UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

IMPROVEMENT OF LIGNOCELLULOSE BIOCONVERSION IN CLOSTRIDIUM CELLULOLYTICUM AND IDENTIFICATION OF ACTIVE LIGNOCELLULOSE DEGRADERS IN TEMPERATE GRASSLAND

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

XUANYU TAO
Norman, Oklahoma
2020

IMPROVEMENT OF LIGNOCELLULOSE BIOCONVERSION IN CLOSTRIDIUM CELLULOLYTICUM AND IDENTIFICATION OF ACTIVE LIGNOCELLULOSE DEGRADERS IN TEMPERATE GRASSLAND

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF MICROBIOLOGY AND PLANT BIOLOGY

BY THE COMMITTEE CONSISTING OF

Dr. Jizhong Zhou, Chair

Dr. Hairong Song

Dr. Lee R. Krumholz

Dr. Heather R. McCarthy

Dr. Chongle Pan

Dedication

To Dan

My companion, friend, and supporter,

with whom my happiness doubled and melancholy halved

Acknowledgement

The journey of pursuing a doctoral degree in a country which is thousands of miles far away from my home country is truly challenging for me. However, this is a wonderful experience that I will never forget. During the 5 years, I received kind help and inspiring support from different people, without which it would not be possible for me to complete the Ph. D program smoothly. I mention them here to express my sincere gratitude for their help in so many ways and I am grateful to all who helped and supported me to live and study in OU.

First of all, I would like to express my great thanks to my advisor, Dr. Jizhong Zhou for his patience, diligence, insight, and motivation throughout my time at OU. Along the way, I really appreciated his financial support and granting me the chance to do the research which I am interested in. Finally, I was equipped with cutting-edge skills and knowledge in both pure culture and microbial communities. I would also like to convey my deep appreciation to my advisory committee members, Dr. Lee Krumholz, Dr. Heather McCarthy, Dr. Chongle Pan, and Dr. Hairong Song, for supporting and leading me towards the pursuit of better science and a better life. I am thankful to Dr. Krumholz for his help, support, and practical advice on my study, especially for my first year living in OU, and for teaching me a lot in geomicrobiology and writing proposals. Discussions with you were never long, but genuinely encouraging. I am grateful to Dr. McCarthy for her expertise in plant biology and helping me a lot during my graduation time. Your help made it possible for me to graduate in Fall, 2020. I also would like to thank Dr. Chongle Pan who has been my committee member for only a short time but giving me helpful advice on my study and enlightening me in the perspective of

bioinformatics. Last but not least, I would like to thank Dr. Hairong Song for introducing me to the statistical world for the first time, for guiding me on how to study statistics and use it as a tool, and for encouraging me during this journey.

Secondly, I would like to thank many of my friends and colleagues at the Institute for Environmental Genomics. I especially thank Dr. Tao Xu, my friend and partner, for encouraging and collaborating with me in almost all of my pure culture studies. Without your help, I cannot complete the program in five years. I thank Dr. Megan Kempher and Mr. Colin Bates for helping me, especially for improving my English and my writing. With their help, I quickly accommodated to the new life in OU. I thank Dr. Jiajie Feng for encouraging and collaborating with me in all my microbial community studies. I thank Drs. Aifen Zhou, Daliang Ning, Jialiang Kuang, Gangsheng Wang, Liyou Wu, Xishu Zhou, Joy D. Van Nostrand, Yujia Qin, Mengting Yuan, Xue Guo, Zhengsheng Yu, Jiantao Liu, Miss. Ying Fu and Miss. Rice Lindsay for helping me with various academic issues. Without you and every other helpful colleague, this dissertation could not be finished.

I also must thank my collaborators from other institutes. I thank Dr. Yunfeng Yang for guiding and teaching me on how to discover good scientific questions and writing a qualified paper. I want to thank Dr. Yongxing He for encouraging me to pursue my dreams, for insightful communications on the project I worked on, and for helping me a lot in biochemistry and mass spectrum. I thank Miss. Siping Zhang, a Ph. D student from Dr. Yongxing He's lab, for her help in some biochemical experiments. At last, I also want to thank Dr. Nannan Zhang for her help in crystallization.

Finally, I dedicate my deepest appreciation to my dear parents, Mr. Xingchun Tao and

Ms. Min Qiu. Although they are physically away from me during my five-year study in OU, they always support and encourage me to move forward, especially when coming across difficulties. You are my greatest treasures forever. Completing this unique journey is never possible without them.

Table of contents

List of Tables

List of Figures

Abstract

As the most abundant sustainable carbon resource on the earth, lignocellulose is considered as one of the most promising feedstocks to produce biofuels, which can mitigate the environmental issues brought by burning of fossil fuels. However, the lignocellulosic biofuels face grand challenges on multiple fronts, including low-cost technology for utilization of cellulose and highly efficient conversion from cellulose to biofuels by high-yield microorganisms. Therefore, these challenges are calling for engineering designs to increase efficiency and reduce the costs. As a model mesophilic clostridial species for studying lignocellulose degradation, *Clostridium cellulolyticum* can perform one-step bioconversion of lignocellulose to biofuels and is considered as a potential candidate for future industrial biofuels productions. However, the efficiency of lignocellulose bioconversion in *C. cellulolyticum* is not high enough, which impedes its further application in industries. Thus, the major aim of this dissertation is to engineer the *C. cellulolyticum* by CRISPR-Cas9 editing method to improve its lignocellulose bioconversion efficiency. In addition to being feedstock for biofuels productions, lignocellulose is the primary carbon input in the natural soil and the microbial decomposition of the lignocellulose is an important global carbon sink. With current global warming, both photosynthesis rate by plants and the carbon decomposition rate by microorganisms can be enhanced but may not equally. As a result, whether warming can cause a positive feedback for C exchange between the terrestrial and atmosphere is unclear. Thus, another major aim of this dissertation is to identify the active

lignocellulose bacteria and understand their lignocellulose degradation mechanisms responding to warming.

In *C. cellulolyticum*, a unique extracellular multi-enzyme complex named cellulosome plays the most important role in degrading the cellulose. The cellulosome has great commercial values and can be used for consolidated bio-saccharification. However, the function for a cellulosomal component named X2 in *C. cellulolyticum* was still unclear, which limited our understanding for the cellulose degradation mechanisms by the cellulosome and its future commercial application. To have a better understanding of the *in-vivo* biological function of X2 modules, we employed CRISPR-Cas9 editing to create dual X2 modules mutant (△X2-NC) by deleting the conserved motif (NGNT) of X2 modules. Compared to the wild type strain, the degradation efficiency and saccharification ability in the △X2-NC were decreased. Further, the *in vivo* adhesion assay and the *in vitro* enzymatic assay found that the biological function of the X2 module was associated with the binding affinity between the cells and its cellulose substrate. This study provides new perspectives on engineering cellulolytic bacteria or modification of commercial cellulases for industrial application.

Major cellulosomal components are encoded by a 26 kb *cip-cel* gene operon named *cip-cel*. Two major large transcripts were detected when *C. cellulolyticum* was grown on cellulose. However, the abundance of 3'- transcript is much lower than the 5'-transcript. To increase the expression of the 3'- transcript of the *cip-cel* operon, we employed CRISPR-Cas9 editing system to insert a synthetic promoter (P4) and an endogenous promoter (P2) within *cip-cel* operon in *Clostridium cellulolyticum*. Both engineered strains increased the transcript abundance of downstream polycistronic

genes and enhanced *in vitro* cellulolytic activities of isolated cellulosomes. Compared to the control strain, both engineered strains could degrade more cellulose and demonstrated a greater growth rate and a higher cell biomass yield. Our strategy, editing regulatory elements of catabolic gene clusters, provides new perspectives on improving cellulose bioconversion in microbes.

Earlier studies have found that the accumulation of cellobiose could inhibit both cell growth and cellulase productions in the cellulolytic clostridia bacteria, such as *Clostridium thermocellum* and *C. cellulolyticum*, which would further decrease the efficiency of cellulose bioconversion. To overcome it, two strategies were applied to release the carbon catabolite repression caused by cellobiose. First, an exogenous β-glucosidase gene from *C.cellulovorans* was integrated into the upstream of the lactate hydrogenase gene (*ccel_2485*) of *C. cellulolyticum* genome for enhancing the enzymatic bioconversion of cellobiose to glucose. We found that the engineered strain could degrade 12% more cellulose than the WT at the final time point, accompanied with 25% more ethanol production. Second, the regulator for carbon catabolite repression (CCR) was inactivated in *C. cellulolyticum*. However, the mutant could not utilize the cellulose anymore, indicating that inactivation of CCR regulator is not an effective strategy for releasing the repression of cellobiose. Together, the integration of the exogenous β-glucosidase gene in the genome provides new perspectives on improving cellulose bioconversion in *C. cellulolyticum*, and also provides a new potential site in the genome of *C. cellulolyticum* for future integration and engineering. Finally, microbial decomposition of soil organic carbon (SOC), which are mainly derived from lignocellulose, has a strong impact on future atmospheric greenhouse gas

concentrations, which serve as important feedbacks to climate warming. However, the underlying decomposition mechanisms remain poorly understood. In order to understand the microbial mechanisms of lignocellulose decomposition and how the active microbes respond to current global warming, we identified active taxa responsible for carbon (C) degradation in temperate grassland subjected to experimental warming. Using a stable-isotope probing incubation experiment with ^{13}C-labeled straw to simulate grassland litter, a total of 56 active amplicon sequence variants (ASVs) were detected only in the warmed samples. Many ASVs belonged to fast-growing bacteria such as *α-Proteobacteria*, *Bacillales*, *Actinobacteria*, and *Bacteroidetes*, which were further verified by our observation that their relative abundances were increased ($p <$ 0.050) by warming over consecutive seven years. Interestingly, warming increased the phylogenetic diversity of active bacterial communities and β-diversity among active bacterial communities. The carbon-degrading potentials of the active bacterial communities were also stimulated by warming. In summary, these results should provide essential support to future field and global scale simulations and enable more accurate predictions of feedbacks between climate change and carbon cycling.

Overall, this dissertation provides valuable insights into engineering *C. cellulolyticum* for improving its lignocellulose bioconversion efficiency. Our strategies can be applied for engineering other clostridial cellulolytic bacteria, such as *C.cellulovorans* and *C.thermocellum*, to improve their lignocellulose bioconversion efficiency. Additionally, the newly identified active lignocellulose degraders in temperate grassland may also provide new insights into finding new industrial potential strain to produce lignocellulosic biofuels.

Keywords: Lignocellulose; Cellulosome; Lignocellulose/Cellulose degradation; Microbial engineering; Bioconversion; stable isotope probing; active communities

Chapter 1: **Introduction**

1.1 Reasons for production of sustainable biofuels

With the fast development of society and the global economy, current global energy mainly relies on fossil fuels including petroleum, natural gas and coal, which have finite reserves on the earth (Jahirul, Rasul et al. 2012). Based on the BP statistical review of world 2020, the primary energy source in the world is still the fossil fuels, which accounts for 84.32% of total energy supply (Looney 2020). However, combustion of fossil fuels can release large amounts of CO_2 into the atmosphere, which is thought to be directly associated with global warming and further aggravates environmental issues (Naik, Goud et al. 2010, Liao, Mi et al. 2016). In 2014, the global CO_2 emissions from fossil fuels were 9.8 Gt and emission rates may double by 2050 if current upward trend continues (Liao, Mi et al. 2016). Considering the limitation of fossil fuels and severe environmental issues brought by burning of fossil fuels, sustainable and environmental-friendly energy sources for societies and industries have become urgent in recent years (Naik, Goud et al. 2010).

The broad definition of the biofuels is the fuels which are produced from bio-based materials including biomass, landfilled gas and even CO_2 by biological or non-biological processes (Liao, Mi et al. 2016). As the liquid biofuels have a higher energy destiny and are more compatible with current infrastructure, they are superior compared to other renewable energy forms including solar-based electricity and biogas (Liao, Mi et al. 2016). Currently, the most common liquid biofuels are the bioethanol and biodiesel generated from food crops, which is also called the first-generation biofuels (Bothast and Schlicher 2005, Naik, Goud et al. 2010, dos Santos Bernardes 2011). Since

the feedstocks (plants) used for production of biofuels use CO_2 for their growth, those biofuels can reduce the production of CO_2 and combustion of biofuels is thought to be carbon neutral and eco-friendly (Naik, Goud et al. 2010). Although now the production of the first-generation biofuels is commercial, it still brings up some debates to scientists. The main debate for the first-generation biofuels is that balance between food and fuels. With limited land and water usages for food production, it impacts the global food market and raises the food price, which finally increases the cost of the first-generation biofuels (LAURSEN 2005, Naik, Goud et al. 2010). Additionally, with the increasing production of the first-generation biofuels, whether those biofuels can keep the carbon balances and are friendly to the environment is questionable (Naik, Goud et al. 2010). Therefore, to have a better carbon balance and more economic production of biofuels, the second-generation biofuels were proposed (Simpson-Holley, Higson et al. 2007). Compared to the first-generation biofuels, the second-generation biofuels are produced from non-edible lignocellulosic materials (plant biomass). The lignocellulose is considered the most abundant sustainable carbon resource on the earth, and the annual production of lignocellulosic biomass is estimated to be over 200 billion tons in the world (Michelin and Maria de Lourdes 2013). In addition, the plant biomass is generated by the photosynthetic CO_2 fixation, and therefore the biofuels produced from lignocellulose are still carbon neutral and eco-friendly energy sources. Although the second generation biofuels are not commercial at this time, it is anticipated that the second generation biofuels do not compete with the food and have better performance on reducing the production of CO_2 compared to the first generation biofuels (Naik, Goud et al. 2010).

1.2 Key challenges and new insights for bioconversion of lignocellulose to biofuels

For the plant biomass, over 67% of its dry mass is in the form of lignocellulose (Kuhad and Singh 1993). The lignocellulose is composed of three major components: cellulose, hemicellulose and lignin (Kumar and Sharma 2017). As the main constituent of the cell wall in the plant, the cellulose is a long-chain linear polysaccharide of repeating D-glucose units linked by β-1,4-glycosidic bonds, which is the most abundant organic polymer in the cell wall of plant cells and is the major energy and carbon storage material in plants. Sometimes, the cellulose can represent up to 50% of the cell wall in some types of plants (Dhyani and Bhaskar 2018, Liu, Li et al. 2020). As the second most abundant polymer and a connecting link between cellulose and lignin, hemicellulose is a heterogenous group commonly consisting of mainly pentoses, such as xylose, arabinose and mannose, and some hexoses, such as glucose and galactose. Compared to cellulose and hemicellulose, lignin is most complicated and chemically recalcitrant, which lacks a defined primary structure and is composed of aromatic and heterogeneous crosslinked polymers with different bonds including hydroxyl bonds and methoxy bonds. Lignin fills up the empty spaces between cellulose and hemicellulose in the cell wall and provides structural support in plant cell walls (Liao, Mi et al. 2016, Dhyani and Bhaskar 2018). Since the lignin is resistant to degradation by microorganisms, it is necessary to destroy it (delignification) which can enhance the hydrolysis of polysaccharides to fermentable sugars by relative hydrolysis enzymes. The conversion of lignocellulose can be processed by both chemical and biological methods. Compared to the chemical strategy, the biological conversion is more

preferred as it meets environmentally friendly requirements (Liu, Li et al. 2020). To convert the lignocellulose into fermentable sugars, two types of enzymes are required: cellulases and hemicellulases (Gupta, Kubicek et al. 2016). Endoglucanase, exoglucanase and β-glucosidase are the major components of the cellulases and play very important roles in hydrolysis of the cellulose in nature (Morag, Halevy et al. 1991). Xylanases, as the most important hemicellulase, is indispensable for hydrolysis of hemicellulose molecules during bioconversion of the lignocellulose (Hemsworth, Johnston et al. 2015). In addition to the above two types of enzymes, peroxidases and laccases (ligninases) can also involve in the bioconversion process to remove the lignin and increase the efficiency of saccharification (Hemsworth, Johnston et al. 2015, Liu, Li et al. 2020). Now, it is agreed that a major limitation of commercial production of second-generation biofuels is the cost and efficiency of the hydrolysis enzymes mentioned above (Lynd, Weimer et al. 2002, Gruno, Väljamäe et al. 2004).

Normally, the bioconversion of lignocellulose contains three major steps, which are enzyme production, enzymatic hydrolysis, and fermentation processes. Based on production time of hydrolysis enzymes, the bioconversion of lignocellulose strategies can be categorized into two types: on-site and off-site saccharification (Liu, Li et al. 2020). If the production of hydrolysis enzymes is coupled with a hydrolysis process, the strategy is called on-site saccharification. If the production of hydrolysis enzymes is prior to the hydrolysis process, and prepared under specific conditions for future saccharification systems, this strategy is called off-site saccharification.

The off-site saccharification strategies include separate hydrolysis and fermentation (SHF), simultaneous saccharification and fermentation (SSF), and nonisothermal

simultaneous saccharification and fermentation (NSSF) and simultaneous saccharification and co-fermentation (SSCF) (Lynd, Weimer et al. 2002). For SHF, the production of hydrolysis enzymes, enzymatic hydrolysis and fermentation these three steps are completed in independent units. The advantage for SHF is that the enzymatic hydrolysis step and fermentation step can have their own best optimal temperature. However, with the accumulation of fermentable sugars, the hydrolysis enzymatic activity is inhibited in the process unit, which is a major problem for the SHF (Philippidis, Smith et al. 1993, Gruno, Väljamäe et al. 2004). Compared to the SHF, the SSF, which combines saccharification (enzymatic hydrolysis) and fermentation steps together, is better as it has lower cost and contamination risk (Galbe and Zacchi 2002, Öhgren, Bura et al. 2007). However, the SSF still faces challenges that most fermentation microorganisms have a lower optimum temperature than the hydrolysis enzymes. To overcome this, NSSF was designed in which the saccharification and fermentation are processed at separate bioreactors at different temperatures but still completed simultaneously, which has high yield of ethanol and less enzyme input. To further increase the fermentation efficiency, SSCF sets out to co-fermentation of the hexose and pentose in a single bioreactor, which is evolved from the SSF (Chandrakant and Bisaria 1998). For SSCF, the microorganisms can ferment hexose and pentose to biofuels at same time. In fact, the SSCF strategy has been applied for commercial biofuels production (Olofsson, Palmqvist et al. 2010, Tang, Zhao et al. 2011). It was reported that if the sugar cost is more than $100 per ton sugar, the second-generation biofuels are not feasible (Chandel, Albarelli et al. 2019). However, the cellulase production is about $10 per kg, which finally causes the cost of sugar production to be

as high as $250 (Klein‐Marcuschamer, Oleskowicz‐Popiel et al. 2012, Liu, Li et al. 2020). Although many efforts have been made to enhance the production efficiency of the cellulase and reduce the cost, the cost of production of cellulase is still the major problem which restricts the development of the bioconversion of lignocellulose for off-site strategies (Taha, Foda et al. 2016, Lynd, Liang et al. 2017, Liu, Li et al. 2020).

In order to further reduce the cost of cellulase production and increase the efficiency of bioconversion of lignocellulose, the on-site saccharification strategies, which include consolidated bioprocessing (CBP) and consolidated bio-saccharification (CBS), were proposed. The concept of CBP was first proposed by Lynd's group in 2002, in which steps for hydrolase production, saccharification and fermentation were combined into a single step and occur in a single reactor (Lynd, Weimer et al. 2002, Lynd, Van Zyl et al. 2005). Compared to those off-site approaches, as CBP integrates the enzyme production step, the CBP has the potential of lower production cost and higher conversion efficiency (Lynd, Van Zyl et al. 2005). Therefore, the ideal CBP-microorganism should have specific traits including highly efficient cellulase production, highly cellulose degradation rate and high yield of interesting end-fermentation products (Liu, Li et al. 2020). However, there is no such CBP-enabling microorganism found in nature. To achieve the CBP, Lynd et al proposed two strategies, the natural strategy and recombinant strategy (Lynd, Van Zyl et al. 2005). The natural strategy involves engineering the naturally cellulolytic microorganisms to improve their ability in formation of the end-fermentation products, and the recombinant strategy involves expressing heterologous cellulolytic systems in non-cellulolytic microorganisms which exhibit superior fermentation ability. Although *E. coli* and *Saccharomyces cerevisiae*

have been reported to convert cellulose and xylan to ethanol by recombinant engineering (Bokinsky, Peralta-Yahya et al. 2011, Fan, Zhang et al. 2012, Matano, Hasunuma et al. 2012), the recombinant strategy is challenging. First, the huge burden of expression of the cellulolytic systems limits the growth of the recombinant cells (Olson, McBride et al. 2012). Second, it has been reported that recombinant cells still need additional nutrients for growth, which increases the cost (Liu, Li et al. 2020). Therefore, the nature strategy is more feasible for the practical application. Based on the CBP strategy, consolidated bio-saccharification was proposed, in which the fermentable sugars or products produced by the natural cellulolytic microorganisms, especially for the cellulosome-producing bacteria, could be further converted to biofuels by another microorganism or methods (Zhang, Liu et al. 2017, Liu, Liu et al. 2019, Liu, Li et al. 2020). Compared to the CBP strategy, the CBS pays more attention to the production of fermentable sugars other than end metabolites, and therefore the ideal CBS-enabling microorganisms should have very high efficiency in converting the cellulose to fermentable sugars (Liu, Li et al. 2020). Although the on-site approaches including CBP and CBS seems to overcome the defects in off-site approaches, we still have a long way to go before the final industrial application (Balan 2014, Liao, Mi et al. 2016), such as low-efficiency substrate utilization, carbon catabolite repression to fermentable sugars and microbial growth inhibition.

1.3 Model mesophilic cellulolytic bacteria: *Clostridium cellulolyticum*

In nature, the truly cellulolytic bacteria (more than 80%) are mainly distributed in two orders, one aerobic and one anaerobic: *Actinomycetales* (Actinobacteria) and *Clostridiales* (Firmicutes) (Schwarz 2001). Compared to the cellulolytic aerobes,

anaerobic cellulolytic bacteria have few advantages for application in industry, such as no need for air agitation and a lower chance of contamination (Lynd, Van Zyl et al. 2005)(32).

As a non-ruminal cellulolytic mesophilic bacterium, *Clostridium cellulolyticum* strain H10 is isolated from decayed grass in France (Petitdemange, Caillet et al. 1984). The most optimum growth temperature for H10 is at 34℃ with a range from 25-45 ℃. When facing an unfavorable or harsh environment, H10 can produce spherical terminal spores (diameter: 1.5 μm), which can resist temperature as high as 100 ℃ for 30 min (Petitdemange, Caillet et al. 1984, Li, Xu et al. 2014). In addition to the cellulose, H10 can use a variety of soluble simple sugars: (i) it was observed that H10 can grow well with cellobiose, D-glucose, xylose, L-arabinose and fructose as sole carbon source; (ii) H10 grows not well with galactose, mannose or ribose as the sole carbon source (Petitdemange, Caillet et al. 1984). Other sugars, including sucrose, lactose, glycerol or glycogen cannot be utilized by H10. It was found that H10 preferred to use the cellobiose than other sugars (Petitdemange, Caillet et al. 1984).

In fact, the most attractive trait of the *C. cellulolyticum* strain H10 is that it can directly convert the lignocellulose to the valuable biofuels: ethanol and H2, which meets the basic requirement for being a CBP-enabling bacterium. Like other aerobic cellulolytic bacteria, *C. cellulolyticum* can secrete numerous individual and extracellular cellulases (90 putative glycoside hydrolases) to hydrolyze the cellulose (Blouzard, Coutinho et al. 2010). However, in *C. cellulolyticum* , a unique extracellular multi-enzyme complex named cellulosome can also be secreted to hydrolyze the cellulose, which is more efficient in hydrolyzing the cellulose than the individual cellulase (Schwarz 2001). In *C.*

cellulolyticum , most major components of the cellulosome are encoded by genes located in a large cluster spanning 26 kb, named *cip-cel* operon (Gal, Pages et al. 1997). The first gene in this operon is a scaffolding gene named *cipC*, which contains one cellulose-binding domain (CBD), two hydrophilic domains (HLDs) and eight cohesion domains. The CipC does not have any catalytic activity and only carries out the scaffolding function for the cellulosome. The CBD allows the binding between the cellulose and cellulosome, and the eight cohesions permit the binding between CipC and cellulases with dockerin domains. The HLD is also called X2-module. However, the function of the X2-module in the cellulosome is still elusive (Desvaux 2005). After the *cipC* gene, 11 genes are following (*cel48F- cel8C-cel9G-cel9E-orfX-cel9H-cel9J-man5K-cel9M-rgl11Y-cel5N*), most of which encode the cellulases except for the *orfX* and *man5K* genes. In addition to the genes located in the *cip-cel* gene operon, some cellulases, including cel5A and cel5D, which are encoded by genes located outside of the *cip-cel* operon can also bind to the CipC scaffolding protein. However, genes located in the *cip-cel* operon are essential for the formation of the cellulosome and its ability in efficiently degrading crystalline cellulose (Maamar, Valette et al. 2004, Xu, Li et al. 2014). Among cellulosome-producing bacteria, only the cellulosomes from *C. thermocellu*m and *C. cellulovorans* have been studied well. In these bacteria, protuberances are found at the cell surface by scanning electron microscopes, which are considered as the cellulosomes. In *C. cellulolyticum*, we observed obvious protuberances on the cell surface in our engineered strains (Tao, Xu et al. 2020). However, the mechanism on how the cellulosome binds to the cell surface is still unclear in *C. cellulolyticum*.

Considering the *C. cellulolyticum* can directly convert the lignocellulose to biofuels and is a cellulosome producing bacteria, *C. cellulolyticum* is considered as a good candidate for CBP/CBS candidate (Liu, Li et al. 2020). In order to further improve its efficiency on bioconversion of cellulose to biofuels, previous studies tried to inactivate the lactate dehydrogenase gene and sporulation gene for blocking the production of lactate and impediment of entering to the stationary phase (Quast, Pruesse et al. 2012, Li, Xu et al. 2014). However, due to the limitation of the traditional method for genome manipulation, deeply metabolic engineering could not be realized at that time (Xu, Li et al. 2015)(50). With the development of the CRISPR-Cas9 system (Xu, Li et al. 2015)(50), more strategies can be applied and tried in *C. cellulolyticum* including editing the responsive element of the *cip-cel* operon or integration of an exogenous gene into the genome. On the other hand, the regulatory mechanisms for cellulose utilization in *C. cellulolyticum* are still unclear. By using the transcriptional fusion approach, only one promoter was found to regulate the expression of *cip-cel* operon (Abdou, Boileau et al. 2008, Xu, Huang et al. 2015). More interestingly, a catabolite-responsive element was also found near the promoter region of the *cip-cel* operon, indicating that the regulator (CcpA) of carbon catabolite repression (CCR) may involve in the regulation of the expression of the *cip-cel* operon (Abdou, Boileau et al. 2008, Xu, Huang et al. 2013, Xu, Huang et al. 2015). Therefore, new engineering strategies, including (i) promoter insertion for enhancing the expression of *cip-cel* operon, (ii) expression of heterogeneous genes for improving cellulose degradation efficiency and (iii) inactivation of the CCR regulator for releasing the carbon repression, could be applied to *C. cellulolyticum* for improving the efficiency of lignocellulose bioconversion.

1.4 Microbial degradation of lignocellulose in soil

In the entire biosphere, microbial communities carry out integral and unique roles in decomposition of soil organic carbon (SOC) and balance of the global carbon cycle. All of SOC found in the soil is primarily derived from plant biomass (lignocellulose) (Kuzyakov and Domanski 2000), and the decomposition of the plant biomass is an important global carbon sink (Myneni, Dong et al. 2001). In natural soil, the decomposition of lignocellulose is influenced by the structure and function of active microbial communities, and the structure and function information is important parameters for current terrestrial carbon cycling models (Strickland and Rousk 2010, Cleveland, Reed et al. 2014). However, due to the limitations of culture-dependent methods and the complexity of soil communities, the composition and ecology of those active lignocellulose decomposers in the soil are still elusive. To identify the active lignocellulose decomposers and resolve, a culture-independent method, such as stable isotope probing (SIP), should be adopted, which is better in reflecting the real in-situ conditions.

On the other hand, in the past 30 years, the earth surface mean temperature has increased 0.2 °C per decade (Hansen, Ruedy et al. 2010), which is a result from anthropogenic emission of greenhouse gases including the burning of fossil fuels (IPCC 2007). Due to the warming, the microbial respiration is increased since the temperature is a primary driver of metabolic rates (Brown, Gillooly et al. 2004, Xue, Yuan et al. 2016, Guo, Feng et al. 2018). If the respiration rate increases more than the photosynthesis rate, more greenhouse gases will be released to the atmosphere and intensify the warming, causing a positive feedback of warming. In fact, previous

studies have shown that soil carbon is highly vulnerable and responds rapidly to the warming (Zhou, Xue et al. 2012, Xue, Yuan et al. 2016, Tao, Feng et al. 2020) and therefore a mechanistic understanding on how microbes respond to climate warming can provide new insights into assess the future C balance and model prediction. However, the microbial mechanisms of SOM decomposition under warming remain elusive since most previous work mainly focused on the relationships between temperature and CO2 emissions (Cheng, Zhang et al. 2017).

Given that the SOC in the soil is primarily composed of lignocellulose, and the microbial mechanisms of SOM decomposition under warming are very important for future carbon prediction, SIP microcosm-based experiments are the best choice to investigate the composition and degradative potential of hemicellulose, cellulose and lignin-degrading populations in the soil and provide new insight into understanding the microbial mechanisms of SOM decomposition under warming.

1.5 Focus and objectives of this work

As discussed above, the utilization of lignocellulosic biofuels will have a positive effect on energy sustainability and environment protection. However, the cost for production of lignocellulosic biofuels is still higher than the conventional fuels. In order to make lignocellulosic biofuels competitive with fossil fuels, it is imperative for microbiologists and engineers to improve the efficiency of the biofuels production and save the cost. As a model organism of mesophilic cellulolytic Clostridia and a CBP/CBS-enabling candidate, *C. cellulolyticum* has the potential to be engineered as a real industrial strain and can also bring instructional significance and application values for future industrial biofuels production. However, due to the limitation of traditional genome manipulation

methods, functional characterization of important cellulosomal components in *C. cellulolyticum* and engineering for *C. cellulolyticum* have been hindered. With the development and application of bacterial CRISPR-Cas9 system, this study aimed to : (i) in-vivo functional characterization of the hydrophilic domain in the cellulosome of *C. cellulolyticum* ; (ii) insertion of strong promoters into the regulation region of *cip-cel* operon for improving lignocellulose bioconversion efficiency; (iii) integration of exogenous gene into the genome of *C. cellulolyticum* for improving lignocellulose degradation efficiency; (iv) inactivation of the CCR regulator for releasing the repression of cellobiose accumulation and further improving lignocellulose degradation efficiency.

On the other hand, the microbial degradation of lignocellulose is not only important for the biofuels production but also plays an important role in global C cycling. With current global warming, whether the C exchange between terrestrial and atmosphere is positive or negative feedback to warming is still debating. Therefore, understanding the identities of active lignocellulose degraders and their degradation mechanisms under warming can provide new insights into predictions of feedback between climate change and carbon cycling. Additionally, identification of active lignocellulose degraders may also find new strains in degrading the lignocellulose, which may also provide new insight into future biofuels production. Applying the cutting-edge method, the quantitative stable isotope probing, this study also aimed to identify active lignocellulose degraders in grassland soil and how those active degraders respond to global warming.

Chapter 2 presents the functional characterization of the hydrophilic domains named X2-1 and X2-2 in the cellulosome of *C. cellulolyticum* H10. First, the X2-1 and X2-2 domains were inactivated by deleting the conserved motifs of X2 modules through CRISPR-Cas9 editing method. Second, both X2 mutants were characterized at the phenotypic, physiological and RNA levels. Then, in-vitro and in-vivo enzymatic assays were performed to further understand and characterize the in-vivo function of X2 modules.

Chapter 3 presents the insertion of two different constitutively active promoters in front of the *cel9H* gene, by which the transcription of the six downstream genes (*cel9H-cel9J-man5K-cel9M-rgl11Y-cel5N*) can be increased independently of the *cip-cel* gene operon. First, the P2 and P4 promoters were inserted within the *cip-cel* gene operon by CRISPR-Cas9 editing method respectively. Second, the efficiencies of cellulose bioconversion were evaluated for both engineered strains. Finally, this strategy was also applied for previously engineered strain for further increasing its cellulolytic ability.

Chapter 4 presents another two strategies for improving the cellulose hydrolysis abilities in *C. cellulolyticum* . The first one was that the integration of an exogenous β-glucosidase gene from *C.cellulovorans* into the genome of *C. cellulolyticum* for improving its enzymatic conversion of cellobiose to glucose; the second one was that inactivation of CCR regulators to release the carbon catabolite repression caused by the accumulation of cellobiose. Finally, the cellulose hydrolysis efficiencies were evaluated for both engineered strain (the β-glucosidase gene integration) and the mutant strain (inactivation of CCR regulator).

Chapter 5 focused on identification of active bacteria responsible for lignocellulose degradation in temperate grassland and how they respond to current global warming. To begin with, the 0-15 cm fresh soil was collected from a warming field and incubated with ^{13}C-labeled straw which simulated grassland litter for seven days in the lab. Then, coupling with the sequencing and qPCR, the identities of active degraders were determined in both warmed and un-warmed soil samples. Finally, a potential microbial mechanism of lignocellulose decomposition under warming was proposed by analyzing the compositions of the active communities, determination of the abundance of C-degrading genes and structural equation modelling analysis.

Chapter 2 **In vivo functional characterization of X2 modules in Clostridium cellulolyticum CipC scaffoldin**

2.1 Abstract:

As an important component of the cellulosome or free cellulase, the X2 module is widely distributed in cellulolytic *Clostridia* or other *Frimicutes* bacteria. Although some studies inferred the functions of the X2 module based on the structure analysis or in vitro biochemical assays, the in vivo biological functions of the X2 module are still elusive. Here, using *Clostridium cellulolyticum* as the model strain, we employed CRISPR-Cas9 editing to create dual X2 modules mutant (△X2-NC) by deleting the conserved motif (NGNT) of X2 modules. The deletion of the conserved motif of X2 modules inactivated the functions of the X2 modules but could keep the structure integrity and the basic functions of CipC protein. Compared to the wild type strain (WT), there were no obvious changes for △X2-NC grown on cellobiose. However, the △X2-NC spent three times longer to enter into the exponential phase than the WT and could only degrade ~ 80% of the cellulose at final, which was 95% for the WT. Additionally, the released total soluble sugars in △X2-NC was decreased by 63% compared to the WT. The in vivo adhesion assay found that the average relative adhesion ability in △X2-NC was decreased by 70% when compared to WT, indicating that the biological function of the X2 module was associated with the binding affinity between the cells and its cellulose substrate. Together, our findings highlighted the in vivo biological role of X2 module in increasing cellulose degradation efficiency by

enhancing the binding affinity between cells and cellulose, which provides new perspectives on increasing cellulose degradation activity in cellulolytic microbes.

2.2 Introduction:

As a sustainable and carbon-neural energy source, biofuels are one of the best alternative energy forms for replacing fossil fuels (Farrell, Plevin et al. 2006, Naik, Goud et al. 2010). Compared to other biomass resources including food grade sources, municipal solid waste and algae, the non-edible lignocellulosic feedstock is a more promising source of material for biofuel production as it is cheap, abundant and renewable (Naik, Goud et al. 2010, Liao, Mi et al. 2016). Although there are many platforms for bioconversion of lignocellulose into biofuels, such as nonisothermal simultaneous saccharification and simultaneous saccharification and co-fermentation, the common barrier for these bioconversion processes is the high cost for cellulase production (Balan 2014, Lynd 2017). Therefore, a new bioconversion process named consolidated bioprocessing (CBP), in which a single microorganism is used for hydrolase production, saccharification and fermentation in a single step in one bioreactor, is considered as a promising technology for bioenergy production at a lower cost (Lynd, Weimer et al. 2002, Lynd, Van Zyl et al. 2005, Wyman and Davison 2012). As a model mesophilic clostridial species, *Clostridium cellulolyticum* can hydrolyze cellulose or hemicellulose and ferment these hydrolysis products to ethanol and other organic acids, which enables it as a potential candidate strain for CBP (Desvaux 2005). Like other cellulolytic *Clostridia*, *C. cellulolyticum* possesses an extracellular enzymatic complex termed cellulosome, which make *C. cellulolyticum* degrade crystalline cellulose more efficiently (Gal, Pages et al. 1997, Shoham, Lamed et al.

1999). The pivotal component of the cellulosome in *C. cellulolyticum* is a scaffolding protein/integrating protein encoded by the *cipC* gene, to which up to 8 different catalytic cellulases can bind. Without catalytic activity, the CipC scaffoldin contains eight cohesion modules (type I), one carbohydrate binding module (CBM) and two X2 modules (Figure S2.1) (Gal, Pages et al. 1997, Schwarz 2001). The cohesion module allows the binding from cellulases with type I dockerin (Fierobe, Pagès et al. 1999). The CBM belonging to the family IIIa (CBM3a) carries out the binding between the entire cellulosome and cellulosic substrate (Boraston, Bolam et al. 2004). X2 modules are two hydrophilic domains, each of which contains ~100 amino acid residues (Mosbah, Belaïch et al. 2000). In addition to existing in *cipC* gene in *C. cellulolyticum*, the X2 module is also found in other scaffoldin proteins from other cellulosome-producing bacteria, such as *Clostridium thermocellum* (Lamed and Bayer 1988), *Clostridium cellulovorans* (Doi, Goldstein et al. 1994), and even existed in free cellulase enzyme from some non-closrtidial cellulolytic bacteria, such as *Lachnoclostridium phytofermentans* (Ravachol, Borne et al. 2015, Vita, Borne et al. 2019)18,19 and *Paenibacillus polymyxa* (Pasari, Adlakha et al. 2017). These suggest that the X2 modules are widely spread and may serve an important role in the biodegradation of lignocellulosic biomass. Based on the structure and *in vitro* biochemistry assays of X2 module, it belongs to the immunoglobulin superfamily and has been predicted to be associated with localization of cellulosome, binding to cellulose based and increasing enzymatic activities for free cellulase (Mosbah, Belaïch et al. 2000, Kosugi, Amano et al. 2004, Chanal, Mingardon et al. 2011, Ravachol, Borne et al. 2015, Pasari, Adlakha et al. 2017). However, little is known about its real in vivo biological functions and

significances.

