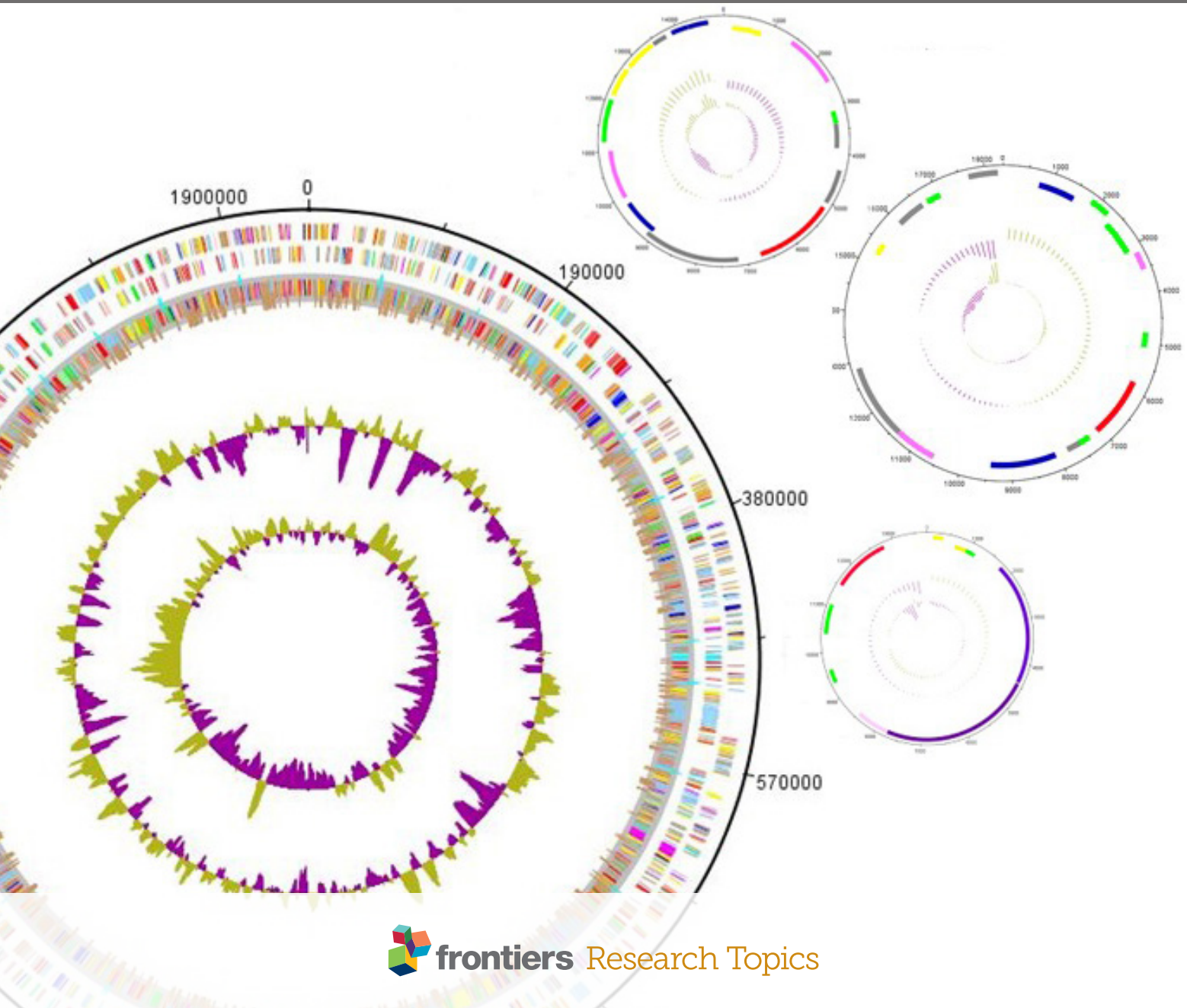


# GENETIC AND GENOME-WIDE INSIGHTS INTO MICROBES STUDIED FOR BIOENERGY

EDITED BY : Katherine M. Pappas, Ed Louis, Nigel Minton,  
Biswarup Mukhopadhyay and Shane Yang  
PUBLISHED IN : Frontiers in Microbiology





