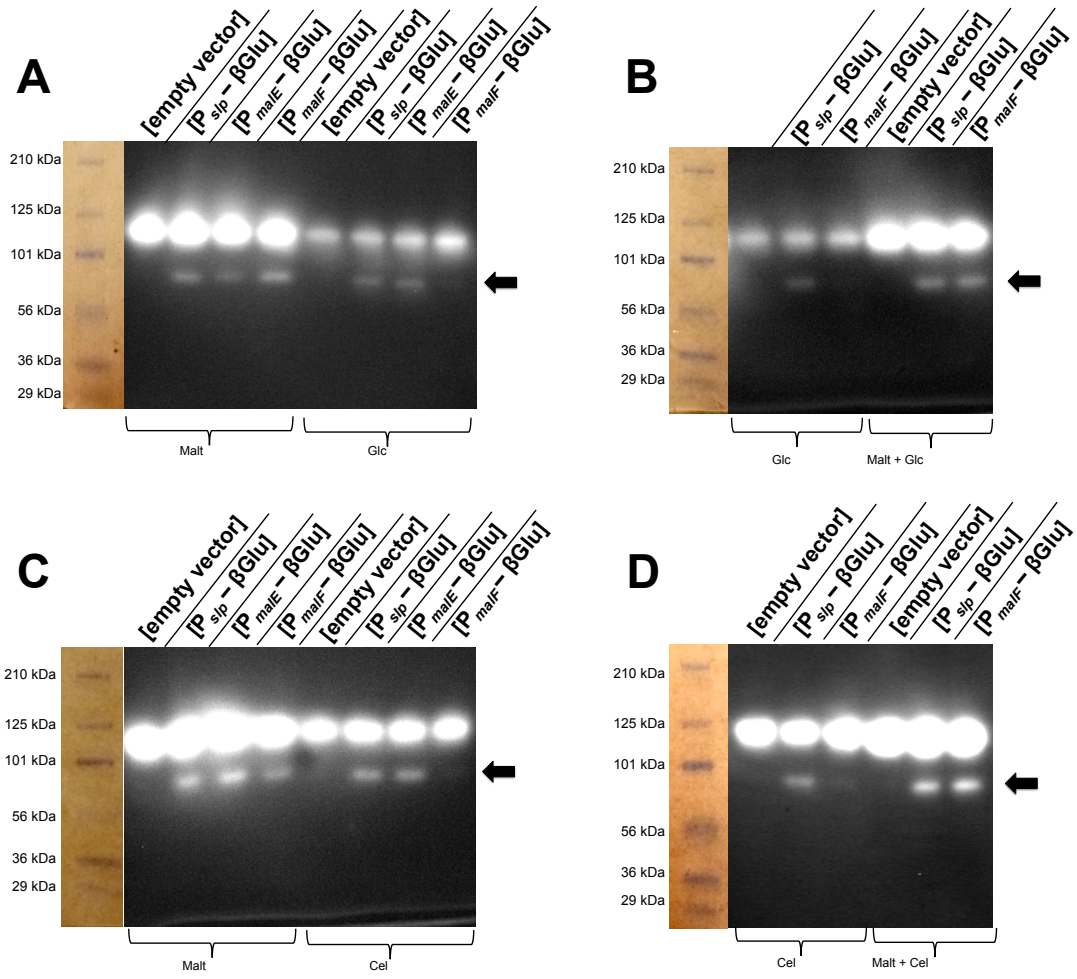


Figure 5.3. Reporter gene assays for chosen promoters. Zymogram expression results for P_{slp}, P_{malE}, and P_{malF} in response to growth on various sugars. The arrow in each panel indicates the expected molecular weight for the heterologous β-glucosidase encoded by *A. cellulolyticus* ORF Acel_0133. **(A)** Cells were grown on either 0.5% (w/v) maltose or 0.5% (w/v) glucose. **(B)** Cells were grown on either 0.5% (w/v) glucose or 0.25% (w/v) glucose + 0.25% (w/v) maltose. **(C)** Cells were grown on either 0.5% (w/v) maltose or 0.5% (w/v) cellobiose. **(D)** Cells were grown on either 0.5% (w/v) glucose or 0.25% (w/v) glucose + 0.25% (w/v) maltose.

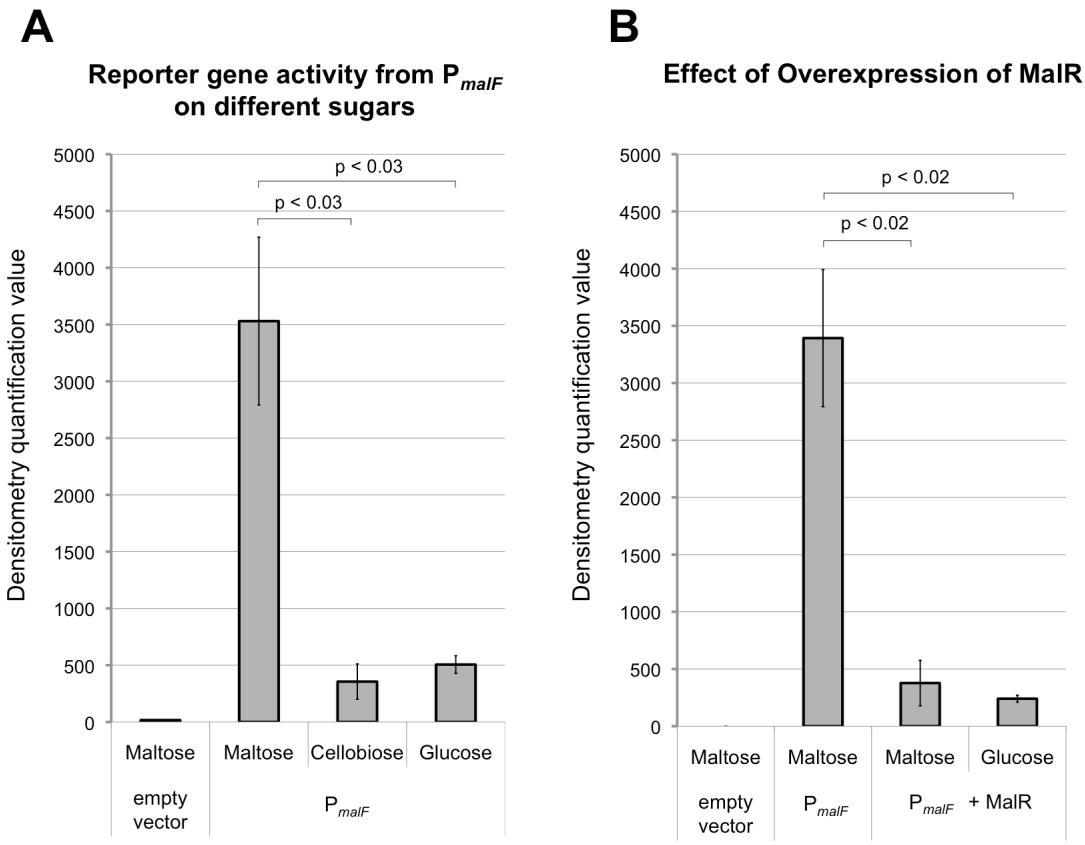


Figure 5.4. Quantitative results from reporter gene zymograms. Units of expression are determined by densitometry quantification using ImageJ. Each panel represents values measured from a single zymogram. Each sample besides the empty vector negative control was performed in biological triplicate. **A)** Expression of the β-glucosidase reporter gene from P_{malF} in cells grown on either 0.5% (w/v) maltose, cellobiose, or glucose. **B)** Expression of the β-glucosidase reporter gene from P_{malF} with or without an overexpression cassette for *malR* (ORF Athe_2575) on the same plasmid. Cells were grown on either 0.5% (w/v) maltose or 0.5% (w/v) glucose.

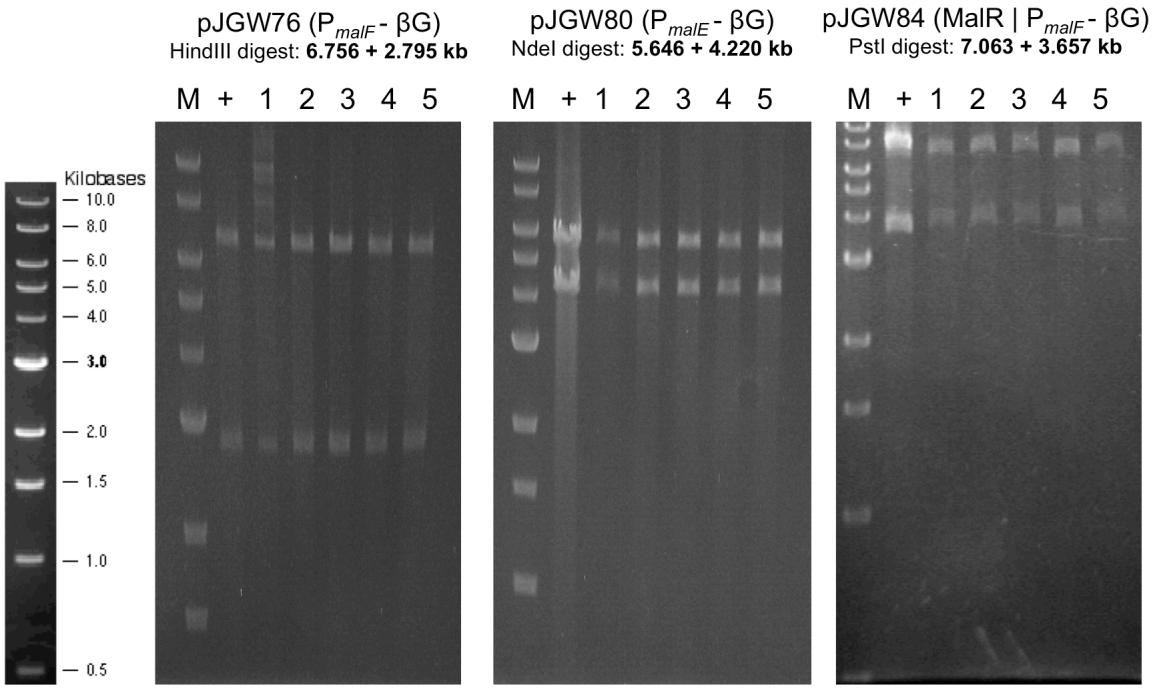


Figure 5.5. Back-transformation of plasmids from *C. bescii* into *E. coli*.

+: pure plasmid before transformation into *C. bescii*, isolated from *E. coli* DH5α.

1-5: individual *E. coli* back-transformant colonies obtained by electroporation of *E. coli* with total DNA isolated from *C. bescii*.

Table 5.1. Strains and plasmids used in Chapter 5.

Strain or plasmid	Genotype / phenotype	Source
Caldicellulosiruptor bescii		
JWCB018	*C. bescii* DSM 6725 Δ*pyrFA* Δ*cbeI ldh::*ISCbe4	[185]
JWCB082	JWCB018 + pJGW07 (*ura*$^{+}$/5-FOA^{S})	Kim et al., submitted
JWCB134	JWCB018 + pSKW18 (*ura*$^{+}$/5-FOA^{S})	This work
JWCB138	JWCB018 + pJGW80 (*ura*$^{+}$/5-FOA^{S})	This work
JWCB139	JWCB018 + pJGW76 (*ura*$^{+}$/5-FOA^{S})	This work
JWCB141	JWCB018 + pJGW84 (*ura*$^{+}$/5-FOA^{S})	This work
Escherichia coli		
JW292	DH5α containing pDCW89 (Apr^{R})	[177]
JW313	DH5α containing pDCW140 (Apr^{R})	[117]
JW402	DH5α containing pJGW07 (Apr^{R})	[219]
JW461	DH5α containing pJGW74 (Apr^{R})	This work
JW464	DH5α containing pJGW76 (Apr^{R})	This work
JW467	DH5α containing pJGW80 (Apr^{R})	This work
JW469	DH5α containing pJGW84 (Apr^{R})	This work
JW527	DH5α containing pSKW18 (Apr^{R})	Kim et al., submitted
Plasmids		
pDCW89	*E. coli/C. bescii* shuttle vector ($P_{Cbes2105}$ – *C. bescii pyrF*, Apr^{R})	[177]
pJGW07	*E. coli/C. bescii* shuttle vector ($P_{Cbes2105}$ – *C. thermocellum pyrF*, Apr^{R})	[219]
pSKW18	pJGW07 with P_{slp} — β-Glu	Kim et al., submitted
pDCW140	*C. bescii* integrating vector ($P_{Cbes2105}$ – *C. bescii pyrF*, Apr^{R})	[117]
pJGW74	pSC101-based cloning vector with P_{slp} — MalR (Apr$^{R)}$	This work
pJGW76	pJGW07 with P_{malF} — β-Glu	This work
pJGW80	pJGW07 with P_{malE} — β-Glu	This work
pJGW84	pJGW07 with P_{malF} — β-Glu and P_{slp} — MalR	This work

Table 5.2. Primers used in Chapter 5.