As an unknown functional module, it is challenging to study X2 modules because they have no catalytic activities, which brings difficulties for in vitro biochemical tests, physical and biochemical analyses. With the development of CRISPR-Cas9 nickase-based genome editing method (Xu, Li et al. 2015), we are now able to edit certain genes which can not be realized by traditional methods. Here, using the *C. cellulolyticum* as the model strain, we aimed to understand the in vivo biological functions of the X2 modules of the cipC gene in *C. cellulolyticum*. By alignment and Cas9n-based editing method (Xu, Li et al. 2015), we found and deleted a conserved motif from both X2 modules in *C. cellulolyticum* and successfully created a dual mutated X2 variant. We systematically characterized the X2 variants from transcriptional and physiological aspects. Based on the phenotypic results and previous predicted functions, we also designed the in vivo and in vitro enzymatic assays to determine the biological functions of X2 modules in *C. cellulolyticum*. Our results demonstrate that the conserved motif of X2 module is a good site for studying the in vivo functions of the X2 modules and the in vivo biological functions of X2 modules can influence the cellulose degradation efficiencies in *C. cellulolyticum* by changing the binding affinity between cells and their cellulose substrate.

2.3 Method & Materials

2.3.1 Bacterial strains and plasmid construction

Strains and plasmids used in this study are listed in Table 2.1. The △X2-N and △X2-C variants were constructed in our previous study (Xu, Li et al. 2015). The dual X2 modules mutant (△X2-NC) was constructed using the pCas9n-X2-N-delete-donor to

delete the conserved motif of the X2-N module in △X2-C. The plasmid for inactivation of the X2-N module in △X2-C was constructed as previously reported (Xu, Li et al. 2015). In brief, for each X2 module, its corresponding variant was created (Xu, Li et al. 2015) by deleting the 12-bp DNA sequence exactly coding for the conserved Asn-Gly-Asn-Thr motif. To do so, the 23-bp target site that Cas9 nickase acted on was partially overlapped with the deletion region such that the customized donor template can direct the nick repair in the genome to make intended changes (Figure 2.1a). Finally, the N-terminal and C-terminal X2 modules (X2-N and X2-C) were mutated precisely in the chromosome, generating △X2-N and △X2-C variants respectively. Sequential mutation of the N-terminal X2 module in △X2-C variant created a dual mutated X2 variant, named △X2-NC (Figure 2.1a). All mutations were verified by amplicon sequencing (Figure 2.1b).

E. coli DH5α strain (Invitrogen,Carlsbad, CA) was used for cloning. Rosetta™(DE3) (Invitrogen, Carlsbad, CA) was used for protein expression. Sequences of X2-C module and CBM3a, amplified from the genome of WT strain, were separately cloned into pET-28a vectors with an C-terminal 6 × His tag respectively, generating corresponding vectors, pET-X2-C and pET- CBM3a. The pET-△X2-C vector was generated from the pET-X2-C by Q5® Site-Directed Mutagenesis Kit (New England Biolabs, Ipswich, MA).

Table 2. 1 List of plasmids and strains used in this study.

Strain or plasmid	**Phenotype or genotype**	**Source or reference**
Strains		

C. cellulolyticum H10	ATCC 35319	(Xu et al., 2015)
△X2-N	Deletion of NGNT in X2-N module	(Xu et al., 2015)
△X2-C	Deletion of NGNT in X2-C module	(Xu et al., 2015)
△X2-NC	Deletion of NGNT in X2-C & X2-N module	This study
C. thermocellum LQR1	ATCC 35609	ATCC
Plasmids		
pCas9n-X2-C-delete-donor	Cmp^r in *E.coli* and Tmp^r in *C. cellulolyticum* H10	(Xu et al., 2015)
pCas9n-X2-N-delete-donor	Cmp^r in *E.coli* and Tmp^r in *C. cellulolyticum* H10	(Xu et al., 2015)
pET- CBM3a	Kan^r in *E.coli*	This study
pET- X2-C	Kan^r in *E.coli*	This study
pET-△X2-C	Kan^r in *E.coli*	This study

2.3.2 Media and culture conditions

LB medium with 35 μg/ml chloramphenicol or 50 μg/ml kanamycin was used for pCas9n-X2-N-delete-donor cloning or pET-X2C/pET-△X2C/pET-CBM3a cloning. Complex modified VM medium supplemented with 2.0 g/L yeast extract was used for reviving and transformation of △X2-C (Quast, Pruesse et al. 2012). Defined modified VM medium containing necessary vitamin solution and mineral solution was used for growth determination and omics experiments (Higashide, Li et al. 2011). All strains (△X2-N, △X2-C, △X2-NC and WT) were cultured with 5 g/L cellobiose or 10 g/L Avicel PH101 crystalline cellulose (Sigma-Aldrich) at 34°C anaerobically depending on the experiment. Solid VM medium with 1.0% (w/v) of Bacto agar (VWR) and 15 μg/ml thiamphenicol was used for selecting the △X2-NC mutant.

2.3.3 Transformation and verification of cellulosome structure integrity in mutant strains

The pCas9n-X2N-donor was firstly methylated by MspI Methyltransferase (New England Biolabs, Ipswich, MA) and then transformed into △X2-C by electroporation as described previously (Quast, Pruesse et al. 2012). The pET-X2N, pET-△X2N and pET-CBM3a were transformed into Rosetta™(DE3) according to the manufacturer's instructions. For each strain, the cellulosome fraction was isolated from 200 ml cellulose-grown culture at mid-exponential growth phase as previously reported (Tao, Xu et al. 2020).

2.3.4 Measurement of cell growth and remaining cellulose

For cellobiose as the carbon source, the growth of all strains (△X2-N, △X2-C, △X2-NC and WT) were determined by the optical density (OD) at 600 nm; for cellulose as the carbon source, the growth of all strains were estimated by total protein measurement as previously described (Quast, Pruesse et al. 2012). Each strain had three biological replicates. The remaining cellulose and released soluble sugar in the medium were measured by the phenol-sulfuric acid method (Hemme, Fields et al. 2011).

2.3.5 Cell-cellulose adhesion assay

All strains (△X2-N, △X2-C, △X2-NC and WT) were grown on 5 g/L cellobiose to same OD_{600}. For each strain, two groups were set up. One group was the experimental group, in which the cells were incubated with the “whatman” filter paper (cellulose) on a tube rotator for 1 h. The other group was as the control group, in which there were only cells rotating for 1 h. In the experimental group, the cells could attach to the filter paper through cellulosomes on the cell surface. After 1 h incubation, the OD_{600} of

planktonic phase in each group was determined. The difference of OD_{600} between experimental and control groups for each strain represented the strength of the adhesion capability in this strain. Each group contained three biological replicates. The relative cell adhesion capability for each strain was normalized by dividing by the strength of the adhesion capability in WT strain.

2.3.6 Microarray analysis

All strains (△X2-N, △X2-C, △X2-NC and WT) were grown in the defined VM medium with 10 g/L cellulose. Each strain was collected at mid-exponential growth phase and the collected cell pellets were immediately frozen with the liquid nitrogen. The total RNA extraction, microarray hybridization and microarray data analysis were performed as previously described (Tao, Xu et al. 2020). Each strain contained four biological replicates.

2.3.7 Expression and purification of X2-N, ⊿X2-N and CBM3a

For expression the X2-N or △X2-N module, the Rosetta™(DE3) containing corresponding vector was grown to an OD_{600} of 1.0-1.2 and then induced with 0.5mM isopropyl-d-1-thiogalactopyranoside (IPTG) for an additional 8 h at 37 °C. Then, the cells were collected by centrifugation and were resuspended in the lysis buffer containing 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 10 mM imidazole. After disruption by sonication and centrifugation at 12, 000 g for 30 min, the supernatant was loaded to Ni2+-nitrilotriacetate affinity resin (Qiagen, Hilden, Germany) equilibrated with 20 mm Tris-HCl pH 8.0, 150 mm NaCl. The X2-N or △X2-N protein was eluted with 20 mm Tris-HCl pH 8.0, 150 mm NaCl, 350 mm imidazole and further purified by buffer exchange to dilute the imidazole. For expressing the CBM3a module, the

Rosetta™(DE3) containing the pET-CBM3a was grown to an OD_{600} of 0.6-0.8 and then induced with 0.2mM isopropyl-d-1-thiogalactopyranoside (IPTG) for an additional 20 h at 16 °C. The remaining steps were the same as above. The structure of X2-N (PDB: 1EHX) was used as the template for construction of the structure of X2-C module by homology modeling in SWISS-MODEL (swissmodel.expasy.org).

2.3.8 The X2 module-cell wall binding assay

The total volumes of 20 ml of *E.coli* grown on LB medium, *Clostridium thermocellum* and *C. cellulolyticum* grown 5 g/L cellobiose VM medium were collected and centrifuged respectively. For each strain, the cell pellets were treated with NaN3/Ca2+ and washed three times with 50 mM Phosphate-buffered saline (PBS) buffer (Sigma Aldrich). Then, 50 μg purified X2-N, △X2-N and CBM3a proteins were incubated with the pellets of each strain respectively in the PBS buffer at 4 °C for 12 h. After centrifugation, the pellet of each group was resuspended in SDS-PAGE Protein Loading Buffer (Thermo Fisher) and boiled for 10 mins. Finally, the 6x-His Tag antibody (R930-25, Thermo Fisher) was used to detect the binding between proteins and cell wall by western blotting.

2.3.9 The X2 module-cellobiose binding assay

The total amount of 50 μg of X2-N, CBM3a and BSA (Thermo Fisher) were separately incubated with 20 mg Avicel PH101 crystalline cellulose in the reaction buffer containing 20 mM Tris-HCl pH 8.0 at 4 °C for 12 h. After centrifugation, the pellet from each group was washed with the reaction buffer for three times. Then the supernatant and the pellets were mixed with the SDS-PAGE Protein Loading Buffer and

boiled for 10 mins. Finally, SDS-PAGE gel was run to detect the binding between proteins and cellulose.

2.3.10 Isothermal Titration Calorimetry (ITC) assay

Titration calorimetry measurement was performed with a VP-ITC calorimeter (MicroCal, Northampton, MA) based on previously described (Duff Jr, Grubbs et al. 2011, Brown, Lord et al. 2013). In brief, 70 µM X2-N protein was titrated into 10 µM CBM3a protein with injection volume of 70 µL and constant stirring at 25 °C. The buffer containing 20 mM Tris–HCl pH 8.0 and 100 mM NaCl was used for the assay. Before the titration from X2-N to the CBM3a, two more titration runs, including titration with buffer alone (background) and titration from X2-N protein into the buffer (control), were performed as background and control groups. A one-site binding model was used for the analysis of binding interaction and calculation of association constant and the final calorimetric data analysis was carried out with Origin 8.0 software (MicroCal).

2.4 Results

2.4.1 Non-catalytic X2 modules of CipC scaffoldin contain a highly conserved Asn-Gly-Asn-Thr motif by in silico analysis

In order to study the in vivo biological function of X2 modules of the CipC scaffoldin protein, both whole sequences encoding these two X2 modules should be deleted from the *cipC* gene sequence. However, previous study found that the loss of only one X2 module sequence could lead to the failure of the cellulosome assembly in *C. cellulolyticum* (Maamar, Valette et al. 2004). As a result, we could not delete the whole X2 module sequence as the traditional way and needed to find out a conserved motif of

X2 module as the target site for deletion. The standards for the conserved motif was that: (i) the size of the sequence should be short enough which could ensure the integrity of the CipC scaffoldin protein; (ii) deletion of the sequence shouldn't cause the frameshift mutation of the *cipC* gene. By aligning the X2-like modules among different bacteria, we found a highly conserved and short motif (Asn-Gly-Asn-Thr) in the X2 module (Figure S2.2). Based on the structures of X2-N (PDB: 1EHX) and X2-C, the conserved motif (Asn-Gly-Asn-Thr) was located in an exposed loop in *C. cellulolyticum* (Figure 2.2). When deleting the conserved motif, the loop region was directly shortened, accompanied with a wider groove formed as demonstrated by the structure modeling. (Figure 2.2), indicating that the functions might also change accompanied with the structure change. Therefore, it was worthwhile to delete the conserved motifs in both X2-N and X2-C modules in *C. cellulolyticum* and examine whether the motif was important in functioning of X2 module and CipC scaffoldin.

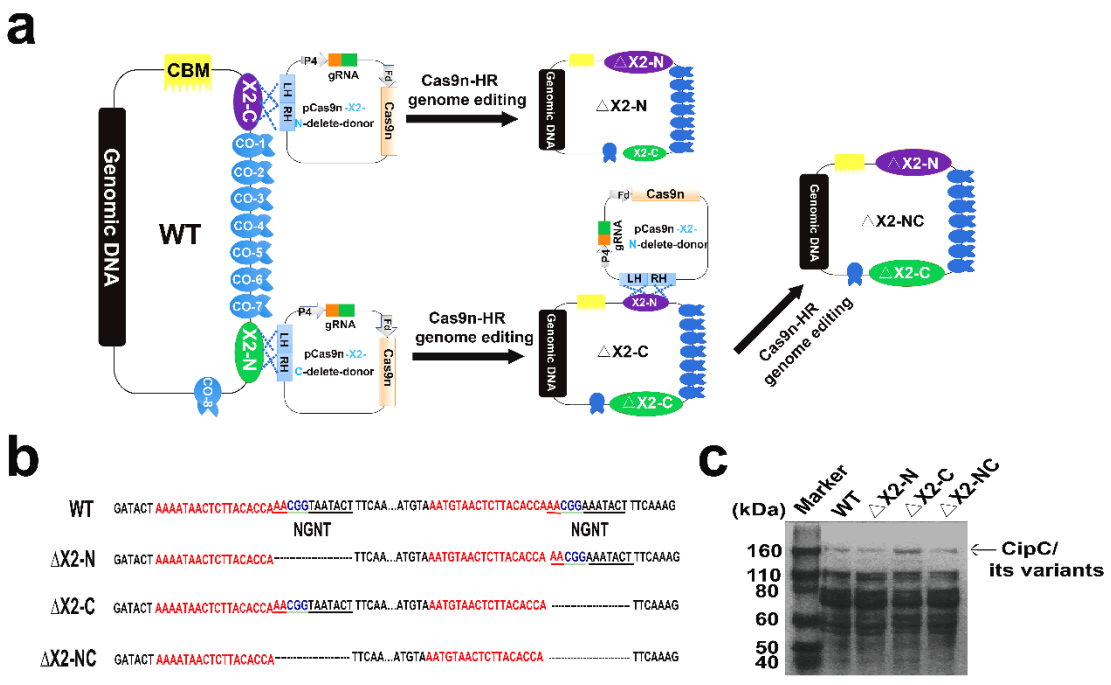

Figure 2. 1 Precise deletion of the conserved motif from the X2 modules could keep the structure integrity of the CipC protein. (a), an overview strategy for constructing the dual X2 module mutant by the Cas9 nickase-based genome editing. Both plasmids pCas9n-X2-C-delete-donor and pCas9n-X2-N-delete-donor were used for the △X2-NC mutant construction. (b), DNA sequence showing the deletion of conserved motif from X2 modules in the cipC gene. (c), SDS-PAGE analysis of cellulosomes extracted from WT and all mutant strains (15 μg protein/lane).

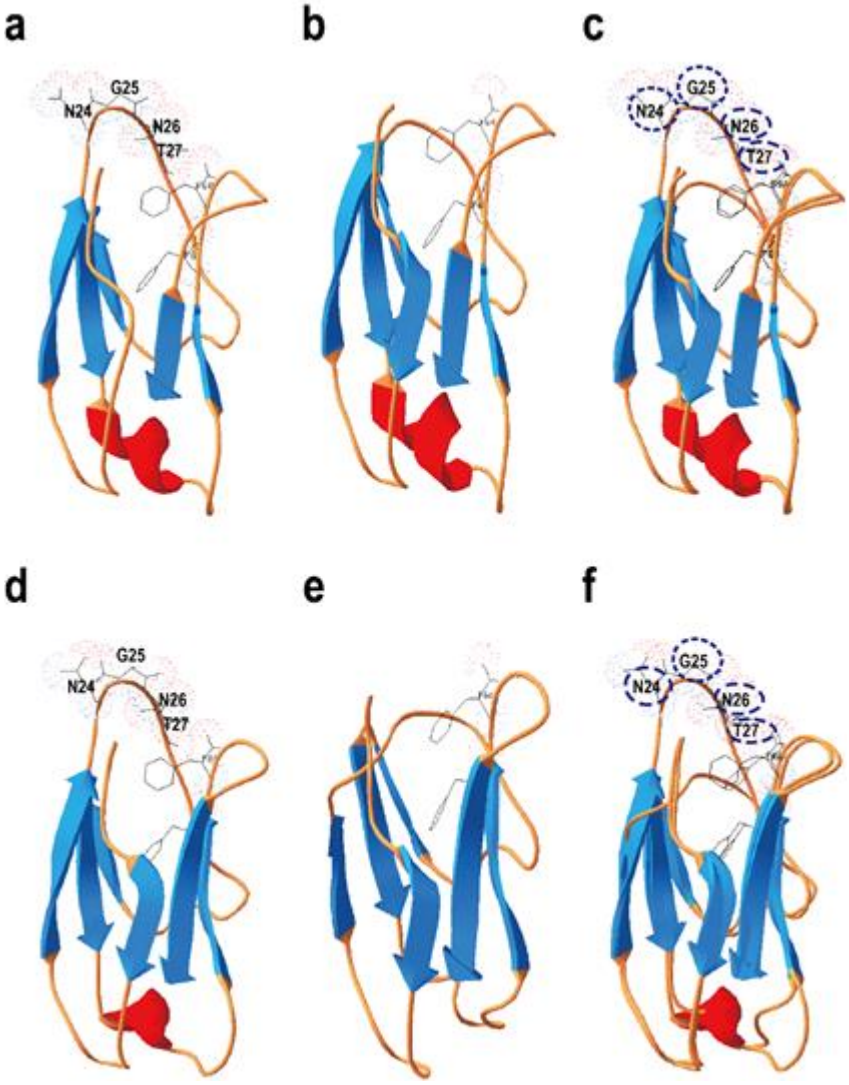

Figure 2. 2 Deletion of the conserved motif (NGNT) could lead to the conformation change of the X2 module. (a), the structure of the X2-N module protein; (b), the structure of △X2-N module in which the conserved NGNT residues were deleted. (c), structures overlapping between X2-N and △X2-N modules. (d) the structure of the X2-C module protein; (e) the structure of △X2-C module in which the conserved NGNT residues were deleted; (f) structures overlapping between X2-C and △X2-C modules.

2.4.2 Single and dual All X2 module modifications in CipC variants diminished physiological performances on cellulose

After successfully obtaining the dual X2 mutant (△X2-NC), cellulosome fractions from cellulose-grown WT and mutant strains were extracted to confirm that the deletion of the conserved motif would not destroy the structure and integrity of the cellulosomes. From the SDS-PAGE gel analysis, both single and double X2 variants presented the same gel pattern and band intensity as the wild type (WT) control strain (Figure 2.1C), indicating that the integrity and basic functions of cellulosome were not impeded in the mutant strains.

Based on the location of the X2 modules and structure modeling analysis, we assumed that X2 modules should be associated with cellulosome performance. Firstly, the cellobiose, as the soluble sugar model, was chosen as the carbon source to determine whether the inactivation of X2 module can influence the soluble sugar utilization. As a result, △X2-N, △X2-C, △X2-NC and WT did not present significant differences in cell growth on 5g/L cellobiose (Figure 2.3a); however, when grown on 10 g/L cellulose, △X2-N, △X2-C and △X2-NC took a much longer adaptation stage than WT before starting up massive cell propagation at a similar growth rate (Figure 2.3b). It suggested that the importance of X2 modules depends on insoluble cellulose substrate, rather than soluble cellobiose. Compared to WT, the release of soluble sugars in the △X2-N, △X2-C and △X2-NC were decreased by 28%, 40% and 63% respectively. Additionally, the amount of residual cellulose in △X2-NC was 15% higher than WT and approximately 10% higher than △X2-N and △X2-C, which indicated the inactivation of X2 modules can reduce the efficiency of cellulose degradation for cellulosome (Figure 2.3c, d &

Figure S2.3). This was consistent with previous in vitro study that the X modules in *C.cellulolovorans* could increase the cellulose degradation efficiency of the cellulosome (Kosugi, Amano et al. 2004). These results together suggested that the deletion of conserved motifs of X2 modules indeed disrupted the function of X2 modules and both X2 modules contributed to physiological performances on cellulose.

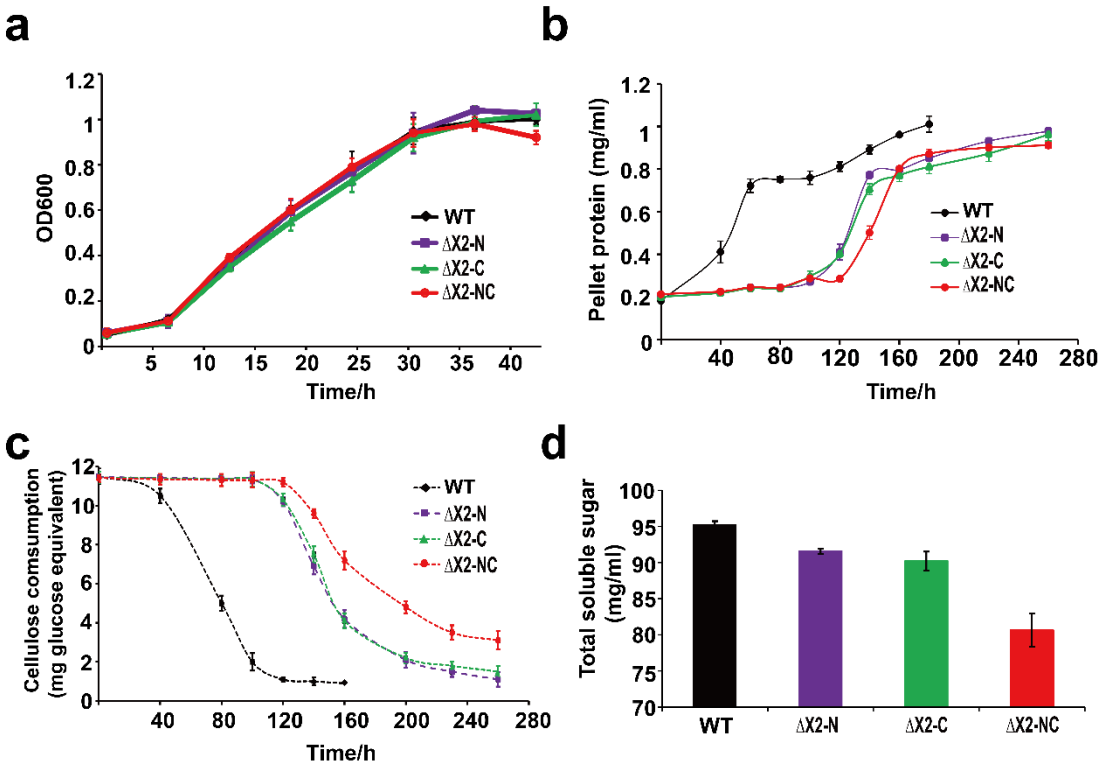

Figure 2. 3 Disruption of X2 modules increased the lag phase and decreased the cellulose degradation efficiency when mutants grown on cellulose. (a), growth profiles of WT, △X2-N, △X2-C and △X2-NC grown on cellobiose. (b), growth profiles of WT, △X2-N, △X2-C and △X2-NC grown on cellulose. (c), cellulose degradation profiles of WT, △X2-N, △X2-C and △X2-NC. (d), released total soluble sugar in supernatant of medium at final time point for each strain. Data are presented as the mean of three biological replicates and error bars represent standard deviation (SD).

2.4.3 Disrupted X2 modules reduced can affect cell affinity (adhesion) to the cellulose

Considering previous study found sequence insertion in the cipC gene could induce a polar effect and further abolishment of the expression of all other genes localized in the

cip-cel operon downstream of cipC 30, we performed the microarray-based transcriptomic assay to determine whether the deficient X2 module could decrease the expression of other genes, especially for those located on the cip-cel gene cluster. Based on our microarray analysis, it was found that there was no differential expression of any genes between WT and X2 module mutants grown on cellulose during the exponential phase. Therefore, the disruption of the X2 module cannot cause the low efficiency of cellulose utilization by regulating other genes' expressions.

Based on above phenotypical experiment results and microarray analysis, we speculated that the possible mechanism was that X2 modules are key factors in responsible for (i) increasing the binding affinity of cellulosomes to the cellulose, (ii) increasing the localization and adhesion of cellulosomes to the cell surfaces or (iii) increasing both binding affinity to cellulose and adhesion to cell surfaces. Therefore, in order to determine whether the disruption of X2 can influence the binding affinity between cells and cellulose, the cell-cellulose adhesion assay was performed. Cell adhesion was measured by monitoring the OD_{600} of planktonic parts after incubating with/without Whatman filter paper. With cells collected at the early exponential stage, we found that the relative cell adhesion capability between the cell and cellulose was severely reduced in △X2-NC mutant, which was 50% of that in WT. For single X2 mutant stains, the △X2-C decreased cell adhesion capability more than △X2-N (Figure 2.4a). When with cells at the late exponential stage, cell adhesion capacity was further decreased in △X2-C and △X2-NC, which was only 16% of that in WT for the △X2-NC. Therefore, X2 modules, especially the X2-C module, play an important role in cell adhesion to cellulose, which was consistent with our hypothesis.

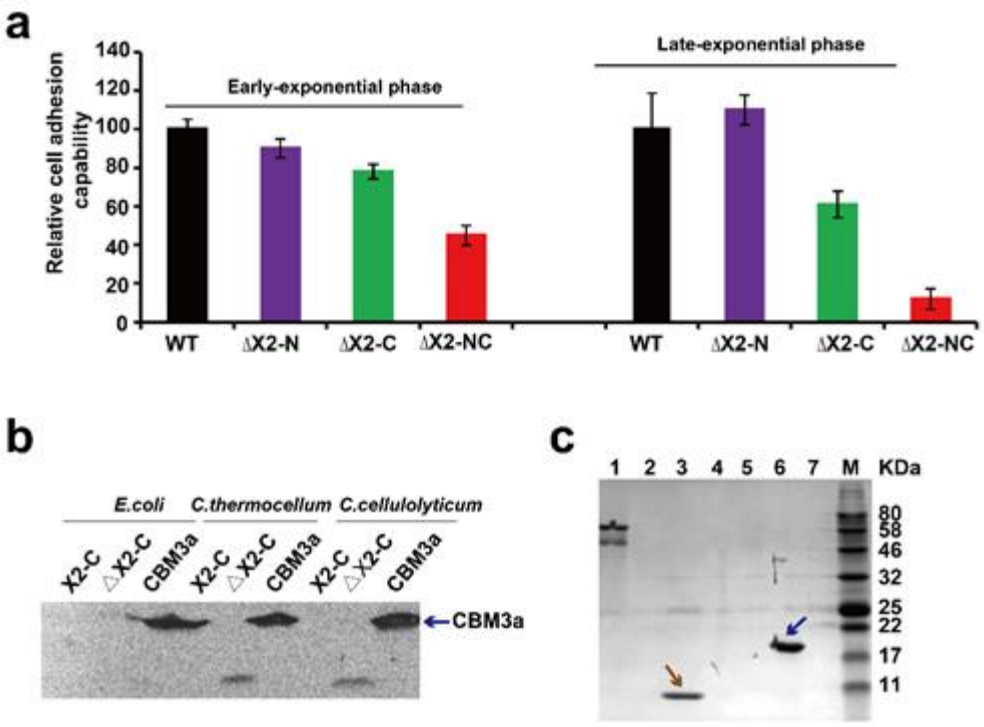

Figure 2. 4 The in vivo function of the X2 module was related with binding affinity between cells and cellulose. (a), the relative cell adhesion capability between cells and cellulose for each strain in early-exponential and late-exponential phase. Data are presented as the mean of three biological replicates and error bars represent SD. (b), binding of X2-C, △X2-C and CBM3a proteins to the cell surfaces of E.coli, *C.thermocellum* or *C. cellulolyticum* respectively. CBM3a was detected in all three strains (as blue arrow indicated), indicating it could bind to the cell surface for both Gram-negative and Gram-positive bacteria. The X2-C could not be detected for all of them, indicating it cannot directly bind to the cell surface. The weak band of △X2-C was detected in *C.thermocellum* and *C. cellulolyticum* , indicating it had a weak binding affinity with the cell surface of the Gram-positive bacteria. (c), binding of X2-C, CBM3a and BSA protein to crystalline cellulose. The CBM3a protein and the BSA protein were used as the positive and negative control respectively. The CBM3a was detected in the cellulose pellet (as blue arrow indicated) and X2-C was only detected in the supernatant fraction as same as the BSA negative control (as brown arrow indicated), indicating that X2-C can't directly bind to the cellulose. Lane 1, BSA+Cellulose in supernatant fraction; Lane 2, CBM+Cellulose in supernatant fraction; Lane 3, X2-C+Cellulose in supernatant fraction (Lane 3); Lane 4, blank; Lane 5, BSA+Cellulose in cellulose-containing pellet; Lane 6, CBM+Cellulose in cellulose-containing pellet; Lane 7, X2-C+Cellulose in cellulose-containing pellet.

2.4.4 X2 modules can't directly bind to the cell surface wall of C. cellulolyticum

As the X2 module could influence the binding affinity between cells and cellulose, the *in-vitro* X2 module-cell wall binding assay was performed to further detect how the X2 module influenced the binding affinity between them. Given that the efficiencies of cellulose saccharification, soluble sugar release and the relative adhesion ability in mutant X2-C were less than the mutant X2-N, we purified X2-2 protein as the X2 module model protein for studying *in vitro*. In this binding assay, we used the *E.coli* and *C. thermocellum* cells as the control group. Based on the western blotting results, we did not find the band of X2 in any of the three strains, indicating that X2 module can't directly bind to the cell surfaces for all three of them (Figure 4B). However, the band of CBM3a was found in all three strains, indicating that it could bind to the polysaccharide of the cell walls (Figure 2.4b). More interesting, we found a very weak band of ΔX2 protein in the group of *C. thermocellum* and *C. cellulolyticum*, indicating that the ΔX2 protein could weakly bind to the cell surface of *C. thermocellum* and *C. cellulolyticum* and might further enhanced the localization of cellulosomes on the surface of the X2 mutant.

2.4.5 X2 module cannot directly bind to the cellulose in C. cellulolyticum

Since the X2 module could not directly bind to the cell wall, the *in-vitro* X2 module-cellulose binding assay was also performed to verify whether the X2 module in *C. cellulolyticum* can directly bind to the cellulose or not. Three groups including BSA incubated with cellulose (negative control), the CBM3a incubated with cellulose (positive control) and the X2 incubated with cellulose were set up. After incubation and centrifugation, same as the negative control, X2 protein was only found in the

supernatant instead of cellulose pellets, indicating that the X2 protein can't directly bind to the cellulose (Figure 2.4c).

In sum, the X2 module in the CipC protein cannot directly bind to the cellulose or cell surface of the cell. However, the disruption of the X2 module in CipC protein could decrease the cellulose utilization efficiency and severely reduce the binding affinity between cells and cellulose.

2.5 Discussion

The function of the X2 module has been studied for many years (Mosbah, Belaïch et al. 2000, Kosugi, Amano et al. 2004, Pasari, Adlakha et al. 2017), and almost all studies on determining the functions of the X2 module were based on *in vitro* biochemical assays. However, the *in vivo* function of the X2 module is still elusive. In *C. cellulolyticum*, two X2 modules are located within the *cipC* gene and the nucleotide sequences between them are very similar (65% for the pairwise nucleotide sequence identity). Due to the limitation of the traditional genome editing, it was very difficult to inactivate both X2 modules and keep the integrity of function and structure of the CipC scaffoldin protein simultaneously in *C. cellulolyticum* by traditional genome editing. Fortunately, with the development and application of the CRISPR-Cas9 based genome editing method, we found a conserved site in X2 modules for deletion and successfully constructed △X2-NC double mutant. To our knowledge, *in vivo* determination of the functions of the X2 module has never been reported in any strain. In a previous study, Hédia Maamar et al. found that the disruption of *cipC* gene in *C. cellulolyticum* could hardly affect the growth of soluble sugar but can lead to barely growth on cellulose (Maamar, Valette et al. 2004). Compared with their study, we also found that the inactivation of X2 modules

did not influence the strain grown on cellobiose (Figure 2.3a). Although, the inactivation of X2 modules could cause a longer lag phase and decrease the efficiency of cellulose degradation, it couldn't severely affect the strain growing on cellulose as the disruption of *cipC* gene did, even the maximum cell biomass between WT and △X2-NC were similar (Figure 2.3b). All of these indicated that the deletion of the conserved motif in X2 modules didn't affect the basic functions of *cipC* gene and the functional loss of X2 modules can directly influence certain function of cellulosomes, by which it indirectly caused lower efficiency of cellulosome in degrading cellulose.

Our microarray analysis confirmed the disruption of the X2 module in cipC gene could not influence any changes for gene expressions. On the other hand, the results of the adhesion assay indicated that the function of X2 modules was related with binding affinity between the cells and cellulose (Figure 2.4a). It is known that the cellulosome on the surface of the cells is the bond for the adhesion between the cells and cellulose(Gal, Pages et al. 1997, Schwarz 2001), although the mechanism of the localization of the cellulosomes on the surface of C. cellulolyticum is unclear (Desvaux 2005). Therefore, the lower binding affinity between cells and cellulose caused by inactivation of X2 modules should be attributed to the functional change of cellulosomes, which was consistent with our assumption.

From the cell adhesion assay (Kosugi, Amano et al. 2004)(Figure 2.4a), we hypothesized three possible mechanisms for X2 modules in regulating the binding affinity between cells and cellulose. Although previous studies indicated that the X2 module might directly bind to the cell wall , our in-vitro proteins-cell wall binding assay indicated that the X2 module protein can not directly bind to the cell wall (Figure 4B).

More interestingly, the △X2 module protein had a weak binding with the cell wall, which might promote the binding between the cellulosomes and cell surface. In other words, the X2 mutant strain might have more numbers of cellulosomes on its cell surface. Based on the structure analysis of the X2 modules (Mosbah, Belaïch et al. 2000), the surface of the X2 module is predominantly covered by hydrophilic amino acid and only contains a hydrophobic shallow groove, which could explain why the X2 module can't directly bind to the cell wall. When the conserved motif (NGNT) was deleted, the hydrophobic shallow groove became wider as the structure modeling indicated (Figure 1), which was more like the characteristic of the typical CBM modules where most hydrophobic residues are protruded outside (Luís, Venditto et al. 2013, Pasari, Adlakha et al. 2017) and might make the △X2 module weakly bind to the cell wall like. This possible mechanism needs to be further investigated in the future work.

In fact, previous study indicated that the localization of the cellulosomes (localized on the surface/ free living) cannot significantly influence the cellulose degradation efficiency of cellulosomes (Xu, Resch et al. 2016). If the function of X2 modules was only related to the localization of cellulosomes, we should not observe the significant decrease of cellulose degradation efficiency and release of soluble sugars in the △X2-NC strain compared to the WT (Figure 3C & D). Meanwhile, the HMMER analysis found that many X2 modules existed in free cellulases, which also indicated that the main function of X2 modules should not be related with localization of cellulosomes. On the other hand, we did not observe the X2 module from *C. Cellulolyticum* could bind to the cellulose directly (Figure 4C). For CBM modules, most hydrophobic

residues are on the surface and are protruded to the solvent for promoting carbohydrate polymers binding (Luís, Venditto et al. 2013). In contrast, few hydrophobic residues are on the surface of the X2 module, and all the polar residues are exposed to the solvent (Mosbah, Belaïch et al. 2000), which indicated the reason why the X2 module cannot directly bind to the cellulose. Therefore, the only possible reasonable explanation for the cell adhesion assay was that the X2 module cannot directly bind to the cellulose but may promote the binding between the cellulosomes and cellulose.

Given that the CBM3a module of the CipC scaffolding protein is for binding to the cellulose, we speculated that the in vivo function of the X2 module was realized by promoting the binding between CBM3a domain and cellulose. Meanwhile, our ITC assay found there was a weak interaction between the X2 module and the CBM3a (Figure S4), which also suggested that the X2 module might interact with CBM3a and promote its binding function. In fact, previous in vitro assay has already found that the CBM3-X2 module had a better cellulose binding affinity to crystalline cellulose compared to CBM3 module alone, indicating the X2 module could indeed promote the binding between the CBM module and cellulose.

Given that the CBM3a module of the CipC scaffolding protein is for binding to the cellulose, we speculated that the in vivo function of the X2 module was realized by promoting the binding between CBM3a domain and cellulose. Meanwhile, our ITC assay found there was a weak interaction between the X2 module and the CBM3a (Figure S4), which also suggested that the X2 module might interact with CBM3a and promote its binding function. In fact, previous in vitro assay has already found that the CBM3-X2 module had a better cellulose binding affinity to crystalline cellulose compared to CBM3

module alone, indicating the X2 module could indeed promote the binding between the CBM module and cellulose (Pasari, Adlakha et al. 2017). In addition to that, X2 module was thought to be associated with cellulase activity, which rendered the Cel9A cellulase from L. phytofermentansis is significantly more efficient on crystalline cellulose than any of the known cellulases from the *C.cellulolyticum* (Ravachol, Borne et al. 2015). Some cellulases engineering studies also pointed out that the integration of the CBM with the X2 module into some cellulases can enhance the avicelase activities, such as Cel48F and Cel9G (Mingardon, Chanal et al. 2007, Vita, Borne et al. 2019). All of these may explain why we observed that the phenotype of WT grown on cellulose was better than that in △X2-NC and why the X2 modules are always next to the CBM3a

2.6 Conclusion

In summary, precise deletion of the NGNT conserved sequences of the X module by Cas9-based genome editing method was demonstrated to be a useful strategy for studying the function of the X module *in vivo* and keeping the integrity of the structure and functions for cellulosomes. This strategy can be applied to study the function of X2 modules or other interesting modules within certain proteins in other bacteria with similar cellulases/cellulosome-producing systems. We found that the inactivation of the X2 modules in *C. cellulolyticum* can indeed influence the cellulose utilization efficiency, and the *in vivo* function of the X2 module was also determined, which was associated with binding affinity between cellulosomes and cellulose. Given that the X2 modules are widely spread in cellulolytic bacteria and play important roles on cellulose degradation , all of these findings can provide new perspectives on engineering those

potential CBP bacteria in improving their cellulose degradation efficiencies or modifying the commercial cellulases in improving their hydrolysis efficiencies.