Primer	Sequence (5'→3')
DC283	AGA GGTACC ACC GTG AGC ATT CTG GAC AGGT
DC576	AGA CTG CAG CTC ACC AAA CCT CCT TGT ATG AT
JG168	TGT TGT GGA TCC TGG AAG AAC TTG AAA GCA GGC T
JG201	ACA ACA AGA TCT ATG ACA CAA ATC GAA GAG CGC
JG202	TCT TCT GGA TCC TTA TAA CGT TTG AGC TTT TCC GCA
JG203	ACA ACA CTG CAG ATG GCA ACA ATA AAA GAA ATA GCT AAA GA
JG204	TCT TCT GCA TGC CTG TGT ACA TTC TTT TTA ATC CAC G
JG206	AGT AGT AGA TCT TTT CCT CCT TTG AAA TAG AAT GTG CA
JG209	TCT TCT GCA TGC GAC CCA GTG ATT TGG TAG TAA
JG211	ACA ACA AGA TCT AAC AAC CTC CCT TAA GTT TAT TTA ATA TGC
JG217	AGA AGA GGT ACC ACA GTT TGA TTA CAG TTT AGT CAG AGC T
JG220	TGT TGT ATC GAT TTA TAA CGT TTG AGC TTT TCC GCA
JG221	AGA AGA ATC GAT ACC AGC CTA ACT TCG ATC ATT G
SK044	AGA GCA TGC ATT CCC ATG AGC CCA CGA ACA GT
SK051	AGA GGG CCC ATG ACA CAA ATC GAA GAG CGC

CONCLUSION

Though there has been documented success with genetic tool development in several thermophiles, certain areas require more attention. In this work, we first wanted to establish whether genetic methodologies from *C. bescii* could be applied other organisms in the *Caldicellulosiruptor* genus. We demonstrated that restriction-modification systems affected DNA transformation, but that they were more important in *C. bescii* than in *C. hydrothermalis*. Therefore, the restriction barrier to transformation must be empirically investigated when developing genetic systems in new organisms,.

The thermophilic replicon pBAS2 would be useful if it could be applied to other organisms, particularly cellulolytic thermophiles. However, it was unclear how stable and useful the *C. bescii* pBAS2 replicon would be in other organisms, both within and outside of the *Caldicellulosiruptor* genus. We showed stable plasmid replication in two thermophiles—one in the same genus, and one in a completely different genus but still a member of the *Clostridia*. It was an exciting result that vectors based on this plasmid were maintained at multiple copy in other organisms, as this facilitates high-level gene expression. It is also vital that the plasmid was structurally stable in both organisms using the tested vector configurations.

Even though replicative stability was observed, it is clear from our work and the work of others that *Clostridium thermocellum* recombines plasmids into

its chromosome, and requires only short regions of homology. Could this be overcome, and would altering recombination increase plasmid stability? Our attempts at eliminating recombination were not successful because of the particular replication mechanism of pBAS2, but there is hope that this strain might prove useful with different or new thermophilic replicons. Nevertheless, we have still learned something about thermophilic plasmid replication that can be applied to future work.

The study and development of all of these methods and molecular phenomena—transformation, plasmid replication, DNA restriction, DNA recombination—are interesting from a basic biology standpoint. Still, manipulative tools for stable, thermophilic gene expression is the end goal. Constitutive promoters have been applied for homologous expression in *C. bescii,* but the development of a reporter gene system to harness native dynamic promoters was an important milestone in our engineering of this organism. The ability to dissect the responses of *C. bescii* to different carbon sources and different growth conditions will inform us about plant biomass utilization in this organism and likely other organisms. The promoters will also be useful for future engineering work in *C. bescii*, and possibly *C. thermocellum* and *Thermoanaerobacter* strains.

These results as a whole are steps forward in the three organisms described, but also for engineering in cellulolytic thermophiles on the whole, and for the development of genetic methodologies for bacteria with industrially promising metabolisms.

REFERENCES

1. Voet, D. and J.G. Voet. Chapter 11: Sugars and Polysaccharides, in *Biochemistry (4th ed.)*. 2011, John Wiley & Sons Inc.: Hoboken, NJ. 359-385.

2. Albersheim, P., A. Darvill, K. Roberts, R. Sederoff, and A. Staehelin. *Plant Cell Walls*. 2011, New York: Garland Sci.

3. Forster, P., V. Ramaswamy, P. Artaxo, T. Berntsen, R. Betts, D.W. Fahey, J. Haywood, J. Lean, D.C. Lowe, G. Myhre, J. Nganga, R. Prinn, G. Raga, M. Schulz, and R. Van Dorlan, *Changes in Atmospheric Constituents and in Radiative Forcing*, in *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, S. Solomon, et al., Editors. 2007, Cambridge University Press: Cambridge, United Kingdon and New York, NY, USA. p. 129-234.

4. Cherubini, F. The biorefinery concept: Using biomass instead of oil for producing energy and chemicals*. Energy Conversion and Management*, 2010. **51**(7): 1412-1421.

5. Fulton, L.M., L.R. Lynd, A. Körner, and N. Greene. The need for biofuels as part of a low carbon energy future*. Biofuels, Bioproducts & Biorefining*, 2015. **9**: 476-483.

6. Lynd, L.R., M. Sow, A. Chimphango, L. Cortez, C.H. Cruz, M. Elmissiry, M. Laser, I.A. Mayaki, M. Moraes, L. Nogueira, G.M. Wolfaardt, J. Woods, and W.H. van Zyl. Bioenergy and African transformation*. Biotechnology for Biofuels*, 2015. **8**: 18.

7. Jordan, N., G. Boody, W. Broussard, J.D. Glover, D. Keeney, B.H. McCown, G. McIsaac, M. Muller, H. Murray, J. Neal, C. Pansing, R.E. Turner, K. Warner, and D. Wyse. Environment. Sustainable development of the agricultural bio-economy*. Science*, 2007. **316**(5831): 1570-1571.

8. Wang, M., J. Han, J.B. Dunn, H. Cai, and A. Elgowainy. Well-to-wheels energy use and greenhouse gas emissions of ethanol from corn, sugarcane and cellulosic biomass for US use. *Environmental Research Letters*, 2012. **7**(4): 045905.

9. RFA, *Ethanol Industry Outlook*. 2017, Renewable Fuels Association: Washington, DC. USA.

10. Orts, W.J. and C.M. McMahan. Biorefinery developments for advanced biofuels from a sustainable array of biomass feedstocks: Survey of recent biomass conversion research from Agricultural Research Service. *BioEnergy Research*, 2016. **9**(2): 430-446.

11. Lynd, L.R., X. Liang, M.J. Biddy, A. Allee, H. Cai, T. Foust, M.E. Himmel, M.S. Laser, M. Wang, and C.E. Wyman. Cellulosic ethanol: status and innovation. *Current Opinion in Biotechnology*, 2017. **45**: 202-211.

12. Jordan, D.B., M.J. Bowman, J.D. Braker, B.S. Dien, R.E. Hector, C.C. Lee, J.A. Mertens, and K. Wagschal. Plant cell walls to ethanol. *Biochemical Journal*, 2012. **442**(2): 241-252.

13. Himmel, M.E., S.Y. Ding, D.K. Johnson, W.S. Adney, M.R. Nimlos, J.W. Brady, and T.D. Foust. Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science*, 2007. **315**(5813): 804-807.

14. Lange, J.P. Lignocellulose conversion: an introduction to chemistry, process and economics. *Biofuels, Bioproducts & Biorefining*, 2007. **1**: 39-48.

15. Dunn, J.B., S. Mueller, H. Kwon, and M.Q. Wang. Land-use change and greenhouse gas emissions from corn and cellulosic ethanol. *Biotechnology for Biofuels*, 2013. **6**: 51.

16. Davis, S.C., W.J. Parton, S.J. Grosso, C. Keough, E. Marx, P.R. Adler, and E.H. DeLucia. Impact of second-generation biofuel agriculture on greenhouse-gas emissions in the corn-growing regions of the US. *Frontiers in Ecology and the Environment*, 2012. **10**(2): 69-74.

17. Qin, Z., Q. Zhuang, and M. Chen. Impacts of land use change due to biofuel crops on carbon balance, bioenergy production, and agricultural

yield, in the conterminous United States. *GCB Bioenergy*, 2012. **4**(3): 277-288.

18. Kantar, M.B., C.E. Tyl, K.M. Dorn, X. Zhang, J.M. Jungers, J.M. Kaser, R.R. Schendel, J.O. Eckberg, B.C. Runck, M. Bunzel, N.R. Jordan, R.M. Stupar, D.M. Marks, J.A. Anderson, G.A. Johnson, C.C. Sheaffer, T.C. Schoenfuss, B. Ismail, G.E. Heimpel, and D.L. Wyse. Perennial grain and oilseed crops. *Annual Review of Plant Biology*, 2016. **67**(1): 703-729.

19. Werling, B.P., T.L. Dickson, R. Isaacs, H. Gaines, C. Gratton, K.L. Gross, H. Liere, C.M. Malmstrom, T.D. Meehan, L. Ruan, B.A. Robertson, G.P. Robertson, T.M. Schmidt, A.C. Schrotenboer, T.K. Teal, J.K. Wilson, and D.A. Landis. Perennial grasslands enhance biodiversity and multiple ecosystem services in bioenergy landscapes. *Proceedings of the National Academy of Sciences USA*, 2014. **111**(4): 1652-1657.

20. Brown, R.A., N.J. Rosenberg, C.J. Hays, W.E. Easterling, and L.O. Mearns. Potential production and environmental effects of switchgrass and traditional crops under current and greenhouse-altered climate in the central United States: a simulation study. *Agriculture, Ecosystems & Environment*, 2000. **78**(1): 31-47.

21. Roddy, D.J. Biomass in a petrochemical world. *Interface Focus*, 2013. **3**: 20120038.

22. Akhtar, J., A. Idris, and R. Abd. Aziz. Recent advances in production of succinic acid from lignocellulosic biomass. *Applied Microbiology and Biotechnology*, 2013. **98**(3): 987-1000.

23. Bozell, J.J., L. Moens, D.C. Elliott, Y. Wang, G.G. Neuenscwander, S.W. Fitzpatrick, R.J. Bilski, and J.L. Jarnefeld. Production of levulinic acid and use as a platform chemical for derived products. *Resources*, 2000. **28**: 227-239.

24. Song, Y., Q. Nguyen, S. Wi, J. Yang, and H.-J. Bae. Strategy for dual production of bioethanol and d-psicose as value-added products from cruciferous vegetable residue. *Bioresource Technology*, 2017. **223**: 34-39.

25. *DuPont Cellulosic Ethanol Corn Stover Harvest Program*. 2017 [cited 2017 2 June].

26. Clomburg, J.M., A.M. Crumbley, and R. Gonzalez. Industrial biomanufacturing: The future of chemical production. *Science*, 2017. **355**: 38.

27. Klein-Marcuschamer, D., P. Oleskowicz-Popiel, B.A. Simmons, and H.W. Blanch. The challenge of enzyme cost in the production of lignocellulosic biofuels. *Biotechnology and Bioengineering*, 2012. **109**(4): 1083-1087.

28. Lynd, L.R., M.S. Laser, D. Bransby, B.E. Dale, B. Davison, R. Hamilton, M. Himmel, M. Keller, J.D. McMillan, J. Sheehan, and C.E. Wyman. How biotech can transform biofuels. *Nature Biotechnology*, 2008. **26**(2): 169-172.

29. Kroon-Batenburg, L.M. and J. Kroon. The crystal and molecular structures of cellulose I and II. *Glycoconjugate Journal*, 1997. **14**(5): 677-690.

30. Atalla, R.S., M.F. Crowley, M.E. Himmel, and R.H. Atalla. Irreversible transformations of native celluloses, upon exposure to elevated temperatures. *Carbohydrate Polymers*, 2014. **100**: 2-8.

31. Sawatdeenarunat, C., D. Nguyen, K.C. Surendra, S. Shrestha, K. Rajendran, H. Oechsner, L. Xie, and S. Khanal. Anaerobic biorefinery: Current status, challenges, and opportunities. *Bioresource Technology*, 2016. **215**: 304-313.

32. Mohnen, D., M. Bar-Peled, and C. Somerville. Cell wall polysaccharide synthesis, in *Biomass Recalcitrance: Deconstructing the Plant Cell Wall for Bioenergy*, M.E. Himmel, Editor. 2008, Blackwell Publishing: Oxford. 94-187.

33. Hatfield, R.D., D.M. Rancour, and J.M. Marita. Grass cell walls: A story of cross-linking. *Frontiers in Plant Science*, 2017. **7**: 2056.

34. Hatfield, R.D., J. Ralph, and J.H. Grabber. Cell wall cross-linking by ferulates and diferulates in grasses. *Journal of the Science of Food and Agriculture*, 1999. **79**: 403-407.

35. Iiyama, K., T.B.T. Lam, and B.A. Stone. Phenolic acid bridges between polysaccharides and lignin in wheat internodes. *Phytochemistry*, 1990. **29**: 733-737.

36. Shevchenko, S.M. and G.W. Bailey. The mystery of the lignin-carbohydrate complex: a computational approach. *Journal of Molecular Structure (Theochem)*, 1996. **364**: 197-208.