Chapter 3 **Precise promoter integration improves cellulose bioconversion and thermotolerance in *Clostridium cellulolyticum***

3.1 Abstract

Lignocellulose has been used for production of sustainable biofuels and value-added chemicals. However, the low-efficiency bioconversion of lignocellulose greatly contributes to a high production cost. Here, we employed CRISPR-Cas9 editing to improve cellulose degradation efficiency by editing a regulatory element of the *cip-cel* gene cluster in *Clostridium cellulolyticum*. Insertion of a synthetic promoter (P4) and an endogenous promoter (P2) in the *mspI*-deficient parental strain (Δ2866) created chromosomal integrants, P4-2866 and P2-2866, respectively. Both engineered strains increased the transcript abundance of downstream polycistronic genes and enhanced in vitro cellulolytic activities of isolated cellulosomes. A high cellulose load of 20 g/L suppressed cellulose degradation in the parental strain in the first 150 h fermentation; whereas P4-2866 and P2-2866 hydrolyzed 29% and 53% of the cellulose, respectively. Both engineered strains also demonstrated a greater growth rate and a higher cell biomass yield. Interestingly, the Δ2866 parental strain demonstrated better thermotolerance than the wildtype strain, and promoter insertion further enhanced thermotolerance. Similar improvements in cell growth and cellulose degradation were reproduced by promoter insertion in the wildtype strain and a lactate production-

defective mutant (LM). P2 insertion in LM increased ethanol titer by 65%. Together, the editing of regulatory elements of catabolic gene clusters provides new perspectives on improving cellulose bioconversion in microbes.

Keywords: Promoter integration; Cellulosome; Cellulose degradation; Microbial engineering

3.2 Introduction:

Cellulose, as the most abundant renewable bioresource on earth, can be used to produce sustainable valuable products (Jarvis 2003, Zhang, Himmel et al. 2006). In order to save cost and improve efficiency for utilization of cellulose, a promising strategy named consolidated bioprocessing (CBP) was proposed, which involves a single microorganism for cellulase production, saccharification, and fermentation in a single step in one bioreactor (Lynd, Cushman et al. 1991, Zhang, Himmel et al. 2006). Although there are some CBP-enabling microorganisms able to perform both cellulose hydrolysis and sugar fermentation simultaneously, they need to be engineered with improved cellulose degradation, greater resistance to abiotic and biotic factors, and higher production efficiency (Lynd, Van Zyl et al. 2005, Quast, Pruesse et al. 2012)(. Efficient conversion of cellulose to fermentable sugars is key to reducing the cost during production (Zhang, Himmel et al. 2006, Quast, Pruesse et al. 2012, Liao, Mi et al. 2016).

Clostridium cellulolyticum is a model mesophilic clostridial species, able to convert lignocellulose to ethanol and organic acids (Desvaux 2005, Quast, Pruesse et al. 2012, Li, Xu et al. 2014). Like other cellulose-degrading *Clostridia*, it forms extracellular enzymatic complexes, termed cellulosomes, that degrade crystalline cellulose with

greater efficiency than free or non-organized enzymes as a result of the proximal synergism of enzyme reactions (Gal, Pages et al. 1997, Desvaux 2005). The cellulosome of *C. cellulolyticum* contains a cell surface-anchored scaffoldin and a diversity of carbohydrate-active enzymes. Major cellulosomal components are encoded by a 26 kb *cip-cel* gene cluster. This cluster has 11 genes (*cipC-cel48F-cel8C-cel9G-cel9E-orfX-cel9H-cel9J-man5K-cel9M-rgl11Y-cel5N*) (Desvaux 2005), driven by a single promoter (Maamar, Abdou et al. 2006). There were two major large transcripts detected when *C. cellulolyticum* was grown on cellulose (Maamar, Abdou et al. 2006). The most abundant transcript covers the first five genes (*cipC-cel48F-cel8C-cel9G-cel9E*); whereas the other transcript is much lower in abundance and only contains the remaining downstream genes (*cel9H-cel9J-man5K-cel9M-rgl11Y-cel5N*). This difference in transcript abundance can be attributed to site-specific RNA processing and differential resistance of processed RNAs to RNase-mediated degradation (Xu, Huang et al. 2015). As Cel9H, Cel9J, Cel9M, and Cel5N, are endoglucanases that are critical for cellulose solubilization (Blouzard, Coutinho et al. 2010), we hypothesized that improvement in cellulose hydrolysis could be achieved by the manipulation of their expression levels. However, thus far, it has been technically difficult to engineer transcriptional regulatory elements in the native genome of *C. cellulolyticum* (Xu, Li et al. 2015).

With the development of Cas9 nickase-based genome editing in Clostridia (Xu, Li et al. 2015), we are now able to tune gene expression by manipulating regulatory systems. Here, we aimed to improve cellulose degradation by inserting constitutively active promoters in front of the *cel9H* gene, by which the transcription of the six downstream

genes (*cel9H-cel9J-man5K-cel9M-rgl11Y-cel5N*) can be increased independently of the upstream genes. We systematically characterized the resulting engineered strains from transcriptional, enzymatic, physiological, and morphological aspects. Additionally, since a previous study demonstrated that a 5°C increase of the fermentation temperature could greatly reduce the cost for fuel ethanol production (Abdel-Banat, Hoshida et al. 2010), and enzymatic assays indicated that the activity of cellulases could be increased when the temperature was increased (Mingardon, Bagert et al. 2011), we also examined the fitness and thermotolerance of our engineered strains at an elevated temperature. Additionally, we applied the same strategy to other engineered *C. cellulolyticum* strains to test whether their cellulolytic activity and ethanol production could be further improved. Our results demonstrated that a precise promoter insertion is an efficient strategy to modulate the transcriptional abundance of catabolic gene clusters for improving efficiencies of cellulose degradation and end-product formation.

3.3 Materials and Methods

3.3.1 Bacterial strains and plasmid construction

Strains, plasmids, and primers used in this study were listed in Supplementary Table S3.1 and S3.2, respectively. pCas9n-P4inser-donor and pCas9n-P2inser-donor, for inserting the promoter in the *cip-cel* gene cluster, were constructed as described before (Xu, Li et al. 2015). First, the predicted P2 promoter, left and right homologous arms, were amplified from wild-type genomic DNA of *C. cellulolyticum* and purified separately. The P4 promoter was synthesized as part of the 9HP4LR and 9HP4RF primers which were used for the amplification of left and right homologous arms. The P4::gRNA cassette containing the 20-bp protospacer was amplified from pCR/8w p4-4

prom4 and pMS-RNA by 3FF, PM9HGRF, PM9HGRR, and 2RR primers. The linear backbone was digested from pFdCas9n-p4-pyrF_w/2kbΔ by KpnI and PvuI. These fragments were assembled using Gibson assembly and the assembled product was transformed into E. coli for colony screening and confirmed via Sanger sequencing (Oklahoma Medical Research Foundation). The plasmid for promoter insertion in *C. cellulolyticum* was cured as previous reported (Li, Tschaplinski et al. 2012).

3.3.2 Media and culture conditions

E. coli DH5α strain (Invitrogen) was used for cloning and grown at 37°C in LB with 35 μg/ml chloramphenicol. Complex modified VM medium supplemented with 2.0 g/L yeast extract was used for general growth and transformation of *C. cellulolyticum* H10 (Quast, Pruesse et al. 2012). Defined modified VM medium containing necessary vitamin solution and mineral solution was used for fermentation and omics experiments (Higashide, Li et al. 2011). *C. cellulolyticum* H10 was cultured with 5 g/L cellobiose or 20 g/L Avicel PH101 crystalline cellulose (Sigma-Aldrich) at 34°C or 40°C anaerobically depending on the experiment. Solid VM medium with 1.0% (w/v) of Bacto agar (VWR) and 15 μg/ml thiamphenicol was used for developing *C. cellulolyticum* colonies.

3.3.3 Transformation

The wild-type *C. cellulolyticum* and Δ2866 parental strain were transformed with corresponding plasmids by electroporation as described previously (Jennert, Tardif et al. 2000, Quast, Pruesse et al. 2012). For transformation of the wildtype *C. cellulolyticum* strain, plasmid DNA was methylated using MspI Methyltransferase (New England

Biolabs, Ipswich, MA) and then purified prior to transformation (Quast, Pruesse et al. 2012).

3.3.4 Microarray analysis

All *C. cellulolyticum* strains (i.e., Δ2866, P2-2866, P4-2866 and wildtype) were cultivated in defined VM medium with 20 g/L cellulose or 5 g/L cellobiose. Six biological replicates of each strain were collected at mid-exponential growth phase. All samples were centrifuged at 4°C and 5000 × g for 10 min and cell pellets were immediately frozen with liquid nitrogen and then stored at -70°C for further use. Total RNA was extracted using TRIzol (Invitrogen) and purified using NucleoSpin RNAII kit (Macherey-Nagel) according to the manufacturer's instructions. RNA integrity was examined on agarose gels; RNA purity and concentration were measured with a NanoDrop spectrophotometer.

For microarray hybridization, 13,098 probes (50 nt in length) were designed to cover 94% of the protein encoding genes in *C. cellulolyticum* (Agilent). For each RNA sample, 0.6 µg of total RNA was reverse transcribed to Cyanine-3 labeled cDNA using Reverse Transcriptase III (Invitrogen) using Cyanine 3-labeled dUTP (Thermo Fisher). Genomic DNA (gDNA) was extracted from the control strain using GenElute bacterial genomic DNA kit (Sigma Aldrich) and was labeled by incorporating Cyanine 5-labeled dUTP with Klenow DNA polymerase (New England Biolabs). 1.5 µg of gDNA was used in each gDNA labeling reaction which was used for eight hybridizations. All labeled cDNA and gDNA were purified with QIAquick PCR purification reagents (Qiagen) and SpinSmart columns (Denville Scientific Inc), and then lyophilized for later use. Labeled cDNA and gDNA were mixed in the hybridization master buffer

(Agilent) containing 8% formamide, followed with denaturation at 95°C for 3 min, incubation at 37°C for 30 min and finally loaded onto an array. Hybridization was carried out at 67°C and 20 rpm for 22 h. Slides were washed and then scanned using a NimbleGen MS200 scanner (Roche) with the following settings: two-channel scanning, 2 μm scanning resolution, 100% laser strength, 30% gain percentage. Using Agilent Feature Extraction version 11.5, all digital images were manually checked to confirm gridding quality and raw data was extracted.

Microarray data analysis was performed using the limma package in R (Ritchie, Phipson et al. 2015). First, probes with both qualified green and red signals were screened (single-to-noise ratio>2, signal-to-background ratio >1.3, coefficient of variation<0.8, minimal gMeansigal>150, and minimal rMeansignal>50) (He and Zhou 2008). Second, the mean signal of each probe was applied to background correction by subtraction, within-array normalization by loess, and then between-array normalization by quantile. Third, using the normalized data, gene probes with significantly different expression levels were identified using limma's linear model and then evaluated by empirical Bayes methods. The expression level of each gene was calculated by averaging the values of qualified probes only if half or more probes of this gene were qualified. In this study, differentially expressed genes (DEGs) refer to genes with a log2 fold-change above 1 (or below -1) and an adjusted p value < 0.05. Venn diagram graphs were generated with the online tool (http://bioinformatics.psb.ugent.be/webtools/Venn/). Blast2go was used for GO enrichment analysis with Fisher's exact test ($p<0.01$, two sided) (Conesa, Götz et al. 2005). PCA analysis was conducted with the prcomp function in R using the normalized

data. The functional enrichment analysis was performed using ClueGO v 2.5.5 and CluePedia v 1.5.5 (Bindea, Mlecnik et al. 2009, Bindea, Galon et al. 2013). The significance of the terms and groups was calculated using ClueGo. The P value was calculated using a two-sided minimal-likelihood test and corrected using the Benjamini-Hochberg method. The final enrichment network was visualized using Cytoscape v3.7.2.

3.3.5 Measurements of FPase, Avicelase, CMCase and xylanase activities

The activities of Avicelase and CMCase were determined as described previously (Zhang, Hong et al. 2009). The cellulosome fraction was isolated from 200 ml cellulose-grown culture at mid-exponential growth phase. Cell cultures were filtered with glassfiber paper to remove media. Retained cellulose-associated components were firstly washed three times on the glassfiber with 50 ml of 50 mM PBS buffer, then washed three times with 50 ml of 25 mM PBS buffer. Finally, cellulose-associated cellulases, mainly cellulosomes, were eluted with 20 ml of sterile water and centrifuged to discard insoluble material before applying to a protein concentrator with a 5 kDa ultracentrifuge membrane. Protein samples were mixed with an equal amount of 40 mM Tris-maleate buffer (with 4 mM Ca2+) with 50 % glycerol and then stored at -20°C. Protein concentration was determined by a BCA assay kit (Thermo Fisher Scientific). Total cellulase activity was monitored using a filter paper assay. For each reaction, 1 ml Tris-maleate buffer was pre-incubated with a 25 mg filter paper disc at 37°C for 20 min before adding a mixture of 200 μg protein and 5 μl of diluted β-glucosidase (1:250) to initiate the reaction. Reaction controls were generated by using the same amount of boiled protein mixtures. Reactions were carried out at 37°C and 200 μl of reaction

products were sampled at 0, 3, 8.5, 22 h. After centrifugation, the concentration of released reducing sugars in the supernatant was determined by the DNS method (Miller 1959). Avicelase was measured while shaking with 2% Avicel that was washed with Tris-maleate buffer. CMCase activity and Xylanase activity were measured with 1.25% CMC and beechwood xylan, respectively, with 10 μg protein used in assays. 120 μl of DNS reagent was mixed with 40 μl reaction products in PCR strip tubes and incubated at 98°C for 10 min in a thermocycler. A 140 μl aliquot of each sample was transferred to a 96 well plate and the absorbance at 544 nM was measured using an Optima plate reader.

3.3.6 Scanning electron microscopy

All *C. cellulolyticum* strains (Δ2866, P2-2866 and P4-2866) were cultivated in defined VM medium with 20 g/L cellulose or 5g/L cellobiose. At the mid-exponential growth phase, cell pellets from each strain were collected by centrifugation at 5,000 g for fixation. Cell pellets were immersed in 4% glutaraldehyde in 0.1 M PBS buffer (pH 7.4). After incubation at 4°C overnight, all samples were washed three times with PBS, and immersed in 1% OsO4 at 4°C for 1 h. After the post-fixation, de-ionized water was used to remove all traces of fixative and buffer solutions. Dehydration was performed by immersing all washed samples for 10 minutes sequentially in 25%, 50%, 70%, 85%, 95%, and 100% ethanol to remove all traces of water. Samples were air dried and critical point dry for 10 minutes after dehydration. Samples then were mounted onto SEM stubs and a dab of silver was applied. Finally, all samples were coated with ~5 nm AuPd. The number of cell surface protuberances was quantified on each strain (Δ2866, P2-2866, and P4-2866) when grown on cellulose or cellobiose. Cells for protuberance

counting were randomly chosen from different views of SEM images. Statistic differences in the number of protuberances between parent strain and engineered strains were determined by the one-way ANOVA with a permutation test (PERM-ANOVA) using the lmPerm package in R (Venables and Ripley 2013).

3.3.7 Genomic DNA sequencing

Genomic DNA of P4-2866 and P2-2866 was extracted using the GenEluteTM Bacterial Genomic DNA Kit. Libraries for gDNA sequencing were constructed using a KAPA DNA Library Preparation kit (Kapa Biosystems) following instructions and then sequenced using Illumina MiSeq platform by the Miseq Reagent Kit V2 (2 × 250 bp).

3.3.8 Measurement of cell growth, fermentation products, and remaining cellulose

All *C. cellulolyticum* strains were revived in complex VM medium with 5 g/L cellobiose and then transferred to defined VM media. The cellobiose-grown cultures at an OD600 of 0.5-0.6 were used to inoculate 50 ml of defined VM media containing 20 g/L Avicel PH101 crystalline cellulose. Each strain had three biological replicates. During growth, 1 mL of cell culture was sampled at each time point and then stored at -80°C for future determination of cell biomass, metabolites, and remaining cellulose. Growth curves for each sample grown on cellulose was estimated by total protein measurement (Quast, Pruesse et al. 2012). High-performance liquid chromatography (HPLC) was used to measure the major fermentation products (lactate, acetate, and ethanol) and soluble sugars (cellobiose and glucose) in the supernatant of spent medium (Quast, Pruesse et al. 2012). The specific rate of product formation was calculated as previously described (Desvaux, Guedon et al. 2000). The remaining cellulose in the medium was measured by the phenol-sulfuric acid method (Hemme, Fields et al. 2011).

3.4 Results

3.4.1 Targeted promoter insertion in the cip-cel gene cluster enhanced the cellulolytic activity of isolated cellulosomal complexes

To test whether promoter insertion into the *cip-cel* gene cluster would promote cellulose degradation, we selected two constitutive promoters, a synthetic promoter named P4 and an endogenous promoter named P2. P4 promoter is a mini promoter from our previous study (Xu, Li et al. 2015), which was used to drive the gRNA cassette expression in our Cas9 genetic editing system for *C. cellulolyticum*. P2 is a predicted promoter of the Ccel_2112 gene, which encodes an extracellular solute-binding protein family 1 and was previously shown to be transcribed at a higher level in cells grown with cellulose and corn over cellobiose (Xu, Huang et al. 2015). For each promoter insertion, we constructed an all-in-one vector for Cas9 nickase-based genome editing, which requires co-expression of Cas9 nickase and a customized gRNA, as well as homologous regions that sandwich the promoter sequence (Figure 3.1A). The corresponding engineered strains, P4-2866 and P2-2866, were generated in the parental strain (Δ2866) which is a *mspI*-deficient mutant that was originally engineered to improve transformation efficiency by allowing non-methylated DNA transformation (Cui, Hong et al. 2012). We performed whole genome resequencing to confirm precise promoter insertion without any off-target insertion in the genome.

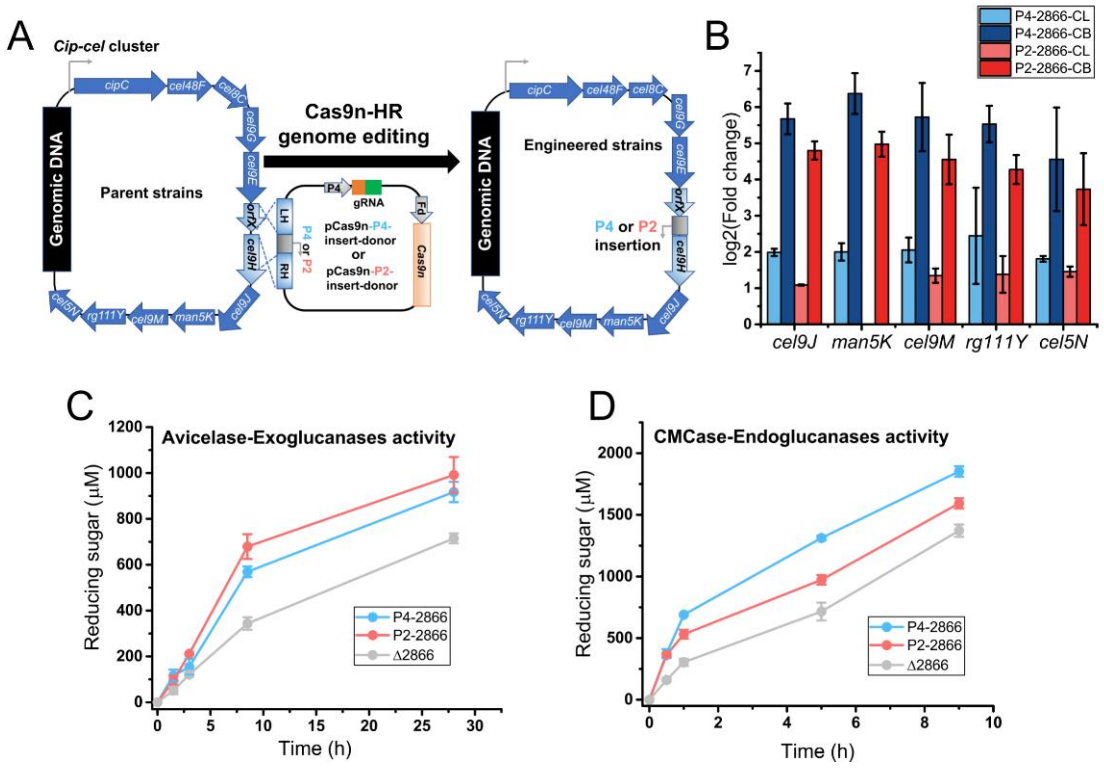

Figure 3. 1 One-step promoter integration increased expression of downstream genes in the *cip-cel* gene cluster and improved in vitro cellulolytic features. (A) An overview of the Cas9 nickase-based genome editing in *C. cellulolyticum*. Plasmids pCas9n-P4insert-donor and pCas9n-P2insert-donor were used for the synthetic P4 (blue) and predicted P2 (red) promoter insertion in the *cip-cel* gene cluster between *orfX* and *cel9H* in the genetic background of Δ2866, wild-type (WT), and lactate production defective strain (LM). (B) Promoter integration increased the transcription of downstream polycistronic genes (cel9H-cel9J-man5K-cel9M-rgl11Y-cel5N). Fold change was determined between either P4-2866 and Δ2866 or P2-2866 and Δ2866. All strains were grown on a defined VM medium with 20 g/L cellulose (CL) or 5 g/L cellobiose (CB). (C) In vitro enzymatic assay measuring the activity of exoglucanases of Avicelase in isolated cellulosomes from P4-2866, P2-2866, and Δ2866. (D) In vitro enzymatic assay measuring the activity of endoglucanases of CMCase in isolated cellulosomes from P4-2866, P2-2866 and Δ2866.

Microarray-based transcriptomic analysis revealed a large increase in the transcript abundance of downstream polycistronic genes driven by either promoter. Compared with Δ2866, when grown on cellulose, cellulase genes downstream of the inserted promoter including *cel9J*, *cel9M*, and *cel5N*, were dramatically upregulated in both P4-2866 and P2-2866, but to a greater extent in P4-2866 (Figure 3.1B). A similar trend was

also observed when grown on cellobiose suggesting that the P4 promoter led to greater expression than the P2 promoter (Figure 3.1B). However, the transcript levels of downstream genes in both promoter integrants were not significantly different between cellulose and cellobiose, indicating the promoters were not cellulose-specific. In addition, promoter insertion also affected the expression of other genes located across the genome and the P4 promoter presented a more profound impact than the P2 promoter, irrespective of using cellobiose or cellulose as a carbon source (Figure S3.1A). The PCA analysis of transcriptomics showed that P4-2866, P2-2866, and Δ2866 strains were separated when grown on cellulose but not on cellobiose (Figure S3.1B). It indicates a cellulose-dependent effect on many other genes is associated with the promoter activity.

To test if promoter insertion altered cellulolytic activity, we isolated cellulosome fractions from cultures grown on cellulose and analyzed cellulosomal composition and enzyme activity. SDS-PAGE analysis showed that the engineered strains significantly changed cellulosomal components, which is in agreement with the observed increased transcriptional abundance (Figure S3.2). It is known that efficient cellulose degradation relies on synergistic reactions of multiple enzymes, such as endoglucanase, exoglucanase, and β-glucosidase (Asztalos, Daniels et al. 2012). We found cellulosomes from both engineered strains dramatically improved the activity of exoglucanases and endoglucanases when compared with isolated cellulosomes from the Δ2866 parental strain (Figure 3.1C & 3.1D). Therefore, targeted promoter insertion enhanced gene expression at both transcriptional and enzymatic levels.

3.4.2 Engineered strains presented notable changes in cell morphology

During screening of engineered strains, we observed that P4-2866 formed visible cotton-like aggregates when statically cultured on cellobiose. This morphology was distinct from planktonic cell suspensions of other strains (wildtype, P2-2866, and Δ2866). Scanning electron microscopy revealed that, irrespective of using cellobiose or cellulose as the carbon source, P4-2866 had an altered cell arrangement changing from typical single rods to cell chains (Figure 3.2 & Figure S3.3). The same morphological change was observed with other randomly picked P4-2866 integrants. Whole genome sequencing confirmed that no other mutations were present and that the change in morphology could be attributed to the insertion of the P4 promoter. Neither engineered strains presented a substantial change in cell diameter or length relative to Δ2866. Cellulosomes are further organized to form protuberances on cell surface as observed in *Clostridium thermocellum* and *Clostridium cellulovorans* (Bayer and Lamed 1986, Tamaru, Miyake et al. 2010, Tachaapaikoon, Kosugi et al. 2012); however, it remains inconclusive in C. *cellulolyticum* (Desvaux 2005, Ferdinand, Borne et al. 2013). By closely examining SEM images, we found P2-2866 and P4-2866 displayed 7-fold more protuberances than Δ2866 on cellulose and 5-fold more on cellobiose (Figure 3.2 & Figure S3.3). It is possible that promoter insertion or increased expression of cellulases influences the organization and localization of cellulosome complexes on cell surface, and further studies will be necessary to confirm this.

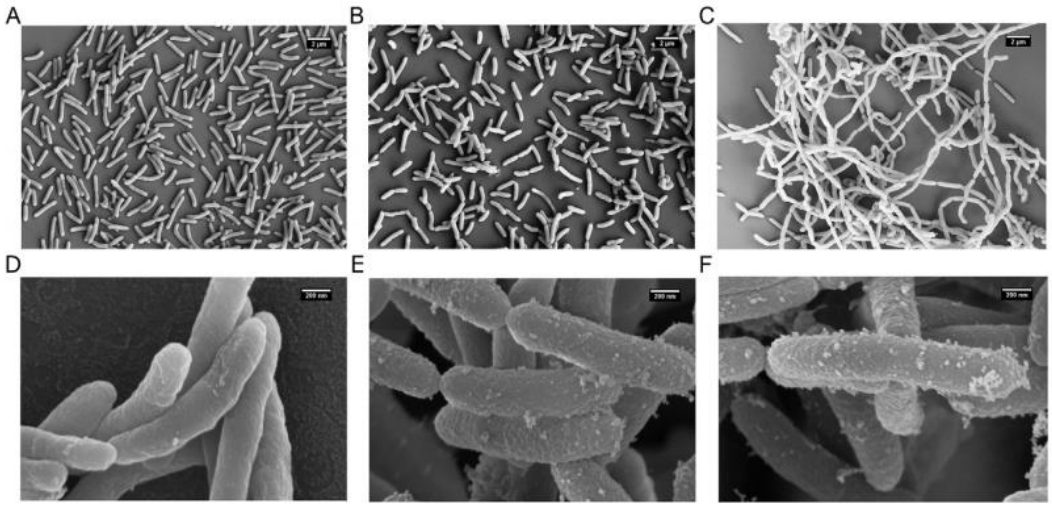

Figure 3. 2 Promoter integrants altered cell morphology. SEM images of (A) Δ2866 grown on 5 g/L cellobiose; (B) P2-2866 grown on 5 g/L cellobiose; (C) P4-2866 grown on 5 g/L cellobiose; (D) Δ2866 grown on 20 g/L cellulose; (E) P2-2866 grown on 20 g/L cellulose; (F) P4-2866 grown on 20 g/L cellulose.

3.4.3 Engineered strains improved conversion of cellulose to end products

We compared cell growth and metabolism of P4-2866, P2-2866, and Δ2866 in defined VM media with 20 g/L Avicel cellulose. P4-2866 and P2-2866 not only had shorter doubling times, which was two-fold faster than Δ2866, but both strains also had increased cell biomass of 60% and 52%, respectively (Table 3.1 & Figure 3.3A). At 150 h, P2-2866 and P4-2866 entered stationary phase and degraded 53% and 29% of cellulose, respectively; whereas Δ2866 used very little cellulose, indicative of strong growth suppression (Figure 3.3B). Although the engineered strains did not enhance the final titers of major end products (lactate, acetate, and ethanol) at 300 h (Figure S3.4A, 3.4B & 3.4C), they demonstrated an efficient product formation within a shorter fermentation time (Figure 3C), which is in line with their faster growth profiles. Also, P2-2866 accumulated more soluble sugars than Δ2866 (Figure S3.4D & 3.4E). More specifically, glucose and cellobiose were increased by 52% and 38%, respectively.

Table 3. 1 Doubling time (h) of strains grown on 20 g/L cellulose at 34 °C or 40 °C.

Strains	34 °C	40 °C
P4-2866	19.3 ± 1.6	23.9 ± 1.8
P2-2866	17.3 ± 0.3	25.8 ± 3.9
Δ2866	34.7 ± 3.6	53.4 ± 1.9
WT	17.8 ± 3.2	64.7 ± 9.6
P2-WT	13.7 ± 0.4	N/A
P4-LM	33.8 ± 3.4	N/A
P2--LM	28.8 ± 2.3	N/A
LM	38.0 ± 3.8	N/A

As the rate of an enzyme (cellulase)-catalyzed reaction increases as the temperature is elevated within an acceptable range (Mingardon, Bagert et al. 2011), we attempted to test the effect of growth temperature on cellulose conversion for our engineered strains. Given that the upper limit temperature for *C. cellulolyticum* growth is 45°C (Desvaux 2005), we increased the growth temperature from 34°C to 40°C. In general, both engineered strains and Δ2866 had significantly increased doubling times when grown at 40°C, along with a slightly reduced cell biomass (Table 3.1 & Figure 3.3D). Surprisingly, at the entry of stationary phase (150 h), P4-2866 degraded up to 51% cellulose, which is a dramatic improvement when compared to 29% at 34°C (Figure 3.3B & 3.3E). The elevated growth temperature did not affect the ability of P2-2866 to degrade cellulose but almost abolished the ability of Δ2866 to use cellulose (Figure 3.3E) within 150 h fermentation. Both engineered strains maintained an efficient formation of end products at 40°C, with much higher yields than the parental strain within 150 h (Figure 3.3F); P4-2866 even had a better metabolic performance than P2-2866 (Figure 3.3F, S3.5A, S3.5B & S3.5C). Interestingly, lactate titers of P4-2866 and

P2-2866 were increased by 10% and 23% when the growth temperature was elevated from 34°C to 40°C (Figure S3.4B & S3.5B), suggesting that a higher growth temperature promoted lactate formation in the engineered strains. In addition, P2-2866 accumulated more soluble sugars than Δ2866 at 40°C, with glucose and cellobiose increased by 55% and 45% respectively (Figure S3.5D & S3.5E). It is consistent with the above observations at 34°C.

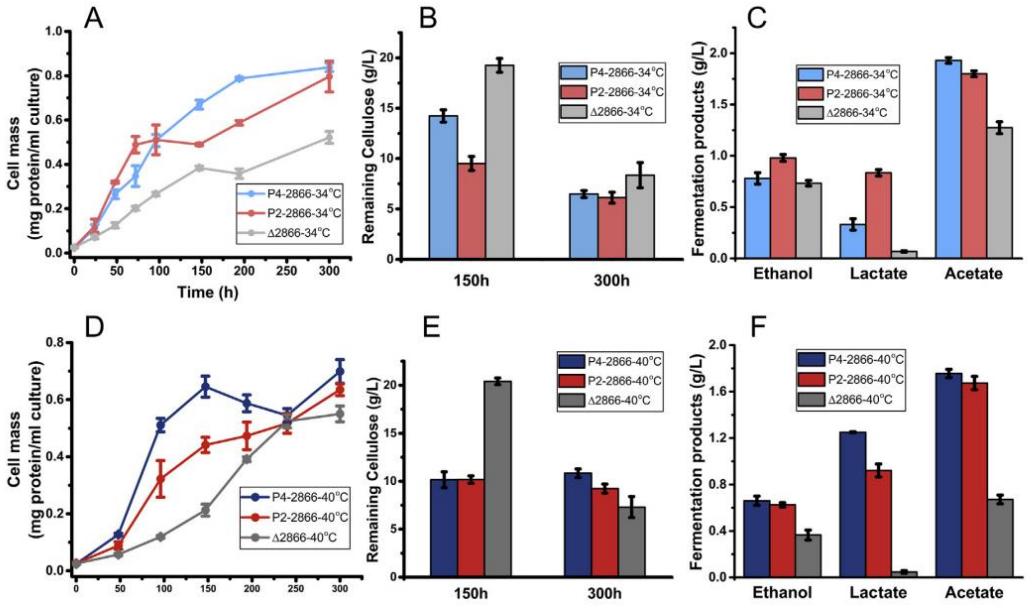

Figure 3. 3 Promoter integrants improved cell growth and the conversion efficiency of cellulose. Growth profiles of P4-2866, P2-2866 and Δ2866 grown at 34 °C (A) and 40 °C (D). Residual cellulose after 150 h or 300 h fermentation at 34 °C (B) and 40 °C (E). Titers of primary products after 150 h fermentation at 34 °C (C) and 40 °C (F). Data are presented as the mean of three biological replicates and error bars represent SD.

3.4.4 The Δ2866 parental strain conferred the strongest thermotolerance for cellulose degradation

We also compared cellulose degradation between the engineered strains and WT when grown at 34°C and 40°C. In general, WT outperformed P4-2866, P2-2866, and Δ2866 in cellulose degradation, ethanol production, and accumulation of soluble sugars at

34°C (Figure 3.3, 3.4A, 3.4B & S3.6). However, acetate production was lower in WT (Figure S6). At 40°C, it is notable that the elevated growth temperature severely inhibited WT growth on 20 g/L cellulose, with a 71% decrease in cell mass yield and only 16% cellulose used in 300 h fermentation when compared to growth at 34°C (Figure 3.4A & 3.4B) and was accompanied by low production of all primary end products and little soluble sugars accumulated (Figure S3.6). In contrast, Δ2866 and the engineered strains demonstrated nearly normal cell growth at 40°C and even slightly improved cellulose utilization (Figure 3.3, 3.4A & 3.4B). These distinct differences suggested that *mspI* disruption in Δ2866 may have improved the thermotolerance of C. *cellulolyticum* when utilizing cellulose.

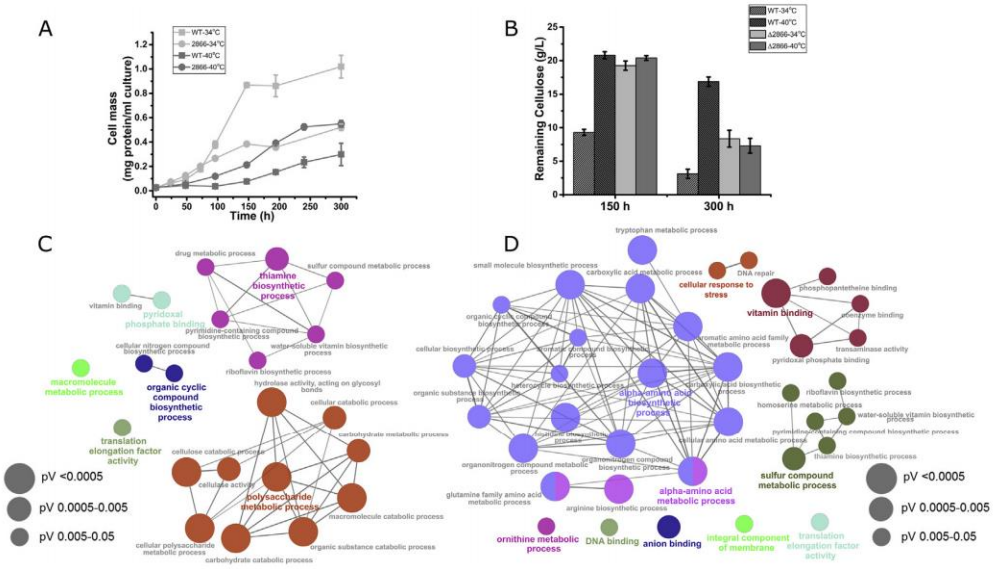

Figure 3. 4 Δ2866 outperformed WT in growth and cellulose degradation at an elevated temperature. (A) Growth profiles of WT and Δ2866 grown on 20 g/L cellulose at 34 °C and 40 °C. (B) The amount of residual cellulose after 150 h or 300 h fermentation for WT and Δ2866 at 34 °C and 40 °C. (C and D) Enrichment map of gene ontology (GO) terms in differentially expressed genes (DEGs) between Δ2866 and WT when grown on 20 g/L cellulose at 34 °C (C) and 40 °C (D). Only GO terms with Bonferroni-Hochberg corrected pV < 0.05 are displayed. Term enrichment significance is represented by circle size and the leading group term was based on the highest significance.

To better understand the altered thermotolerance, we monitored transcriptional changes between Δ2866 and WT at both 34°C and 40°C. A functional enrichment analysis of the differentially expressed genes (DEGs) found 22 GO terms significantly enriched between these two strains at 34°C, half of which were associated with polysaccharide metabolism, such as cellulose catabolic process and cellulase activity (Figure 3.4C). It is notable that 83% of the DEGs involved in polysaccharide metabolic processes were downregulated in Δ2866 at 34°C, especially those located in the *cip-cel* and *xyl-doc* gene clusters (Figure 3.4C & Figure S3.7). This may explain why Δ2866 had a lower efficiency of cellulose degradation than WT at 34°C. In comparison, there were 38 GO terms significantly enriched between Δ2866 and WT at 40°C and these were mainly involved in cellular response to stress, amino acid metabolism, and vitamin biosynthetic processes. 11 of the 13 DEGs associated with stress response and DNA repair processes were upregulated in Δ2866 at 40°C. In addition, the heat shock protein Hsp20 gene (Ccel_2938) had 3.4 folds increased expression in Δ2866. These upregulated genes in Δ2866 at 40°C presumably conferred better thermotolerance for Δ2866.

In response to the elevated growth temperature, WT and Δ2866 had transcriptional changes in 276 and 135 genes, respectively. There were no GO terms significantly enriched for the DEGs (34°C compared to 40°C) in WT (Figure S3.8); whereas DEGs (34°C compared to 40°C) in Δ2866 had a significant enrichment in bacterial-type flagellar filament, flagellum-dependent cell motility, bacterial-type flagellum assembly, and ABC transporter complex (Figure S3.9). This suggested that the loss of the *mspI* gene in Δ2866 made bacterial flagellum associated functions more susceptible to growth temperature.

3.4.5 Promoter insertion in other genetic backgrounds also improved microbial physiology

As the WT strain outperformed Δ2866 in cellulose degradation and ethanol production at 34°C (Figure 3.4B & S3.6), we attempted to insert P4 and P2 promoters into the WT strain at the same locus. However, we only successfully produced the P2 integrant (P2-WT). Consistently, we found P2-WT had a shortened doubling time and an increased maximum cell biomass compared to WT when grown on 20 g/L cellulose at 34°C (Table 3.1 & Figure 3.5A). P2-WT had a higher cellulose degradation efficiency than WT with 50% of cellulose was degraded by P2-WT versus 15% by WT within 69 h of fermentation (Figure 3.5B). Consistently, higher titers of end products were produced by P2-WT within 69 h fermentation (Figure 3.5C). As for the final titers, P2-WT produced more acetate but less ethanol than WT, indicating that P2-WT preferentially produced acetate as opposed to ethanol (Figure S3.10). Moreover, P2-WT accumulated more soluble sugars than WT as glucose and cellobiose were increased by 35% and 6 %, respectively in P2-WT after 300 h fermentation (Figure S3.10).

The P4 and P2 promoters were also inserted into our previous LM (Δ*ldh/mdh*) strain which does not produce lactate but produces more ethanol than WT (Li et al., 2012b). Similarly, the resulting strains, P4-LM and P2-LM, had a shorter doubling time, increased cell biomass, and more efficient cellulose degradation than the parental LM (Table 3.1 & Figure 5D, 5E). After 336 h fermentation, ethanol production by P4-LM and P2-LM was increased by 22% and 68%, respectively (Figure 5F); like LM, both P4-LM and P2-LM did not accumulate glucose and cellobiose. In sum, all of our engineered strains, regardless of genetic background, consistently demonstrated that

promoter insertion improved growth rate, maximum cell biomass, efficiencies of cellulose degradation and end-product formation.

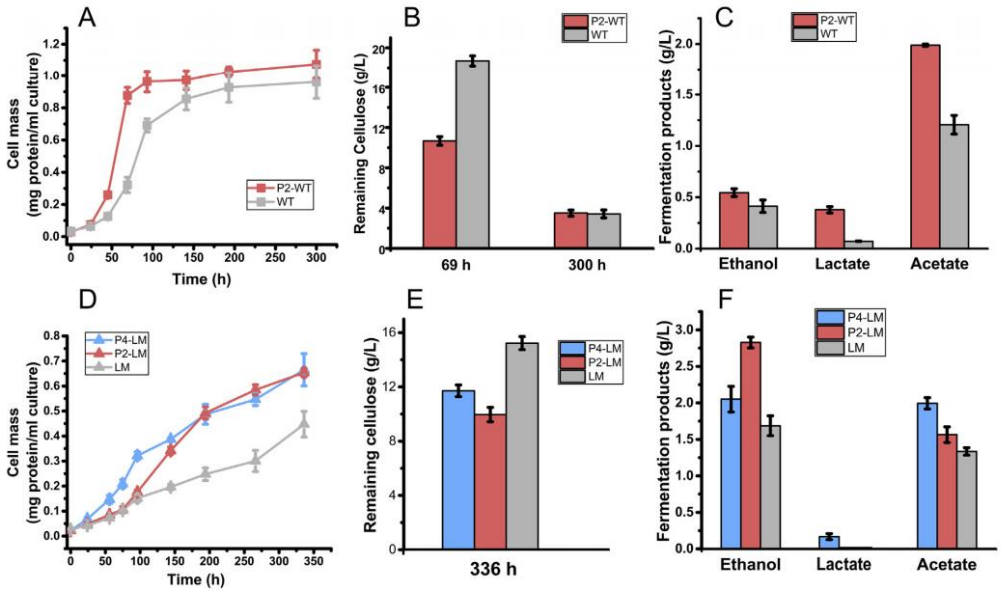

Figure 3. 5 Targeted promoter insertion in WT and LM enhanced cell growth and the conversion efficiency of cellulose. (A) Growth profiles of P2-WT and WT. (B) Residual cellulose after 69 h and 300 h fermentation by P2-WT and WT. (C) Fermentation products profiles after 69 h fermentation with P2-WT and WT. (D) Growth profiles of P4-LM, P2-LM, and LM. (E) Residual cellulose after 336 h fermentation with P4-LM, P2-LM and LM. (F) Fermentation products profiles at the end time point (336 h) for P4-LM, P2-LM, and LM.

3.5 Discussion

C. cellulolyticum is a CBP candidate and has been engineered for efficient cellulose degradation and biosynthesis of valuable products by both the introduction of heterologous genes and targeted gene inactivation (Guedon, Desvaux et al. 2002, Quast, Pruesse et al. 2012, Li, Xu et al. 2014). However, traditional genome editing approaches suffer from various technical limitations that affect our ability to edit *C. cellulolyticum* for metabolic engineering purposes. To our knowledge, increasing native cellulase expression by rewiring regulatory elements has never been reported in cellulolytic Clostridia. In this study, chromosomal integration of promoters (P4 or P2) improved

physiological features of Δ2866, WT, and LM strains, including growth rate, maximum cell biomass, cellulose degradation efficiency, and efficiency of formation of end products. Although engineered strains, such as P4-2866, P2-2866, and P2-WT, did not improve the production rates of all end-fermentation products, such as ethanol (Table 3.2), their higher growth rate and increased cell biomass dramatically enhanced the efficiency of cellulose bioconversion (Figure 3.3C, 3.3F & 3.5C), which therefore shortens the duration of fermentation and promotes equipment effectiveness.

Table 3. 2 Average values[a] of specific rates of product formation for strains grown in defined VM medium with 20 g/L cellulose at 34 °C

	q value (μmol [g of cells]$^{-1}$ h^{-1}) for:		
Strains	**EtOH**	**Lactate**	**Acetate**
P4-2866	185.50 ± 1.58	33.85 ± 2.60	322.41 ± 9.17
P2-2866	234.23 ± 7.32	79.34 ± 2.23	429.52 ± 7.25
Δ2866	319.37 ± 6.91	31.26 ± 0.92	356.63 ± 10.30
P2-WT	240.54 ± 6.70	66.65 ± 5.32	272.98 ± 27.40
WT	398.00 ± 14.01	38.08 ± 0.67	237.86 ± 11.80
P4-LM	252.13 ± 19.16	18.75 ± 6.12	231.60 ± 3.64
P2-LM	324.71 ± 21.60	0.00	183.13 ± 10.92
LM	226.51 ± 16.79	0.00	133.93 ± 12.70

[a]Average of three different experiments.

Gene expression analyses with microarray indicated that the P4 promoter had a stronger activity than the P2 promoter (Figure 3.1B). In addition to increasing the expression of downstream cellulase genes, the P4 promoter, which is only 60 bp in length, caused many more transcriptional changes than the P2 promoter and a staggering change in cell morphology (the formation of cell chains), irrespective of carbon source (Figure 3.2, Figure S3.1 & S3.3). SEM observations indicated that daughter cells formed but failed

to properly segregate. This could have resulted from the down-regulation of two cell wall hydrolase genes in cellobiose-grown P4-2866, Ccel_2890 (2.5-fold decrease) and Ccel_2941 (2-fold decrease). The transcriptional change of Ccel_2491 remained when P4-2866 was grown on cellulose. Other genes involved in cell wall biogenesis were also downregulated in P4-2866, especially when grown on cellobiose. How the insertion of the P4 promoter in the *cip-cel* gene cluster influenced the expression of the cell wall genes will be further investigated in the future. A recent study found that self-induced mechanical stress can trigger the formation of biofilm in E. coli (Chu, Kilic et al. 2018). It is possible that the stronger promoter activity may have caused a stressful burden during RNA transcription and protein synthesis, which then led to the formation of cell chains in P4-2866. Although P2-2866 presented faster cell growth and more efficient cellulose bioconversion than P4-2866 at 34°C, P4-2866 outperformed P2-2866 at an elevated temperature (40°C). Whether the formation of cell chain in P4-2866 contributes to its thermotolerance is an interesting question that we will investigate in future experiments.

Although we were unable to generate a P4 integrant in the WT background, which is most likely due to a heavy translational burden on cell resources, the resulting P2-WT strain demonstrated a faster growth rate and cellulose degradation resulted in higher concentrations of soluble sugars accumulated when compared to WT. For instance, glucose and cellobiose were 0.68 g/L and 2.14 g/L in P2-WT while only 0.15 g/L and 0.92 g/L in WT after 141 h fermentation (Figure S3.10D & S3.10E). The efficient accumulation of soluble sugars could be utilized in the future for co-culturing with sugar consuming bacteria (e.g. *Clostridium beijerinckii*) to diversify end products and

promote carbon utilization. In addition, we found P2-WT preferentially produced more acetate accompanied with a lower production of ethanol when compared to WT (Figure S3.10C). This is in line with the previous observation that increasing cellobiose diverts most carbon flux towards acetate formation (Giallo, Gaudin et al. 1983, Payot, Guedon et al. 1998, Guedon, Payot et al. 1999).

Although previous in vitro experiments indicated that a higher temperature (30-60°C) could increase the activity of cellulases from *C. cellulolyticum* (Mingardon, Bagert et al. 2011), we did not observe significant improvement by the WT strain in cell-based cellulose utilization experiments at an elevated temperature. Δ2866 and its derived engineered strains (P4-2866 and P2-2866) grew much better than WT at an elevated temperature. Previous studies have shown that heat shock proteins are involved in temperature resistance (Arsène, Tomoyasu et al. 2000, Chhabra, He et al. 2006), and it has been reported that E.coli cells expressing HSP20 protein from *S. solfataricus* gained greater thermotolerance in response to temperature stress (50°C) (Li, Yang et al. 2012). The increased expression of the hsp20 gene (Ccel_2938) in Δ2866 may have contributed to better thermotolerance. In addition to heat shock proteins, it has also been reported that thermotolerant genes involved in membrane formation/stabilization, DNA repair, and transmembrane transportation, were required for growth of Acetobacter tropicalis (Soemphol, Deeraksa et al. 2011), *E. coli* (Murata, Fujimoto et al. 2011) and *Zymomonas mobilis* (Charoensuk, Sakurada et al. 2017) at a high temperature. Our functional enrichment analysis also found that GO terms, such as cellular response to stress, DNA repair, and integral components of the membrane, were significantly enriched between Δ2866 and WT at 40°C (Figure 3.4D). 85% of the DEGs associated

with DNA repair, recombination repair, and base-excision repair, were upregulated in Δ2866 and 50% of the DEGs in integral membrane component category had increased expression. Therefore, these DEGs could be involved in repair of any heat-induced DNA damage and stabilization of the membrane, which may help to confer the improved thermotolerance observed for Δ2866 compared to WT. Although it is believed that the restriction-modification (RM) system in bacteria plays a role in stress response (Vasu and Nagaraja 2013), exactly how the loss of the MpsI endonuclease influenced the expression of the heat shock protein gene and other thermotolerant genes needs to be further investigated in future studies. Considering the Ccel_2866 gene, encoding the MspI endonuclease, belongs to the RM system in *C. cellulolyticum*, *mspI* inactivation might allow for the existence or accumulation of unmethylated CCGG DNA islands across the genome. Changes in methylation status could affect RNA transcription and explain the broad transcriptional impacts on many genes, some of which are associated with carbon metabolism and stress response. This possibility will be further investigated in future work.

In conclusion, one-step targeted promoter insertion in the *cip-cel* gene cluster was demonstrated to be a useful strategy for improving cellulose utilization and end-product formation, which can be applied in other bacteria with similar cellulosome-producing or other biosynthetic systems. We found that inactivation of the *mspI* gene improved thermotolerance in *C. cellulolyticum*, which was further enhanced by the insertion of the P4 and P2 promoters. This feature is an advantage for heat-producing microbial fermentation and high temperature-demanding production that usually requires a strong thermotolerance (Hendriks and Zeeman 2009, Abdel-Banat, Hoshida et al.

2010).Together, these findings provide new perspectives on how to reduce the cost of industrial fermentation processes and enhance the efficiency of cellulolytic bacteria on cellulose conversion.

Chapter 4 **Improvement of cellulose bioconversion in *Clostridium cellulolyticum* via targeted integration of β-glucosidase and inactivation of pleiotropic regulator CcpA**

4.1 Abstract

As the most abundant renewable feedstock on earth, lignocellulose has been used for production of sustainable biofuels and value-added chemicals. However, the low-efficiency bioconversion of lignocellulose greatly contributes to a high production cost. Here, two strategies were applied to improve the cellulose bioconversion in *Clostridium cellulolyticum*. Firstly, we employed CRISPR-Cas9 editing to improve cellulose degradation efficiency by integration of a exogenous β-glucosidase gene from *C.cellulovorans* to the front of the lactate hydrogenase gene (*ccel_2485*) of *C. cellulolyticum* genome. The β-glucosidase activity in the medium was increased by 2.5-fold in the engineered strain after integration. After incubation with 20 g/L cellulose, the engineered strain could degrade 12% more cellulose than the WT at the final time point, accompanied with 32% more ethanol production. Secondly, the *ccpA* gene in *C. cellulolyticum* was inactivated by CRISPR-Cas9 editing for releasing carbon catabolite caused by the cellobiose. To our surprise, the mutant even could not utilize the cellulose any more, indicating that the "*ccpA*" gene is necessary for the basic carbon metabolisms in *C. cellulolyticum* , and the inactivation of *ccpA* gene is not a good strategy applying in the *C. cellulolyticum* . Together, the integration of the exogenous β-glucosidase gene in the genome provides new perspectives on improving cellulose bioconversion in *C. cellulolyticum*, and also provides a new potential site in the genome of *C. cellulolyticum* for future integration and engineering.

4.2 Introduction

In 2014, as nonrenewable energy sources, fossil fuels accounted for 81% of the world's total primary energy supply ((2016) 2016). In the U.S., 67% of electricity was generated from fossil fuels in 2015 (FAQ: 2016). The tremendous reliance on fossil fuels contributes to many problems including a global energy shortage, environmental issues, and energy security (IEA 2009). Thus, it is necessary to reduce reliance on fossil fuels and develop green, carbon neutral, renewable energy to replace fossil fuels in the future. Based on the compatibility with current infrastructure and energy density, liquid biofuels are superior to other renewable energy forms, such as solar electricity or biogas, and could be used for the replacement of fossil fuels (Liao, Mi et al. 2016). Compared with other materials, non-edible lignocellulosic materials are the most promising feedstock as natural and reliable renewable resource for biofuels production (Lynd, Weimer et al. 2002, Doi 2008, Liao, Mi et al. 2016)[5-7]. In industry, there are two major steps for lignocellulose bioconversion: hydrolysis of lignocellulose into soluble sugars (e.g., pentoses, hexoses, and oligosaccharides) and fermentation of those soluble sugars into biofuels and commodity chemicals. The newest concept of lignocellulose bioconversion processes is named consolidated bioprocessing (CBP), which combines these two major steps into a single step in one bioreactor (Lynd, Weimer et al. 2002). Large scale implementation of CBP could offer the potential of lowering production costs, minimizing production energy, and increasing conversion efficiency rates (Lynd, Weimer et al. 2002, Balan 2014). Although the concept of CBP is promising for biofuels production, it still faces grand challenges on multiple fronts, including low-cost

technology for utilization of cellulose and highly efficient conversion from cellulose to biofuels by high-yield microorganisms (Lynd, Van Zyl et al. 2005, Liao, Mi et al. 2016, Lynd 2017). Therefore, these challenges are calling for engineering designs to increase efficiency and reduce the costs.

As the cellulose is recalcitrant to enzymatic degradation due to its crystallinity, more than 80% of the truly cellulolytic bacteria are distributed in two orders, *Actinomycetales* and *Clostridiales* (Koeck, Pechtl et al. 2014). Compared to the aerobic *Actinomycetales*, most cellulolytic bacteria in *Clostridiales* have a unique extracellular multi-enzyme complex, named cellulosome, which is more efficient in degrading the cellulose compared to the individual cellulases(Schwarz 2001). After the synergistic degradation of different cellulases in the cellulosome, the major soluble end products from hydrolysis of cellulose is the disaccharide, cellobiose(Desvaux 2005). Then the cellobiose can be directly transported into the cell by ABC transporter or phosphoenolpyruvate:carbohydrate phosphotransferase system (PTS), and further hydrolyzed into glucose by periplasmic β-glucosidases. However, earlier studies have found that the accumulation of cellobiose could inhibit both cell growth and cellulase production in the cellulolytic clostridia bacteria, such as *Clostridium thermocellum* and *Clostridium cellulolyticum* (Desvaux, Guedon et al. 2000, Abdou, Boileau et al. 2008, Gefen, Anbar et al. 2012, Xu, Huang et al. 2013), which would further decrease the efficiency of cellulose bioconversion. To overcome it, previous work has shown that the addition of the β-glucosidase (cellobiose degrading enzyme) could enhance the converting efficiency from cellobiose to the non-inhibitory glucose, which finally improved the cellulose degradation efficiencies in *C. thermocellum* (Kadam and

Demain 1989, Lamed, Kenig et al. 1991, Gefen, Anbar et al. 2012). In fact, previous studies hypothesized that a carbon catabolite repression (CCR) mechanism, which is a regulatory phenomenon by which the expression of functions for the use of secondary carbon sources and the activities of the corresponding enzymes are reduced in the presence of a preferred carbon source, involving cellobiose or other soluble sugars might contribute to the regulation of cellulase genes (Zhang and Lynd 2005, Abdou, Boileau et al. 2008). In *Clostridium acetobutylicum*, inactivation of the *ccpA* gene, which is the major player for the CCR, could enhance the xylose utilization (Ren, Gu et al. 2010).

Here, two engineering strategies were applied in the *C. cellulolyticum*, a model mesophilic CBP strain, for releasing the inhibition of cellobiose and increasing its cellulose degradation efficiency. Firstly, a reported β-glucosidase gene (Jeng, Wang et al. 2011) with high efficiency was integrated into the genome of *C. cellulolyticum* by the CRISPR-Cas9 system to test whether the cellulolytic activity and ethanol production could be further improved in the engineered strain. Secondly, by silico analysis, the "*ccpA*" gene was inactivated in *C. cellulolyticum* and was supposed to release or eliminate the CCR caused by the cellobiose. Our results demonstrated that the integration of β-glucosidase into the genome of *C. cellulolyticum* is an efficient strategy to increase the converting from cellobiose to glucose, and further improve efficiencies of cellulose degradation and end-product formation. However, not only could the inactivation of *ccpA* gene not increase the cellulose degradation efficiency, but it could influence the utilization of cellulose.

4.3 Materials and Methods

4.3.1 Bacterial strains and plasmid construction

Strains, plasmids, and primers used in this study were listed in Supplementary Table S4.1 and S4.2, respectively. pCas9n-CcBglAinser-donor, pCas9n-Δ1005-donor and pCas9n-Δ1438-donor for integration of β-glucosidase in front of *ccel_2485* gene, inactivation of *ccel_1005* gene and *ccel_1438* gene were constructed by Gibson assembly. Briefly, left and right homologous arms for each vector were amplified from wild-type genomic DNA of *C. cellulolyticum* and purified separately. The *CcBglA* gene was amplified from the plasmid pEr-bglA using primers 2485BglAf and 2485 BglAR. The P4::gRNA cassette containing the 20-bp protospacer was amplified from pCR/8w p4-4 prom4 and pMS-RNA by 3F and 2RR primers. The linear backbone was digested from pFdCas9n-p4-pyrF_w/2kbΔ by KpnI and PvuI. Finally, these fragments were assembled using Gibson assembly and the assembled product was transformed into *E. coli* for colony screening and confirmed via Sanger sequencing (Oklahoma Medical Research Foundation). The vectors for above transformants were cured as previously reported (Quast, Pruesse et al. 2012).

4.3.2 Transformation

The wild-type *C. cellulolyticum* strain was transformed with above corresponding plasmids by electroporation as described previously (Lamed, Kenig et al. 1991, Jennert, Tardif et al. 2000). Before the transformation, each plasmid was methylated using MspI Methyltransferase (New England Biolabs, Ipswich, MA) and then purified.

4.3.3 Mutant verification and measurements of β-glucosidase activity

Genomic DNA of WT-CcBglA, Δ1005 and Δ1438 were extracted using the GenElute™ Bacterial Genomic DNA Kit. Primers 2485LF and ID2485R were used for verification of WT-CcBglA; 1005_SF and 1005_SR were used for verification of Δ1005; 1438_SF and 1438_SR were used for verification of Δ1438. The activity of β-glucosidase was measured as previously reported using the β-glucosidase Activity Assay Kit (Sigma Aldrich) following instructions.

4.3.4 Measurement of cell growth, fermentation products, and remaining cellulose

All *C. cellulolyticum* strains including engineered strain and mutant strains were revived in complex VM medium with 5 g/L cellobiose and then transferred to defined VM media. The cellobiose-grown cultures at an OD_{600} of 0.5-0.6 were used to inoculate 50 ml of defined VM media containing 20 g/L Avicel PH101 crystalline cellulose. Each strain had three biological replicates. During growth, 1 mL of cell culture was sampled at each time point and then stored at -80°C for future determination of cell biomass, metabolites, and remaining cellulose. Growth curves for each sample grown on cellulose was estimated by total protein measurement (Quast, Pruesse et al. 2012). High-performance liquid chromatography (HPLC) was used to measure the major fermentation products (lactate, acetate, and ethanol) and soluble sugars (cellobiose and glucose) in the supernatant of spent medium (Quast, Pruesse et al. 2012). The specific rate of product formation was calculated as previous described (Desvaux, Guedon et al. 2000). The remaining cellulose in the medium was measured by the phenol-sulfuric acid method (Hemme, Fields et al. 2011).

4.3.5 RT-PCR and microarray analysis

WT-CcBglA and WT were grown in the defined VM medium with 20 g/L cellulose. Each strain was collected at mid-exponential growth phase and the collected cell pellets were immediately frozen with the liquid nitrogen. The total RNA extraction, microarray hybridization and microarray data analysis were performed as previously described (Tao, Xu et al. 2020). Primers BglA-RTF and BglA-RTR were used for verification of the expression of the *CcBglA* gene in WT-CcBglA by RT-PCR. Each strain contained four biological replicates.

4.4 Results and Discussion

4.4.1 Targeted integration of β-glucosidase into the genome of C. cellulolyticum .

In *C. cellulolyticum*, *ccel_0374* is supposed to encode β-glucosidase based on the annotation analysis. However, no previous studies proved that *ccel_0374* could indeed encode functional β-glucosidase and the efficiency of the β-glucosidase encoded by *ccel_0374* is elusive. Therefore, integration of an exogenous β-glucosidase gene with clear function and high efficiency into *C. cellulolyticum* is a suitable strategy for enhancing the enzymatic conversion from cellobiose to glucose. In terms of the activities of different β-glucosidases from previous study(Jeng, Wang et al. 2011), the β-glucosidase from *C.cellulovorans* has the strongest activity and the optimal temperature for it is 30-50 ℃, which is also compatible with the growth temperature of *C. cellulolyticum* . On the other hand, considering that the cellobiose is normally accumulated outside the cell (Desvaux 2005), the exogenous β-glucosidase should be secreted into extracellular space. Therefore, we slightly modified the *C.cellulovorans* β-

glucosidase (CcBglA) gene with a signal peptide of Cel9E from *C. cellulolyticum* , which could help secretion of the CcBglA outside the cell.

Based on previous RNA sequencing results for *C. cellulolyticum* responding to cellulose (Xu, Huang et al. 2015), seven genes in *C. cellulolyticum* belong to monocistron and their promoters were response to the cellulose (Table S4.3). Among them, ccel*_1380 and ccel_1432*, encoding hypothetical protein and with relatively high cellulose response ratio, were chosen as the potential sites for integration of *CcBglA* gene. In addition to them, *ccel_2485* encoding the L-lactate dehydrogenase was also chosen as the potential site, since previous study found inactivation of the *ccel_2485* could enhance the ethanol production and a non rho-independent terminator was predicted downstream of the gene (Quast, Pruesse et al. 2012). For each potential site with *CcBglA* gene integration, we constructed an all-in-one vector for Cas9 nickase-based genome editing, which helped to locate and insert the *CcBglA* gene upstream of the corresponding potential site (*ccel_1380, ccel_1432* or *ccel_2485*) (Figure 4.1A). However, we only successfully integrated the *CcBglA* gene upstream of the *ccel_2485*, which was confirmed by PCR and Sanger sequencing (Figure 4.1B). The corresponding engineered strain was named WT-CcBglA.

The reverse-transcription PCR (RT-PCR) revealed that the exogenous *CcBglA* gene was successfully expressed (Figure 4.1C) and the β-glucosidase activity assay also found that WT-CcBglA had a 2.5 times greater β-glucosidase activity in the supernatant of the growth medium compared to WT (Figure 4.1 D). All of these indicated that the expression of *CcBglA* gene was normal and the secretory CcBglA could perform proper function outside the cell.

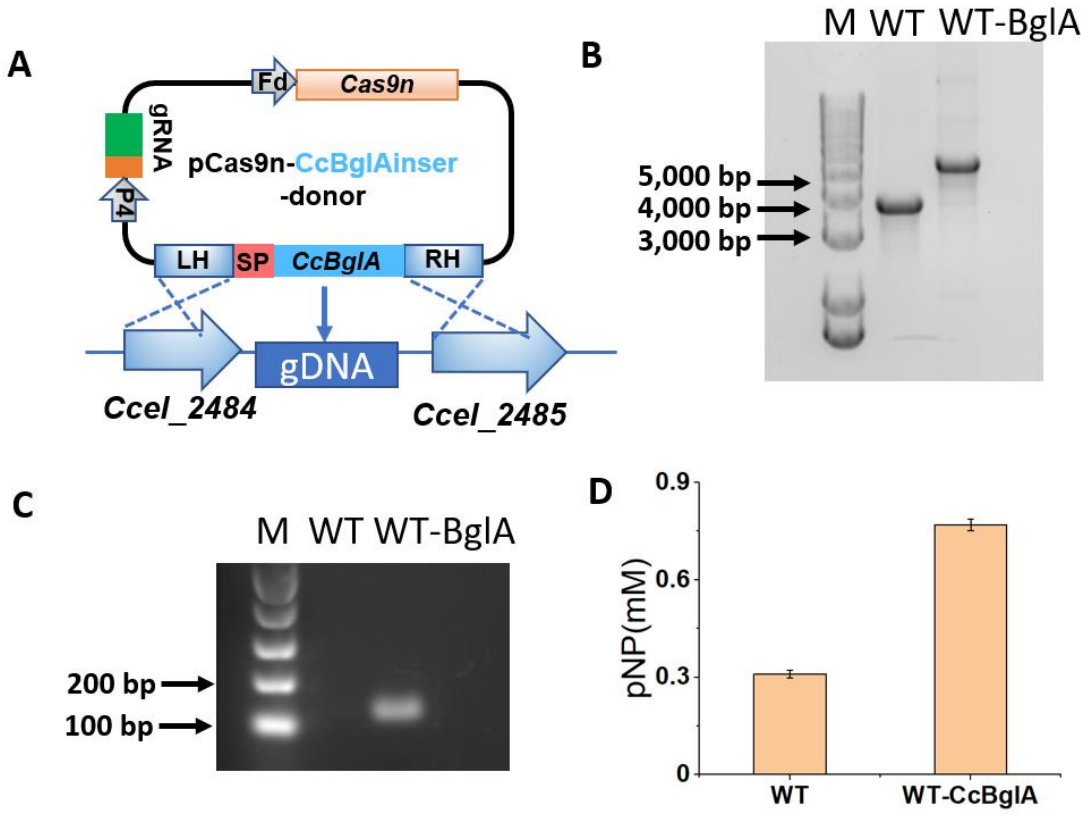

Figure 4. 1 One-step promoter integration increased the extracellular β-glucosidase activity. (A) An overview of the Cas9 nickase-based genome editing in *C. cellulolyticum.* Plasmid pCas9n-CcBglAinser-donor was used for the integration of β-glucosidase gene from *C.cellulovorans* to the genome of *C. cellulolyticum*. SP, signal peptide; CcBglA, *C.cellulovorans* β-glucosidase. (B) PCR identification of WT-CcBglA. (C) RT-PCR identification of expression of *C.cellulovorans* β-glucosidase gene in *C. cellulolyticum*, which was incubated with 20 g/L cellulose in the defined VM medium. (D) *In vitro* enzymatic assay measuring the extracellular β-glucosidase activity in WT and WT-CcBglA

4.4.2 The conversion of cellulose to end products was improved in engineered strain WT-CcBglA

To determine whether the newly engineered strain (WT-CcBglA) was better in degradation of cellulose compared to the WT, we inoculated the WT-CcBglA and WT into the defined VM medium containing 20 g/L cellulose. In general, WT-CcBglA outperformed WT in cellulose degradation and ethanol production. Although WT-CcBglA had longer doubling times (21 h in WT-CcBglA and 18 h in WT) and the

maximum cell biomass in WT-CcBglA was less than WT (Figure 4.2A), the WT-CcBglA could degrade 12% more cellulose than the WT at final time point (Figure 4.2B), indicating that the cellulose degradation efficiency was better in the engineered strain. For the end fermentation products, the production of ethanol was increased by 32% compared to the WT, and the productions of the lactate and acetate were decreased in WT-CcBglA (Figure 4.2C). Also, WT-CcBglA accumulated more glucose than WT. More specifically, glucose was increased by 18% (Figure 4.2C).

As the *ccel_2485* encodes the lactate dehydrogenase gene and the integration of CcBglA in *C. cellulolyticum* disrupted the expression of *ccel_2485*, it was not surprised that the production of lactate decreased by 50% in WT-CcBglA compared to WT, which was consistent with previous study (Li, Tschaplinski et al. 2012). The microarray results also confirmed that the expression of ccel_2485 in WT-CcBglA was decreased by 70% compared to WT. As a result, the carbon flow had to divert to acetate and ethanol production pathways. On the other hand, coupled with increased ethanol production, the acetate production was slightly decreased by 10% in WT-CcBglA (Figure 4.2C). In fact, the acetate production pathway in *C. cellulolyticum* can produce more ATPs than other two pathways based on previous study (Desvaux, Guedon et al. 2000). Therefore, less production of acetate in WT-CcBglA indicated less ATPs formation, which may be the reason why WT-CcBglA had a longer doubling time and less maximum cell biomass. On the other hand, disruption of lactate also could cause the imbalance between NAD+ and NADH as the lactate production is a process for NADH consumption (Quast, Pruesse et al. 2012), which further causes expression changes of other genes. Based on the microarray analysis, we found that genes involved

in DNA topological change and double stranded DNA binding were significantly enriched, which may be caused by the imbalance of NAD^+ and NADH (Li, Tschaplinski et al. 2012).

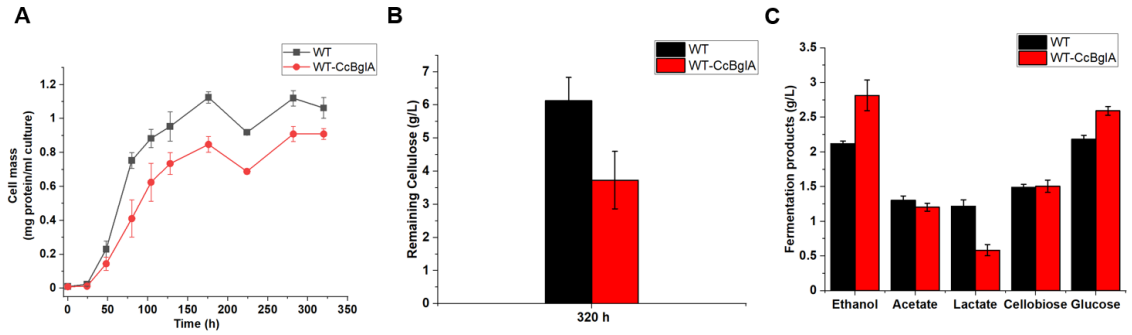

Figure 4. 2 Integration of the *CcBglA* gene improved the conversion efficiency of cellulose and ethanol production. Growth profiles of WT and WT-CcBglA grown at defined VM medium containing 20 g/L cellulose at 34 °C (A) . Residual cellulose for WT and WT-CcBglA at the end fermentation time at 34 °C (B). Titers of primary products and soluble sugars for WT and WT-CcBglA at the end fermentation time at 34 °C (C). Data are presented as the mean of three biological replicates and error bars represent SD.

Although there was no significant change for the accumulation of cellobiose, more glucose was accumulated in engineered strain, indicating that more cellulose was degraded in engineered strain, which was consistent with our phenotype result (Figure 4.2 B & C). Meanwhile, we didn't observe any genes belonging to the *cip-cel* operon were changed, indicating that the enhanced degradation efficiency of cellulose was not caused by the increased abundance of cellulases but resulted from synergistic degradation of the cellulosomes and addition of CcBglA. Additionally, it also indicated that the accumulation of cellobiose caused repression of cellulase genes may not exist in *C. cellulolyticum*.

4.4.3 Distribution of PTS components in C. cellulolyticum

The CCR is one kind of carbon metabolism regulations carried out by PTS (Deutscher, Aké et al. 2014). There are three major components in the PTS, which are enzyme I, enzyme II complex and HPr protein. In canonical CCR mechanisms in *Bacillus subtilis*, HPr and its upstream regulator (HPrK) and downstream transcription factor (CcpA) are major players in regulating the carbon catabolite repression (Görke and Stülke 2008). Based on the *in-silico* analysis in *C. cellulolyticum*, we did not find any ortholog of the EI or EII complex. An HPr ortholog was found in *C. cellulolyticum*, which is encoded by the *ccel_0806* gene. However, the key His15 in canonical HPr is substituted to Asp15 in the HPr of *C. cellulolyticum*. Secondly, an HprK ortholog (Ccel_2293) was found with high sequence identity to well-characterized HprK proteins in *C. cellulolyticum* and it contains a conserved nucleotide-binding motif and a downstream signature sequence (Galinier, Kravanja et al. 1998, Hanson, Steinhauer et al. 2002). Finally, five transcriptional regulators belonging to the LacI family in *C. cellulolyticum* are the candidate homologs of the CcpA protein. Among them, Ccel_1005 has a sequence identity of 34% and similarity of 55% to that of B. subtilis CcpA (Xu, Huang et al. 2013), which was thought to play the role of CcpA in *C. cellulolyticum* (Figure 4.3A).

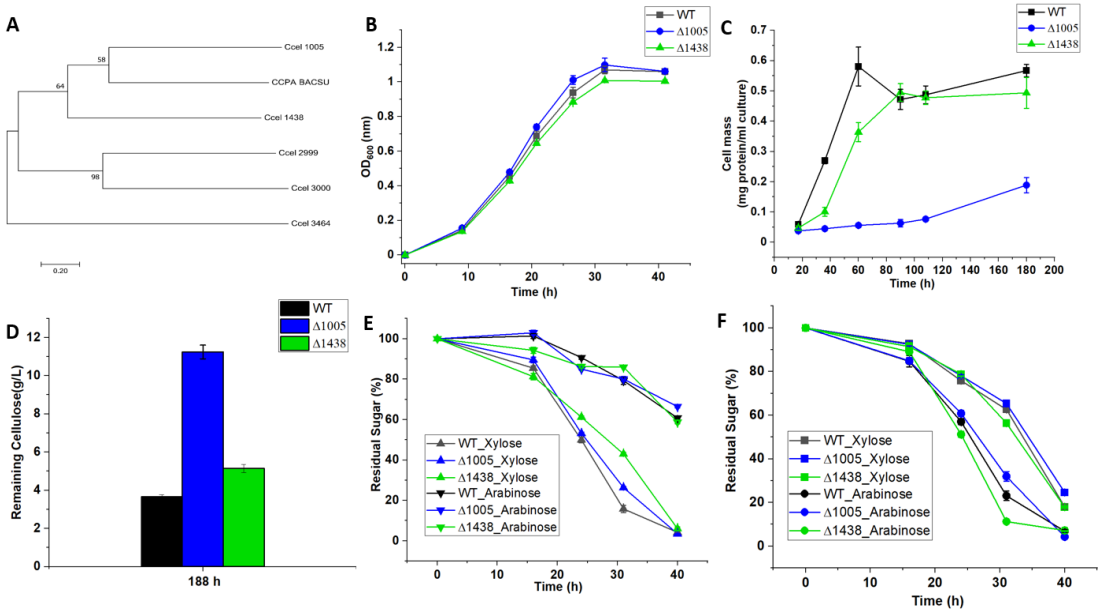

Figure 4. 3 Inactivation of the *ccpA* (*Ccel_1005*) in *C. cellulolyticum* disrupted the utilization of cellulose. A, molecular phylogenetic analysis between LacI family proteins in *C. cellulolyticum* (Ccel_1005, Ccel_1438, Ccel_2999, Ccel_3000 and Ccel_3484) and CcpA protein in *Bacillus subtilis* (CCPA BACSU); B, growth profiles of WT, Δ1005 and Δ1438 grown at defined VM medium containing 4 g/L cellobiose, 4 g/L glucose, 3 g/L xylose and 2 g/L arabinose at 34 °C; C, growth profiles of WT, Δ1005 and Δ1438 grown at defined VM medium containing 10 g/L cellulose at 34 °C; D, residual cellulose for WT, Δ1005 and Δ1438 at the end fermentation time at 34 °C; E, xylose and arabinose consumption profiles of WT, Δ1005 and Δ1438 grown at defined VM medium containing 4 g/L cellobiose, 4 g/L glucose, 3 g/L xylose and 2 g/L arabinose at 34 °C; F, cellobiose and glucose consumption profiles of WT, Δ1005 and Δ1438 grown at defined VM medium containing 4 g/L cellobiose, 4 g/L glucose, 3 g/L xylose and 2 g/L arabinose at 34 °C. All data are presented as the mean of three biological replicates and error bars represent SD.

4.4.4 Mutagenesis and characterization of LacI regulator genes in C. cellulolyticum .

Since previous studies have found that the inactivation of *ccpA* gene could release the repression of glucose in *B. subtilis* and *C. acetobutylicum* (Görke and Stülke 2008, Ren, Gu et al. 2010), and the accumulation of cellobiose was thought to cause the repression of *cip-cel* operon through CCR, we were planning to inactivate the ccel_1005 gene for releasing the repression and increase the expressions of *cip-cel* genes. We constructed an all-in-one vector for Cas9 nickase-based genome editing, which requires co-

expression of Cas9 nickase and a customized gRNA, as well as homologous regions that inserted 7 bp restriction enzyme site at the 5' of *ccel_1005* gene to cause frameshift mutation. The corresponding mutant strain is named Δ1005, which was confirmed by PCR amplification and sequencing. Before we determined the cellulose degradation ability for the Δ1005, we inoculated it to the defined VM medium containing carbon mixture to determine whether the inactivation could influence its growth on soluble carbons. However, we did not observe any differences for the growth of carbon mixture between the Δ1005 mutant and WT (Figure 4.3B, 4.3E & 4.3F).