37. Chundawat, S.P., G.T. Beckham, M.E. Himmel, and B.E. Dale. Deconstruction of lignocellulosic biomass to fuels and chemicals. *Annual Review of Chemical and Biomolecular Engineering*, 2011. **2**: 121-145.

38. Pingali, S.V., V.S. Urban, W.T. Heller, J. McGaughey, H. O'Neill, M. Foston, D.A. Myles, A. Ragauskas, and B.R. Evans. Breakdown of cell wall nanostructure in dilute acid pretreated biomass. *Biomacromolecules*, 2010. **11**(9): 2329-2335.

39. Ragauskas, A.J., G.T. Beckham, M.J. Biddy, R. Chandra, F. Chen, M.F. Davis, B.H. Davison, R.A. Dixon, P. Gilna, M. Keller, P. Langan, A.K. Naskar, J.N. Saddler, T.J. Tschaplinski, G.A. Tuskan, and C.E. Wyman. Lignin valorization: improving lignin processing in the biorefinery. *Science*, 2014. **344**(6185): 1246843.

40. Frei, M. Lignin: characterization of a multifaceted crop component. *The Scientific World Journal*, 2013. **2013**: 436517.

41. Holladay, J.E., J.F. White, J.J. Bozell, and D. Johnson. Top value added chemicals from biomass-Volume II, Results of screening for potential candidates from biorefinery lignin. *Report No. PNNL-16983*, 2007. **Evaluation II**: 87.

42. Colquhoun, I.J., M.-C. Ralet, J.-F. Thibault, C.B. Faulds, and G. Williamson. Structure identification of feruloylated oligosaccharides from sugar-beet pulp by NMR spectroscopy. *Carbohydrate Research*, 1994. **263**(243-256).

43. Fry, S.C. Feruloylated pectins from the primary cell wall: their structure and possible functions. *Planta*, 1983. **147**: 111-123.

44. Ridley, B.L., M.A. O'Neill, and D. Mohnen. Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. *Phytochemistry*, 2001. **57**(6): 929-967.

45. Mohnen, D. Pectin structure and biosynthesis. *Current Opinion in Plant Biology*, 2008. **11**(3): 266-277.

46. Caffall, K. and D. Mohnen. The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydrate Research*, 2009. **344**(14): 1879-1900.

47. Tan, L., S. Eberhard, S. Pattathil, C. Warder, J. Glushka, C. Yuan, Z. Hao, X. Zhu, U. Avci, J.S. Miller, D. Baldwin, C. Pham, R. Orlando, A. Darvill, M.G. Hahn, M.J. Kieliszewski, and D. Mohnen. An Arabidopsis cell wall proteoglycan consists of pectin and arabinoxylan covalently linked to an arabinogalactan protein. *Plant Cell*, 2013. **25**(1): 270-287.

48. Atmodjo, M.A., Z. Hao, and D. Mohnen. Evolving views of pectin biosynthesis. *Annual Review of Plant Biology*, 2013. **64**: 747-779.

49. Biswal, A.K., Z. Hao, S. Pattathil, X. Yang, K. Winkeler, C. Collins, S.S. Mohanty, E.A. Richardson, I. Gelineo-Albersheim, K. Hunt, D. Ryno, R.W. Sykes, G.B. Turner, A. Ziebell, E. Gjersing, W. Lukowitz, M.F. Davis, S.R. Decker, M.G. Hahn, and D. Mohnen. Downregulation of GAUT12 in Populus deltoides by RNA silencing results in reduced recalcitrance, increased growth and reduced xylan and pectin in a woody biofuel feedstock. *Biotechnology for Biofuels*, 2015. **8**: 41.

50. Wang, T., Y. Park, D.J. Cosgrove, and M. Hong. Cellulose-pectin spatial contacts are inherent to never-dried Arabidopsis primary cell walls: Evidence from solid-state nuclear magnetic resonance. *Plant Physiology*, 2015. **168**(3): 871-884.

51. Blumer-Schuette, S.E., S.D. Brown, K.B. Sander, E.A. Bayer, I. Kataeva, J.V. Zurawski, J.M. Conway, M.W.W. Adams, and R.M. Kelly. Thermophilic lignocellulose deconstruction. *FEMS Microbiology Reviews*, 2014. **38**(3): 393-448.

52. Tan, L., S. Eberhard, S. Pattathil, C. Warder, J. Glushka, C. Yuan, Z. Hao, X. Zhu, U. Avci, J.S. Miller, D. Baldwin, C. Pham, R. Orlando, A. Darvill, M.G. Hahn, M.J. Kieliszewski, and D. Mohnen. An Arabidopsis cell wall proteoglycan consists of pectin and arabinoxylan covalently linked to an arabinogalactan protein. *The Plant Cell*, 2013. **25**(1): 270-287.

53. Willis, J.D., J.A. Smith, M. Mazarei, J.-Y.Y. Zhang, G.B. Turner, S.R. Decker, R.W. Sykes, C.R. Poovaiah, H.L. Baxter, D.G. Mann, M.F. Davis, M.K. Udvardi, M.J. Peña, J. Backe, M. Bar-Peled, and C.N. Stewart Jr. Downregulation of a UDP-arabinomutase gene in switchgrass (*Panicum virgatum* L.) results in increased cell wall lignin while reducing arabinose-glycans. *Frontiers in Plant Science*, 2016. **7**: 1580.

54. Kumar, R., M. Tabatabaei, K. Karimi, and I. Horváth. Recent updates on lignocellulosic biomass derived ethanol - A review. *Biofuel Research Journal*, 2016. **3**(1): 347-356.

55. Ding, S.Y., Y.S. Liu, Y. Zeng, M.E. Himmel, J.O. Baker, and E.A. Bayer. How does plant cell wall nanoscale architecture correlate with enzymatic digestibility? *Science*, 2012. **338**(6110): 1055-1060.

56. Studer, M.H., J.D. DeMartini, M.F. Davis, R.W. Sykes, B.H. Davison, M. Keller, G.A. Tuskan, and C.E. Wyman. Lignin content in natural Pupulas variants affects sugar release. *Proceedings of the National Academy of Sciences USA*, 2011. **108**(15): 6300-6305.

57. Davison, B.H., S.R. Drescher, G.A. Tuskan, M.F. Davis, and N.P. Nghiem. Variation of S/G ratio and lignin content in a Pupulus family influences the release of xylose by dilute acid hydrolysis. *Applied Biochemistry and Biotechnology*, 2006. **130**: 427-435.

58. Chen, F. and R.A. Dixon. Lignin modification improves fermentable sugar yields for biofuel production. *Nature Biotechnology*, 2007. **25**(7): 759-761.

59. Fu, C., J.R. Mielenz, X. Xiao, Y. Ge, C.Y. Hamilton, M. Rodriguez, Jr., F. Chen, M. Foston, A.J. Ragauskas, J. Bouton, R.A. Dixon, and Z.Y. Wang. Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. *Proceedings of the National Academy of Sciences USA*, 2011. **108**(9): 3803-3808.

60. Baxter, H.L., M. Mazarei, N. Labbe, L.M. Kline, Q. Cheng, M.T. Windham, D.G.J. Mann, C. Fu, A. Ziebell, R.W. Sykes, M. Rodriguez, M.F. Davis, J.R. Mielenz, R.A. Dixon, Z.-Y. Wang, and N.C. Stewart Jr. Two-year field analysis of reduced recalcitrance transgenic switchgrass. *Plant Biotechnology Journal*, 2014. **12**(7): 914-924.

61. Eudes, A., A. George, P. Mukerjee, J.S. Kim, B. Pollet, P.I. Benke, F. Yang, P. Mitra, L. Sun, O.P.P. Cetinkol, S. Chabout, G. Mouille, L. Soubigou-Taconnat, S. Balzergue, S. Singh, B.M. Holmes, A. Mukhopadhyay, J.D. Keasling, B.A. Simmons, C. Lapierre, J. Ralph, and D. Loqué. Biosynthesis and incorporation of side-chain-truncated lignin monomers to reduce lignin polymerization and enhance saccharification. *Plant Biotechnology Journal*, 2012. **10**(5): 609-620.

62. Wilkerson, C.G., S.D. Mansfield, F. Lu, S. Withers, J.-Y. Park, S.D. Karlen, E. Gonzales-Vigil, D. Padmakshan, F. Unda, J. Rencoret, and J. Ralph. Monolignol ferulate transferase introduces chemically labile linkages into the lignin backbone. *Science*, 2014. **344**(6179): 90-93.

63. Urbanowicz, B.R., M.J. Peña, S. Ratnaparkhe, U. Avci, J. Backe, H.F. Steet, M. Foston, H. Li, M.A. O'Neill, A.J. Ragauskas, A.G. Darvill, C. Wyman, H.J. Gilbert, and W.S. York. 4-O-methylation of glucuronic acid in Arabidopsis glucuronoxylan is catalyzed by a domain of unknown function family 579 protein. *Proceedings of the National Academy of Sciences USA*, 2012. **109**(35): 14253-14258.

64. Yennamalli, R.M., A.J. Rader, A.J. Kenny, J.D. Wolt, and T.Z. Sen. Endoglucanases: insights into thermostability for biofuel applications. *Biotechnology for Biofuels*, 2013. **6**(1): 136.

65. Brunecky, R., M.J. Selig, T.B. Vinzant, M.E. Himmel, D. Lee, M.J. Blaylock, and S.R. Decker. In planta expression of *A. cellulolyticus* Cel5A endocellulase reduces cell wall recalcitrance in tobacco and maize. *Biotechnology for Biofuels*, 2011. **4**: 1.

66. Teymouri, F., H. Alizadeh, L. Laureano-Perez, B. Dale, and M. Sticklen. Effects of ammonia fiber explosion treatment on activity of endoglucanase from *Acidothermus cellulolyticus* in transgenic plant. *Applied Biochemistry and Biotechnology*, 2004. **113-116**: 1183-1191.

67. McKendry, P. Energy production from biomass (part 2): conversion technologies. *Bioresource Technology*, 2002. **83**(1): 47-54.

68. Li, M., N. Luo, and Y. Lu. Biomass Energy Technological Paradigm (BETP): Trends in this sector. *Sustainability*, 2017. **9**(4): 567.

69. Hendriks, A.T.W.M. and G. Zeeman. Pretreatments to enhance the digestibility of lignocellulosic biomass. *Bioresource Technology*, 2009. **100**(1): 10-18.

70. Alvira, P., E. Tomás-Pejó, M. Ballesteros, and M.J. Negro. Pretreatment technologies for an efficient bioethanol production process based on enzymatic hydrolysis: a review. *Bioresource Technology*, 2010. **101**: 4851-4861.

71. Kumar, P., D.M. Barrett, M.J. Delwiche, and P. Stroeve. Methods for pretreatment of lignocellulosic biomass for efficient hydrolysis and biofuel production. *Industrial & Engineering Chemistry Research*, 2009. **48**(8): 3713-3729.

72. Chundawat, S.P.S., B.S. Donohoe, L. da Costa Sousa, T. Elder, U.P. Agarwal, F. Lu, J. Ralph, M.E. Himmel, V. Balan, and B.E. Dale. Multi-scale visualization and characterization of lignocellulosic plant cell wall deconstruction during thermochemical pretreatment. *Energy & Environmental Science*, 2011. **4**: 973.

73. Alonso, D., S.H. Hakim, S. Zhou, W. Won, O. Hosseinaei, J. Tao, V. Garcia-Negron, A. Motagamwala, M.A. Mellmer, K. Huang, C.J. Houtman, N. Labbé, D.P. Harper, C. Maravelias, T. Runge, and J.A. Dumesic. Increasing the revenue from lignocellulosic biomass: Maximizing feedstock utilization. *Science Advances*, 2017. **3**(5): e1603301.

74. Cai, C.M., T. Zhang, R. Kumar, and C.E. Wyman. THF co-solvent enhances hydrocarbon fuel precursor yields from lignocellulosic biomass. *Green Chemistry*, 2013. **15**(11): 3140.

75. Nguyen, T.Y., C.M. Cai, R. Kumar, and C.E. Wyman. Co-solvent pretreatment reduces costly enzyme requirements for high sugar and ethanol yields from lignocellulosic biomass. *ChemSusChem*, 2015. **8**(10): 1716-1725.

76. Sant'Ana da Silva, A., L.R. Vasconcelos de Sa, E.C.G. Aguieiras, M. Fernandes de Souza, R.S.S. Teixeira, M.C. Cammarota, E.P.S. Bon, D.M.G. Freire, and V.S. Ferreira-Leitao. Productive chain of biofuels and industrial biocatalysis: Two important opportunities for Brazilian sustainable development, in *Biotechnology of Microbial Enzymes:*

Production, Biocatalysis and Industrial Applications, G. Brahmachari, Editor. 2017, Elsevier: London. 545-572.

77. Horn, S.J., G. Vaaje-Kolstad, B. Westereng, and V.G.H. Eijsink. Novel enzymes for the degradation of cellulose. *Biotechnology for Biofuels*, 2012. **5**(1): 45.

78. Cannella, D. and H. Jørgensen. Do new cellulolytic enzyme preparations affect the industrial strategies for high solids lignocellulosic ethanol production? *Biotechnology and Bioengineering*, 2014. **111**(1): 59-68.