4.4.5 LacI repressor is essential for cellulose degradation in C. cellulolyticum.

To test whether the inactivation of *ccpA* could enhance the degradation of cellulose through releasing the repression, the Δ1005 mutant and WT were inoculated into the defined VM medium containing 20 g/L cellulose. To our surprise, the Δ1005 mutant could not utilize the cellulose as the carbon source anymore (Figure 4.3C & 4.3D). After a complementary plasmid which could express the native *ccel_1005* gene was transformed into the Δ1005 mutant, the Δ1005 mutant containing the complementary plasmid could be recovered with the ability in utilization of cellulose. All of these indicated that the *ccel_1005* gene was essential and necessary for cellulose degradation in *C. cellulolyticum.*

Considering there are five LacI family members in *C. cellulolyticum* and the other gene, and *ccel_1438,* was the second most similar *ccpA* gene among the five LacI family genes, we also inactivated the *ccel_1438* gene by the all-in-one vector for Cas9 nickase-based genome editing in *C. cellulolyticum* . The corresponding mutant strain was named Δ1438. Compared to Δ1005, the soluble sugar consumption rates were changed in

Δ1438, especially for the arabinose consumption rate, which was significantly increased by 20% (Figure 4.3E & 4.3F), indicating that the 1438 protein is more like the previously reported CcpA protein. However, after determining the Δ1438 on cellulose, although the Δ1438 did not lose the function of cellulose utilization as the Δ1005, the degradation efficiency of cellulose was decreased by 40% in Δ1438 (Figure 4.3C & 4.3D). All of these indicated that the Ccel_1438 protein was very important in maintaining the stability of carbon metabolisms in *C. cellulolyticum*, and the disruption of Ccel_1438 protein may cause the disorder of the carbon metabolisms in *C. cellulolyticum*.

4.5 Conclusion

Integration of CcBglA into the genome of *C. cellulolyticum* is a great strategy for improving the cellulose degradation efficiency. Additionally, we also found a good potential target site (*ccel_2485*) for integration of exogenous genes into the genome of *C. cellulolyticum* , which could avoid the use of plasmid for exogenous gene expression and could ensure the increase of ethanol production. The only defect for the WT-CcBglA was the relatively slower growth rate after disruption of the ccel_2485. In our previous study (Tao, Xu et al. 2020), we inserted a Pm2112 promoter into the *cip-cel* gene operon, which could help to enhance the growth rate and increase maximum cell biomass. Therefore, to overcome the defect of WT-CcBglA, we may apply the "promoter insertion" strategy into the WT-CcBglA in the future. On the other hand, the "CcpA inactivation" strategy was failed in this study, which indicated that the LacI members were very important in regulation of carbon metabolism in *C. cellulolyticum*. In fact, the PTS components found in *C. cellulolyticum* were totally different from those

reported in *B.subtilis*, which also indicated that the CCR mechanisms should be different from the *B.subtilis*. Therefore, simply inactivation of "*ccpA*" might not have the same phenomena as previously reported (Ren, Gu et al. 2010). Together, these findings provide new perspectives on how to engineer the cellulolytic bacteria for enhancing their abilities in cellulose degradation and conversion.

Chapter 5 **Long-term warming likely stimulates soil microbial respiration due to enlarged and activated carbon degraders**

5.1 Abstract

Global climate and ecosystem models for accurate climate projections require explicit consideration of the physiology and composition of active microorganisms that convert soil organic carbon (SOC) to atmospheric CO_2. However, it is extremely challenging because of the complex interplay between microbial community composition, activity, CO_2 emission, and global warming. Here, we identified active taxa responsible for carbon (C) degradation in a temperate grassland subjected to 7-year experimental warming. In a quantitative stable-isotope probing experiment with the ^{13}C-labeled grass straw, we found that soil respiration was increased by 13.3% in warmed soil samples, with a 14.4% higher priming effect on native SOC. Warming increased the total abundance of active C degraders by 58.2%. A total of 56 C-degrading phylotypes were detected only in the warmed samples, which mainly belong to *α-Proteobacteria*, *Bacillales*, *Actinobacteria*, and *Bacteroidetes*. To verify it, we demonstrate that their relative abundances were increased ($p < 0.050$) by warming over 7-years, based on yearly measurements of bacterial communities by 16S rRNA gene sequencing. In addition to higher phylogenetic diversity of C-degrading communities, warming increased β-diversity among biological replicates, showing phylogenetic diversification. Microbial activities of C degradation were also higher, concomitant with increased abundance of functional genes responsible for C degradation. Being a rare integrative study with multifacet of composition, capacity, activity, and function analyses, we predict that grassland soil C loss by warming can be higher than currently anticipated.

5.2 Introduction

Global warming is occurring at an average warming rate of 0.2 °C per decade, which becomes a major scientific and political issue worldwide (Masson-Delmotte, Zhai et al. 2018). A current explanation suggests that recent global warming is a result of the positive energy imbalance in the radiative budget of the Earth system caused by anthropogenic emission of greenhouse gases (Zhou, Wan et al. 2007). In a warmer world, the decomposition rate of SOC is increased, which causes a major loss of terrestrial C to the atmosphere, and finally contributes to the increase in atmospheric CO_2, hence leading to positive feedback to warming (Pachauri and Reisinger 2008). In high-latitude regions of the northern hemisphere, it has been confirmed and predicted in many ecosystems that the increased temperature could enhance the positive feedbacks through the increased decomposition of SOC (Schuur, Bockheim et al. 2008, Schuur, McGuire et al. 2015, Xue, Yuan et al. 2016). However, other ecosystems, such as temperate grasslands, were much less reported to respond similarly in terms of the feedback type and the magnitude of responses to warming (Lenton, Held et al. 2008). Containing ~12% of the terrestrial organic matter, the grassland C pool is mainly composed of grass litter and rhizodeposits, and regulated by microbial communities (Schlesinger 1977, Shahzad, Chenu et al. 2012). Since temperature is a primary driver of metabolic rates and biochemical processes (Brown, Gillooly et al. 2004, Guo, Zhou et al. 2019), in temperate grassland, several previous studies have found that warming could stimulate microbial C degradation, and some C-degrading genes involved in hemicellulose and cellulose degradation were increased under warming. (Zhou, Xue et al. 2012, Stegen, Lin et al. 2013). However, microbes are often dormant in the natural

environment (Jones and Lennon 2010). Failure to consider dormancy in microbially controlled C degradation can result in inaccurate explanation and prediction for microbial mechanisms of SOC degradation and future SOC dynamics, especially for microbes in field conditions that are influenced by environmental factors including soil temperature and moisture (Wang, Mayes et al. 2014, Wang, Jagadamma et al. 2015). Incorporated with functional analysis of active microbial community, a previous study using the stable isotope probing (SIP) microcosm-based experiments found that warming could indeed increase the abundance of C-degrading genes in active bacterial community in temperate grassland soil (Cheng, Zhang et al. 2017). However, due to the limitation of the traditional SIP (Hungate, Mau et al. 2015), the specific identities of those active degraders and their biomass were not determined, which are very important microbial parameters for improving the accuracy of microbial SOC degradation model. Therefore, the microbial mechanisms of SOC degradation under warming in temperate grassland have not been fully characterized. Meanwhile, warming was also shown to increase the divergent succession of soil microbial communities in temperate grassland soil (Guo, Feng et al. 2018), indicating that the composition of active degraders in warmed soil may also be different from those in un-warmed soil.

On the other hand, microbial degradation of native soil organic C can be induced by the fresh C input, termed as the priming effect (Mau, Dijkstra et al. 2018). Our previous study found that the priming effect could be enhanced by warming in tundra soil in high-latitude regions (Tao, Feng et al. 2020). However, whether the priming effect can also be enhanced by warming remains elusive for the temperate grassland ecosystem.

An updated SIP method named quantitative stable isotope probing (qSIP) has been developed recently (Hungate, Mau et al. 2015), enabling quantification of C assimilation rate and abundance of active taxa responsible for utilizing isotopic C. To address the above knowledge gaps, we employed the qSIP technique combined with Biolog EcoPlates and a functional gene microarray named GeoChip to investigate how climate warming affects active SOC-degrading community in temperate grassland ecosystem and further understand the microbial mechanisms of SOC degradation responding to warming. Our overarching hypothesis is that long-term warming would stimulate the degradation of both fresh C and native soil organic C (SOC) by shifting the abundance, composition, and enhancing C-degrading capacities of active bacterial communities. Our findings can yield new insights into improving the accuracy of microbially controlled SOC degradation models and further understanding the microbial mechanisms of SOC degradation under warming.

5.3 Materials and Methods

5.3.1 Site description and field measurements.

The in situ warming experiment was carried out in the tallgrass prairie of Kessler Atmospheric and Ecological Field Station (KAEFS) in McClain County, Oklahoma, USA (34° 58' 44" N, 97° 31' 15" W). Detailed site information was described in our previous study (Guo, Feng et al. 2018). The soil type of this study site is the Port-Pulaski-Keokuk complex, which is a well-drained soil formed in loamy sediment on flood plains (Stegen, Lin et al. 2013). The soil texture is loam with 51% of sand, 35% of silt, and 13% of clay, with a soil bulk density of 1.2 g cm−3 (Stegen, Lin et al. 2013).

The soil has a high available water holding capacity (37%), neutral pH, and a deep (about 70 cm), moderately penetrable root zone (Guo, Feng et al. 2018). Constantan-copper thermocouples wired to a Campbell Scientific CR10X datalogger (Campbell Scientific Inc., Logan, UT, USA) were used to measure and record soil temperature every 15 min at the soil depth of 7.5, 20, 45 and 75 cm at the center of each plot. To analyze the micro-climate of soil samples, we calculated the annual average temperature data of 2016 (the sampling year) at a depth of 7.5 cm. Soil volumetric water content at a depth of 0–15 cm was measured every month by placing a portable time-domain reflectometer (SoilMoisture Equipment Corp., Goleta, CA, USA) in three randomly selected locations of each plot. Then the annual average soil moisture of 2016 was calculated. Above ground plant biomass was measured in the peak growing season (September 2016) by a modified pin-touch method, as described previously (Xu, Sherry et al. 2013).

5.3.2 Soil sample preparation and geochemical factor measurements.

Eight soil samples used in this study were collected in September 2016 from the 0–15 cm depth of 4 warmed plots and 4 control plots (i.e., four biological replicates for warming or control). Visible roots longer than 0.25 cm and stones were removed from soil by 2-mm-mesh metal sieves (Hogentogler Co. Inc., Columbia, MD, USA), and thoroughly homogenized by manual mixing. All samples were then analyzed for soil geochemistry by the Soil, Water, and Forage Analytical Laboratory at Oklahoma State University (Stillwater, OK, USA). Organic C and total N contents in soil were determined by a dry combustion C and N analyzer (LECO Corp., St. Joseph, MI, USA).

Soil pH was measured at a water-to-soil mass ratio of 2.5:1 by an Accumet XL15 pH meter with a calibrated combined glass electrode (Accumet Engineering Inc., Westford, MA, USA). Detailed information for environmental factors in 2016 was listed in Table S5.1.

5.3.3 SIP incubation and priming effect calculation.

^{13}C- and ^{12}C-straw of common wild oat (Avena fatua was provided by Dr. Mary K. Firestone, University of California, Berkeley, CA, USA), was used as stable-isotope probe substrate to simulate deposition of grass litter to the soil. The ^{13}C atom% of the ^{13}C-straw was 75.1%, as determined by the Stable Isotope Facility, University of California, Davis, CA, USA. Three incubation groups, i.e., (1) with 0.1 g of ^{13}C-straw in 5 g of soil (2% w/w) as isotopic treatment, (2) with 0.1 g of ^{12}C-straw in 5 g of soil (2% w/w) as isotopic control, and (3) with 5 g of soil as the background, were set up for both in situ warmed and control samples. To obtain homogenized soil samples, those three groups were thoroughly stirred with steel spoons. Each replicate was sealed in a 25-ml light proof bottle and incubated at 25 °C. The duration of the incubation was set to be seven days to minimize the cross-feeding among microbes for increasing reliability in identification and the impact of the incubation on microbial community structures.

Headspace gas was collected daily into 12-ml evacuated vials (Labco Ltd., Lampeter, UK), after which the bottles were opened and refreshed for 30 min in a clean bench with the maximal flow of wind. To avoid gas contamination from the atmosphere, we diluted sampled gas by injecting 10 ml of N_2 gas into each vial, generating a positive

pressure to the atmosphere. $^{12}CO2$ and $^{13}CO2$ concentration were measured at the Stable Isotope Facility, University of California, Davis, California, USA, and calculated as previously described (Tao, Feng et al. 2020). Briefly, the percentage of the CO_2-C deriving from ^{13}C-straw was calculated as:

$$\%C_{substrate} = \frac{\delta_C - \delta_T}{\delta_C - \delta_L} \times 100\%$$

where δ_C is the $\delta^{13}C$ value of respired CO2 from the soil with no straw, δ_T is the $\delta^{13}C$ value of respired CO2 from the soil with ^{13}C-straw, and δ_L is the $\delta^{13}C$ value of ^{12}C-straw. The amount of SOC primed by straw was calculated as total soil respiration after straw addition minus the amount of C respired from straw, and then minus the amount of C respired from the soil with no straw.

5.3.4 Soil DNA extraction.

After a 7-day incubation, soil DNA was extracted with the liquid nitrogen grinding method (Zhou, Bruns et al. 1996), followed by PowerMax Soil DNA Isolation Kit (MO BIO Laboratories, Inc., Carlsbad, CA, USA) according to the manufacturer's protocol. DNA quality was assessed based on spectrometry absorbance at wavelengths of 230 nm, 260 nm, and 280 nm by a NanoDrop ND-1000 Spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). The absorbance ratios of 260/280 nm were larger than 1.8 and 260/230 nm were about 1.7, which were considered to be of excellent DNA quality. DNA was quantified by PicoGreen using a FLUOstar OPTIMA fluorescence plate reader (BMG LabTech, Jena, Germany), which showed that DNA concentrations were 49.1 ± 2.6 ng/µl, with no difference ($p > 0.050$) among treatments or isotopic treatments. Soil DNA was stored at -80 °C before further analyses.

5.3.5 Density-gradient ultracentrifugation of soil DNA.

To reveal the effect of ^{13}C-straw incubation on soil DNA density, we performed density-gradient ultracentrifugation (Tao, Feng et al. 2020). Briefly, we centrifuged 5.1 ml of a solution composed of 3.6 µg of soil DNA (the minimum of total DNA amount in all samples), 1.90 g ml-1 cesium chloride (MP Biomedicals, Santa Ana, CA, USA), and a gradient buffer (1 mM EDTA, 0.1 M KCl, and 0.1 M Tris-HCl), reaching a final density of 1.725 g ml-1. The solution was sealed in a polyallomer centrifuge tube (cat. No. 342412, Beckman Coulter, Brea, CA, USA) with a cordless tube topper, and centrifuged on a Vti 65.2 rotor of an Optima L-XP ultracentrifuge (Beckman Coulter, Brea, CA, USA) at 177,000 g and 20 °C for 48 hours. The solution from each centrifuged tube was then separated into twenty-four fractions (14 drops per fraction). The buoyant density of each fraction was determined by an AR200 digital refractometer (Reichert, Depew, NY, USA). DNA in each fraction was then precipitated with 20 µg of glycogen and two volumes of PEG solution (30% PEG 6000 and 1.6 M NaCl), washed with 70% ethanol, and resuspended in 35 µl of ultrapure water.

5.3.6 qPCR of 16S rRNA genes.

qPCR was used to determine the absolute abundances of 16S rRNA genes in each fraction. Universal primers 515F (5'-GTGCCAGCMGCCGCGGTAA-3') and 806R (5'-GGACTACHVGGGTWTCTAAT-3') were used to target the V4 region of 16S rRNA genes. qPCR was performed in triplicate 20-µl reactions containing 10 µl of SsoAdvanced Universal SYBR Green Supermix (Bio-Rad, Hercules, CA, USA), 350

nM of each primer and 1 µl of template, using a thermocycler program of 35 cycles of 95 °C for 20 sec., 53 °C for 25 sec. and 72 °C for 30 sec. on an IQ5 Multicolor Real-time PCR Detection System (Bio-Rad, Hercules, CA, USA). Gene abundances (copy numbers) were determined by a standard curve generated with the 16S rRNA gene segment on the TA cloning vector within E. coli JM109 cells (Promega, Madison, WI, USA).

5.3.7 Amplicon sequencing of 16S rRNA genes.

A two-step PCR was performed to generate amplicon libraries of 16S rRNA genes (Tao, Feng et al. 2020). Briefly, the first step of the V4 region of 16S rRNA genes was amplified by the universal primers 515F and 806R in triplicate 25-µl reactions containing 2.5 µl of 10× AccuPrime PCR buffer containing dNTPs (Invitrogen, Grand Island, NY, USA), 0.2 µl of AccuPrime High-Fidelity Taq Polymerase, 1 µl of 10 µM forward and reverse primers, and 10 ng of template DNA. The thermocycler program was as follows: 94 °C for 1 min., 10 cycles of 94 °C for 20 sec., 53 °C for 25 sec. and 68 °C for 45 sec., followed by a final extension at 68 °C for 10 min. Bead purification was performed to retrieve amplicons generated by the first step, using AMPure XP magnetic particles (Agencourt Bioscience Corp., Beverly, MA, USA) with a 1:1 volume to the reactions. The second step of PCR also used triplicate 25-µl reactions comprised of 2.5 µl of 10× AccuPrime PCR buffer containing dNTPs, 0.2 µl of AccuPrime High-Fidelity Taq Polymerase, 1 µl of 10 µM 515F and 806R primer combined with the Illumina adaptor sequence (a pad and a linker of two bases, and a unique barcode sequence on the reverse primer), and 15 µl of the purified PCR product of the first step.

The thermal cycling condition was the same as the first step except for a cycle number of 20. Triplicate PCR products from the second step were combined, examined for DNA band of 16S rRNA genes by agarose gel electrophoresis, and quantified by PicoGreen.

PCR products from all fractions were pooled at equal molarity and sequenced in the same MiSeq run (Zhang, Gao et al. 2017). First, raw sequence reads underwent PhiX removal, followed by assignment to corresponding samples according to barcodes with 0 mismatches, and trimming of primers using a pipeline built on the Galaxy platform (http://zhoulab5.rccc.ou.edu:8080/). Next, high-resolution amplicon sequence variants (ASVs) with filtered sequencing errors were identified from the reads using the DADA2 procedure (Callahan, McMurdie et al. 2016) with the "dada2 (v.1.12)" package on R software (v.3.5.2). The ASV table was randomly resampled based on the abundance of 16S rRNA genes normalized by microbial biomass, using qPCR as previously described (Tao, Feng et al. 2020). Lastly, a representative sequence of each ASV was annotated through SILVA ribosomal RNA gene database (v. 132) with a confidence score of 50% (Quast, Pruesse et al. 2012). The 16S rRNA gene copy numbers were annotated through the RDP classifier (Wang, Garrity et al. 2007).

5.3.8 Identification of active degraders of straw.

Active degraders of straw were identified by qSIP, a cutting-edge, quantitative stable-isotope probing technology (Hungate, Mau et al. 2015) with minor modifications, which can fully identify and determine active degraders and their biomass compared to the traditional SIP. Briefly, the weighted density based on the absolute abundance of each

ASV, calculated by combining 16S rRNA genes sequencing and qPCR data, was determined for soil samples with ^{13}C- or ^{12}C-straw. The density shift (difference of density) of the ASV between ^{12}C-straw samples and ^{13}C-straw samples was calculated for all four biological replicates, and a 90% confidence interval (CI) was calculated for the density shift using the bootstrap method with the "boot (v.1.3-22)" package in R software. The ASV was considered as active degrader if the lower bound of the CI was above zero.

5.3.9 Determination of functional potentials by GeoChip microarray.

The functional potentials of active communities were determined by GeoChip 5.0S (Tao, Feng et al. 2020). Briefly, four fractions of each ^{13}C-straw sample were selected and regarded as representative for active communities if 16S rRNA genes of the corresponding ^{12}C-straw samples at the same density fraction were close to zero (Figure S5.1). Approximately 50 ng of DNA separated from ^{13}C-fractions in warmed or control samples were amplified using a Templiphi kit (GE Healthcare, Little Chalfont, UK). Then, 2 µg of amplified DNA was labeled with a fluorescent dye (Cy-3) dUTP using random primers and Klenow fragment of DNA polymerase I at 37 °C for 6 hrs, followed by heating at 95 °C for 3 min. Labeled DNA for each sample was purified with QIAquick PCR purification reagents (Qiagen Inc., Hilden, Germany) and SpinSmart columns (Thomas Scientific Inc., Swedesboro, NJ, USA), dried in a SpeedVac at 45 °C for 45 min., and resuspended in 43.1 µl of hybridization buffer containing 27.5 µl of 2× HI-RPM hybridization buffer, 5.5 µl of 10× CGH blocking agent, 2.4 µl of cot-1 DNA, 2.2 µl of universal standard and 5.5 µl of formamide. DNA

was hybridized with GeoChip 5.0S (60 K) in an SL incubator (Shel Lab, Cornelius, OR, USA) at 67 °C plus 10% formamide and 20 rpm for 24 hours. GeoChip arrays were washed and scanned by an MS 200 Microarray Scanner (Roche, Basel, Switzerland) at 532 nm and 635 nm. Raw signal intensities were processed by an online pipeline (http://ieg.ou.edu/microarray/) as previously described (Yang, Wu et al. 2013). The response ratio of signal intensities to warming was calculated as ln(Iwarming/Icontrol), in which Iwarming is the signal intensity of C-degrading genes in warmed samples and Icontrol is the signal intensity of C-degrading genes in control samples.

5.3.10 Determination of carbohydrates utilization capacity by Biolog EcoPlates.

Biolog EcoPlates (Biolog Inc., Hayward, CA, USA) containing 31 different labile C sources and one control without C source were used to assess carbohydrate utilization capacity of soil microbial communities before the SIP incubation. For each soil sample, 0.5 g of soil was mixed with 45 ml of 0.85% NaCl solution, shaken for 20 min at 180 rpm, and settled at 4 °C for 30 min. Subsequently, 1.5 ml of supernatant was mixed with 13.5 ml of distilled water and added onto Biolog EcoPlates with 100 µl of supernatant per well. The Biolog EcoPlates were incubated for 4.5 days using a Biolog Omnilog PM incubator (Torcon Instruments Inc., Torrance, CA, USA) at 25 °C. Color changes of the wells were transmitted to absorbance-time curves. The area under the curves was calculated to assess the utilization capacities of various C sources (Guckert, Carr et al. 1996).

5.3.11 Statistical and phylogenetic analyses.

Most statistical analyses were performed in R software (version 3.5.2). The difference among 16S rRNA gene abundances was determined by the one-way ANOVA with a permutation test (Perm-ANOVA) using the "lmPerm" package. The linear mixed-effects model (LMM) in the "lme4" package was used to determine warming effects on the relative abundance of active ASVs during the Year 2010–2016. Based on the LMM, the means of relative abundances for the active ASVs were least-squares means (estimated marginal means) produced by the "emmeans" function in the "emmeans" package (Guo, Feng et al. 2018). The difference between respiration and the priming effect was determined by repeated-measure ANOVA in the "vegan (v.2.3-2)" package. Linear models were used to detect correlations among microbial communities and C fluxes in the "stats (v.3.5.2)" package, which was subsequently tested for significance by permutation tests in the "lmPerm (v.2.1.0)" package. Mean values and standard errors of the mean are calculated. Unless otherwise stated, values of $p \leqslant 0.050$ were considered to be significant.

The maximum likelihood phylogenetic tree was constructed with the representative sequence for each active ASV. Cultured species of >99.6% 16S rRNA gene nucleotide identity with the 6 top abundant active ASVs was obtained from BLASTn on NCBI (blast.ncbi.nlm.nih.gov/BlastAlign.cgi) and anchored into the tree as reference species. MEGA 6.05 (Hall 2013) was used to construct the phylogenetic tree with MUSCLE alignment, maximum likelihood method, and a bootstrap of 1,000 times. Visualization of the tree was generated by iTOL (itol.embl.de/) (Life 2011). Beta nearest taxon index (βNTI) was calculated by an online pipeline (ieg3.rccc.ou.edu:8080/) as the

phylogenetic β-diversity among samples (Stegen, Lin et al. 2012). Phylogenetic diversity was calculated by the "AllenH" function in the "entropart" package.

The C assimilation rate for each active degrader was calculated as described previously (Hungate, Mau et al. 2015). The variance partitioning analysis was performed to assess the relative proportion of variance in the C assimilation rate of the active bacterial community, using the "lme" function in the "nlme" package followed by the "varcomp" function in the "ape" package (Morrissey, Mau et al. 2019). To calculate the relative importance of environmental factors in determining active bacterial phylogenetic diversity, we performed the model selection analysis using the "glmulti" package. The importance score of a certain factor was calculated as the sum of the Akaike weights for all models containing this factor, with a threshold value of 0.8 as described previously (Terrer, Vicca et al. 2016).

To establish structural relationships among environmental factors and bacterial community structure, structural equation modeling (SEM) analysis was performed with Amos 24.0 software package (Small Waters Corp., Chicago, IL, USA). A chi-square test of model fit was adopted to determine whether the proposed model was supported by the data. Three other widely used indices of model fit were also used to evaluate the model fit, including comparative fit index (CFI), Tucker Lewis index (TLI) and root mean square error of approximation (RMSEA), wherein the good models have a CFI and TLI value of >0.95 and an RMSEA value of <0.05 (Byrne 2010).

5.4 Results

5.4.1 Warming stimulated soil respiration and induced the priming effect during SIP.

To assess the C-degrading activities of soil microbial communities in warmed and un-warmed soil samples, we calculated soil respiration (microbial respiration) and the priming effect on SOC during the 7-day SIP incubation. For the overall soil respiration during 7-day incubation, soil respiration of warmed samples was higher (p = 0.027) than control samples by 13.3% (Figure 5.1), which was consistent with a previous study that more CO_2 was released from warmed soil samples during the 9-week incubation. As we expected, the overall priming effect on SOC of warmed samples was also higher (p = 0.040) than control samples by 14.4%, indicating that warming could also enhance the priming effect in temperate grassland (Figure 5.1). For the individual respiration and priming effect, there were significant differences between warning and control in the last four days instead of the first three days, suggesting delayed stimulation of SOC degradation by warming (p < 0.050).

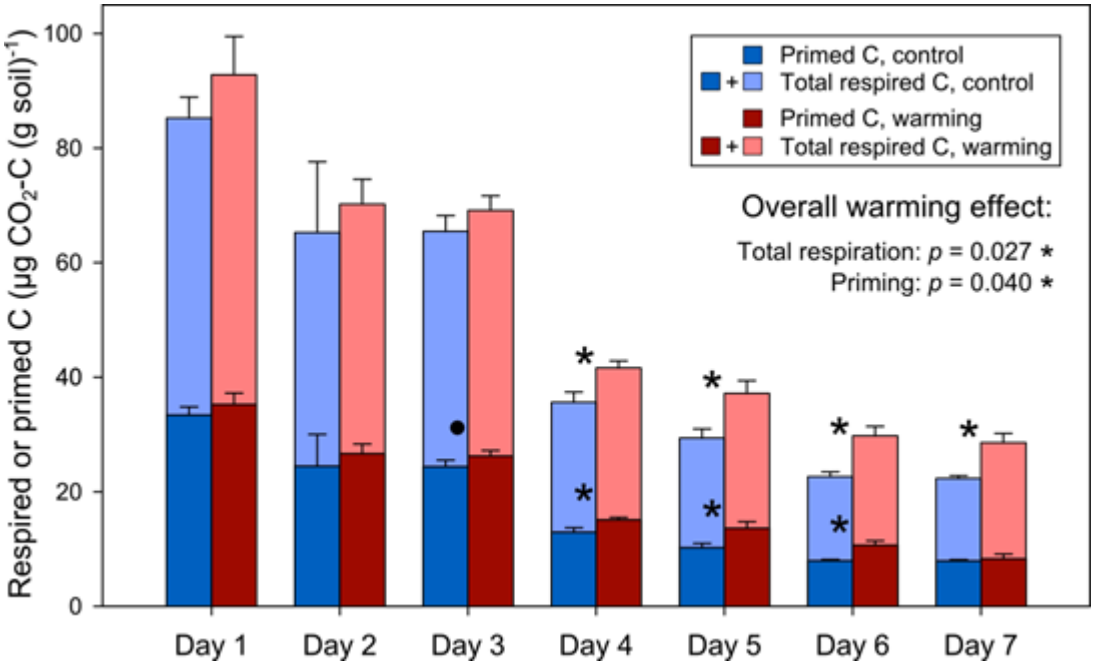

Figure 5. 1 Daily soil respiration and priming effect during the 7-day incubation with plant litter. The columns represent the average ± standard error of 4 replicates of warmed or control samples. Significances are indicated using as $0.050 < p \leq 0.100$ and * as $0.010 < p \leq 0.050$ for total respiration or priming effect, as determined by permutation ANOVA (individual days) or repeated measures ANOVA (overall warming effect).

5.4.2 Warming enlarged active bacterial abundance and restructured community composition.

Compared to soil incubated with ^{12}C-straw for seven days, a considerable amount of 16S rRNA gene abundance in soil incubated with ^{13}C-straw was shifted to heavier densities (Figure S5.1), indicating an efficient SIP labeling. qPCR experiments of bacterial 16S rRNA genes showed that warming increased the abundance of active community by 58.2% (p = 0.076, Figure 5.2a) and total bacterial community by 44.9% (p = 0.091, Figure 5.2b).

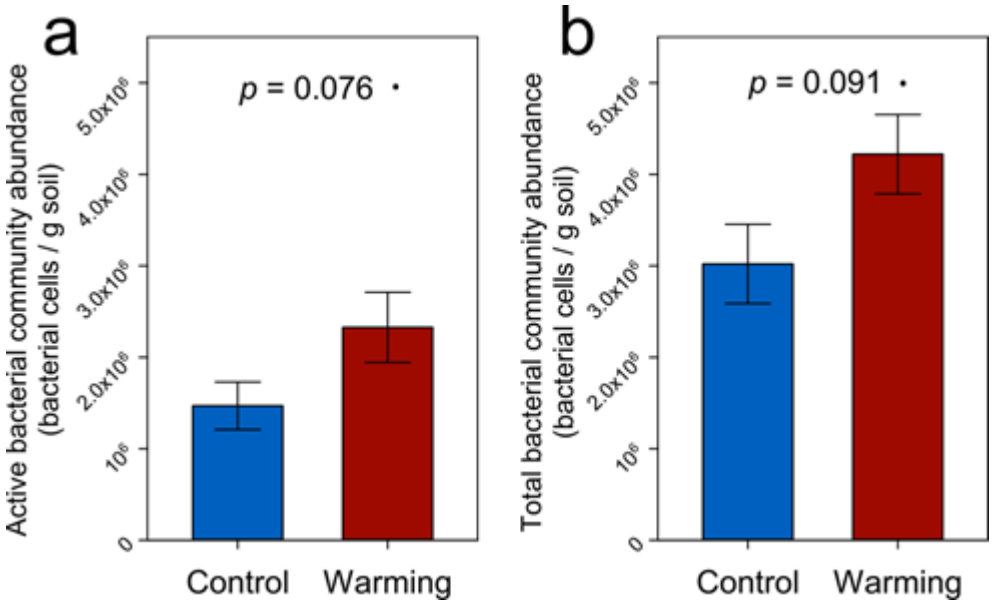

Figure 5. 2 Absolute cell abundances of (a) active bacterial community and (b) total bacterial community of in situ warmed or control soils, after the 7-day incubation with plant litter, as normalized by 16S rRNA gene copy numbers. The columns represent the average ± standard error of 4 replicates of warmed or control samples. Significances are indicated using · as $0.050 < p \leq 0.100$ as determined by permutation ANOVA

A total of 7,945 ASVs of 16S rRNA gene amplicons were identified across all samples and fractions. Among them, only 147 ASVs were identified as active C degraders (Figure 5.3). Those ASVs accounted for 63.2% of total bacterial abundance, indicating that active C degraders are extraordinarily abundant in soil samples. More than half of the active ASVs belonged to the phylum *Proteobacteria*, followed by 31 *Actinobacteria*, 11 *Bacteroidetes*, and 10 *Firmicutes* ASVs. Thirty-eight ASVs were active in both warmed and control samples, accounting for 33.9% of total bacterial abundance. In contrast, 53 ASVs were only active in control samples, most of which belonged to *Proteobacteria*, especially *β*- and *r-Proteobacteria*. Fifty-six ASVs were only active in warmed samples, which included 20 ASVs belonging to *α*-*Proteobacteria*, 9 ASVs belonging to *Bacillales* of *Firmicutes*, 9 ASVs belonging to *Actinobacteria*, 8 ASVs belonging to *Bacteroidetes*, 4 ASVs belonging to γ-*Proteobacteria*, 4 ASVs belonging to unclassified bacteria, 1 ASV belonging to β-*Proteobacteria*, and 1 ASV belonging to *Thaumarchaeota*. Based on variance partitioning of the nested linear model analysis (Hungate, Mau et al. 2015), warming explains 74.2% of the variation in C assimilation rate of the active bacterial community (Table S5.2), indicating that the C assimilation rate was highly responsive to warming. In contrast, microbial taxonomy at the order level explains the rest 25.8%. Since it was exclusively attributed to warming (43.0%, Table S5.2), the variation in C assimilation rate in warming was determined by one or more bacterial orders such as *Bacillales*. More interestingly, the abundance of two active *Bacillales* ASVs (ASV 58 and ASV 70), both of which were only detected in warmed samples, positively and strongly

correlated with soil respiration (R^2 = 0.467, p = 0.016 for ASV 58, and R^2 = 0.586, p = 0.037 for ASV 70, Figure S5.2a).

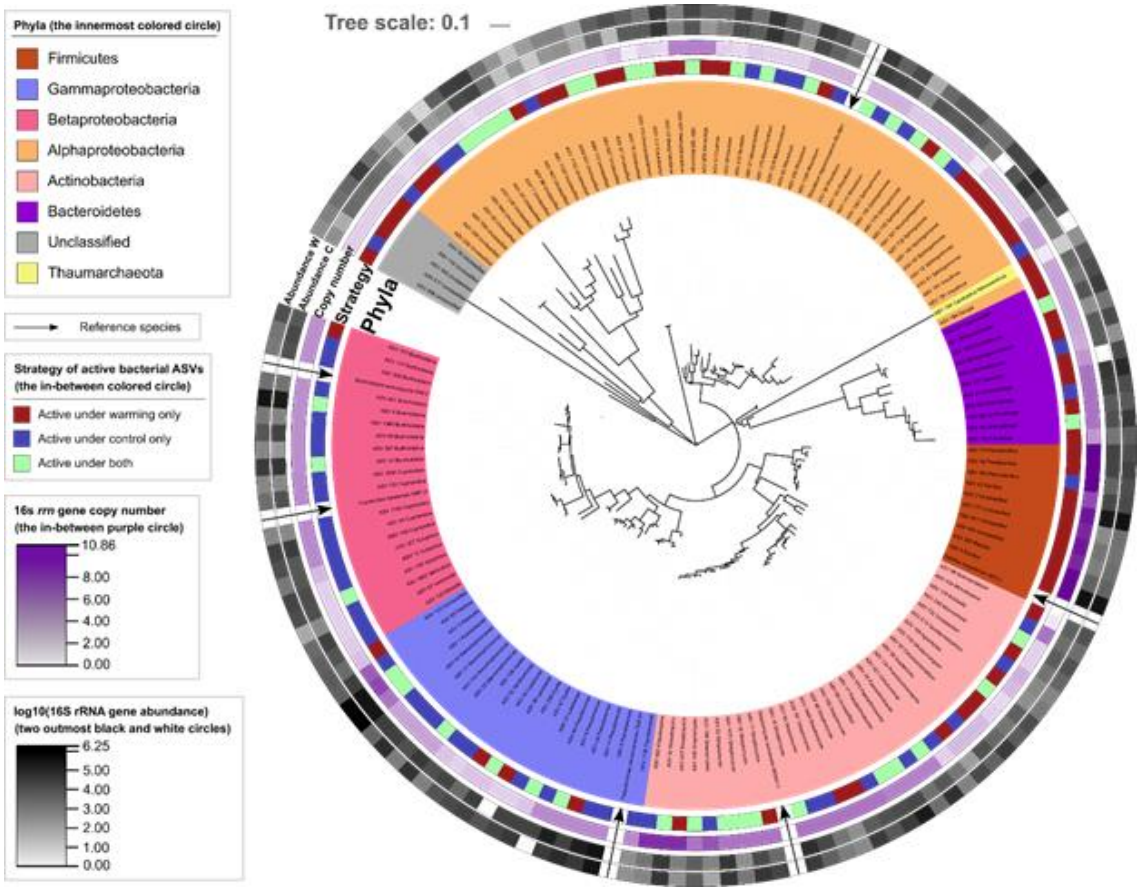

Figure 5. 3 The maximum-likelihood phylogenetic tree of active bacterial ASVs (degraders). Abbreviations: W, warmed samples; C, control samples.

Both phylogenetic α-diversity and weighted βNTI (i.e., the phylogenetic β-diversity) of the active bacterial community was increased by warming (p = 0.049, Figure 4a; p = 0.034, Figure 5.4c), revealing a more diversified pattern. In sharp contrast, warming changed neither the phylogenetic α-diversity nor the weighted βNTI of the entire bacterial communities (both active and inactive bacterial communities) (p = 0.606, Figure 5.4b; p = 0.273, Figure 5.4d). Based on model selection analysis, warming treatment was the most important factor in regulating the phylogenetic diversity of active bacterial communities between warmed and control samples (Figure 5.4e).