79. Wilson, D.B. Cellulases and biofuels. *Current Opinion in Biotechnology*, 2009. **20**(3): 295-299.

80. Nguyen, T., C.M. Cai, O. Osman, R. Kumar, and C.E. Wyman. CELF pretreatment of corn stover boosts ethanol titers and yields from high solids SSF with low enzyme loadings. *Green Chemistry*, 2015. **18**(6): 1581-1589.

81. Kang, K., D.-P. Chung, Y. Kim, B.-W. Chung, and G.-W. Choi. High-titer ethanol production from simultaneous saccharification and fermentation using a continuous feeding system. *Fuel*, 2015. **145**: 18-24.

82. Yang, S., Q. Fei, Y. Zhang, L.M. Contreras, S.M. Utturkar, S.D. Brown, M.E. Himmel, and M. Zhang. *Zymomonas mobilis* as a model system for production of biofuels and biochemicals. *Microbial Biotechnology*, 2016. **9**(6): 699-717.

83. Chundawat, S.P., R. Vismeh, L.N. Sharma, J.F. Humpula, L. da Costa Sousa, C.K. Chambliss, A.D. Jones, V. Balan, and B.E. Dale. Multifaceted characterization of cell wall decomposition products formed during ammonia fiber expansion (AFEX) and dilute acid based pretreatments. *Bioresource Technology*, 2010. **101**(21): 8429-8438.

84. Allen, S.A., W. Clark, J.M. McCafery, Z. Cai, A. Lanctot, P.J. Slininger, Z.L. Liu, and S.W. Gorsich. Furfural induces reactive oxygen species accumulation and cellular damage in *Saccharomyces cerevisiae.* *Biotechnology for Biofuels*, 2010. **3**: 2.

85. Modig, T., G. Liden, and M.J. Taherzadeh. Inhibition effects of furfural on alcohol dehydrogenase, aldehyde dehydrogenase and pyruvate dehydrogenase. *Biochemical Journal*, 2002. **363**: 769-776.

86. Qing, Q., B. Yang, and C.E. Wyman. Xylooligomers are strong inhibitors of cellulose hydrolysis by enzymes. *Bioresource Technology*, 2010. **101**(24): 9624-9630.

87. Ximenes, E., Y. Kim, N. Mosier, and B. Dien. Inhibition of cellulases by phenols. *Enzyme and Microbial Technology*, 2010. **46**: 170-176.

88. Brunecky, R., M. Alahuhta, Q. Xu, B.S. Donohoe, M.F. Crowley, I.A. Kataeva, S.-J. Yang, M.G. Resch, M.W.W. Adams, V.V. Lunin, M.E. Himmel, and Y.J. Bomble. Revealing nature's cellulase diversity: The digestion mechanism of *Caldicellulosiruptor bescii* CelA. *Science*, 2013. **342**(6165): 1513-1516.

89. Yarbrough, J.M., R. Zhang, A. Mittal, T. Wall, Y.J. Bomble, S.R. Decker, M.E. Himmel, and P.N. Ciesielski. Multi-functional cellulolytic enzymes outperform processive fungal cellulases for co-production of nanocellulose and biofuels. *ACS Nano*, 2017. **11**: 3101-3109.

90. Yang, S.J., I. Kataeva, S.D. Hamilton-Brehm, N.L. Engle, T.J. Tschaplinski, C. Doeppke, M. Davis, J. Westpheling, and M.W. Adams. Efficient degradation of lignocellulosic plant biomass, without pretreatment, by the thermophilic anaerobe "*Anaerocellum thermophilum*" DSM 6725. *Applied and Environmental Microbiology*, 2009. **75**(14): 4762-4769.

91. Yamada, R., T. Hasunuma, and A. Kondo. Endowing non-cellulolytic microorganisms with cellulolytic activity aiming for consolidated bioprocessing. *Biotechnology Advances*, 2013. **31**(6): 754-763.

92. Yang, P., H. Zhang, and S. Jiang. Construction of recombinant *sestc Saccharomyces cerevisiae* for consolidated bioprocessing, cellulase characterization, and ethanol production by in situ fermentation. *3 Biotech*, 2016. **6**: 192.

93. Fan, L.-H., Z.-J. Zhang, S. Mei, Y.-Y. Lu, M. Li, Z.-Y. Wang, J.-G. Yang, S.-T. Yang, and T.-W. Tan. Engineering yeast with bifunctional

minicellulosome and cellodextrin pathway for co-utilization of cellulose-mixed sugars. *Biotechnology for Biofuels*, 2016. **9**: 137.

94. Ryu, S. and M.N. Karim. A whole cell biocatalyst for cellulosic ethanol production from dilute acid-pretreated corn stover hydrolyzates. *Applied Microbiology and Biotechnology*, 2011. **91**: 529-542.

95. Vishnivetskaya, T.A., S.D. Hamilton-Brehm, M. Podar, J.J. Mosher, A.V. Palumbo, T.J. Phelps, M. Keller, and J.G. Elkins. Community analysis of plant biomass-degrading microorganisms from Obsidian Pool, Yellowstone National Park. *Microbial Ecology*, 2014.

96. Egorova, K. and G. Antranikian. Industrial relevance of thermophilic Archaea. *Current opinion in Microbiology*, 2005. **8**(6): 649-655.

97. Skinner, K.A. and T.D. Leathers. Bacterial contaminants of fuel ethanol production. *Journal of Industrial Microbiology and Biotechnology*, 2004. **31**(9): 401-408.

98. Demain, A.L., M. Newcomb, and D.J.H. Wu. Cellulase, clostridia, and ethanol. *Microbiology and Molecular Biology Reviews*, 2005. **69**(1): 124-154.

99. Alfani, F., A. Gallifuoco, A. Saporosi, A. Spera, and M. Cantarella. Comparison of SHF and SSF processes for the bioconversion of steam-exploded wheat straw. *Journal of Industrial Microbiology and Biotechnology*, 2000. **25**(4): 184-192.

100. Blumer-Schuette, S.E., R.J. Giannone, J.V. Zurawski, I. Ozdemir, Q. Ma, Y. Yin, Y. Xu, I. Kataeva, F.L. Poole, 2nd, M.W. Adams, S.D. Hamilton-Brehm, J.G. Elkins, F.W. Larimer, M.L. Land, L.J. Hauser, R.W. Cottingham, R.L. Hettich, and R.M. Kelly. Caldicellulosiruptor core and pangenomes reveal determinants for noncellulosomal thermophilic deconstruction of plant biomass. *J Bacteriol*, 2012. **194**(15): 4015-4028.

101. Basen, M., A.M. Rhaesa, I. Kataeva, C.J. Prybol, I.M. Scott, F.L. Poole, and M.W. Adams. Degradation of high loads of crystalline cellulose and of unpretreated plant biomass by the thermophilic bacterium *Caldicellulosiruptor bescii. Bioresour Technol*, 2014. **152**: 384-392.

102. Blumer-Schuette, S.E., D.L. Lewis, and R.M. Kelly. Phylogenetic, microbiological, and glycoside hydrolase diversities within the extremely thermophilic, plant biomass-degrading genus Caldicellulosiruptor. *Appl Environ Microbiol*, 2010. **76**(24): 8084-8092.

103. Cantarel, B.L., P.M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, and B. Henrissat. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res*, 2009. **37**(Database issue): D233-238.

104. Dam, P., I. Kataeva, S.J. Yang, F. Zhou, Y. Yin, W. Chou, F.L. Poole, 2nd, J. Westpheling, R. Hettich, R. Giannone, D.L. Lewis, R. Kelly, H.J. Gilbert, B. Henrissat, Y. Xu, and M.W. Adams. Insights into plant biomass conversion from the genome of the anaerobic thermophilic bacterium *Caldicellulosiruptor bescii* DSM 6725. *Nucleic Acids Res*, 2011. **39**(8): 3240-3254.

105. Vanfossen, A.L., M.R. Verhaart, S.M. Kengen, and R.M. Kelly. Carbohydrate utilization patterns for the extremely thermophilic bacterium Caldicellulosiruptor saccharolyticus reveal broad growth substrate preferences. *Applied and Environmental Microbiology*, 2009. **75**(24): 7718-7724.

106. Zurawski, J., S. Blumer-Schuette, J. Conway, and R. Kelly. The extremely thermophilic genus *Caldicellulosiruptor*: physiological and genomic characteristics for complex carbohydrate conversion to molecular hydrogen., in *Microbial BioEnergy: Hydrogen Production, Advances in Photosynthesis and Respiration*, D.Z.a.R.D. Philippis, Editor. 2014. 177-195.

107. Rainey, F.A., A.M. Donnison, P.H. Janssen, D. Saul, A. Rodrigo, P.L. Bergquist, R.M. Daniel, E. Stackebrandt, and H.W. Morgan. Description of *Caldicellulosiruptor saccharolyticus* gen. nov., sp. nov: An obligately anaerobic, extremely thermophilic, cellulolytic bacterium. *FEMS Microbiology Letters*, 1994. **120**(3): 263-266.

108. Yang, S.J., I. Kataeva, J. Wiegel, Y. Yin, P. Dam, Y. Xu, J. Westpheling, and M.W. Adams. Classification of 'A*naerocellum thermophilum*' strain DSM 6725 as *Caldicellulosiruptor bescii* sp. nov. *International Journal of Systematic and Evolutionary Microbiology*, 2010. **60**(9): 2011-2015.

109. McBee, R.H. The Characteristics of *Clostridium thermocellum. Journal of Bacteriology*, 1954. **67**(4): 505-506.

110. Lynd, L.R., P.J. Weimer, W.H. van Zyl, and I.S. Pretorius. Microbial cellulose utilization: Fundamentals and biotechnology. *Microbiology and Molecular Biology Reviews*, 2002. **66**(3): 506-577.

111. Paye, J.M.D., A. Guseva, S.K. Hammer, E. Gjersing, M.F. Davis, B.H. Davison, J. Olstad, B.S. Donohoe, T. Nguyen, C.E. Wyman, S. Pattathil, M.G. Hahn, and L.R. Lynd. Biological lignocellulose solubilization: comparative evaluation of biocatalysts and enhancement via cotreatment. *Biotechnology for Biofuels*, 2016. **9**: 8.

112. Argyros, D.A., S.A. Tripathi, T.F. Barrett, S.R. Rogers, L.F. Feinberg, D.G. Olson, J.M. Foden, B.B. Miller, L.R. Lynd, D.A. Hogsett, and N.C. Caiazza. High ethanol titers from cellulose by using metabolically engineered thermophilic, anaerobic microbes. *Applied and Environmental Microbiology*, 2011. **77**(23): 8288-8294.

113. Lamed, R. and E.A. Bayer. The Cellulosome of *Clostridium thermocellum*, in *Advances in Applied Microbiology*, I.L. Allen, Editor. 1988, Academic Press. 1-46.

114. Bayer, E.A., Y. Shoham, and R. Lamed. Lignocellulose-Decomposing Bacteria and Their Enzyme Systems, in *The Prokaryotes - Prokaryotic Physiology and Biochemistry*, E. Rosenberg, Editor. 2013, Springer-Verlag: Berlin Heidelberg. 215-266.

115. Xu, Q., M.G. Resch, K. Podkaminer, S. Yang, J.O. Baker, B.S. Donohoe, C. Wilson, D.M. Klingeman, D.G. Olson, S.R. Decker, R.J. Giannone, R.L. Hettich, S.D. Brown, L.R. Lynd, E.A. Bayer, M.E. Himmel, and Y.J. Bomble. Dramatic performance of *Clostridium thermocellum* explained by its wide range of cellulase modalities. *Science Advances*, 2016. **2**: e1501254.

116. Shaw, J.A., F.E. Jenney, M. Adams, and L.R. Lynd. End-product pathways in the xylose fermenting bacterium, *Thermoanaerobacterium saccharolyticum. Enzyme and Microbial Technology*, 2008. **42**(6): 453-458.

117. Chung, D., M. Cha, A.M. Guss, and J. Westpheling. Direct conversion of plant biomass to ethanol by engineered *Caldicellulosiruptor bescii. Proceedings of the National Academy of Sciences of the United States of America*, 2014. **111**(24): 8931-8936.

118. Herrero, A.A. and R.F. Gomez. Development of ethanol tolerance in *Clostridium thermocellum*: effect of growth temperature*. Applied and Environmental Microbiology*, 1980. **40**: 571-577.

119. Kristensen, J.B., C. Felby, and H. Jørgensen. Yield-determining factors in high-solids enzymatic hydrolysis of lignocellulose*. Biotechnology for Biofuels*, 2009. **2**: 11.

120. Shaw, A.J., K.K. Podkaminer, S.G. Desai, J.S. Bardsley, S.R. Rogers, P.G. Thorne, D.A. Hogsett, and L.R. Lynd. Metabolic engineering of a thermophilic bacterium to produce ethanol at high yield*. Proceedings of the National Academy of Sciences USA*, 2008. **105**(37): 13769-13774.