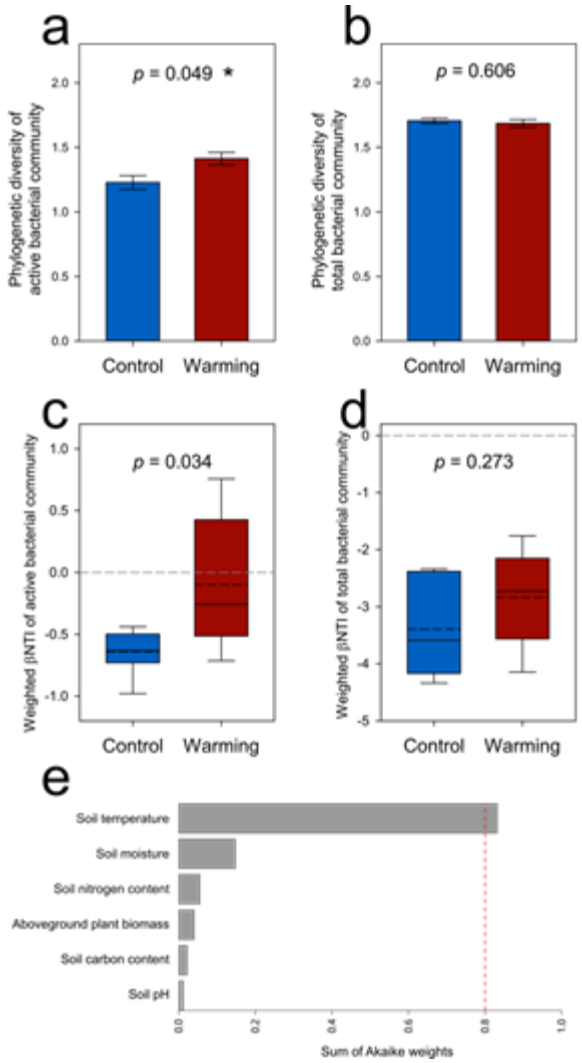

Figure 5. 4 The phylogenetic diversity of (a) active bacterial communities and (b) total bacterial communities, the weighted βNTI (phylogenetic β-diversity) among (c) active bacterial communities and (d) total bacterial communities, and (e) the relative importance of edaphic factors regulating phylogenetic diversity as determined by model selection analysis. The columns in (a) and (b) represent the average ± standard error of 4 replicates of warmed or control samples. In (c) and (d), the average values are shown as black dashed lines in the boxes, and each box was plotted from 6 pairwise differences among the 4 warmed samples or control samples. Significance is indicated using * as $0.010 < p \leq 0.050$ as determined permutation ANOVA.

To verify whether the 56 ASVs only active in warmed samples were induced by warming during 2010–2016, we analyzed their relative abundances in soil samples collected on a yearly basis. A total of 48 ASVs, mostly belonging to *α-Proteobacteria*, *Bacillales*, *Actinobacteria*, and *Bacteroidetes*, were found in the *in-situ* soil samples

from 2010 to 2016. The mean relative abundances of those active ASVs were increased by 27%–205% under warming (linear mixed model analysis, Figure 5.5), which was consistent with our SIP experimental results showing that they were major responders of warming.

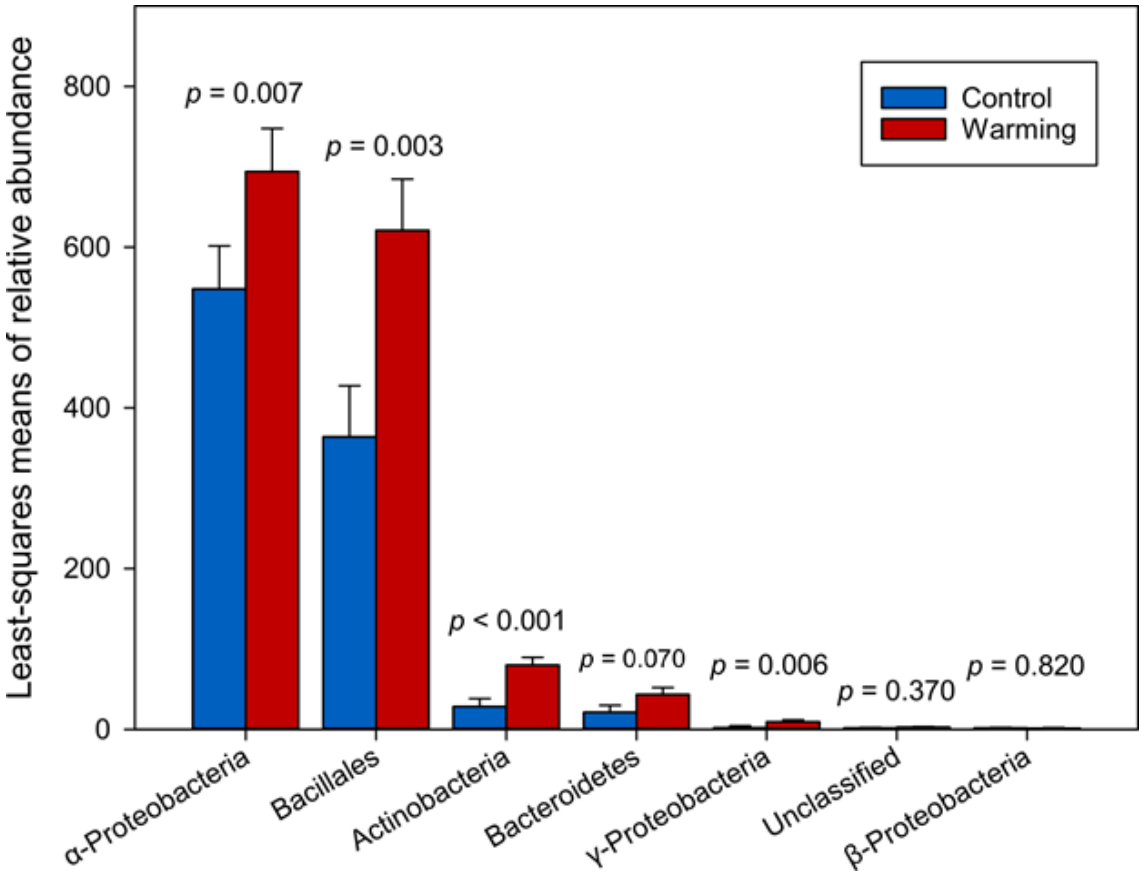

Figure 5. 5 Yearly means of relative abundance of *α-Proteobacteria*, *Bacillales*, *Actinobacteria*, *Bacteroidetes*, *γ-Proteobacteria*, unclassified bacteria, and *β-Proteobacteria*. The significance is determined by the linear mixed-effects model (LMM). The mean values are least-squares means produced by LMM. Each column represents average ± standard error of $n = 4$ field biological replicates of *in situ* warming or control over seven yearly repeated measures during 2010–2016.

5.4.3 Warming enhanced C-degrading potentials and activities of active communities.

To examine whether higher microbial respiration in warmed samples arises from changes in functional genes associated with C degradation, especially for hemicellulose, cellulose and lignin, relative abundances of related functional genes in active

communities were quantified by GeoChip 5.0. Among a total of 45 C-degrading genes detected by GeoChip, 15 genes associated with both chemically labile and recalcitrant C degradation were increased by warming in relative abundance (Figure 5.6a). As cellulose, hemicellulose and lignin are the most abundant components of oat straw(Mood, Golfeshan et al. 2013), all of the four functional genes encoding xylanase, mannanase, xylose isomerase, or an L-arabinose operon comprised of L-arabinose isomerase, ribulokinase, and L-ribulose-5-phosphate 4-epimerase were increased in relative abundance by warming (Figure 5.6a). Notably, the total abundance of hemicellulose-degrading genes of active communities strongly correlated with respiration derived from ^{13}C-straw ($R^2 = 0.523$, $p = 0.031$, Figure S5.2b). Similarly, the gene *axe* encoding acetyl esterase for cellulose degradation and the gene encoding phenol oxidase for lignin degradation was also increased in relative abundance by warming. Therefore, the higher C-degrading potentials and activities were consistent with higher total CO_2 flux ratio and greater priming effect under warming (Figure 5.1). Among the C-degradation genes belonging to active *Firmicutes* species, *xylA* encoding xylose isomerase, *pme* encoding pectin methylesterase, and *vdh* encoding vanillin dehydrogenase were also increased in relative abundance by warming (Figure S5.3). Biolog EcoPlates were used to assess the capacity of carbohydrate utilization of microbial communities in soil samples before SIP incubation. Consistent with the increase of the *xylA* gene abundance ($p < 0.001$, Figure 6a), the microbial utilization capacity of xylose, a major component of hemicellulose, showed a 77.4% increase by warming ($p = 0.039$, Figure S5.4). D,L-Glycerol-3-Phosphate, a precursor producing

triacylglycerol in oil-producing organisms, was also increased for 79.8% by warming (p = 0.011).

By validating correlations between abundance or functions of active microbial communities and soil temperature, structural equation modeling (SEM) was applied to further explain possible mechanisms for increased microbial respiration in warmed soil samples (Figure 5.6b). The overall interactive model was well fitted (χ^2 = 0.169, df = 1, p = 0.681). First, soil temperature positively correlated with the absolute abundance of the active bacterial community ($p < 0.001$); second, the active bacterial abundance positively correlated with the abundance of hemicellulose-degrading genes (p = 0.009); finally, the abundance of hemicellulose-degrading genes positively correlated with soil respiration (p = 0.065). All these positive correlations unveiled a possible mechanism between warming and microbial SOC degradation that (i) the increased soil temperature could increase active bacterial biomass; (ii) the increased abundances of C-degrading genes, such as hemicellulose degradation genes, could improve the degradation of SOC and finally cause more CO_2 release from the soil.

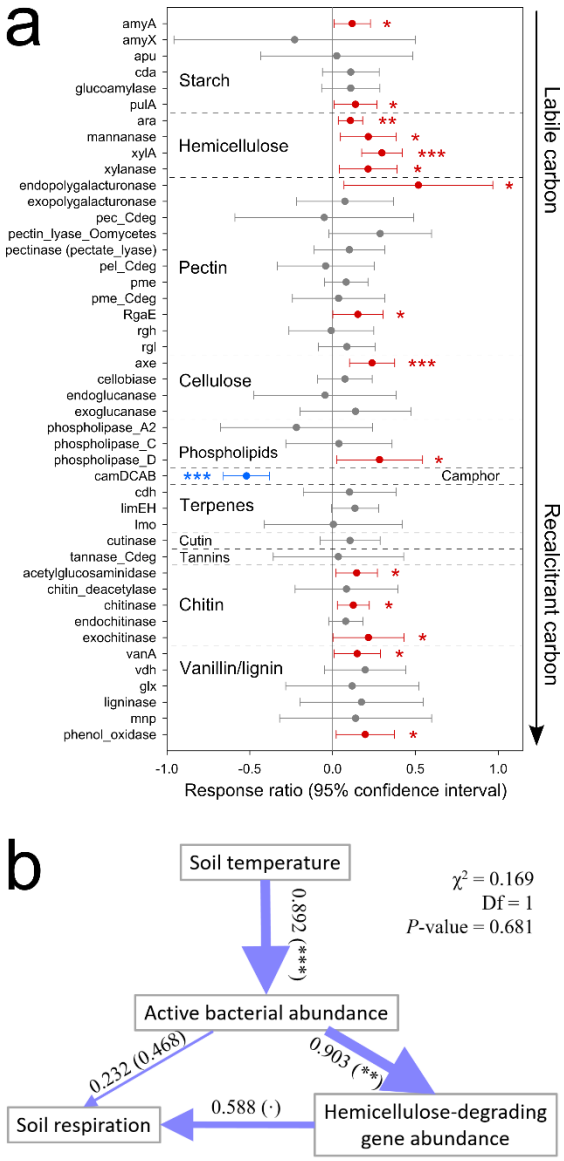

Figure 5. 6 Response ratios of GeoChip signal intensities of active community C-degrading genes to the in situ warming treatment (a) and structural equation modeling (SEM) of soil temperature, active bacterial biomass, C-degrading genes and soil respiration (b). In (a), the symbols represent the average ± standard error of 4 replicates of warmed or control samples; red symbols represent significantly positive response ratios, and blue symbols represent significantly negative response ratios; significances are indicated using * as $0.010 < p \leq 0.050$, ** as $0.001 < p \leq 0.010$, and *** as $p \leq 0.001$ as determined by confidence intervals. In (b), blue arrows indicate positive relationships; numbers adjacent to arrows are standardized path coefficients (co-variation coefficients) which are proportional to the thickness of arrow lines, with P-values in the brackets; significances are indicated using · as $0.050 < p \leq 0.100$, * as $0.010 < p \leq 0.050$, ** as $0.001 < p \leq 0.010$, and *** as $p \leq 0.001$.

5.5 Discussion

Soil respiration is expected to increase with warming, following an empirical exponential function between soil respiration and temperature (Brown, Gillooly et al. 2004). However, *in situ* measurements of soil respiration under the experimental warming treatment has been reported to increase (Rustad, Campbell et al. 2001, Melillo, Steudler et al. 2002, Emmett, Beier et al. 2004, Zhou, Wan et al. 2007), decrease (Wan, Xia et al. 2009), or remain unchanged (Wan, Norby et al. 2007), resulting in great uncertainty on whether the terrestrial-atmospheric C exchange was positive or negative feedback to warming. Our SIP-incubation results showed that both soil respiration (p = 0.027) and the priming effect on SOC (p = 0.040) were significantly higher in warmed samples compared to control samples (Figure 5.1), leading to a severe concern that soil C sink in temperate grasslands could be worsened by global warming. Our findings support the global analysis showing that topsoil SOC is vulnerable to microbial degradation under the warmed condition (Crowther, Todd-Brown et al. 2016). Higher soil C emission stimulated by warming (Figure 1), also observed elsewhere (Zhou, Xue et al. 2012, Carey, Tang et al. 2016, Cheng, Zhang et al. 2017), can be well explained by concomitant changes of active communities with enlarged bacterial abundance and restructured community composition (Figure 5.2 & 5.3).

qSIP is a cutting-edge technique for identifying microorganisms linking to certain functions (Hungate, Mau et al. 2015, Morrissey, Mau et al. 2019). Here, we identified a total of 147 active ASVs (Figure 5.3), including well-characterized C degraders such as *Burkholderia*, *Sphingomonas*, and *Bacillus* (Bugg, Ahmad et al. 2011, Amore, Pepe et al. 2013, Tao, Feng et al. 2020). We also detected 29 unclassified active genera,

suggesting that our understanding of soil C-degraders is still limited in temperate grasslands. Among 147 active ASVs, 53 of them were detected only in control samples, while 56 of them were detected only in warmed samples. Therefore, there was a radical compositional change, which could give rise to long-lasting effects on soil C degradation. Similarly, in tundra soils, experimental warming was shown to change active ligninolytic communities from *β-Proteobacteria* to *α-Proteobacteria* (Tao, Feng et al. 2020). More interestingly, we found that warming increased the phylogenetic diversity of the active bacterial communities (Figure 5.4a), which was also consistent with the previous finding that experimental warming in tundra soils could increase the phylogenetic diversity of active bacterial communities (Tao, Feng et al. 2020). A recent study at our study site has revealed the increasingly divergent succession of microbial communities over time by warming (Guo, Feng et al. 2018), which was consistent with our observation that warming increased the phylogenetic β-diversity among active bacterial communities (Figure 5.4c). Meanwhile, model selection analysis verified that soil temperature was the most important factor in determining the phylogenetic diversity of the active community (Figure 5.4e). All these results collectively suggest that the more diverse succession of active bacterial communities by warming might be a generalizable phenomenon.

Bacterial taxonomy at the order level played an important role in explaining the variation of C assimilation rate under warming (Table S5.2), which was consistent with our findings that order of *Bacillales* was highly responsive to warming (Figure 5.3 & Figure 5.5) and positively correlated with soil respiration (Figure S5.2a). On the other hand, *Bacillales* is typical of high 16S rRNA gene copy numbers (Figure 5.3) and

responds rapidly to nutrient input (Klappenbach, Dunbar et al. 2000, Feng, Wang et al. 2020). Therefore, stimulation of *Bacillales* might reflect better nutrient conditions induced by the warming treatment. In fact, *Bacillales* has been well documented as effective degraders of both hemicellulose and cellulose (Amore, Pepe et al. 2013). Furthermore, strain IMT21 of *Bacillus megaterium*, which was of 100% nucleotide identity with *Bacillales* ASV 5 in the active community linkages (Figure 5.2), is efficient in degrading the aromatic compound of dichloroaniline (Yao, Khan et al. 2011). Strain TN41 and TN42 of *Bacillus* sp., possessing more than 99.2% nucleotide identities with ASV 5, are also capable of degrading many aromatic compounds, including phenol, toluene, biphenyl, and naphthalene (Đokić, Narančić et al. 2011). Given that *Bacillales* are also abundant at our *in-situ* warming site (Ning, Yuan et al. 2020), all these findings suggest that *Bacillales* appear to be major C-degrading responders of warming in the temperate grassland.

Warming substantially increased C-degrading potentials of active microbial communities (Figure 5.6a). Compared to other C-degrading genes, all detected functional genes associated with hemicellulose degradation were increased in warmed soil samples. An explanation is that hemicellulose is an abundant component of the oat straw only second to cellulose (~27% w/w) but is more chemically labile than cellulose and lignin (Mood, Golfeshan et al. 2013), and our SIP experiment only lasted for seven days to minimize the possibility of substrate cross-feeding, and therefore higher lability of hemicellulose to degradation was likely a preferred substrate for active degraders compared to cellulose and lignin. Additionally, both the SEM analysis (Figure 5.6b) and Pearson correlation analysis (Figure S5.2b) found the strong positive correlation

between hemicellulose degradation genes and soil respiration, which indicated a contribution of hemicellulose to respiration and well supported faster utilization of xylose in warmed samples (Figure S5.4). We found that warming also promoted the functional potentials to degrading cellulose, chitin, vanillin, and lignin (Figure 5.6a), revealing a possibility that more recalcitrant C could be available to microbial degradation with longer incubation.

In summary, this study has found increased soil respiration and priming effect in warmed soil samples, which were positively linked to higher abundances of active bacterial communities and C-degrading potentials. Our finding of a stronger priming effect by warming is alarming, which reveals a hitherto overlooked mechanism in the temperate grassland that accelerates climate warming. Our results also provide identities of active degraders and the changes of their composition and abundances responding to warming, which filled the gaps for identifications of active communities in temperate grassland and provide new insights into global climate and ecosystem models for accurate projection of feedbacks between climate change and C cycle. Additionally, this study is located in the place well known as the Dust Bowl, which has taught modern society a painful lesson that human welfare depends on looking after the soil (Egan 2006). In that context, identification of active bacterial taxa related to C degradation may provide key targets to help reduce C load in the atmosphere, as below-ground communities are indispensable for achieving key ecosystem functions, such as soil C sequestration.

Chapter 6 **Summary and Outlook**

Although the lignocellulosic biofuels are better than the first-generation biofuels and friendly to the environment, the low efficiency and high cost for lignocellulosic biofuels production make them very difficult to compete with the fossil fuels. By using the cellulolytic model organism *C. cellulolyticum*, this study contributes to alleviating these challenges in two ways: (i), by characterizing the function of key cellulosomal component which may allow the cellulosomes produced from *C. cellulolyticum* to be used as commercial hydrolysis enzymes in the future; (ii), by engineering the *C. cellulolyticum* to improve its lignocellulose bioconversion ability for future application in industrial community. Major achievements are summarized below.

First, this work characterized the function of the X2 module belonging to the cellulosome. In fact, the cellulosome is the most promising CBS-catalyst, which is better than the fungal cellulases (Liu, Li et al. 2020). However, only the cellulosome produced from *C.thermocellum* is applied for CBS-catalyst now as it is the best investigated cellulosome (Schwarz 2001, Liu, Li et al. 2020). The cellulosome from mesophilic cellulolytic bacteria is different from that in thermophilic bacteria. Therefore, understanding the function for each cellulosomal component in *C. cellulolyticum* is very important for the future application of mesophilic cellulosomes in commercial. We found that the inactivation of the X2 modules in *C. cellulolyticum* can indeed influence the cellulose utilization efficiency, and the *in vivo* function of the X2 module was associated with binding affinity between cellulosomes and cellulose. These findings can provide new perspectives into applying the mesophilic cellulosomes as the

CBS-catalyst in industries and modifying the commercial cellulosome in improving their hydrolysis efficiencies.

Second, the difference abundances of *cip-cel* transcripts in *C. cellulolyticum* have been found for many years (Maamar, Abdou et al. 2006, Xu, Huang et al. 2015). However, due to the limitation of traditional genome manipulation methods, it was very difficult to edit the responsive element of the *cip-cel* operon. Using the Cas9-based method, our one-step targeted promoter insertion in the *cip-cel* gene cluster was a useful strategy for improving cellulose utilization and end-product formation, which can be applied in other bacteria with similar cellulosome-producing or other biosynthetic systems, and provides new perspectives on how to reduce the cost of industrial fermentation processes and enhance the efficiency of cellulolytic bacteria on lignocellulose bioconversion.

Third, the accumulation of cellobiose in repressing the expressions of cellulase genes is a potential issue for the commercial application of *C. cellulolyticum* (Xu, Huang et al. 2013). Our strategy, integration of *CcBglA* gene from *C.cellulovorans* into the genome of *C. cellulolyticum* , demonstrated that integration of exogenous *CcBglA* could enhance the cellulose degradation efficiency in *C. cellulolyticum* . Additionally, our results found that the upstream of *ccel_2485* was a good potential integration site for future exogenous genes integration in *C. cellulolyticum* , which can avoid the use of plasmid for exogenous gene expression and could ensure the increase of ethanol production.

Finally, decomposition of lignocellulose by microorganisms not only plays an important role in industrial biofuels production, but also is an important carbon sink in natural soil, which is firmly correlated with carbon cycling between atmosphere and terrestrial.

Given that global climate and ecosystem models for accurate climate projections require explicit consideration of the physiology and composition of active microorganisms that convert soil organic carbon (primarily is lignocellulose) to atmospheric CO_2, we identified active taxa responsible for lignocellulose decomposition in temperate grassland and explained their decomposition mechanism responding to global warming. Additionally, we found that a stronger priming effect was enhanced by warming, which is alarming and reveals a hitherto overlooked mechanism in the temperate grassland that accelerates climate warming. Our results also provide the identities of active degraders and the changes of their composition and abundance responding to warming, which filled the gaps for identifications of active communities in temperate grassland and provide new insights into isolation of new strains for lignocellulosic biofuels production.

In conclusion, being a small part of the huge effort of improving efficiency for biofuels production, our study provided novel insights into engineering the CBP/CBS candidate *C. cellulolyticum* for improving its lignocellulose bioconversion ability. Aforementioned discoveries will direct microbial engineers to develop more feasible processing strategies for commercial application of *C. cellulolyticum* or other candidates. Additionally, the mechanism for active microbial decomposition of lignocellulose under warming will provide more accurate predictions of feedbacks between climate change and C cycling.

Listed below are the individual manuscripts, published or in preparation, in relation to this dissertation. Chapter 3 in this dissertation presented contents in the published

journal articles 1 below. The Publisher of the journal granted the author to re-use these published materials in this dissertation by copyright policy.

Peer reviewed journal papers published:

1. **Tao, Xuanyu**, Tao Xu, Megan L. Kempher, Jiantao Liu, and Jizhong Zhou. "Precise promoter integration improves cellulose bioconversion and thermotolerance in *Clostridium cellulolyticum*." *Metabolic Engineering* (2020).

2. **Tao, Xuanyu**, Jiajie Feng, Yunfeng Yang, Gangsheng Wang, Renmao Tian, Fenliang Fan, Daliang Ning et al. "Winter warming in Alaska accelerates lignin decomposition contributed by *Proteobacteria*." *Microbiome* 8, no. 1 (2020): 1-12.

3. Feng, Jiajie, Cong Wang, Jiesi Lei, Yunfeng Yang, Qingyun Yan, Xishu Zhou, **Xuanyu Tao** et al. "Warming-induced permafrost thaw exacerbates tundra soil carbon decomposition mediated by microbial community." *Microbiome* 8, no. 1 (2020): 1-12.

4. Guo, Xue, Jiajie Feng, Zhou Shi, Xishu Zhou, Mengting Yuan, **Xuanyu Tao**, Lauren Hale et al. "Climate warming leads to divergent succession of grassland microbial communities." *Nature Climate Change* 8, no. 9 (2018): 813-818.

Manuscripts in preparation or submitted:

1. **Tao, Xuanyu**, Tao Xu, Megan L. Kempher, Jiantao Liu, and Jizhong Zhou. "In vivo functional characterization of X2 modules in Clostridium cellulolyticum CipC scaffoldin." (In preparation)

2. **Tao, Xuanyu**, Tao Xu, Megan L. Kempher, Yongxing He, Nannan Zhang. "Non-canonical PTS systems with distinct regulatory processes for carbon metabolisms." (In preparation)

3. **Tao, Xuanyu**, Jiajie Feng, Yunfeng Yang, Gangsheng Wang, Siyang Jian, Xue Guo, Daliang Ning et al. " Long-term warming unlikely pacifies soil microbial respiration due to enlarged and activated carbon degrader." (Submitted)

Appendix A: Supplementary Figures

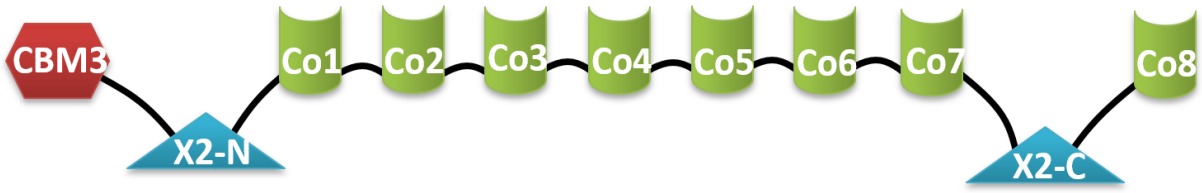

Figure S2.1 Scheme of the modular structure of the CipC scaffoldin in *C. cellulolyticum*. The CipC contains one carbohydrate binding module, eight cohesion modules and two X modules.

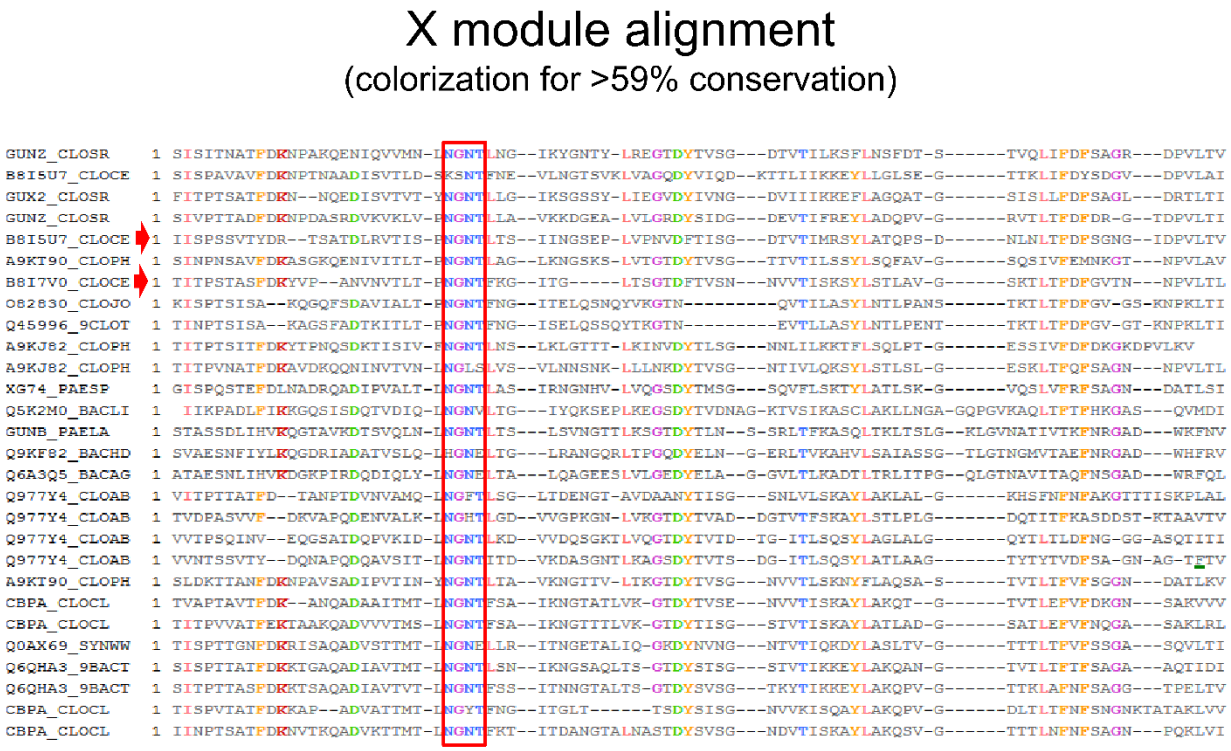

Figure S2.2 Alignment of the X2 modules among different species. The red box indicated the conserved motif of the X2 modules, and the red arrow indicated X2-N and X2-C modules in *C. cellulolyticum*.

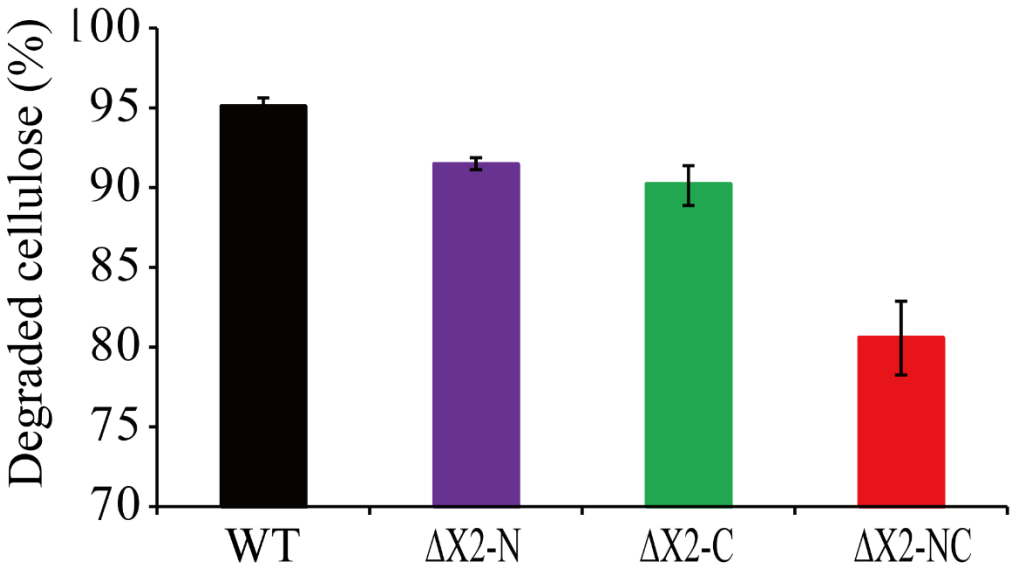

Figure S2.3 Degraded cellulose (%) for each strain at the final time point.

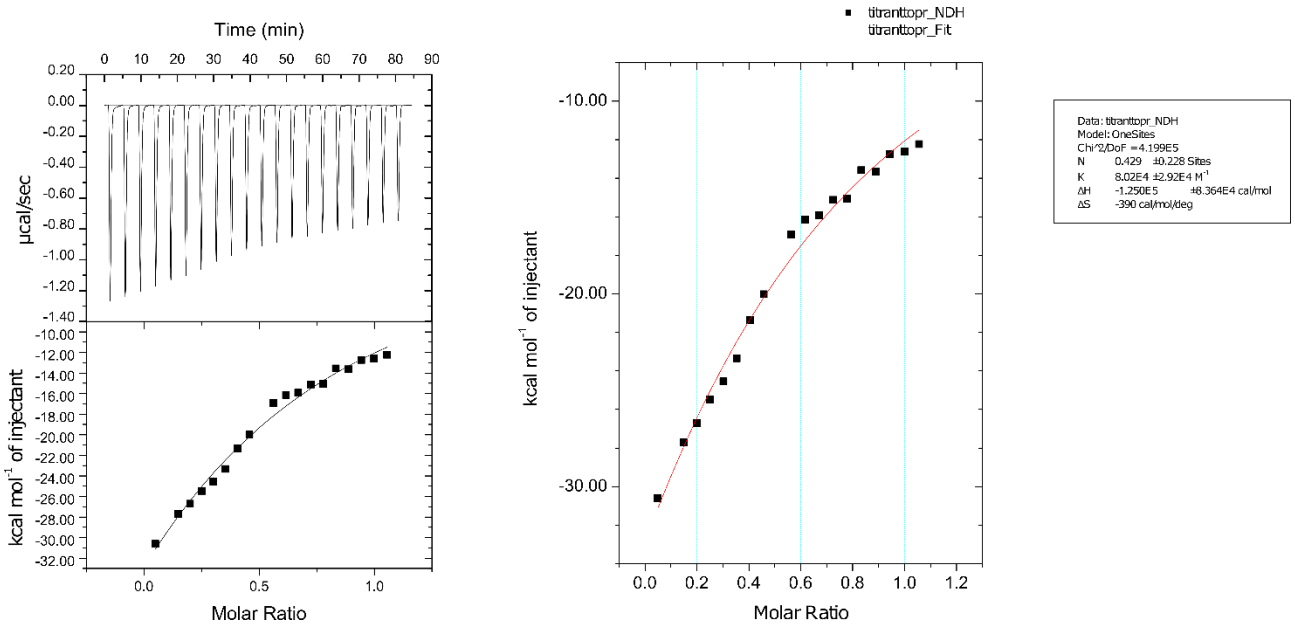

Figure S2.4 Isothermal titration calorimetry (ITC) data for binding interactions between CBM3a and the X2-C module. There was weakly binding between them, and the binding affinity constants is $8.02 \pm 2.92 \times 10^4$ M^{-1}.

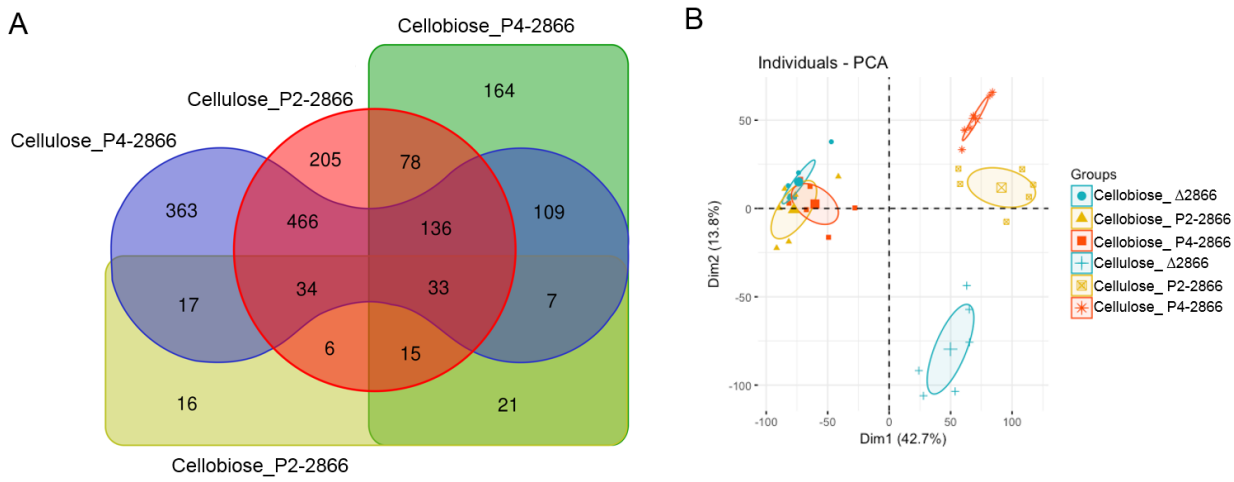

Figure S3.1 Venn diagram (A) and PCA analysis (B) of differentially expressed genes (DEGs) among P4-2866, P2-2866 and Δ2866 grown on 20 g/L cellulose or 5 g/L cellobiose.

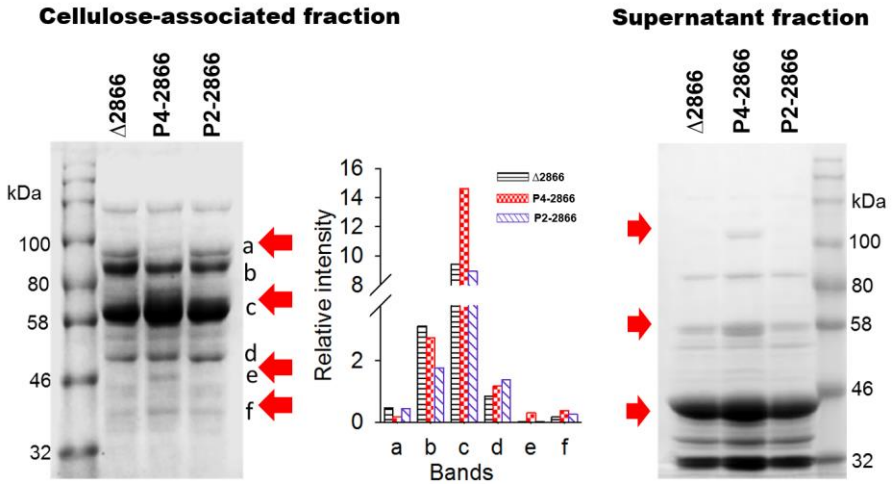

Figure S3.2 Cellulosomal composition of P2-2866, P4-2866 and Δ2866 analyzed by SDS-PAGE. The cellulosomal fraction was isolated from P2-2866, P4-2866 and Δ2866 at mid-exponential growth phase, respectively. Bands, a – f, represented different

components (cellulases) of the cellulosome. Red arrows indicated there are obvious changes of bands (a-f) between engineered strains and parent strain.

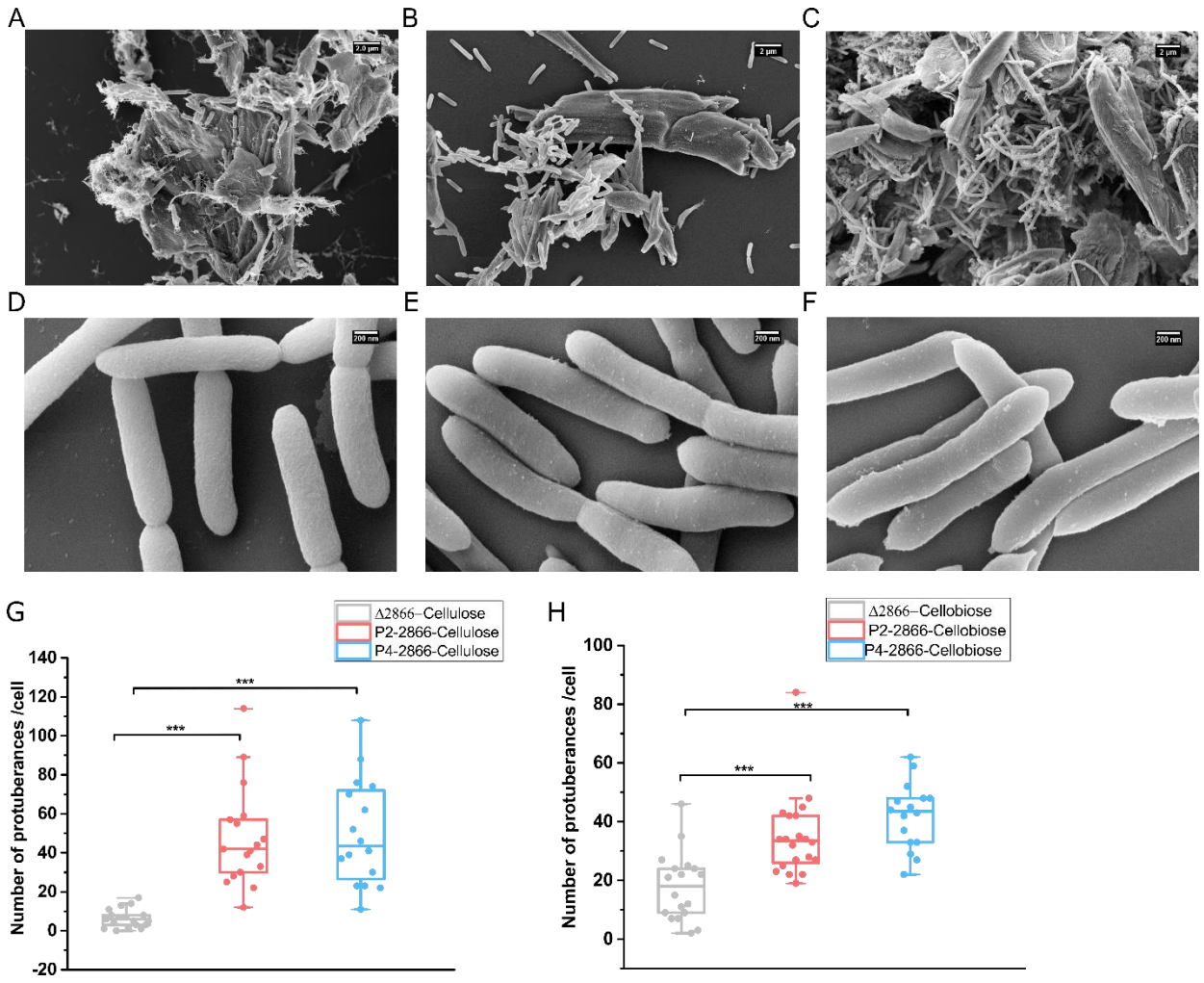

Figure S3.3 The morphology of engineered strains observed by SEM (A-F) and statistical analysis on the difference of the number of protuberances between engineered strains and parent strain. (A) Δ2866 grown on 20 g/L cellulose. (B) P2-2866 grown on 20 g/L cellulose. (C) P4-2866 grown on 20 g/L cellulose. (D) Δ2866 grown on 5 g/L cellobiose. (E) P2-2866 grown on 5 g/L cellobiose. (F) P4-2866 grown on 5 g/L cellobiose. (G) Box chart for number of protuberances of each cell for strains (Δ2866, P2-2866 and P4-2866) grown on 20 g/L cellulose. Significance is indicated by ***, $P<0.001$ as determined by PERM ANOVA (The number of cells for each strain is ~ 17). (H) Box chart for number of protuberances of each cell for strains (Δ2866, P2-2866

and P4-2866) grown on 5 g/L cellobiose. Significance is indicated by ***, P<0.001 as determined by PERM ANOVA (The number of cells for each strain is ~ 17).

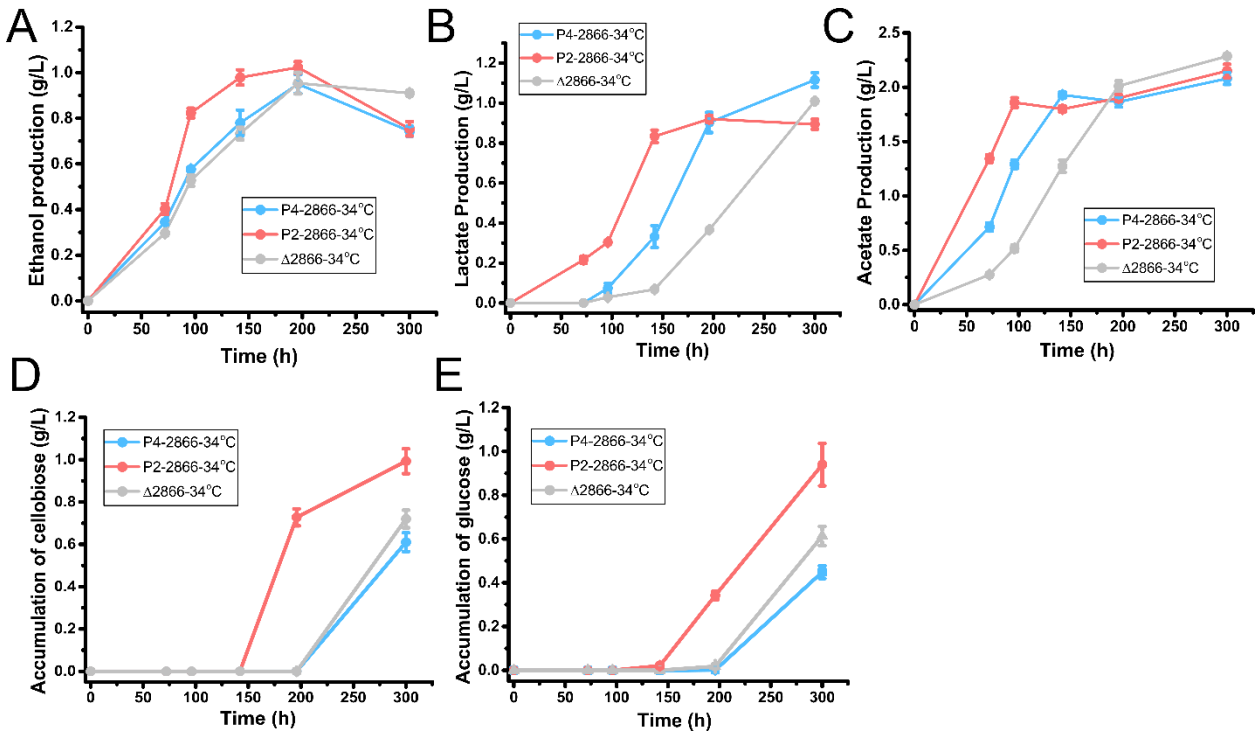

Figure S3.4. Fermentation products and accumulation of soluble sugars profiles for P4-2866, P2-2866, and Δ2866 grown on defined VM medium with 20 g/L cellulose at 34°C. (A) Ethanol production for P4-2866, P2-2866, and Δ2866. (B) Lactate production for P4-2866, P2-2866, and Δ2866. (C) Acetate production for P4-2866, P2-2866, and Δ2866. (D) Accumulation of cellobiose for P4-2866, P2-2866, and Δ2866. (E) Accumulation of glucose for P4-2866, P2-2866, and Δ2866. Data are shown as mean ± standard deviation (n=3 biological replicates).

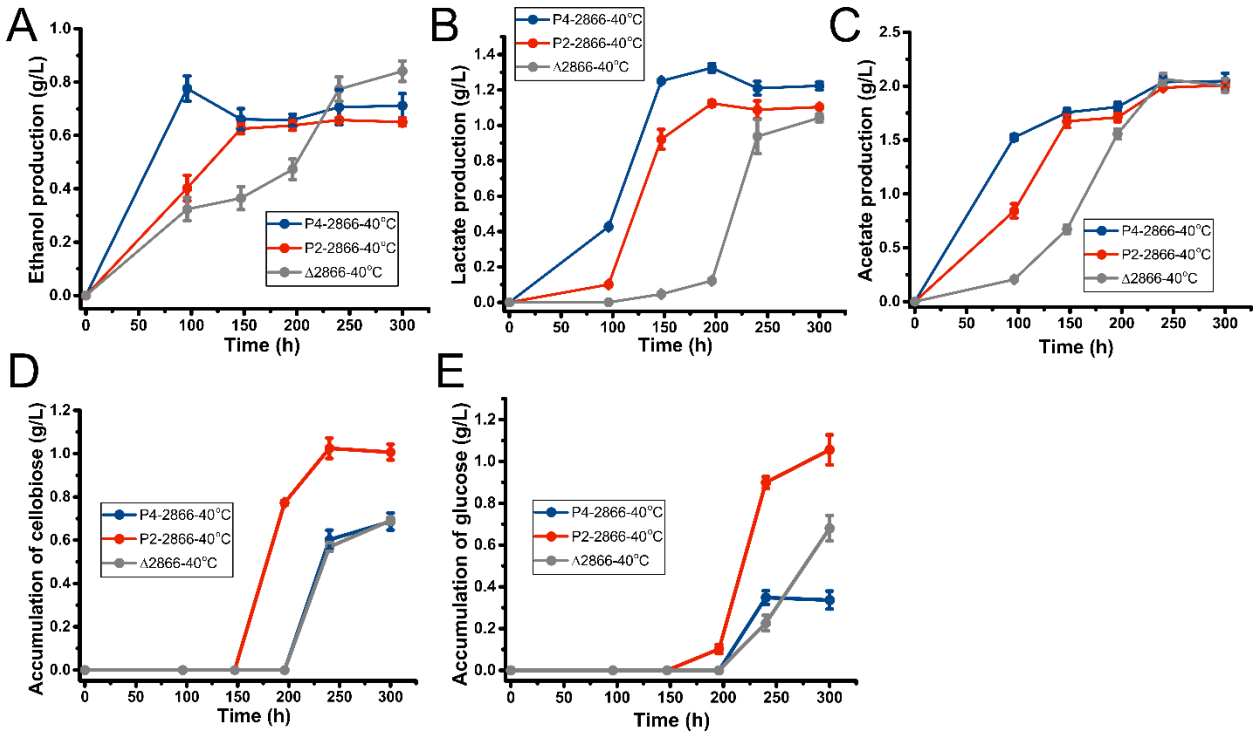

Figure S3.5. Fermentation products and accumulation of soluble sugars profiles for P4-2866, P2-2866, and Δ2866 grown on defined VM medium with 20 g/L cellulose at 40°C. (A) Ethanol production for P4-2866, P2-2866, and Δ2866. (B) Lactate production for P4-2866, P2-2866, and Δ2866. (C) Acetate production for P4-2866, P2-2866, and Δ2866. (D) Accumulation of cellobiose for P4-2866, P2-2866, and Δ2866. (E) Accumulation of glucose for P4-2866, P2-2866, and Δ2866. Data are shown as mean ± standard deviation (n=3 biological replicates).

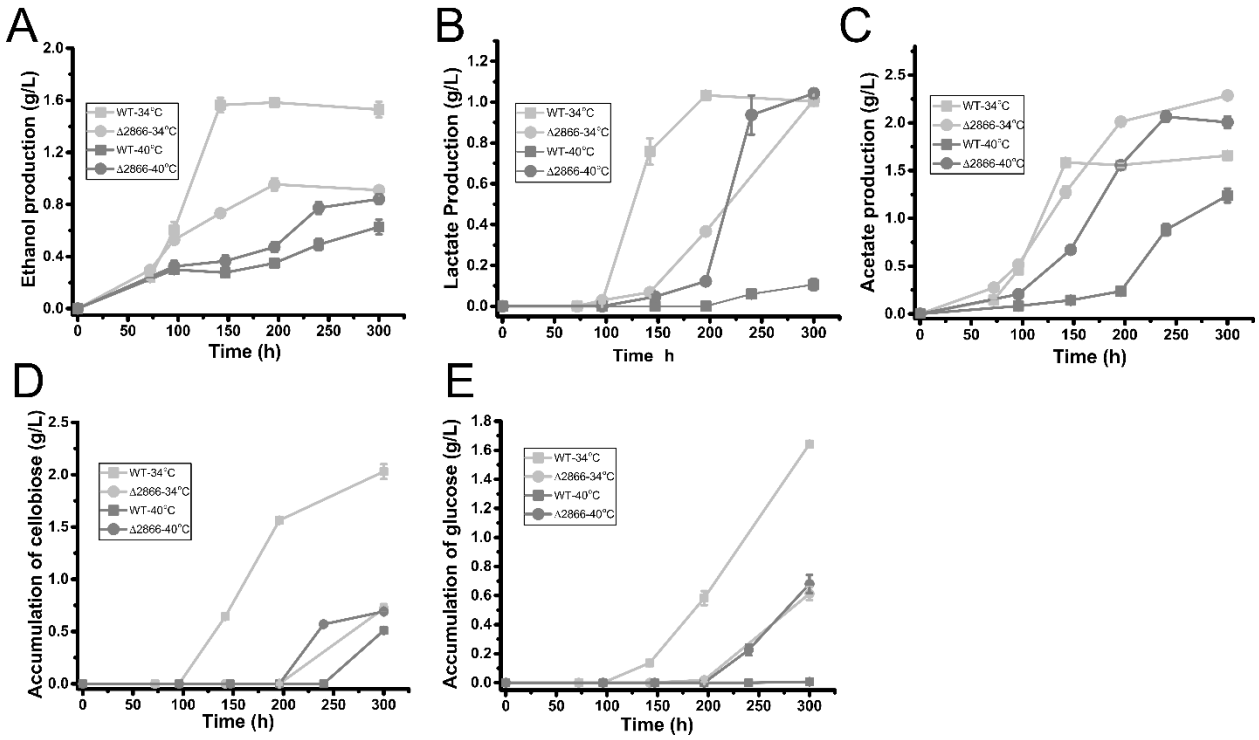

Figure S3.6. Fermentation products and accumulation of soluble sugars profiles for WT and Δ2866 strains grown on defined VM medium with 20 g/L cellulose at 34 °C and 40 °C during 300 h fermentation. (A) Ethanol production for WT and Δ2866. (B) Lactate production for WT and Δ2866. (C) Acetate production for WT and Δ2866. (D) Accumulation of cellobiose for WT and Δ2866. (E) Accumulation of glucose for WT and Δ2866. Data are shown as mean ± standard deviation (n=3 biological replicates).

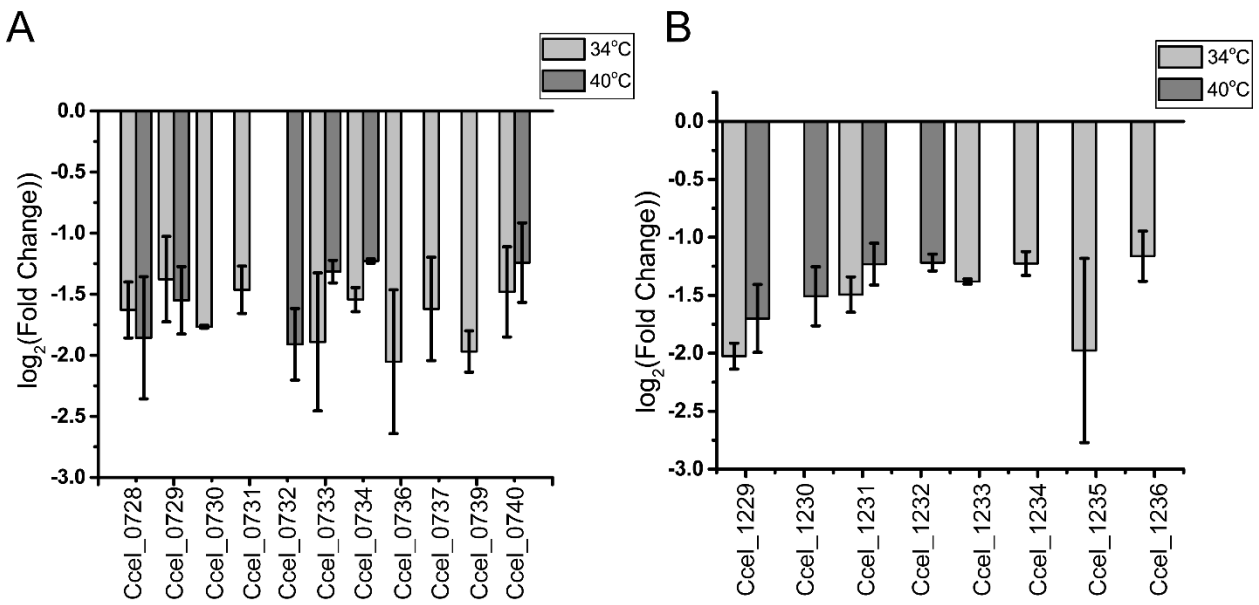

Figure S3.7. Expression changes of genes located in (A) the *cip-cel* cluster (*Ccel_0728-0740*) and (B) *xyl-doc* gene cluster (*Ccel_1229-1242*). Fold change was determined between Δ2866 and WT grown on defined VM medium with 20 g/L cellulose at 34 °C or 40 °C. Data are shown as mean ± standard deviation (n=6 biological replicates).

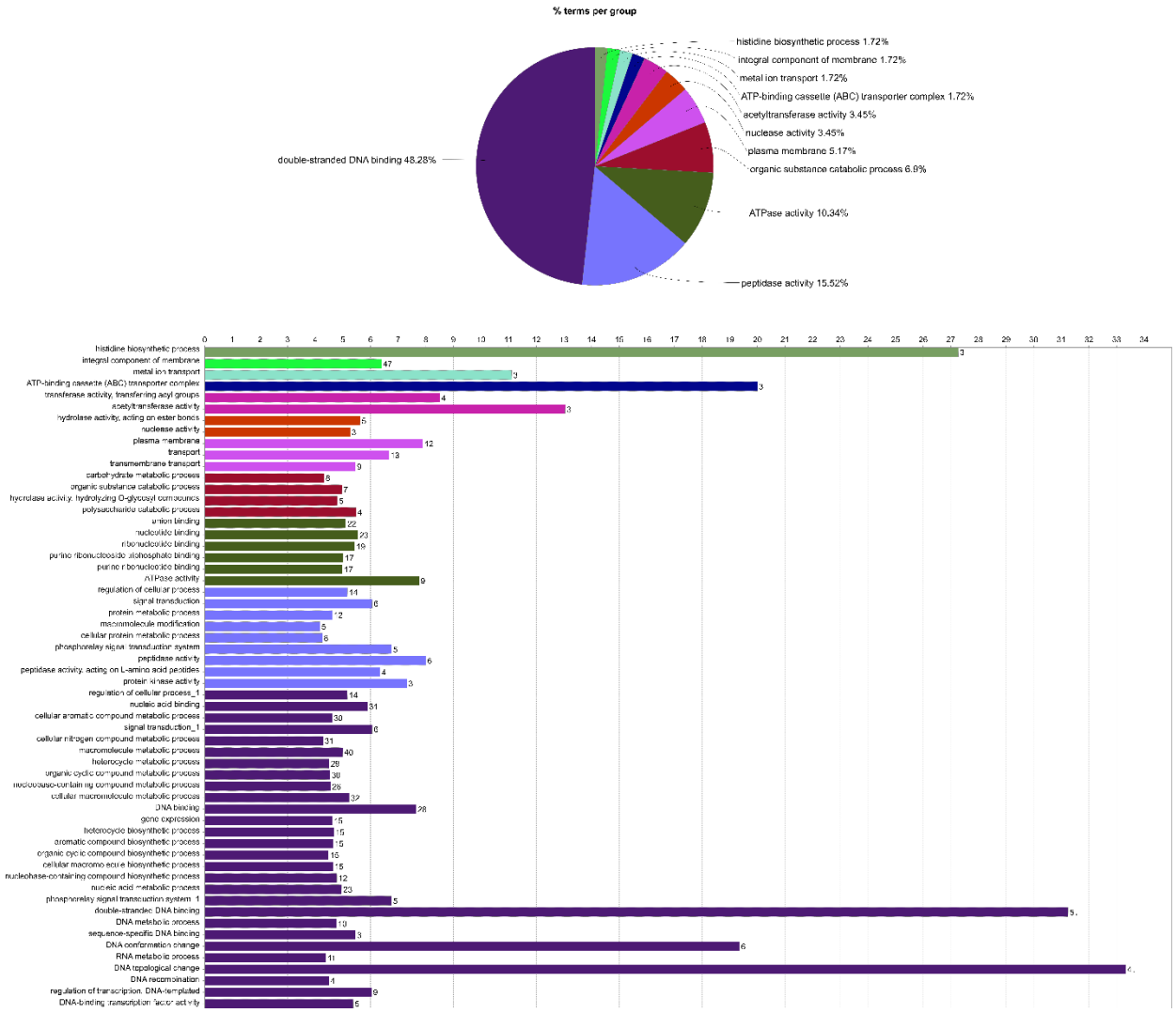

Figure S3.8. Clusters of annotated DEGs associated with representative gene ontology (GO) for WT grown between 34 °C and 40 °C. All GO terms are presented here, and none were significantly enriched for WT.

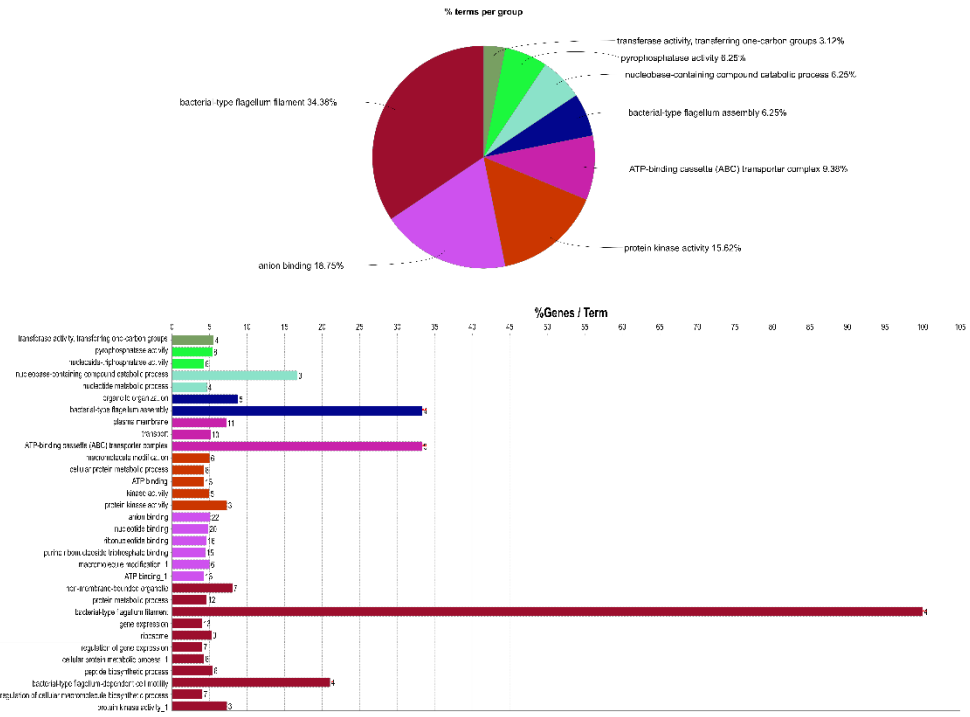

Figure S3.9. Clusters of annotated DEGs associated with representative gene ontology (GO) for Δ2866 grown between 34 °C and 40 °C. All GO terms are presented here. The significantly enriched biological processes are anion/nucleotide binding, bacterial-type flagellum filament, and protein kinase activity. Significance is indicated by *, 0.001<*P* ⩽0.05 and **, P<0.001.

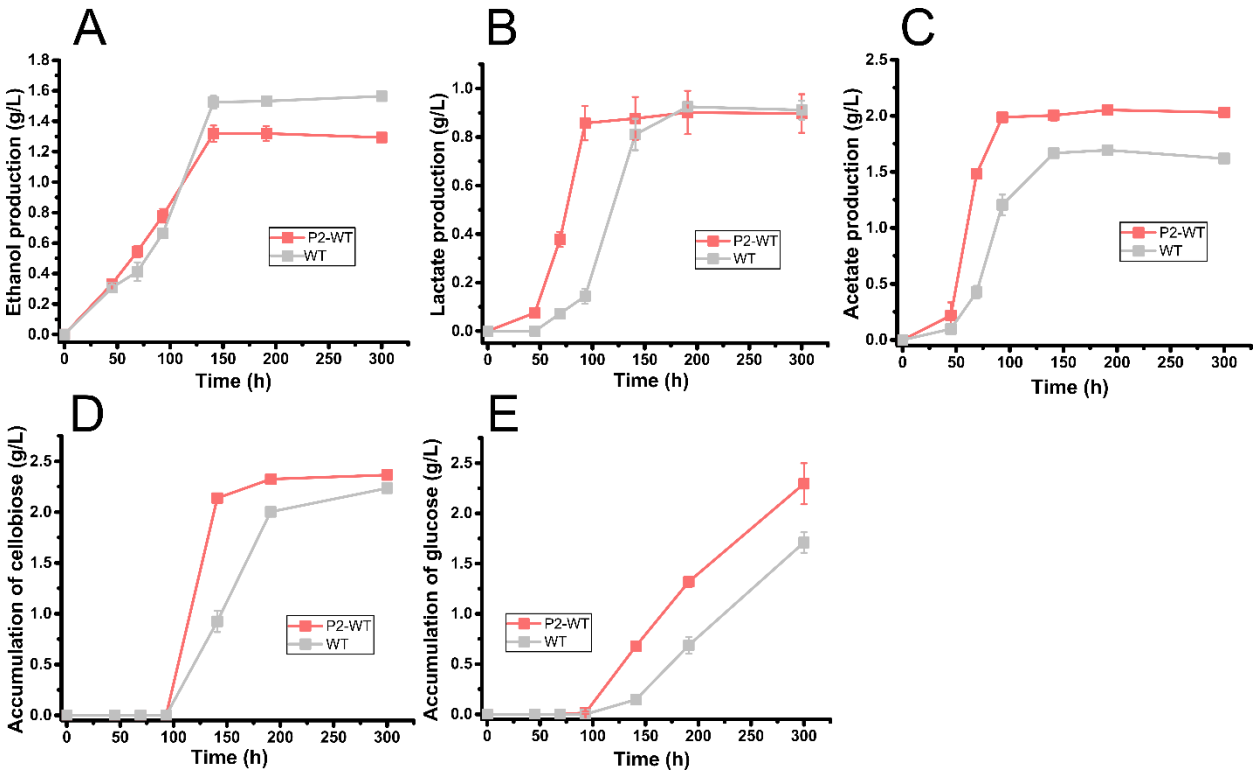

Figure S3.10. Fermentation products and accumulation of soluble sugars profiles for P2-WT and WT strains grown on defined VM medium with 20 g/L cellulose at 34°C. (A) Ethanol production for P2-WT and WT. (B) Lactate production for P2-WT and WT. (C) Acetate production for P2-WT and WT. (D) Accumulation of cellobiose for P2-WT and WT. (E) Accumulation of glucose for P2-WT and WT. Data are shown as mean ± standard deviation (n=3 biological replicates).

.

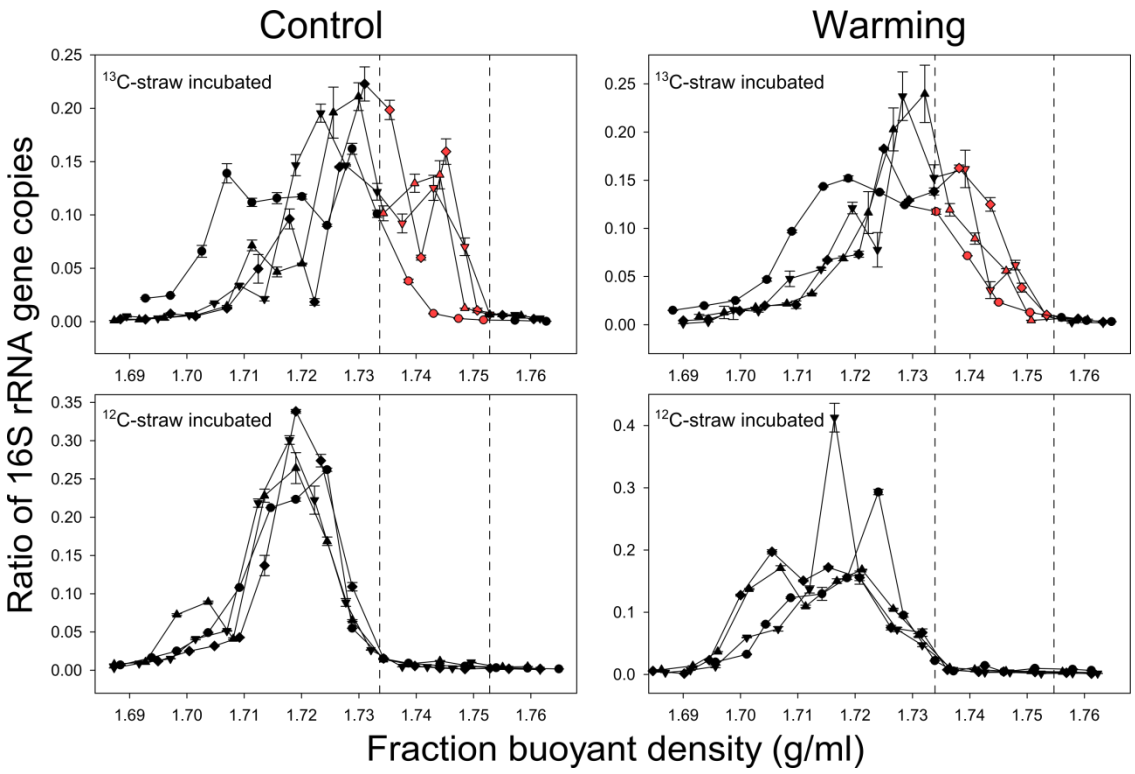

Figure S5.1 Distribution of 16S rRNA gene abundance along with buoyant density. In each pane, symbols of the shapes triangles up, triangles down, diamonds, and circles represent fractions of biological replicate (plot) 1, 2, 3, and 4, respectively. Red symbols in panes of ^{13}C-plant litter represent fractions of active communities, where the corresponding ^{12}C-plant-litter-incubated samples at the same densities were close to zero. Symbols represent the average ± standard error of 3 technical replicates of qPCR.

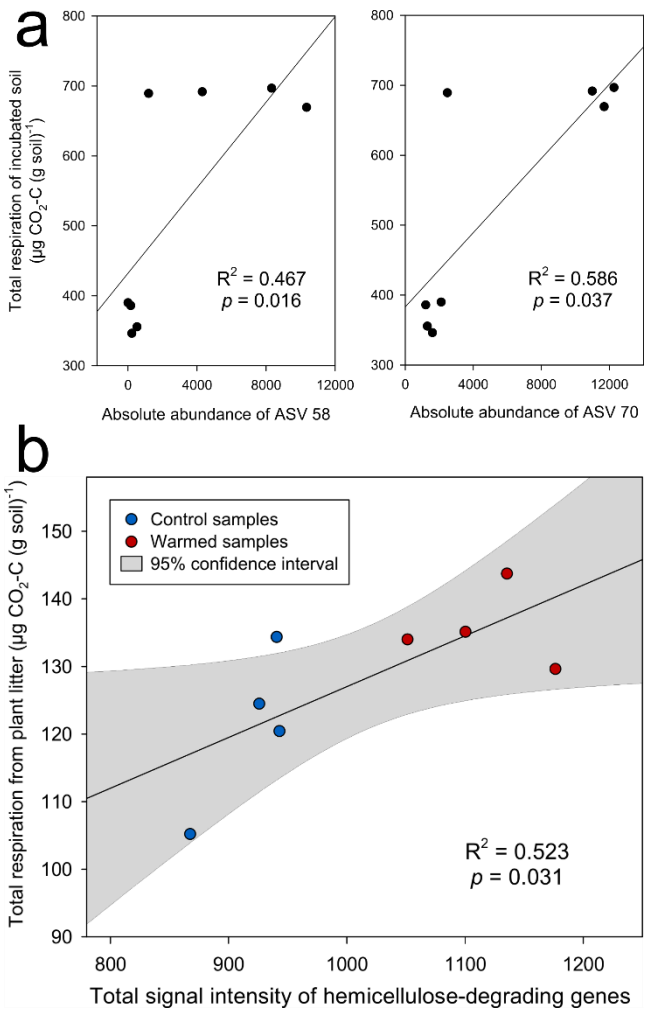

Figure S5.2 Linear regression between absolute abundance of ASV 58 and soil respiration and absolute abundance of ASV 70 and soil respiration (a), and total signal intensity of hemicellulose-degrading genes of active communities revealed by GeoChip and total respiration from the straw during the 7-day incubation (b).

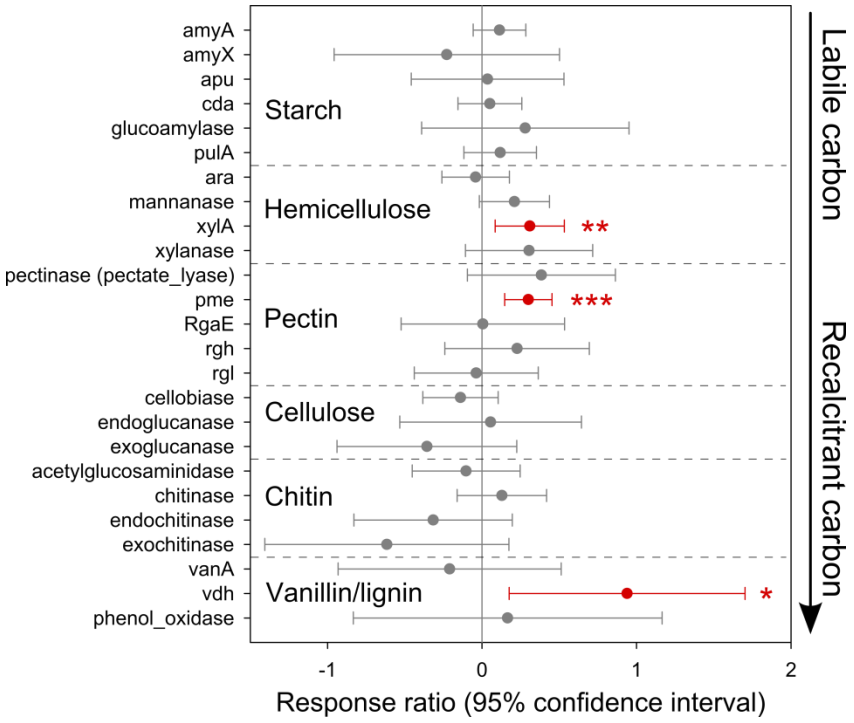

Figure S5.3 Response ratios of GeoChip signal intensities of active *Firmicutes* C-degrading genes to the *in situ* warming treatment. Each symbol represents average ± standard error of $n = 4$ biological replicates of *in situ* warming or control. Red symbols represent significantly positive response ratios. Significances are indicated using * as $0.010 < p \leq 0.050$, ** as $0.001 < p \leq 0.010$, and *** as $p \leq 0.001$ as determined by confidence intervals.

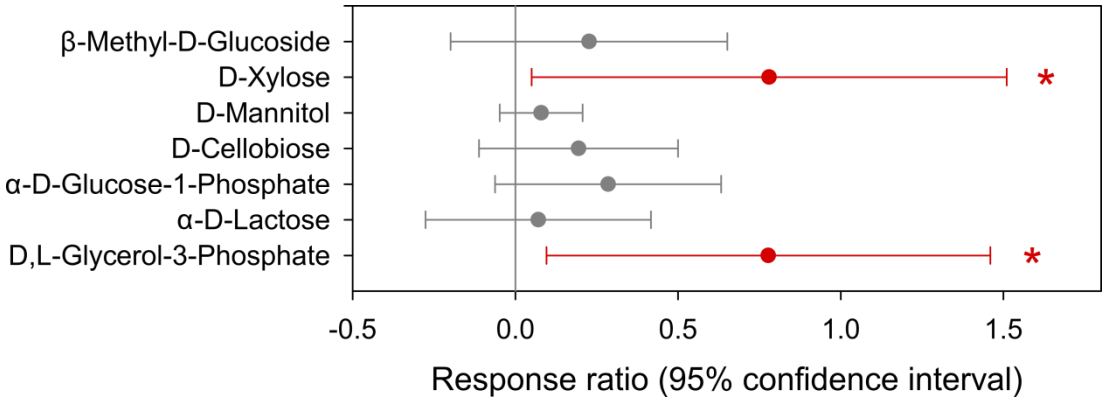

Figure S5.4 Response ratios of carbohydrates utilization capacity of soil microbial community to the *in situ* warming treatment, as determined by Biolog EcoPlates. Each symbol represents average ± standard error of $n = 4$ biological replicates of *in situ* warming or control. Red symbols represent significantly positive response ratios. Significances are indicated using * as $0.010 < p \leq 0.050$ as determined by confidence intervals.

Appendix B: Supplementary Tables

Table S3.1 List of plasmids and strains used in this study.

Strain or plasmid	Phenotype or genotype	Source or reference
Strains		
Wild type of *C. cellulolyticum* H10	ATCC 35319	Xu et al., *AEM*, 2015
Δ2866	Δ*mspI*	Xu et al., *AEM*, 2015
P4-2866	Δ*mspI* & P4 promoter insertion in *cip-cel* gene cluster	This study
P2-2866	Δ*mspI* & Pm2112 promoter insertion in *cip-cel* gene cluster	This study
P2-WT	Pm2112 promoter insertion in *cip-cel* gene cluster	This study
LM	Δ*mdh* & Δ*ldh*	Li et al., *Biotechnology for biofuels*, 2012
P4-LM	Δ*mdh,* Δ*ldh* and P4 promoter insertion in *cip-cel* gene cluster	This study
P2--LM	Δ*mdh,* Δ*ldh* and Pm2112 promoter insertion in *cip-cel* gene cluster	This study
Plasmids		
pFdCas9n-p4-pyrF w/2kbΔ	Cmp^r in *E.coli* and Tmp^r in *C. cellulolyticum* H10	Xu et al., *AEM*, 2015
pMS-RNA	$Spec^r$ in *E.coli*	Xu et al., *AEM*, 2015
pCR/8w p4-4 prom	$Spec^r$ in *E.coli*	Xu et al., *AEM*, 2015
pCas9n-P4inser-donor	Cmp^r in *E.coli* and Tmp^r in *C. cellulolyticum* H10	This study
pCas9n-Pm2112inser-donor	Cmp^r in *E.coli* and Tmp^r in *C. cellulolyticum* H10	This study

Table S3.2 List of primers used for plasmid construction.