121. Herring, C.D., W.R. Kenealy, J.A. Shaw, S.F. Covalla, D.G. Olson, J. Zhang, R.W. Sillers, V. Tsakraklides, J.S. Bardsley, S.R. Rogers, P.G. Thorne, J.P. Johnson, A. Foster, I.D. Shikhare, D.M. Klingeman, S.D. Brown, B.H. Davison, L.R. Lynd, and D.A. Hogsett. Strain and bioprocess improvement of a thermophilic anaerobe for the production of ethanol from wood*. Biotechnology for Biofuels*, 2016. **9**: 125.

122. Tripathi, S.A., D.G. Olson, D.A. Argyros, B.B. Miller, T.F. Barrett, D.M. Murphy, J.D. McCool, A.K. Warner, V.B. Rajgarhia, L.R. Lynd, D.A. Hogsett, and N.C. Caiazza. Development of *pyrF*-based genetic system for targeted gene deletion in *Clostridium thermocellum* and creation of a *pta* mutant*. Applied and Environmental Microbiology*, 2010. **76**(19): 6591-6599.

123. Rydzak, T., L.R. Lynd, and A.M. Guss. Elimination of formate production in *Clostridium thermocellum. Journal of Industrial Microbiology & Biotechnology*, 2015. **42**: 1263-1272.

124. Biswas, R., S. Prabhu, L.R. Lynd, and A.M. Guss. Increase in ethanol yield via elimination of lactate production in an ethanol-tolerant mutant of *Clostridium thermocellum. PLoS ONE*, 2014. **9**(2): e86389.

125. van der Veen, D., J. Lo, S.D. Brown, C.M. Johnson, T.J. Tschaplinski, M. Martin, N.L. Engle, R.A. van den Berg, A.D. Argyros, N.C. Caiazza, A.M. Guss, and L.R. Lynd. Characterization of *Clostridium thermocellum* strains with disrupted fermentation end-product pathways*. Journal of Industrial Microbiology & Biotechnology*, 2013. **40**(7): 725-734.

126. Biswas, R., T. Zheng, D.G. Olson, L.R. Lynd, and A.M. Guss. Elimination of hydrogenase active site assembly blocks H2 production and increases ethanol yield in *Clostridium thermocellum. Biotechnology for Biofuels*, 2015. **8**: 20.

127. Papanek, B., R. Biswas, T. Rydzak, and A.M. Guss. Elimination of metabolic pathways to all traditional fermentation products increases ethanol yields in *Clostridium thermocellum. Metabolic Engineering*, 2015. **32**: 49-54.

128. Tian, L., B. Papanek, D.G. Olson, T. Rydzak, E.K. Holwerda, T. Zheng, J. Zhou, M. Maloney, N. Jiang, R.J. Giannone, R.L. Hettich, A.M. Guss, and L.R. Lynd. Simultaneous achievement of high ethanol yield and titer in Clostridium thermocellum*. Biotechnology for Biofuels*, 2016. **9**: 116.

129. Lin, P.P., L. Mi, A.H. Morioka, K.M. Yoshino, S. Konishi, S.C. Xu, B.A. Papanek, L.A. Riley, A.M. Guss, and J.C. Liao. Consolidated bioprocessing of cellulose to isobutanol using *Clostridium thermocellum. Metabolic Engineering*, 2015. **31**: 44-52.

130. Cha, M., D. Chung, J.G. Elkins, A.M. Guss, and J. Westpheling. Metabolic engineering of *Caldicellulosiruptor bescii* yields increased hydrogen production from lignocellulosic biomass*. Biotechnology for Biofuels*, 2013. **6**: 85.

131. Chung, D., M. Cha, E.N. Snyder, J.G. Elkins, A.M. Guss, and J. Westpheling. Cellulosic ethanol production via consolidated bioprocessing at 75 °C by engineered *Caldicellulosiruptor bescii. Biotechnology for Biofuels*, 2015. **8**: 163.

132. Kim, S.-K., D. Chung, M.E. Himmel, Y.J. Bomble, and J. Westpheling. Heterologous expression of family 10 xylanases from *Acidothermus cellulolyticus* enhances the exoproteome of *Caldicellulosiruptor bescii* and growth on xylan substrates*. Biotechnology for Biofuels*, 2016. **9**: 176.

133. Conway, J.M., W.S. Pierce, J.H. Le, G.W. Harper, J.H. Wright, A.L. Tucker, J.V. Zurawski, L.L. Lee, S.E. Blumer-Schuette, and R.M. Kelly. Multidomain, surface layer-associated glycoside hydrolases contribute to Plant polysaccharide degradation by *Caldicellulosiruptor* species. *Journal of Biological Chemistry*, 2016. **291**(13): 6732-6747.

134. Chung, D., T.J. Verbeke, K.L. Cross, J. Westpheling, and J.G. Elkins. Expression of a heat-stable NADPH-dependent alcohol dehydrogenase in *Caldicellulosiruptor bescii* results in furan aldehyde detoxification. *Biotechnology for Biofuels*, 2015. **8**(1): 1-11.

135. Kim, S.-K., J. Groom, D. Chung, J. Elkins, and J. Westpheling. Expression of a heat-stable NADPH-dependent alcohol dehydrogenase from *Thermoanaerobacter pseudethanolicus* 39E in *Clostridium thermocellum* 1313 results in increased hydroxymethylfurfural resistance. *Biotechnology for Biofuels*, 2017. **10**(1): 66.

136. Young, J., D. Chung, Y.J. Bomble, M.E. Himmel, and J. Westpheling. Deletion of *Caldicellulosiruptor bescii* CelA reveals its crucial role in the deconstruction of lignocellulosic biomass. *Biotechnology for Biofuels*, 2014. **7**: 142.

137. Chung, D., S. Pattathil, A.K. Biswal, M.G. Hahn, D. Mohnen, and J. Westpheling. Deletion of a gene cluster encoding pectin degrading enzymes in Caldicellulosiruptor bescii reveals an important role for pectin in plant biomass recalcitrance. *Biotechnol Biofuels*, 2014. **7**(1): 147.

138. Phillips, G.J. Plasmids as Genetic Tools for Study of Bacterial Gene Function, in *Plasmid Biology*, B.E. Funnell and G.J. Phillips, Editors. 2004, ASM Press: Washington, DC. 567.

139. Kohanski, M.A., D.J. Dwyer, and J.J. Collins. How antibiotics kill bacteria: from targets to networks. *Nature Reviews Microbiology*, 2010. **8**(6): 423-435.

140. Toyn, J.H., P.L. Gunyuzlu, W.H. White, L.A. Thompson, and G.F. Hollis. A counterselection for the tryptophan pathway in yeast: 5-fluoroanthranilic acid resistance. *Yeast* 2000. **16**(6): 553-560.

141. Boeke, J., F. LaCroute, and G. Fink. A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. *Molecular & General Genetics*, 1984. **197**: 345-346.

142. Pritchett, M.A., J.K. Zhang, and W.W. Metcalf. Development of a Markerless Genetic Exchange Method for *Methanosarcina acetivorans C2A* and Its Use in Construction of New Genetic Tools for Methanogenic Archaea*. Applied and Environmental Microbiology*, 2004. **70**(3): 1425-1433.

143. Stout, J.T. and C.T. Caskey. HPRT: gene structure, expression, and mutation*. Annual review of genetics*, 1985. **19**: 127-148.

144. Shao, X., J. Zhou, D.G. Olson, and L.R. Lynd. A markerless gene deletion and integration system for *Thermoanaerobacter ethanolicus. Biotechnology for Biofuels*, 2016. **9**: 100.

145. Gay, P., D. Le Coq, M. Steinmetz, E. Ferrari, and J.A. Hoch. Cloning structural gene sacB, which codes for exoenzyme levansucrase of *Bacillus subtilis*: expression of the gene in *Escherichia coli. Journal of Bacteriology*, 1983. **153**(3): 1424-1431.

146. Yan, X., F. Chu, A.W. Puri, Y. Fu, and M.E. Lidstrom. Electroporation-Based Genetic Manipulation in Type I Methanotrophs*. Applied and Environmental Microbiology*, 2016. **82**(7): 2062-2069.

147. Zhou, S., G. Du, Z. Kang, J. Li, J. Chen, H. Li, and J. Zhou. The application of powerful promoters to enhance gene expression in industrial microorganisms*. World Journal of Microbiology and Biotechnology*, 2016. **33**(2): 23.

148. Wilson, C.M., D.M. Klingeman, C. Schlachter, M.H. Syed, C.-w. Wu, A.M. Guss, and S.D. Brown. LacI Transcriptional Regulatory Networks in *Clostridium thermocellum* DSM1313*. Applied and Environmental Microbiology*, 2017. **83**(5): 16.

149. Mearls, E.B., D.G. Olson, C.D. Herring, and L.R. Lynd. Development of a regulatable plasmid-based gene expression system for *Clostridium thermocellum. Applied Microbiology and Biotechnology*, 2015. **99**(18): 7589-7599.

150. Seo, S., J.-S. Yang, I. Kim, J. Yang, B. Min, S. Kim, and G. Jung. Predictive design of mRNA translation initiation region to control prokaryotic translation efficiency. *Metabolic Engineering*, 2012. **15**: 67-74.

151. Seo, S., J.-S. Yang, H.-S. Cho, J. Yang, S. Kim, J. Park, S. Kim, and G. Jung. Predictive combinatorial design of mRNA translation initiation regions for systematic optimization of gene expression levels. *Scientific Reports*, 2014. **4**(1): 4515.

152. Straub, C.T., B.M. Zeldes, G.J. Schut, M.W.W. Adams, and R.M. Kelly. Extremely thermophilic energy metabolisms: biotechnological prospects. *Current Opinion in Biotechnology*, 2017. **45**: 104-112.

153. Lipscomb, G.L., K. Stirrett, G.J. Schut, F. Yang, F.E. Jenney, R.A. Scott, M.W.W. Adams, and J. Westpheling. Natural competence in the hyperthermophilic archaeon *Pyrococcus furiosus* facilitates genetic manipulation: construction of markerless deletions of genes encoding the two cytoplasmic hydrogenases. *Applied and Environmental Microbiology*, 2011. **77**(7): 2232-2238.

154. Farkas, J., D. Chung, M. DeBarry, M.W. Adams, and J. Westpheling. Defining components of the chromosomal origin of replication of the hyperthermophilic archaeon Pyrococcus furiosus needed for construction of a stable replicating shuttle vector. *Applied and Environmental Microbiology*, 2011. **77**(18): 6343-6349.

155. Farkas, J., K. Stirrett, G.L. Lipscomb, W. Nixon, R.A. Scott, M.W.W. Adams, and J. Westpheling. Recombinogenic properties of *Pyrococcus furiosus* strain COM1 enable rapid selection of targeted mutants. *Applied and Environmental Microbiology*, 2012. **78**(13): 4669-4676.

156. Deng, L., H. Zhu, Z. Chen, Y.X. Liang, and Q. She. Unmarked gene deletion and host-vector system for the hyperthermophilic crenarchaeon *Sulfolobus islandicus. Extremophiles* 2009. **13**(4): 735-746.

157. Kurosawa, N. and D.W. Grogan. Homologous recombination of exogenous DNA with the *Sulfolobus acidocaldarius* genome: properties and uses. *FEMS Microbiology Letters*, 2005. **253**: 141-149.

158. Sato, T., T. Fukui, H. Atomi, and T. Imanaka. Targeted gene disruption by homologous recombination in the *hyperthermophilic archaeon*

Thermococcus kodakaraensis KOD1. *Journal of Bacteriology*, 2003. **185**(1): 210-220.

159. Joshua, C.J., L.D. Perez, and J.D. Keasling. Functional characterization of the origin of replication of pRN1 from *Sulfolobus islandicus* REN1H1. *PLoS ONE*, 2013. **8**(12): e84664.

160. Berkner, S., D. Grogan, and S.V. Albers. Small multicopy, non-integrative shuttle vectors based on the plasmid pRN1 for *Sulfolobus acidocaldarius* and *Sulfolobus solfataricus*, model organisms of the (cren-)archaea. *Nucleic Acids Research*, 2007. **35**(12): e88.

161. Santangelo, T.J., L.u. Cubonová, and J.N. Reeve. Shuttle vector expression in *Thermococcus kodakaraensis*: contributions of cis elements to protein synthesis in a hyperthermophilic archaeon. *Applied and Environmental Microbiology*, 2008. **74**(10): 3099-3104.

162. Vásquez, C., B. González, and R. Vicuña. Plasmids from thermophilic bacteria. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry*, 1984. **78**(3): 507-514.

163. Noll, K.M. and M. Vargas. Recent advances in genetic analyses of hyperthermophilic Archaea and Bacteria. *Archives of Microbiology*, 1997. **168**(2): 73-80.

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

165. Tyurin, M.V., S.G. Desai, and L.R. Lynd. Electrotransformation of *Clostridium thermocellum. Applied and Environmental Microbiology*, 2004. **70**(2): 883-890.

166. Matsumura, M., S. Yasumura, and A. Aiba. Cumulative effect of intragenic amino-acid replacements on the thermostability of a protein. *Nature*, 1986. **323**: 356-358.

167. Matsumura, M. and S. Aiba. Screening for thermostable mutant of kanamycin nucleotidyltransferase by the use of a transformation system

for a thermophile, *Bacillus stearothermophilus. Journal of Biological Chemistry*, 1985. **260**(28): 15298-15303.

168. Mather, M.W. and J.A. Fee. Development of plasmid cloning vectors for *Thermus thermophilus* HB8: Expression of a heterologous, plasmid-borne kanamycin nucleotidyltransferase gene*. Applied and Environmental Microbiology*, 1992. **58**(1): 421-425.

169. Hoseki, J., T. Yano, Y. Koyama, S. Kuramitsu, and H. Kagamiyama. Directed evolution of thermostable kanamycin-resistance gene: a convenient selection marker for *Thermus thermophilus. Journal of Biochemistry*, 1999. **126**(5): 951-956.

170. Lipscomb, G.L., J.M. Conway, S.E. Blumer-Schuette, R.M. Kelly, and M.W.W. Adams. A highly thermostable kanamycin resistance marker expands the tool kit for genetic manipulation of *Caldicellulosiruptor bescii. Applied and Environmental Microbiology*, 2016. **82**(14): 4421-4428.

171. Shaw, A.J., D.A. Hogsett, and L.R. Lynd. Natural competence in *Thermoanaerobacter* and *Thermoanaerobacterium* species*. Applied and Environmental Microbiology*, 2010. **76**(14): 4713-4719.

172. Peng, H., B. Fu, Z. Mao, and W. Shao. Electrotransformation of *Thermoanaerobacter ethanolicus* JW200*. Biotechnology Letters*, 2006. **28**(23): 1913-1917.

173. Olson, D.G. and L.R. Lynd. Transformation of *Clostridium thermocellum* by electroporation*. Methods in Enzymology*, 2012. **510**: 317-330.

174. Olson, D.G., S.A. Tripathi, and R.J. Giannone. Deletion of the Cel48S cellulase from *Clostridium thermocellum. Proceedings of the National Academy of Sciences USA*, 2010. **107**(41): 17727-17732.

175. Rydzak, T., D. Garcia, D.M. Stevenson, M. Sladek, D.M. Klingeman, E.K. Holwerda, D. Amador-Noguez, S.D. Brown, and A.M. Guss. Deletion of Type I glutamine synthetase deregulates nitrogen metabolism and increases ethanol production in *Clostridium thermocellum. Metabolic Engineering*, 2017. **41**: 182-191.

176. Olson, D.G., M. Maloney, A.A. Lanahan, S. Hon, L.J. Hauser, and L.R. Lynd. Identifying promoters for gene expression in *Clostridium thermocellum. Metabolic Engineering Communications*, 2015. **2**: 23-29.

177. Chung, D., M. Cha, J. Farkas, and J. Westpheling. Construction of a stable replicating shuttle vector for *Caldicellulosiruptor* species: use for extending genetic methodologies to other members of this genus*. PLoS ONE*, 2013. **8**(5): e62881.

178. Clausen, A., M.J. Mikkelsen, I. Schröder, and B.K. Ahring. Cloning, sequencing, and sequence analysis of two novel plasmids from the thermophilic anaerobic bacterium *Anaerocellum thermophilum. Plasmid*, 2004. **52**(2): 131-138.

179. Chung, D., J. Young, Y.J. Bomble, T.A. Vander Wall, J. Groom, M.E. Himmel, and J. Westpheling. Homologous Expression of the *Caldicellulosiruptor bescii* CelA Reveals that the Extracellular Protein Is Glycosylated*. PLoS ONE*, 2015. **10**(3): e0119508.

180. Kim, S.K., D. Chung, M.E. Himmel, Y.J. Bomble, and J. Westpheling. Engineering the N-terminal end of CelA results in improved performance and growth of *Caldicellulosiruptor bescii* on crystalline cellulose*. Biotechnology and Bioengineering*, 2017. **114**(5): 945-950.

181. Cha, M., D. Chung, and J. Westpheling. Deletion of a gene cluster for [Ni-Fe] hydrogenase maturation in the anaerobic hyperthermophilic bacterium *Caldicellulosiruptor bescii* identifies its role in hydrogen metabolism*. Applied Microbiology and Biotechnology*, 2016. **100**(4): 1823-1831.

182. Chung, D., J. Young, M. Cha, R. Brunecky, Y.J. Bomble, M.E. Himmel, and J. Westpheling. Expression of the *Acidothermus cellulolyticus* E1 endoglucanase in *Caldicellulosiruptor bescii* enhances its ability to deconstruct crystalline cellulose*. Biotechnology for Biofuels*, 2015. **8**: 113.

183. Chung, D., T.J. Verbeke, K.L. Cross, J. Westpheling, and J.G. Elkins. Expression of a heat-stable NADPH-dependent alcohol dehydrogenase in *Caldicellulosiruptor bescii* results in furan aldehyde detoxification*. Biotechnology for Biofuels*, 2015. **8**: 102.

184. Wilson, G.G. Organization of restriction-modification systems*. Nucleic Acids Res*, 1991. **19**(10): 2539-2566.

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

186. Chung, D., J. Farkas, J.R. Huddleston, E. Olivar, and J. Westpheling. Methylation by a unique alpha-class N4-cytosine methyltransferase is required for DNA transformation of *Caldicellulosiruptor bescii* DSM6725*. PLoS One*, 2012. **7**(8): e43844.

187. Suzuki, S. and N. Kurosawa. Disruption of the gene encoding restriction endonuclease SuaI and development of a host–vector system for the thermoacidophilic archaeon *Sulfolobus acidocaldarius. Extremophiles*, 2016. **20**(2): 139-148.

188. Barakat, A., F. Monlau, J.P. Steyer, and H. Carrere. Effect of lignin-derived and furan compounds found in lignocellulosic hydrolysates on biomethane production*. Bioresour Technol*, 2012. **104**: 90-99.

189. Hsu, T. Ch. 10: Pretreatment of biomass, in *Handbook on Bioethanol: Production and Utilization*, C. Wyman, Editor. 1996, Taylor & Francis: London.

190. Zverlov, V., S. Mahr, K. Riedel, and K. Bronnenmeier. Properties and gene structure of a bifunctional cellulolytic enzyme (CelA) from the extreme thermophile '*Anaerocellum thermophilum*' with separate glycosyl hydrolase family 9 and 48 catalytic domains. *Microbiology*, 1998. **144**(2): 457-465.

191. Chung, D., J. Farkas, and J. Westpheling. Detection of a novel active transposable element in *Caldicellulosiruptor hydrothermalis* and a new search for elements in this genus*. J Ind Microbiol Biotechnol*, 2013. **40**(5): 517-521.

192. Clausen, A., M.J. Mikkelsen, I. Schroder, and B.K. Ahring. Cloning, sequencing, and sequence analysis of two novel plasmids from the thermophilic anaerobic bacterium *Anaerocellum thermophilum. Plasmid*, 2004. **52**(2): 131-138.

193. Chung, D.H., J.R. Huddleston, J. Farkas, and J. Westpheling. Identification and characterization of CbeI, a novel thermostable restriction enzyme from *Caldicellulosiruptor bescii* DSM 6725 and a member of a

new subfamily of HaeIII-like enzymes. *J Ind Microbiol Biotechnol*, 2011. **38**(11): 1867-1877.

194. Farkas, J., D. Chung, M. Cha, J. Copeland, P. Grayeski, and J. Westpheling. Improved growth media and culture techniques for genetic analysis and assessment of biomass utilization by *Caldicellulosiruptor bescii. Journal of Industrial Microbiology & Biotechnology*, 2013. **40**(1): 41-49.

195. O'Sullivan, D. and T. Klaenhammer. Rapid Mini-Prep Isolation of High-Quality Plasmid DNA from *Lacotococcus* and *Lactobacillus* spp. *Applied and Environmental Microbiology*, 1993. **59**(8): 2730-2733.

196. Lee, C., D. Ow, and S. Oh. Quantitative real-time polymerase chain reaction for determination of plasmid copy number in bacteria. *Journal of Microbiological Methods*, 2006. **65**: 258-267.

197. Miroshnichenko, M.L., I.V. Kublanov, N.A. Kostrikina, T.P. Tourova, T.V. Kolganova, N.K. Birkeland, and E.A. Bonch-Osmolovskaya. *Caldicellulosiruptor kronotskyensis* sp. nov. and *Caldicellulosiruptor hydrothermalis* sp. nov., two extremely thermophilic, cellulolytic, anaerobic bacteria from Kamchatka thermal springs. *Int J Syst Evol Microbiol*, 2008. **58**(Pt 6): 1492-1496.

198. Blumer-Schuette, S.E., I. Ozdemir, D. Mistry, S. Lucas, A. Lapidus, J.F. Cheng, L.A. Goodwin, S. Pitluck, M.L. Land, L.J. Hauser, T. Woyke, N. Mikhailova, A. Pati, N.C. Kyrpides, N. Ivanova, J.C. Detter, K. Walston-Davenport, S. Han, M.W. Adams, and R.M. Kelly. Complete genome sequences for the anaerobic, extremely thermophilic plant biomass-degrading bacteria *Caldicellulosiruptor hydrothermalis, Caldicellulosiruptor kristjanssonii, Caldicellulosiruptor kronotskyensis*, *Caldicellulosiruptor owensensis*, and *Caldicellulosiruptor lactoaceticus. J Bacteriol*, 2011. **193**(6): 1483-1484.

199. Donahue, J.P., D.A. Israel, R.M. Peek, M.J. Blaser, and G.G. Miller. Overcoming the restriction barrier to plasmid transformation of *Helicobacter pylori. Molecular microbiology*, 2000. **37**(5): 1066-1074.

200. Johnston, C., B. Martin, G. Fichant, P. Polard, and J.-P.P. Claverys. Bacterial transformation: distribution, shared mechanisms and divergent control. *Nature reviews. Microbiology*, 2014. **12**(3): 181-196.

201. Akinosho, H., K. Yee, D. Close, and A. Ragauskus. The emergence of *Clostridium thermocellum* as a high utility candidate for consolidated bioprocessing applications*. Frontiers in Chemistry*, 2014. **2**: 66.

202. Deng, Y., D.G. Olson, J. Zhou, C.D. Herring, and J.A. Shaw. Redirecting carbon flux through exogenous pyruvate kinase to achieve high ethanol yields in *Clostridium thermocellum. Metabolic Engineering*, 2013. **15**: 151-158.

203. Biswas, R., T. Zheng, D.G. Olson, L.R. Lynd, and A.M. Guss. Elimination of hydrogenase active site assembly blocks H2 production and increases ethanol yield in *Clostridium thermocellum. Biotechnology for Biofuels*, 2015. **8**(1): 20.

204. Papanek, B.A., R. Biswas, T. Rydzak, and A.M. Guss. Elimination of metabolic pathways to all traditional fermentation products increases ethanol yields in *Clostridium thermocellum. Metabolic Engineering*, 2015. **32**: 49-54.

205. Olson, D.G. and L.R. Lynd. Chap. 17: Transformation of *Clostridium thermocellum* by electroporation., in *Methods in Enzymology*. 2012. 317-330.

206. Guss, A.M., D.G. Olson, N.C. Caiazza, and L.R. Lynd. Dcm methylation is detrimental to plasmid transformation in *Clostridium thermocellum. Biotechnology for biofuels*, 2012. **5**(1): 30.

207. Olson, D.G. and L.R. Lynd. Computational design and characterization of a temperature-sensitive plasmid replicon for gram positive thermophiles*. Journal of Biological Engineering*, 2012. **6**(1).

208. Khan, S.A. Plasmid rolling-circle replication: highlights of two decades of research*. Plasmid*, 2005. **53**(2): 126-136.

209. Ehrlich, S.D. Replication and expression of plasmids from *Staphylococcus aureus* in *Bacillus substilis. Proc Natl Acad Sci U S A*, 1977. **74**(4): 1680-1682.

210. Dubnau, D. *Molecular Cloning in Bacillus subtilis*. Experimental Manipulation of Gene Expression, ed. M. Inouye. 1983, New York, London: Academic Press.

211. Luchansky, J.B., P.M. Muriana, and T.R. Klaenhammer. Application of electroporation for transfer of plasmid DNA to *Lactobacillus*, *Lactococcus*, *Leuconostoc*, *Listeria*, *Pediococcus*, *Bacillus*, *Staphylococcus*, *Enterococcus* and *Propionibacterium. Mol Microbiol*, 1988. **2**(5): 637-646.

212. Jannière, L., C. Bruand, and S.D. Ehrlich. Structurally stable *Bacillus subtilis* cloning vectors*. Gene*, 1990. **87**: 53-61.

213. Serwold-Davis, T.M., N. Groman, and M. Rabin. Transformation of *Corynebacterium diphtheriae*, *Corynebacterium ulcerans*, *Corynebacterium glutamicum*, and *Escherichia coli* with the *C. diphtheriae* plasmid pNG2*. Proc Natl Acad Sci U S A*, 1987. **84**(14): 4964-4968.

214. McNamara, P.J. Chapter 5: Genetic Manipulation of Staphylococcus aureus, in *Staphylococcus: Molecular Genetics*, J. Lindsay, Editor. 2008, Caister Academic Press: United Kingdom.