Primer Name	Sequences 5'-3'
PM9HGRF	GTTGATGCATGGAGTGTGTAAATAGTTTTAGAGCTAGAAATAGCAAGT
PM9HGRR	CTAAAACTATTTACACACTCCATGCATCAACTTAATTTTAACTTTAAAAAATAAATTTGT
9HLF	ACTGAATTTTATTATGGTACCCGGGACGGCGATGGAAACAAGGAT
9HLR	TTATTAATTAGTATGTTACACACTCCATGCATGAAT
9HRF	AATAAAAAGGAGGATTTATATGGATAAAATGAAAAGAGTAAGTATATATGCCC
9HRR	CTCATCAATTTGTTGCAACGAGCTGTCAAGCTTTGCATGACCACCC
PM2112F	TCATGCATGGAGTGTGTAACATACTAATTAATAATGCTATAACATGGTG
PM2112R	CATTTTATCCATATAAATCCTCCTTTTTATTTTTGGCATTTTCATATTGAAATGC
9HP4LR	CCTCCTTTTTATTCTTAATTTTAACTTTAAAAAATAAATTTGTCAATATATTATTTACACACTCCATGCATGAAT
9HP4RF	GTTAAAATTAAGAATAAAAAGGAGGATTTATATGGATAAAATGAAAAGAGTAAGTATATATGCCC

Table S4.1 List of plasmids and strains used in this study.

Strain or plasmid	Phenotype or genotype	Source or reference
Strains		
Wild type of *C. cellulolyticum* H10	ATCC 35319	Xu et al., *AEM*, 2015
WT-CcBglA	*ccbglA* gene insertion in upstream of *ccel_2485*	This study
Δ1005	Δ *ccel_1005*	This study
Δ1438	Δ *ccel_1438*	This study
Plasmids		
pFdCas9n-p4-pyrF w/2kbΔ	Cmp^r in *E.coli* and Tmp^r in *C. cellulolyticum* H10	Xu et al., *AEM*, 2015
pMS-RNA	$Spec^r$ in *E.coli*	Xu et al., *AEM*, 2015
pCR/8w p4-4 prom	$Spec^r$ in *E.coli*	Xu et al., *AEM*, 2015
pCas9n-CcBglAinser-donor	Cmp^r in *E.coli* and Tmp^r in *C. cellulolyticum* H10	This study
pCas9n-Δ1005-donor	Cmp^r in *E.coli* and Tmp^r in *C. cellulolyticum* H10	This study
pCas9n-Δ1438-donor	Cmp^r in *E.coli* and Tmp^r in *C. cellulolyticum* H10	This study

Table S4.2 List of primers used for plasmid construction.

Primer Name	Sequences 5'-3'
2485GRF	GTTGTAAATACTAAATGGAAGAGTGTTTTAGAGCTAGAAATAGCAAGT
2485GRR	CTAAAACACTCTTCCATTTAGTATTTACAACTTAATTTTAACTTTAAAAAATAAATTTG
2485LF	GAATTTTATTATGGTACCCGGGTACCCATATATCGGACTGCGGGC
2485LR	TTTCCATTTAGTATTTACTATAAATATACCC
2485BGLAF	TTATAGTAAATACTAAATGGAAAAATTTTAAGGAGGTGTATTTCATATG
2485BGLAR	CATAATCAATATCCTACTGCGTCGGAAAAGTTAAACTGCGA
2485RF	TCCGACGCAGTAGGATATTGATTATGAAAAATAAATCTA
2485RR	CTCATCAATTTGTTGCAACGAGAACCGATGCAGTGGTCGGAGTAG
ID2485R	GGGAATACGTGACTTATTTGGTG
2485SEQR	GTACAATCTCGGAAACTAGTCCGC
1438GRF	GTTGTTTGATATAGCAAGGATGGCGTTTTAGAGCTAGAAATAGCAAGT
1438GRR	CTAAAAC GCCATCCTTGCTATATCAAA CAACTTAATTTTAACTTTAAAAAATAAATTTGT
1438_LF	GAATTTTATTATGGTACCCGGGGGTGCTTGCAGCGACATAAGC
1438LR	CCCGCCTGGTACCATCCTTGCTATATCAAAAATATTGTTCA
1438RF	TATAGCAAGGATGGTACCAGGCGGGAGTATCAAAAACTACCGTTT
1438RR	CTCATCAATTTGTTGCAACGAGCCCAACTCGGATGCCCTTTGTA
1438_SF	GGGAATACGCTTTGTCAGCG
1438_SR	CATTAAGGCGTTCCAGTGCC

Table S4.3 [a]List of potential genome loci for integration in *C. cellulolyticum*.

Gene_ID	Function	Cellulose specific ratio	Operon
Ccel_0164	transcriptional regulator, AbrB family	6.48	orphan
Ccel_0401	putative anti-sigma regulatory factor, serine/threonine protein kinase	3.3	teriminal gene
Ccel_0689	Stage V sporulation protein S	1.8	teriminal gene
Ccel_0902	Rubrerythrin	1.97	orphan
Ccel_1108	hypothetical protein	1.52	orphan
Ccel_1380	hypothetical protein	2.19	
Ccel_1432	hypothetical protein	2.24	orphan
Ccel_2485	L-lactate dehydrogenase	1.8	orphan
Ccel_2901	cold-shock DNA-binding domain protein	2.79	orphan
Ccel_3398	hypothetical protein	0.81	
Ccel_2093	GatB/Yqey domain protein	1.15	teriminal gene
Ccel_2253	preprotein translocase, SecG subunit	1.01	orphan

a, this table is adapted from Xu, Chenggang, et al. *Nature communications* 6.1 (2015): 1-13.

Table S5.1 Annual average data of environmental factors in 2016.

Environmental factors	Control	Warming	*p* value[a]
Soil carbon content (%)	0.798 ± 0.051[b]	0.888 ± 0.060	0.359
Soil nitrogen content (%)	0.088 ± 0.005	0.096 ± 0.006	0.384
Soil pH	7.30 ± 0.13	7.19 ± 0.18	0.666
Soil temperature at 7.5 cm depth (°C)[c]	**17.15 ± 0.19**	**18.88 ± 0.37**	**0.001**
Soil moisture (% v/v)	11.75 ± 1.17	10.02 ± 0.98	0.365
Soil moisture when sampled (% v/v)	24.48 ± 2.60	22.64 ± 1.12	0.4
Aboveground plant biomass (g m^{-2})	204.9 ± 60.6	107.2 ± 22.7	0.239

[a]The *p*-value of permutation ANOVA between warmed and control samples.

[b]Values shown in this table are average ± standard error of $n = 4$ biological replicates.

[c]Bold font represents a significant difference ($p < 0.050$) between warmed and control samples.

Table S5.2 Variance in potential carbon assimilation rate[a] attributable to taxonomy and warming treatment.

Samples	Phylum	Class	Order	Family	Genus	Phylotype	Treatment	Within
Total[b]	<0.1%	<0.1%	25.8%	<0.1%	<0.1%	<0.1%	74.2%	<0.1%
Control	<0.1%	<0.1%	<0.1%	<0.1%	<0.1%	100%	N/A	<0.1%
Warming	<0.1%	<0.1%	43.0%	<0.1%	<0.1%	57%	N/A	<0.1%

[a] Potential carbon assimilation rate is reflected by an active degrader's genomic excess atom fraction (EAF) ^{13}C.

[b]Total means that active degraders in warmed and control samples.

Reference

(2016), I. (2016). "Key World Energy Statistics 2016." International Energy Agency: 6.

Abdel-Banat, B. M., et al. (2010). "High-temperature fermentation: how can processes for ethanol production at high temperatures become superior to the traditional process using mesophilic yeast?" Applied microbiology and biotechnology **85**(4): 861-867.

Abdou, L., et al. (2008). "Transcriptional regulation of the Clostridium cellulolyticum cip-cel operon: a complex mechanism involving a catabolite-responsive element." Journal of bacteriology **190**(5): 1499-1506.

Abdou, L., et al. (2008). "Transcriptional regulation of the Clostridium cellulolyticum cip-cel operon: a complex mechanism involving a catabolite-responsive element." Journal of bacteriology **190**(5): 1499-1506.

Amore, A., et al. (2013). "Cellulolytic Bacillus strains from natural habitats-a review." Chimica Oggi/Chemistry Today **31**: 49-52.

Arsène, F., et al. (2000). "The heat shock response of Escherichia coli." International journal of food microbiology **55**(1-3): 3-9.

Asztalos, A., et al. (2012). "A coarse-grained model for synergistic action of multiple enzymes on cellulose." Biotechnology for biofuels **5**(1): 55.

Balan, V. (2014). "Current challenges in commercially producing biofuels from lignocellulosic biomass." ISRN biotechnology **2014**.

Balan, V. (2014). "Current challenges in commercially producing biofuels from lignocellulosic biomass." International Scholarly Research Notices **2014**.

Bayer, E. A. and R. Lamed (1986). "Ultrastructure of the cell surface cellulosome of Clostridium thermocellum and its interaction with cellulose." Journal of bacteriology **167**(3): 828-836.

Bindea, G., et al. (2013). "CluePedia Cytoscape plugin: pathway insights using integrated experimental and in silico data." Bioinformatics **29**(5): 661-663.

Bindea, G., et al. (2009). "ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks." Bioinformatics **25**(8): 1091-1093.

Blouzard, J. C., et al. (2010). "Modulation of cellulosome composition in Clostridium cellulolyticum: adaptation to the polysaccharide environment revealed by proteomic and carbohydrate‐active enzyme analyses." Proteomics **10**(3): 541-554.

Bokinsky, G., et al. (2011). "Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered Escherichia coli." Proceedings of the National Academy of Sciences **108**(50): 19949-19954.

Boraston, A. B., et al. (2004). "Carbohydrate-binding modules: fine-tuning polysaccharide recognition." Biochemical journal **382**(3): 769-781.

Bothast, R. and M. Schlicher (2005). "Biotechnological processes for conversion of corn into ethanol." Applied microbiology and biotechnology **67**(1): 19-25.

Brown, B. L., et al. (2013). "The Escherichia coli toxin MqsR destabilizes the transcriptional repression complex formed between the antitoxin MqsA and the mqsRA operon promoter." Journal of Biological Chemistry **288**(2): 1286-1294.

Brown, J. H., et al. (2004). "Toward a metabolic theory of ecology." Ecology **85**(7): 1771-1789.

Bugg, T. D., et al. (2011). "The emerging role for bacteria in lignin degradation and bio-product formation." Current opinion in biotechnology **22**(3): 394-400.

Byrne, B. M. (2010). Structural equation modeling with AMOS Basic concepts, applications, and programming (Multivariate Applications Series), New York: Routledge.

Callahan, B. J., et al. (2016). "DADA2: high-resolution sample inference from Illumina amplicon data." Nature methods **13**(7): 581.

Carey, J. C., et al. (2016). "Temperature response of soil respiration largely unaltered with experimental warming." Proceedings of the National Academy of Sciences **113**(48): 13797-13802.

Chanal, A., et al. (2011). "Scaffoldin modules serving as “Cargo” domains to promote the secretion of heterologous cellulosomal cellulases by Clostridium acetobutylicum." Appl. Environ. Microbiol. **77**(17): 6277-6280.

Chandel, A. K., et al. (2019). "Comparative analysis of key technologies for cellulosic ethanol production from Brazilian sugarcane bagasse at a commercial scale." Biofuels, Bioproducts and Biorefining **13**(4): 994-1014.

Chandrakant, P. and V. Bisaria (1998). "Simultaneous bioconversion of cellulose and hemicellulose to ethanol." Critical Reviews in Biotechnology **18**(4): 295-331.

Charoensuk, K., et al. (2017). "Thermotolerant genes essential for survival at a critical high temperature in thermotolerant ethanologenic Zymomonas mobilis TISTR 548." Biotechnology for biofuels **10**(1): 204.

Cheng, L., et al. (2017). "Warming enhances old organic carbon decomposition through altering functional microbial communities." The ISME journal **11**(8): 1825-1835.

Chhabra, S., et al. (2006). "Global analysis of heat shock response in Desulfovibrio vulgaris Hildenborough." Journal of bacteriology **188**(5): 1817-1828.

Chu, E. K., et al. (2018). "Self-induced mechanical stress can trigger biofilm formation in uropathogenic Escherichia coli." Nature communications **9**(1): 4087.

Cleveland, C. C., et al. (2014). "Litter quality versus soil microbial community controls over decomposition: a quantitative analysis." Oecologia **174**(1): 283-294.

Conesa, A., et al. (2005). "Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research." Bioinformatics **21**(18): 3674-3676.

Crowther, T. W., et al. (2016). "Quantifying global soil carbon losses in response to warming." Nature **540**(7631): 104.

Cui, G.-z., et al. (2012). "Targeted gene engineering in Clostridium cellulolyticum H10 without methylation." Journal of microbiological methods **89**(3): 201-208.

Desvaux, M. (2005). "The cellulosome of Clostridium cellulolyticum." Enzyme and Microbial Technology **37**(4): 373-385.

Desvaux, M. (2005). "Clostridium cellulolyticum: model organism of mesophilic cellulolytic clostridia." FEMS microbiology reviews **29**(4): 741-764.

Desvaux, M., et al. (2000). "Cellulose catabolism by Clostridium cellulolyticum growing in batch culture on defined medium." Appl. Environ. Microbiol. **66**(6): 2461-2470.

Deutscher, J., et al. (2014). "The bacterial phosphoenolpyruvate: carbohydrate phosphotransferase system: regulation by protein phosphorylation and phosphorylation-dependent protein-protein interactions." Microbiology and molecular biology reviews **78**(2): 231-256.

Dhyani, V. and T. Bhaskar (2018). "A comprehensive review on the pyrolysis of lignocellulosic biomass." Renewable Energy **129**: 695-716.

Doi, R. H. (2008). "Cellulases of mesophilic microorganisms." Annals of the New York Academy of Sciences **1125**(1): 267-279.

Doi, R. H., et al. (1994). "The Clostridium cellulovorans cellulosome." Critical reviews in microbiology **20**(2): 87-93.

Đokić, L., et al. (2011). "Four Bacillus sp. soil isolates capable of degrading phenol, toluene, biphenyl, naphthalene and other aromatic compounds exhibit different aromatic catabolic potentials." Arch. Biol. Sci., Belgrade **63**(4): 1057-1067.

dos Santos Bernardes, M. A. (2011). "Biofuel Production-Recent Developments and Prospects."

Duff Jr, M. R., et al. (2011). "Isothermal titration calorimetry for measuring macromolecule-ligand affinity." JoVE (Journal of Visualized Experiments)(55): e2796.

Egan, T. (2006). The worst hard time: The untold story of those who survived the great American dust bowl, Houghton Mifflin Harcourt.

Emmett, B. A., et al. (2004). "The response of soil processes to climate change: results from manipulation studies of shrublands across an environmental gradient." Ecosystems **7**(6): 625-637.

Fan, L.-H., et al. (2012). "Self-surface assembly of cellulosomes with two miniscaffoldins on Saccharomyces cerevisiae for cellulosic ethanol production." Proceedings of the National Academy of Sciences **109**(33): 13260-13265.

FAQ:, E. (2016). "What is U.S. electricity generation by energy source? :." U.S. Energy Information Adminstration.

Farrell, A. E., et al. (2006). "Ethanol can contribute to energy and environmental goals." Science **311**(5760): 506-508.

Feng, J., et al. (2020). "Warming-induced permafrost thaw exacerbates tundra soil carbon decomposition mediated by microbial community." Microbiome **8**(1): 1-12.

Ferdinand, P.-H., et al. (2013). "Are cellulosome scaffolding protein CipC and CBM3-containing protein HycP, involved in adherence of Clostridium cellulolyticum to cellulose?" PLoS One **8**(7).

Fierobe, H.-P., et al. (1999). "Cellulosome from Clostridium cellulolyticum: molecular study of the dockerin/cohesin interaction." Biochemistry **38**(39): 12822-12832.

Gal, L., et al. (1997). "Characterization of the cellulolytic complex (cellulosome) produced by Clostridium cellulolyticum." Applied and environmental microbiology **63**(3): 903-909.

Galbe, M. and G. Zacchi (2002). "A review of the production of ethanol from softwood." Applied microbiology and biotechnology **59**(6): 618-628.

Galinier, A., et al. (1998). "New protein kinase and protein phosphatase families mediate signal transduction in bacterial catabolite repression." Proceedings of the National Academy of Sciences **95**(4): 1823-1828.

Gefen, G., et al. (2012). "Enhanced cellulose degradation by targeted integration of a cohesin-fused β-glucosidase into the Clostridium thermocellum cellulosome." Proceedings of the National Academy of Sciences **109**(26): 10298-10303.

Giallo, J., et al. (1983). "Metabolism of glucose and cellobiose by cellulolytic mesophilic Clostridium sp. strain H10." Appl. Environ. Microbiol. **45**(3): 843-849.

Görke, B. and J. Stülke (2008). "Carbon catabolite repression in bacteria: many ways to make the most out of nutrients." Nature Reviews Microbiology **6**(8): 613.

Gruno, M., et al. (2004). "Inhibition of the Trichoderma reesei cellulases by cellobiose is strongly dependent on the nature of the substrate." Biotechnology and bioengineering **86**(5): 503-511.

Guckert, J. B., et al. (1996). "Community analysis by Biolog: curve integration for statistical analysis of activated sludge microbial habitats." Journal of Microbiological Methods **27**(2-3): 183-197.

Guedon, E., et al. (2002). "Improvement of cellulolytic properties of Clostridium cellulolyticum by metabolic engineering." Appl. Environ. Microbiol. **68**(1): 53-58.

Guedon, E., et al. (1999). "Carbon and electron flow in Clostridium cellulolyticum grown in chemostat culture on synthetic medium." Journal of bacteriology **181**(10): 3262-3269.

Guo, X., et al. (2018). "Climate warming leads to divergent succession of grassland microbial communities." Nature Climate Change **8**(9): 813-818.

Guo, X., et al. (2019). "Climate warming accelerates temporal scaling of grassland soil microbial biodiversity." Nature ecology & evolution **3**(4): 612-619.

Gupta, V. K., et al. (2016). "Fungal enzymes for bio-products from sustainable and waste biomass." Trends in biochemical sciences **41**(7): 633-645.

Hall, B. G. (2013). "Building phylogenetic trees from molecular data with MEGA." Molecular biology and evolution **30**(5): 1229-1235.

Hansen, J., et al. (2010). "Global surface temperature change." Reviews of Geophysics **48**(4).

Hanson, K., et al. (2002). "HPr kinase/phosphatase of Bacillus subtilis: expression of the gene and effects of mutations on enzyme activity, growth and carbon catabolite repression." Microbiology **148**(6): 1805-1811.

He, Z. and J. Zhou (2008). "Empirical evaluation of a new method for calculating signal-to-noise ratio for microarray data analysis." Applied and environmental microbiology **74**(10): 2957-2966.

Hemme, C. L., et al. (2011). "Correlation of genomic and physiological traits of Thermoanaerobacter species with biofuel yields." Appl. Environ. Microbiol. **77**(22): 7998-8008.

Hemsworth, G. R., et al. (2015). "Lytic polysaccharide monooxygenases in biomass conversion." Trends in biotechnology **33**(12): 747-761.

Hendriks, A. and G. Zeeman (2009). "Pretreatments to enhance the digestibility of lignocellulosic biomass." Bioresource technology **100**(1): 10-18.

Higashide, W., et al. (2011). "Metabolic engineering of Clostridium cellulolyticum for production of isobutanol from cellulose." Appl. Environ. Microbiol. **77**(8): 2727-2733.

Hungate, B. A., et al. (2015). "Quantitative microbial ecology through stable isotope probing." Appl. Environ. Microbiol. **81**(21): 7570-7581.

IEA (2009). "Bioenergy-a Sustainable and Reliable Energy Source."

Jahirul, M. I., et al. (2012). "Biofuels production through biomass pyrolysis—a technological review." Energies **5**(12): 4952-5001.

Jarvis, M. (2003). "Cellulose stacks up." Nature **426**(6967): 611-612.

Jeng, W.-Y., et al. (2011). "Structural and functional analysis of three β-glucosidases from bacterium Clostridium cellulovorans, fungus Trichoderma reesei and termite Neotermes koshunensis." Journal of structural biology **173**(1): 46-56.

Jennert, K. C., et al. (2000). "Gene transfer to Clostridium cellulolyticum ATCC 35319." Microbiology **146**(12): 3071-3080.

Jones, S. E. and J. T. Lennon (2010). "Dormancy contributes to the maintenance of microbial diversity." Proceedings of the National Academy of Sciences **107**(13): 5881-5886.

Kadam, S. K. and A. L. Demain (1989). "Addition of cloned β-glucosidase enhances the degradation of crystalline cellulose by the Clostridium thermocellum cellulase complex." Biochemical and biophysical research communications **161**(2): 706-711.

Klappenbach, J. A., et al. (2000). "rRNA operon copy number reflects ecological strategies of bacteria." Appl. Environ. Microbiol. **66**(4): 1328-1333.

Klein‐Marcuschamer, D., et al. (2012). "The challenge of enzyme cost in the production of lignocellulosic biofuels." Biotechnology and bioengineering **109**(4): 1083-1087.

Koeck, D. E., et al. (2014). "Genomics of cellulolytic bacteria." Current opinion in biotechnology **29**: 171-183.

Kosugi, A., et al. (2004). "Hydrophilic domains of scaffolding protein CbpA promote glycosyl hydrolase activity and localization of cellulosomes to the cell surface of Clostridium cellulovorans." Journal of bacteriology **186**(19): 6351-6359.

Kuhad, R. C. and A. Singh (1993). "Lignocellulose biotechnology: current and future prospects." Critical Reviews in Biotechnology **13**(2): 151-172.

Kumar, A. K. and S. Sharma (2017). "Recent updates on different methods of pretreatment of lignocellulosic feedstocks: a review." Bioresources and Bioprocessing **4**(1): 7.

Kuzyakov, Y. and G. Domanski (2000). "Carbon input by plants into the soil. Review." Journal of Plant Nutrition and Soil Science **163**(4): 421-431.

Lamed, R. and E. A. Bayer (1988). "The cellulosome of Clostridium thermocellum." Adv. Appl. Microbiol **33**: 1-46.

Lamed, R., et al. (1991). "Efficient cellulose solubilization by a combined cellulosome-β-glucosidase system." Applied biochemistry and biotechnology **27**(2): 173-183.

LAURSEN, W. (2005). "Students take a green initiative." Chemical engineer(774-75): 32-34.

Lenton, T. M., et al. (2008). "Tipping elements in the Earth's climate system." Proceedings of the national Academy of Sciences **105**(6): 1786-1793.

Li, D.-C., et al. (2012). "Thermotolerance and molecular chaperone function of the small heat shock protein HSP20 from hyperthermophilic archaeon, Sulfolobus solfataricus P2." Cell stress and chaperones **17**(1): 103-108.

Li, Y., et al. (2012). "Combined inactivation of the Clostridium cellulolyticum lactate and malate dehydrogenase genes substantially increases ethanol yield from cellulose and switchgrass fermentations." Biotechnology for biofuels **5**(1): 1.

Li, Y., et al. (2014). "Improvement of cellulose catabolism in Clostridium cellulolyticum by sporulation abolishment and carbon alleviation." Biotechnology for biofuels **7**(1): 25.

Liao, J. C., et al. (2016). "Fuelling the future: microbial engineering for the production of sustainable biofuels." Nature Reviews Microbiology **14**(5): 288.

Life, I. T. O. (2011). "v2: online annotation and display of phylogenetic trees made easy Letunic, Ivica; Bork, Peer." Nucleic Acids Research **39**: W475-W478.

Liu, S., et al. (2019). "Construction of consolidated bio-saccharification biocatalyst and process optimization for highly efficient lignocellulose solubilization." Biotechnology for biofuels **12**(1): 35.

Liu, Y.-J., et al. (2020). "Consolidated bio-saccharification: Leading lignocellulose bioconversion into the real world." Biotechnology Advances: 107535.

Looney, B. (2020). Full report–BP statistical review of world energy 2020, BP plc, London.

Luís, A. S., et al. (2013). "Understanding how noncatalytic carbohydrate binding modules can display specificity for xyloglucan." Journal of Biological Chemistry **288**(7): 4799-4809.

Lynd, L. R. (2017). "The grand challenge of cellulosic biofuels." Nature biotechnology **35**(10): 912.

Lynd, L. R., et al. (1991). "Fuel ethanol from cellulosic biomass." Science **251**(4999): 1318-1323.

Lynd, L. R., et al. (2017). "Cellulosic ethanol: status and innovation." Current opinion in biotechnology **45**: 202-211.

Lynd, L. R., et al. (2005). "Consolidated bioprocessing of cellulosic biomass: an update." Current opinion in biotechnology **16**(5): 577-583.

Lynd, L. R., et al. (2002). "Microbial cellulose utilization: fundamentals and biotechnology." Microbiology and molecular biology reviews **66**(3): 506-577.

Maamar, H., et al. (2006). "Transcriptional analysis of the cip-cel gene cluster from Clostridium cellulolyticum." Journal of bacteriology **188**(7): 2614-2624.

Maamar, H., et al. (2004). "Cellulolysis is severely affected in Clostridium cellulolyticum strain cipCMut1." Molecular microbiology **51**(2): 589-598.

Masson-Delmotte, V., et al. (2018). "Global warming of 1.5 C." An IPCC Special Report on the impacts of global warming of **1**.

Matano, Y., et al. (2012). "Display of cellulases on the cell surface of Saccharomyces cerevisiae for high yield ethanol production from high-solid lignocellulosic biomass." Bioresource technology **108**: 128-133.

Mau, R. L., et al. (2018). "Warming induced changes in soil carbon and nitrogen influence priming responses in four ecosystems." Applied Soil Ecology **124**: 110-116.

Melillo, J., et al. (2002). "Soil warming and carbon-cycle feedbacks to the climate system." Science **298**(5601): 2173-2176.

Michelin, M. and T. Maria de Lourdes (2013). "Application of Lignocelulosic Residues in the Production of." Fungal enzymes: 31.

Miller, G. L. (1959). "Use of dinitrosalicylic acid reagent for determination of reducing sugar." Analytical chemistry **31**(3): 426-428.

Mingardon, F., et al. (2011). "Comparison of family 9 cellulases from mesophilic and thermophilic bacteria." Appl. Environ. Microbiol. **77**(4): 1436-1442.

Mingardon, F., et al. (2007). "Exploration of new geometries in cellulosome-like chimeras." Appl. Environ. Microbiol. **73**(22): 7138-7149.

Mood, S. H., et al. (2013). "Lignocellulosic biomass to bioethanol, a comprehensive review with a focus on pretreatment." Renewable and Sustainable Energy Reviews **27**: 77-93.

Morag, E., et al. (1991). "Isolation and properties of a major cellobiohydrolase from the cellulosome of Clostridium thermocellum." Journal of bacteriology **173**(13): 4155-4162.

Morrissey, E. M., et al. (2019). "Evolutionary history constrains microbial traits across environmental variation." Nature Ecology & Evolution **3**(7): 1064-1069.

Mosbah, A., et al. (2000). "Solution structure of the module X2_1 of unknown function of the cellulosomal scaffolding protein CipC of Clostridium cellulolyticum." Journal of molecular biology **304**(2): 201-217.

Murata, M., et al. (2011). "Molecular strategy for survival at a critical high temperature in Eschierichia coli." PLoS One **6**(6): e20063.

Myneni, R. B., et al. (2001). "A large carbon sink in the woody biomass of Northern forests." Proceedings of the National Academy of Sciences **98**(26): 14784-14789.

Naik, S. N., et al. (2010). "Production of first and second generation biofuels: a comprehensive review." Renewable and sustainable energy reviews **14**(2): 578-597.

Ning, D., et al. (2020). "A quantitative framework reveals the ecological drivers of grassland soil microbial community assembly in response to warming." bioRxiv.

Öhgren, K., et al. (2007). "A comparison between simultaneous saccharification and fermentation and separate hydrolysis and fermentation using steam-pretreated corn stover." Process Biochemistry **42**(5): 834-839.

Olofsson, K., et al. (2010). "Improving simultaneous saccharification and co-fermentation of pretreated wheat straw using both enzyme and substrate feeding." Biotechnology for biofuels **3**(1): 17.

Olson, D. G., et al. (2012). "Recent progress in consolidated bioprocessing." Current opinion in biotechnology **23**(3): 396-405.

Pachauri, R. and A. Reisinger (2008). "Climate change 2007. Synthesis report. Contribution of Working Groups I, II and III to the fourth assessment report." Cambridge University Press, Cambridge.

Pasari, N., et al. (2017). "Impact of module-X2 and carbohydrate binding module-3 on the catalytic activity of associated glycoside hydrolases towards plant biomass." Scientific reports **7**(1): 1-15.

Payot, S., et al. (1998). "Metabolism of cellobiose by Clostridium cellulolyticum growing in continuous culture: evidence for decreased NADH reoxidation as a factor limiting growth." Microbiology **144**(2): 375-384.

Petitdemange, E., et al. (1984). "Clostridium cellulolyticum sp. nov., a cellulolytic, mesophilic: species from decayed grass." International Journal of Systematic and Evolutionary Microbiology **34**(2): 155-159.

Philippidis, G. P., et al. (1993). "Study of the enzymatic hydrolysis of cellulose for production of fuel ethanol by the simultaneous saccharification and fermentation process." Biotechnology and bioengineering **41**(9): 846-853.

Quast, C., et al. (2012). "The SILVA ribosomal RNA gene database project: improved data processing and web-based tools." Nucleic acids research **41**(D1): D590-D596.

Ravachol, J., et al. (2015). "Combining free and aggregated cellulolytic systems in the cellulosome-producing bacterium Ruminiclostridium cellulolyticum." Biotechnology for biofuels **8**(1): 114.

Ren, C., et al. (2010). "Identification and inactivation of pleiotropic regulator CcpA to eliminate glucose repression of xylose utilization in Clostridium acetobutylicum." Metabolic engineering **12**(5): 446-454.

Ritchie, M. E., et al. (2015). "limma powers differential expression analyses for RNA-sequencing and microarray studies." Nucleic acids research **43**(7): e47-e47.

Rustad, L., et al. (2001). "A meta-analysis of the response of soil respiration, net nitrogen mineralization, and aboveground plant growth to experimental ecosystem warming." Oecologia **126**(4): 543-562.

Schlesinger, W. H. (1977). "Carbon balance in terrestrial detritus." Annual review of ecology and systematics **8**(1): 51-81.

Schuur, E. A., et al. (2008). "Vulnerability of permafrost carbon to climate change: Implications for the global carbon cycle." BioScience **58**(8): 701-714.

Schuur, E. A., et al. (2015). "Climate change and the permafrost carbon feedback." Nature **520**(7546): 171-179.

Schwarz, W. (2001). "The cellulosome and cellulose degradation by anaerobic bacteria." Applied microbiology and biotechnology **56**(5-6): 634-649.

Shahzad, T., et al. (2012). "Plant clipping decelerates the mineralization of recalcitrant soil organic matter under multiple grassland species." Soil biology and biochemistry **51**: 73-80.

Shoham, Y., et al. (1999). "The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides." Trends in microbiology **7**(7): 275-281.

Simpson-Holley, M., et al. (2007). "Bring on the biorefinery." Chemical engineer(795): 46-48.

Soemphol, W., et al. (2011). "Global analysis of the genes involved in the thermotolerance mechanism of thermotolerant Acetobacter tropicalis SKU1100." Bioscience, biotechnology, and biochemistry: 1108312638-1108312638.

Stegen, J. C., et al. (2013). "Quantifying community assembly processes and identifying features that impose them." The ISME journal **7**(11): 2069-2079.

Stegen, J. C., et al. (2012). "Stochastic and deterministic assembly processes in subsurface microbial communities." The ISME journal **6**(9): 1653.

Strickland, M. S. and J. Rousk (2010). "Considering fungal: bacterial dominance in soils–methods, controls, and ecosystem implications." Soil biology and biochemistry **42**(9): 1385-1395.

Tachaapaikoon, C., et al. (2012). "Isolation and characterization of a new cellulosome-producing Clostridium thermocellum strain." Biodegradation **23**(1): 57-68.

Taha, M., et al. (2016). "Commercial feasibility of lignocellulose biodegradation: possibilities and challenges." Current opinion in biotechnology **38**: 190-197.

Tamaru, Y., et al. (2010). "Comparative genomics of the mesophilic cellulosome‐producing Clostridium cellulovorans and its application to biofuel production via consolidated bioprocessing." Environmental technology **31**(8-9): 889-903.

Tang, Y., et al. (2011). "Simultaneous saccharification and cofermentation of lignocellulosic residues from commercial furfural production and corn kernels using different nutrient media." Biotechnology for biofuels **4**(1): 22.

Tao, X., et al. (2020). "Winter warming in Alaska accelerates lignin decomposition contributed by Proteobacteria." Microbiome **8**(1): 1-12.

Tao, X., et al. (2020). "Precise promoter integration improves cellulose bioconversion and thermotolerance in Clostridium cellulolyticum." Metabolic engineering.

Terrer, C., et al. (2016). "Mycorrhizal association as a primary control of the CO2 fertilization effect." Science **353**(6294): 72-74.

Vasu, K. and V. Nagaraja (2013). "Diverse functions of restriction-modification systems in addition to cellular defense." Microbiol. Mol. Biol. Rev. **77**(1): 53-72.

Venables, W. N. and B. D. Ripley (2013). Modern applied statistics with S-PLUS, Springer Science & Business Media.

Vita, N., et al. (2019). "Turning a potent family‐9 free cellulase into an operational cellulosomal component and vice versa." The FEBS journal **286**(17): 3359-3373.

Wan, S., et al. (2007). "Responses of soil respiration to elevated CO2, air warming, and changing soil water availability in a model old‐field grassland." Global Change Biology **13**(11): 2411-2424.

Wan, S., et al. (2009). "Photosynthetic overcompensation under nocturnal warming enhances grassland carbon sequestration." Ecology **90**(10): 2700-2710.

Wang, G., et al. (2015). "Microbial dormancy improves development and experimental validation of ecosystem model." The ISME journal **9**(1): 226.

Wang, G., et al. (2014). "Representation of dormant and active microbial dynamics for ecosystem modeling." PloS one **9**(2): e89252.

Wang, Q., et al. (2007). "Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy." Applied and environmental microbiology **73**(16): 5261-5267.

Wyman, C. E. and B. H. Davison (2012). Seventeenth Symposium on Biotechnology for Fuels and Chemicals: Proceedings as Volumes 57 and 58 of Applied Biochemistry and Biotechnology, Springer Science & Business Media.

Xu, C., et al. (2015). "Cellulosome stoichiometry in Clostridium cellulolyticum is regulated by selective RNA processing and stabilization." Nature communications **6**: 6900.

Xu, C., et al. (2015). "Cellulosome stoichiometry in Clostridium cellulolyticum is regulated by selective RNA processing and stabilization." Nature communications **6**(1): 1-13.

Xu, C., et al. (2013). "Structure and regulation of the cellulose degradome in Clostridium cellulolyticum." Biotechnology for biofuels **6**(1): 73.

Xu, Q., et al. (2016). "Dramatic performance of Clostridium thermocellum explained by its wide range of cellulase modalities." Science advances **2**(2): e1501254.

Xu, T., et al. (2014). "Dockerin‐containing protease inhibitor protects key cellulosomal cellulases from proteolysis in C lostridium cellulolyticum." Molecular microbiology **91**(4): 694-705.

Xu, T., et al. (2015). "Efficient genome editing in Clostridium cellulolyticum via CRISPR-Cas9 nickase." Applied and environmental microbiology: AEM. 00873-00815.

Xu, T., et al. (2015). "Efficient genome editing in Clostridium cellulolyticum via CRISPR-Cas9 nickase." Applied and environmental microbiology **81**(13): 4423-4431.

Xu, X., et al. (2013). "Net primary productivity and rain‐use efficiency as affected by warming, altered precipitation, and clipping in a mixed‐grass prairie." Global Change Biology **19**(9): 2753-2764.

Xue, K., et al. (2016). "Tundra soil carbon is vulnerable to rapid microbial decomposition under climate warming." Nature Climate Change **6**(6): 595-600.

Yang, Y., et al. (2013). "Responses of the functional structure of soil microbial community to livestock grazing in the Tibetan alpine grassland." Global change biology **19**(2): 637-648.

Yao, X.-F., et al. (2011). "Degradation of dichloroaniline isomers by a newly isolated strain, Bacillus megaterium IMT21." Microbiology **157**(3): 721-726.

Zhang, J., et al. (2017). "Bacteriophage–prokaryote dynamics and interaction within anaerobic digestion processes across time and space." Microbiome **5**(1): 57.

Zhang, J., et al. (2017). "Efficient whole-cell-catalyzing cellulose saccharification using engineered Clostridium thermocellum." Biotechnology for biofuels **10**(1): 1-14.

Zhang, Y.-H. P., et al. (2006). "Outlook for cellulase improvement: screening and selection strategies." Biotechnology advances **24**(5): 452-481.

Zhang, Y.-H. P. and L. R. Lynd (2005). "Regulation of cellulase synthesis in batch and continuous cultures of Clostridium thermocellum." Journal of bacteriology **187**(1): 99-106.

Zhang, Y. P., et al. (2009). Cellulase assays. Biofuels, Springer**:** 213-231.

Zhou, J., et al. (1996). "DNA recovery from soils of diverse composition." Applied and environmental microbiology **62**(2): 316-322.

Zhou, J., et al. (2012). "Microbial mediation of carbon-cycle feedbacks to climate warming." Nature Climate Change **2**(2): 106-110.

Zhou, X., et al. (2007). "Source components and interannual variability of soil CO2 efflux under experimental warming and clipping in a grassland ecosystem." Global Change Biology **13**(4): 761-775.