215. Projan, S.J., M. Monod, C.S. Narayanan, and D. Dubnau. Replication properties of pIM13, a naturally occurring plasmid found in *Bacillus subtilis*, and of its close relative pE5, a plasmid native to *Staphylococcus aureus. Journal of bacteriology*, 1987. **169**(11): 5131-5139.

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

217. Garnier, T. and S.T. Cole. Complete nucleotide sequence and genetic organization of the bacteriocinogenic plasmid, pIP404, from *Clostridium perfringens. Plasmid*, 1988. **19**(2): 134-150.

218. Klapatch, T.R., M.L. Guerinot, and L.R. Lynd. Electrotransformation of *Clostridium thermosaccharolyticum. J Ind Microbiol*, 1996. **16**(6): 342-347.

219. Groom, J., D. Chung, J. Young, and J. Westpheling. Heterologous complementation of a *pyrF* deletion in *Caldicellulosiruptor hydrothermalis*

generates a new host for the analysis of biomass deconstruction. *Biotechnol Biofuels*, 2014. **7**(1): 132.

220. Subramanya, H.S., L.K. Arciszewska, R.A. Baker, L.E. Bird, D.J. Sherratt, and D.B. Wigley. Crystal structure of the site-specific recombinase, XerD. *The EMBO Journal*, 1997. **16**(17): 5178-5187.

221. Colloms, S.D. The topology of plasmid-monomerizing Xer site-specific recombination. *Biochemical Society transactions*, 2013. **41**(2): 589-594.

222. Goujon, M., H. McWilliam, W. Li, F. Valentin, S. Squizzato, J. Paern, and R. Lopez. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Research*, 2010. **38**(Web Server).

223. Sievers, F., A. Wilm, D. Dineen, T.J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J.D. Thompson, and D.G. Higgins. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 2011. **7**(1): 539-539.

224. Darriba, D., G.L. Taboada, R. Doallo, and D. Posada. ProtTest 3: fast selection of best-fit models of protein evolution. *Bioinformatics*, 2011. **27**(8): 1164-1165.

225. Huelsenbeck, J.P. and F. Ronquist. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 2001. **17**(8): 754-755.

226. Whelan, S. and N. Goldman. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol*, 2001. **18**(5): 691-699.

227. Le, S.Q. and O. Gascuel. An improved general amino acid replacement matrix. *Mol Biol Evol*, 2008. **25**(7): 1307-1320.

228. Bailey, T.L. and C. Elkan. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol*, 1994. **2**: 28-36.

229. Angellotti, M.C., S.B. Bhuiyan, G. Chen, and X.F. Wan. CodonO: codon usage bias analysis within and across genomes. *Nucleic Acids Research*, 2007. **35**(Web Server).

230. Kanehisa, M. and S. Goto. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, 2000. **28**(1): 27-30.

231. Kanehisa, M., S. Goto, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe. Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res*, 2014. **42**(Database issue): D199-205.

232. Geer, L.Y., A. Marchler-Bauer, R.C. Geer, L. Han, J. He, S. He, C. Liu, W. Shi, and S.H. Bryant. The NCBI BioSystems database. *Nucleic Acids Res*, 2010. **38**(Database issue): D492-496.

233. Rainey, F.A., N.L. Ward, H.W. Morgan, R. Toalster, and E. Stackebrandt. Phylogenetic analysis of anaerobic thermophilic bacteria: aid for their reclassification. *J Bacteriol*, 1993. **175**(15): 4772-4779.

234. Marchler-Bauer, A., M.K. Derbyshire, N.R. Gonzales, S. Lu, F. Chitsaz, L.Y. Geer, R.C. Geer, J. He, M. Gwadz, D.I. Hurwitz, C.J. Lanczycki, F. Lu, G.H. Marchler, J.S. Song, N. Thanki, Z. Wang, R.A. Yamashita, D. Zhang, C. Zheng, and S.H. Bryant. CDD: NCBI's conserved domain database. *Nucleic Acids Res*, 2015. **43**(Database issue): D222-226.

235. Summers, D.K. and D.J. Sherratt. Resolution of ColE1 dimers requires a DNA sequence implicated in the three-dimensional organization of the *cer* site. *The EMBO journal*, 1988. **7**(3): 851-858.

236. Cornet, F., I. Mortier, J. Patte, and J.M. Louarn. Plasmid pSC101 harbors a recombination site, *psi*, which is able to resolve plasmid multimers and to substitute for the analogous chromosomal *Escherichia coli* site *dif. Journal of bacteriology*, 1994. **176**(11): 3188-3195.

237. Stirling, C.J., G. Szatmari, G. Stewart, M.C. Smith, and D.J. Sherratt. The arginine repressor is essential for plasmid-stabilizing site-specific recombination at the ColE1 *cer* locus. *The EMBO journal*, 1988. **7**(13): 4389-4395.

238. Blake, J.A., N. Ganguly, and D.J. Sherratt. DNA sequence of recombinase-binding sites can determine Xer site-specific recombination outcome. *Molecular microbiology*, 1997. **23**(2): 387-398.

239. Makarova, K.S., A.A. Mironov, and M.S. Gelfand. Conservation of the binding site for the arginine repressor in all bacterial lineages. *Genome biology*, 2001. **2**(4): research0013.0011–0013.0018.

240. Nadir, T.M., B. Annemie Van den, B. Ilse Van den, S. Patrick, L. Yves, L. Anne Marie, M. Gaston, J. John, and C. Mohamed. Arginine residues as stabilizing elements in proteins. *Biochemistry*, 1992. **31**(8): 2239-2253.

241. Barton, L.L. Ch.8 Section 2.1: Thermostability of Proteins., in *Structural and Functional Relationships in Prokaryotes*. 2005, Springer: New York. 351-352.

242. Phillips, L.G., D.M. Whitehead, and J.E. Kinsella. Ch. 2 Section 6: Hydrophobicity and Protein Stability., in *Structure-Function Properties of Food Proteins*. 2013, Academic Press: San Diego.

243. Shintani, M., Z.K. Sanchez, and K. Kimbara. Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy. *Frontiers in microbiology*, 2015. **6**: 242.

244. Sprincova, A., P. Javorsky, and P. Pristas. pSRD191, a new member of RepL replicating plasmid family from *Selenomonas ruminantium. Plasmid*, 2005. **54**(1): 39-47.

245. Olson, D.G., J.E. McBride, A.J. Shaw, and L.R. Lynd. Recent progress in consolidated bioprocessing. *Curr Opin Biotechnol*, 2012. **23**(3): 396-405.

246. Groom, J., D. Chung, D.G. Olson, L.R. Lynd, A.M. Guss, and J. Westpheling. Promiscuous plasmid replication in thermophiles: Use of a novel hyperthermophilic replicon for genetic manipulation of *Clostridium thermocellum* at its optimum growth temperature. *Metabolic Engineering Communications*, 2016. **3**: 30-38.

247. Bianco, P.R. and S.C. Kowalczykowski. RecA Protein. *eLS*, 2005.

248. Clark, A.J. and A.D. Margulies. Isolation and characterization of recombination-deficient mutants of Escherichia coli K12. *Proceedings of the National Academy of Sciences*, 1965. **53**: 451-459.

249. Lenhart, J.S., J.W. Schroeder, B.W. Walsh, and L.A. Simmons. DNA repair and genome maintenance in *Bacillus subtilis. Microbiology and Molecular Biology Reviews*, 2012. **76**(3): 530-564.

250. Castan, P., L. Casares, J. Barbe, and J. Berenguer. Temperature-dependent hypermutational phenotype in *recA* mutants of *Thermus thermophilus* HB27. *Journal of Bacteriology*, 2003. **185**(16): 4901-4907.

251. Capaldo, F.N., G. Ramsey, and S.D. Barbour. Analysis of the growth of recombination-deficient strains of *Escherichia coli* K-12. *Journal of Bacteriology*, 1974. **118**(1): 242-249.

252. Sciochetti, S.A., G.W. Blakely, and P.J. Piggot. Growth phase variation in cell and nucleoid morphology in a *Bacillus subtilis recA* mutant. *Journal of Bacteriology*, 2001. **183**(9): 2963-2968.

253. Vierling, S., T. Weber, and W. Wohlleben. Evidence that an additional mutation is required to tolerate insertional inactivation of the *Streptomyces lividans recA* gene. *Journal of Bacteriology*, 2001. **183**(14): 4374-4381.

254. Zyskind, J.W., A.L. Svitil, W.B. Stine, M.C. Biery, and D.W. Smith. RecA protein of *Escherichia coli* and chromosome partitioning. *Molecular Microbiology*, 1992. **6**(17): 2525-2537.

255. Cox, M.M. Recombination DNA repair in bacteria and the RecA protein. *Progress in Nucleic Acid Research and Molecular Biology*, 1999. **63**: 311-366.

256. Liu, Z., N. Guiliani, C. Appia-Ayme, F. Borne, J. Ratouchniak, and V. Bonnefoy. Construction and characterization of a *recA* mutant of *Thiobacillus ferrooxidans* by marker exchange mutagenesis. *Journal of Bacteriology*, 2000. **182**(8): 2269-2276.

257. Gomez, R.F., B. Snedecor, and B. Mendez. Development of genentic principles in *Clostridium thermocellum*, in *Developments in Industrial*

Microbiology. 1980, Society for Industrial Microbiology: Arlington, Virginia. 87-95.

258. Martin, B., P. Garcia, C. Marie-Pierre, and J.-P.P. Claverys. The *recA* gene of *Streptococcus pneumoniae* is part of a competence-induced operon and controls lysogenic induction. *Molecular Microbiology*, 1995. **15**(2): 367-379.

259. Ceglowski, P., G. Lüder, and J.C. Alonso. Genetic analysis of *recE* activities in *Bacillus subtilis. Molecular & General Genetics*, 1990. **222**: 441-445.

260. Olson, D.G. and L.R. Lynd. Computational design and characterization of a temperature-sensitive plasmid replicon for gram positive thermophiles. *Journal of Biological Engineering*, 2012. **6**: 5.

261. Serebriiskii, I.G. and E.A. Golemis. Uses of lacZ to study gene function: Evaluation of β-Galactosidase assays employed in the yeast two-hybrid system. *Analytical Biochemistry*, 2000. **285**(1): 1-15.

262. Inoue, T. and Y. Sako. Host-Vector Systems in Thermophiles, in *Thermophilic Microbes in Environmental and Industrial Biotechnology: Biotechnology of Thermophiles*, T. Satyanarayana, J. Littlechild, and Y. Kawarabayasi, Editors. 2013, Springer Science+Business Media: Dordrecht. 351-373.

263. Pédelacq, J.-D.D., S. Cabantous, T. Tran, T.C. Terwilliger, and G.S. Waldo. Engineering and characterization of a superfolder green fluorescent protein. *Nature Biotechnology*, 2006. **24**(1): 79-88.

264. Cava, F., M.A. de Pedro, E. Blas-Galindo, G.S. Waldo, L.F. Westblade, and J. Berenguer. Expression and use of superfolder green fluorescent protein at high temperatures in vivo: a tool to study extreme thermophile biology. *Environmental Microbiology*, 2008. **10**(3): 605-613.

265. Drepper, T., T. Eggert, F. Circolone, A. Heck, U. Krauß, J.-K. Guterl, M. Wendorff, A. Losi, W. Gärtner, and K.-E. Jaeger. Reporter proteins for in vivo fluorescence without oxygen. *Nature Biotechnology*, 2007. **25**(4): 443-445.

266. Cantarel, B.L., P.M. Coutinho, C. Rancurel, T. Bernard, V. Lombard, and B. Henrissat. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Research*, 2009. **37**(Database issue): D233-D238.

267. Hon, S., A.A. Lanahan, L. Tian, R.J. Giannone, R.L. Hettich, D.G. Olson, and L.R. Lynd. Development of a plasmid-based expression system in *Clostridium thermocellum* and its use to screen heterologous expression of bifunctional alcohol dehydrogenases (*adhE*s). *Metabolic Engineering Communications*, 2016. **3**: 120-129.

268. Vagner, V., E. Dervyn, and S.D. Ehrlich. A vector for systematic gene inactivation in *Bacillus subtilis. Microbiology*, 1998. **144**(11): 3097-3104.

269. Zhou, S., G. Du, Z. Kang, J. Li, J. Chen, H. Li, and J. Zhou. The application of powerful promoters to enhance gene expression in industrial microorganisms. *World Journal of Microbiology and Biotechnology*, 2017. **33**(2): 23.

270. Bertoldo, C. and G. Antranikian. Amylolytic enzymes from hyperthermophiles. *Methods in Enzymology*, 2001. **330**: 269-289.

271. Webb, A.J., K.A. Homer, and A.H.F. Hosie. Two closely related ABC transporters in *Streptococcus mutans* are involved in disaccharide and/or oligosaccharide uptake. *Journal of Bacteriology*, 2007. **190**(1): 168-178.

272. Webb, A.J., K.A. Homer, and A.H.F. Hosie. A phosphoenolpyruvate-dependent phosphotransferase system is the principal maltose transporter in *Streptococcus mutans. Journal of Bacteriology*, 2007. **189**(8): 3322-3327.

273. Afzal, M., S. Shafeeq, I. Manzoor, and O.P. Kuipers. Maltose-Dependent Transcriptional Regulation of the *mal* Regulon by MalR in *Streptococcus pneumoniae. PLoS One*, 2015. **10**(6): e0127579.

274. Boos, W. and H. Shuman. Maltose/maltodextrin system of *Escherichia coli*: transport, metabolism, and regulation. *Microbiology and Molecular Biology Reviews*, 1998. **62**(1): 204-229.

275. Lacks, S. Genetic regulation of maltosaccharide utilization in Pneumococcus. *Genetics*, 1968. **60**(4): 685-706.

276. Schneider, E. ABC transporters catalyzing carbohydrate uptake. *Research in Microbiology*, 2001. **152**: 303-310.

277. Schlösser, A., T. Kampers, and H. Schrempf. The *Streptomyces* ATP-binding component MsiK assists in cellobiose and maltose transport. *Journal of Bacteriology*, 1997. **179**(6): 2092-2095.

278. Puyet, A. and M. Espinosa. Structure of the maltodextrin-uptake locus of *Streptococcus pneumoniae*: Correlation to the *Escherichia coli* maltose regulon. *Journal of Molecular Biology*, 1993. **230**(3): 800-811.

279. Raibaud, O. and E. Richet. Maltotriose is the inducer of the maltose regulon of *Escherichia coli. Journal of Bacteriology*, 1987. **169**(7): 3059-3061.

280. Reidl, J. and W. Boos. The *malX malY* operon of *Escherichia coli* encodes a novel enzyme II of the phosphotransferase system recognizing glucose and maltose and an enzyme abolishing the endogenous induction of the maltose system. *Journal of Bacteriology*, 1991. **173**(15): 4862-4876.

281. Puyet, A., A.M. Ibanez, and M. Espinosa. Characterization of the *Streptococcus pneumoniae* maltosaccharide regulator MalR, a member of the LacI-GalR family of repressors displaying distinctive genetic features. *Journal of Biological Chemistry*, 1993. **268**(34): 25402-25408.

282. van Wezel, G.P., J. White, P. Young, P.W. Postma, and M.J. Bibb. Substrate induction and glucose repression of maltose utilization by *Streptomyces coelicolor* A3(2) is controlled by *malR*, a member of the *lacI–galR* family of regulatory genes. *Molecular Microbiology*, 1997. **23**(3): 537-549.

283. Andersson, U. and P. Rådström. Physiological function of the maltose operon regulator, MalR, in *Lactococcus lactis. BMC Microbiology*, 2002. **2**: 28.

284. Weickert, M.J. and S. Adhya. A family of bacterial regulators homologous to Gal and Lac repressors. *The Journal of Biological Chemistry*, 1992. **267**(22): 15869-15874.

285. Nieto, C., M. Espinosa, and A. Puyet. The maltose/maltodextrin regulon of *Streptococcus pneumoniae*: Differential promoter regulation by the transcriptional repressor MalR. *The Journal of Biological Chemistry*, 1997. **272**(49): 30860-30865.

286. Shelburne, S.A., P. Sahasrobhajane, B. Suber, D.B. Keith, M.T. Davenport, N. Horstmann, M. Kumaraswami, R.J. Olsen, R.G. Brennan, and J.M. Musser. Niche-specific contribution to streptococcal virulence of a MalR-regulated carbohydrate binding protein. *Molecular Microbiology*, 2011. **81**(2): 500-514.

287. Solovyev, V. and A. Salamov. Automatic Annotation of Microbial Genomes and Metagenomic Sequences, in *Metagenomics and its Applications in Agriculture, Biomedicine and Environmental Studies*, R.W. Li, Editor. 2016, Nova Science Publishers. 61-78.

288. Macke, T.J., D.J. Ecker, R.R. Gutell, D. Gautheret, D.A. Case, and R. Sampath. RNAMotif, an RNA secondary structure definition and search algorithm. *Nucleic Acids Research*, 2001. **29**(22): 4724-4735.

289. Gautheret, D. and A. Lambert. Direct RNA motif definition and identification from multiple sequence alignments using secondary structure profiles. *Journal of Molecular Biology*, 2001. **313**(5): 1003-1011.

290. Bailey, T.L., M. Boden, F.A. Buske, M. Frith, C.E. Grant, L. Clementi, J. Ren, W.W. Li, and W.S. Noble. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*, 2009. **37**: W202–W208.

291. van de Werken, H.J., M.R. Verhaart, A.L. VanFossen, K. Willquist, D.L. Lewis, J.D. Nichols, H.P. Goorissen, E.F. Mongodin, K.E. Nelson, E.W. van Niel, A.J. Stams, D.E. Ward, W.M. de Vos, J. van der Oost, R.M. Kelly, and S.W. Kengen. Hydrogenomics of the extremely thermophilic bacterium *Caldicellulosiruptor saccharolyticus. Applied and Environmental Microbiology*, 2008. **74**(21): 6720-6729.

292. van Wezel, G.P., J. White, M.J. Bibb, and P.W. Postma. The malEFG gene cluster of *Streptomyces coelicolor* A3(2): characterization, disruption

and transcriptional analysis. *Molecular and General Genetics*, 1997. **254**(5): 604-608.

293. Hülsmann, A., R. Lurz, F. Scheffel, and E. Schneider. Maltose and maltodextrin transport in the thermoacidophilic gram-positive bacterium *Alicyclobacillus acidocaldarius* is mediated by a high-affinity transport system that includes a maltose binding protein tolerant to low pH. *Journal of Bacteriology*, 2000. **182**(22): 6292-6301.

294. Barabote, R.D., G. Xie, D.H. Leu, P. Normand, A. Necsulea, V. Daubin, C. Médigue, W.S. Adney, X. Xu, A. Lapidus, R.E. Parales, C. Detter, P. Pujic, D. Bruce, C. Lavire, J.F. Challacombe, T.S. Brettin, and A.M. Berry. Complete genome of the cellulolytic thermophile *Acidothermus cellulolyticus* 11B provides insights into its ecophysiological and evolutionary adaptations. *Genome Research*, 2009. **19**(6): 1033-1043.

295. Swint-Kruse, L. and K.S. Matthews. Allostery in the LacI/GalR family: variations on a theme. *Current Opinion in Microbiology*, 2009. **12**(2): 129-137.

296. Nieto, C., P. de Palencia, P. López, and M. Espinosa. Construction of a tightly regulated plasmid vector for *Streptococcus pneumoniae*: Controlled expression of the green fluorescent protein. *Plasmid*, 2000. **43**(3): 205-213.

297. Mackey, M.C., M. Santillán, M. Tyran-Kamińska, and E.S. Zeron. The Lactose Operon, in *Simple Mathematical Models of Gene Regulatory Dynamics*. 2016, Springer International Publishing: Cham. 73-85.

298. Oehler, S., E.R. Eismann, H. Krämer, and B. Müller-Hill. The three operators of the *lac* operon cooperate in repression. *The EMBO Journal*, 1990. **9**(4): 973-979.

299. Nieto, C., A. Puyet, and M. Espinosa. MalR-mediated regulation of the *Streptococcus pneumoniae malMP* operon at promoter *Pm*: Influence of a proximal divergent promoter region and competition between MalR and RNA Polymerase proteins. *Journal of Biological Chemistry*, 2001. **276**(18): 14946-14954.

300. Gorke, B. and J. Stulke. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Microbiol*, 2008. **6**(8): 613-624.

301. Lochner, A., R.J. Giannone, M. Rodriguez, M.B. Shah, J.R. Mielenz, M. Keller, G. Antranikian, D.E. Graham, and R.L. Hettich. Use of label-free quantitative proteomics to distinguish the secreted cellulolytic systems of *Caldicellulosiruptor bescii* and *Caldicellulosiruptor obsidiansis. Applied and Environmental Microbiology*, 2011. **77**(12): 4042-4054.

302. Jobe, A. and S. Bourgeois. The *lac* repressor-operator interaction. VII. A repressor with unique binding properties: the X86 repressor. *Journal of Molecular Biology*, 1972. **72**(1): 139-152.

303. Revzin, A. and P.H. von Hippel. Direct measurement of association constants for the binding of *Escherichia coli lac* repressor to non-operator DNA. *Biochemistry*, 1977. **16**(22): 4769-4776.

304. Elf, J., G.W. Li, and X.S. Xie. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science*, 2007. **316**: 1191-1194.

APPENDIX A

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(3): e0119508.

Abstract

Members of the bacterial genus *Caldicellulosiruptor* are the most thermophilic cellulolytic microbes described with ability to digest lignocellulosic biomass without conventional pretreatment. The cellulolytic ability of different species varies dramatically and correlates with the presence of the multimodular cellulase CelA, which contains both a glycoside hydrolase family 9 endoglucanase and a glycoside hydrolase family 48 exoglucanase known to be synergistic in their activity, connected by three cellulose-binding domains via linker peptides. This architecture exploits the cellulose surface ablation driven by its general cellulase processivity as well as excavate cavities into the surface of the substrate, revealing a novel paradigm for cellulase activity. We recently reported that a deletion of *celA* in *C. bescii* had a significant effect on its ability to utilize complex biomass. To analyze the structure and function of CelA and its role in biomass deconstruction, we constructed an new expression vector for *C. bescii* and were able, for the first time, to express significant quantities of full-length protein *in vivo* in the native host. The protein, which contains a Histidine tag, was active and

excreted from the cell. Expression of CelA protein with and without its signal sequence allowed comparison of protein retained intracellularly to protein transported extracellularly. Analysis of protein in culture supernatants revealed that the extracellular CelA protein is glycosylated whereas the intracellular CelA is not, suggesting that either protein transport is required for this post-translational modification or that glycosylation is required for protein export. The mechanism and role of protein glycosylation in bacteria is poorly understood and the ability to express CelA *in vivo* in *C. bescii* will allow the study of the mechanism of protein glycosylation in this thermophile. It will also allow the study of glycosylation of CelA itself and its role in the structure and function of this important enzyme in biomass deconstruction.

APPENDIX B

Yu M, Ji L, Neumann DA, Chung D, Groom J, Westpheling J, He C, Schmitz RJ. (2015) Base-resolution detection of N^4-methylcytosine in genomic DNA using 4mC-Tet-assisted-bisulfite sequencing. *Nucleic Acids Research* 43(21): e148.

Abstract

Restriction modification (R-M) systems pose a major barrier to DNA transformation and genetic engineering of bacterial species. Systematic identification of DNA methylation in R-M systems, including N^6-methyladenine (6mA), 5-methylcytosine (5mC) and N^4-methylcytosine (4mC), will enable strategies to make these species genetically tractable. Although single-molecule, real time (SMRT) sequencing technology is capable of detecting 4mC directly for any bacterial species regardless of whether an assembled genome exists or not, it is not as scalable to profiling hundreds to thousands of samples compared with the commonly used next-generation sequencing technologies. Here we present 4mC-Tet-Assisted Bisulfite-sequencing (4mC-TAB-seq), a next-generation sequencing method that rapidly and cost efficiently reveals the genome-wide locations of 4mC for bacterial species with an available assembled reference genome. In 4mC-TAB-seq, both cytosines and 5mCs are read out as thymines, whereas only 4mCs are read out as cytosines, revealing their specific positions throughout the genome. We applied 4mC-TAB-seq to study the methylation of a member of the hyperthermophilc genus, *Caldicellulosiruptor*, in which 4mC-

related restriction is a major barrier to DNA transformation from other species. In combination with MethylC-seq, both 4mC- and 5mC-containing motifs are identified which can assist in rapid and efficient genetic engineering of these bacteria in the future.

APPENDIX C

Kim S, 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. *Biotechnology for Biofuels* 10: 66.

Abstract

Background: Resistance to deconstruction is a major limitation to the use of lignocellulosic biomass as a substrate for the production of fuels and chemicals. Consolidated bioprocessing (CBP), the use of microbes for the simultaneous hydrolysis of lignocellulose into soluble sugars and fermentation of the resulting sugars to products of interest, is a potential solution to this obstacle. The pretreatment of plant biomass, however, releases compounds that are inhibitory to the growth of microbes used for CBP.

Results: Heterologous expression of the *Thermoanaerobacter pseudethanolicus* 39E *bdhA* gene, that encodes an alcohol dehydrogenase, in *Clostridium thermocellum* significantly increased resistance to furan derivatives at concentrations found in acid-pretreated biomass. The mechanism of detoxification of hydroxymethylfurfural (HMF) was shown to be primarily reduction using NADPH as the cofactor. In addition, we report the construction of new expression vectors for homologous and heterologous expression in *C. thermocellum.* These vectors use regulatory signals from both *C. bescii* (the S-

layer promoter) and *C. thermocellum* (the enolase promoter) shown to efficiently drive expression of the BdhA enzyme.

Conclusions: Toxic compounds present in lignocellulose hydrolysates that inhibit cell growth and product formation are obstacles to the commercialization of fuels and chemicals from biomass. Expression of genes that reduce the effect of these inhibitors, such as furan derivatives, will serve to enable commercial processes using plant biomass for the production of fuels and chemicals.