# frontiers

## Frontiers Copyright Statement

© Copyright 2007-2017 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88945-085-5

DOI 10.3389/978-2-88945-085-5

## About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: [researchtopics@frontiersin.org](mailto:researchtopics@frontiersin.org)

# GENETIC AND GENOME-WIDE INSIGHTS INTO MICROBES STUDIED FOR BIOENERGY

Topic Editors:

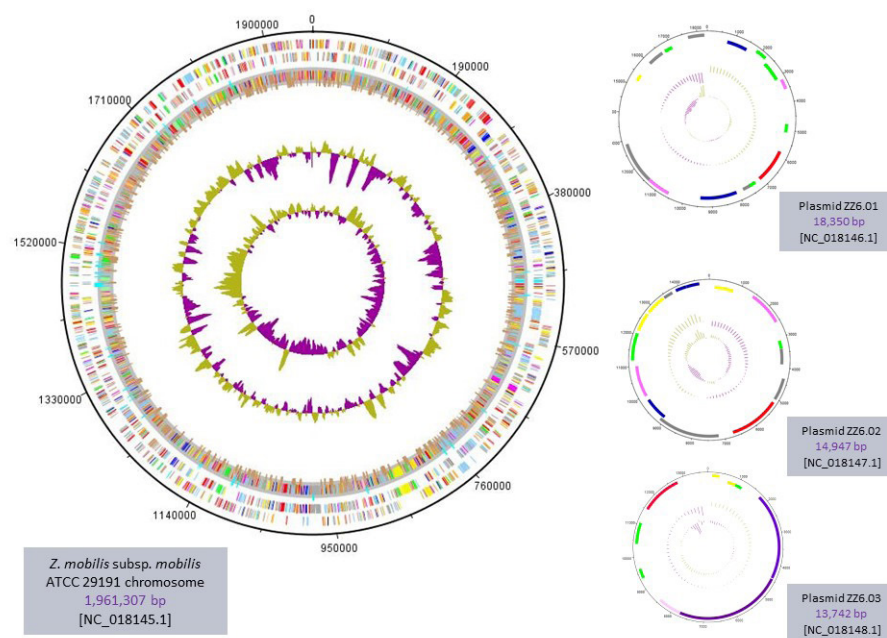
**Katherine M. Pappas**, National & Kapodistrian University of Athens, Greece

**Ed Louis**, University of Leicester, UK

**Nigel Minton**, University of Nottingham, UK

**Biswarup Mukhopadhyay**, VirginiaTech, USA

**Shane Yang**, DOE National Renewable Energy Laboratory, USA



Genetic maps of the ethanologenic *Zymomonas mobilis* strain ATCC 29191 chromosome and plasmids (Desiniotis *et al.* 2012; *J. Bacteriol.* 194: 5966). Image provided by the K. M. Pappas Lab.

The global mandate for safer, cleaner and renewable energy has accelerated research on microbes that convert carbon sources to end-products serving as biofuels of the so-called first, second or third generation – e.g., bioethanol or biodiesel derived from starchy, sugar-rich or oily crops; bioethanol derived from composite lignocellulosic biomass; and biodiesels extracted from oil-producing algae and cyanobacteria, respectively. Recent advances in ‘omics’ applications are beginning to cast light on the biological mechanisms underlying biofuel production. They also unravel mechanisms important for organic solvent or high-added-value chemical production, which, along with those for fuel chemicals, are significant to the broader field of Bioenergy.

The Frontiers in Microbial Physiology Research Topic that led to the current e-book publication, operated from 2013 to 2014 and welcomed articles aiming to better understand the genetic basis behind Bioenergy production. It invited genetic studies of microbes already used or carrying the potential to be used for bioethanol, biobutanol, biodiesel, and fuel gas production, as also of microbes posing as promising new catalysts for alternative bioproducts. Any research focusing on the systems biology of such microbes, gene function and regulation, genetic and/or genomic tool development, metabolic engineering, and synthetic biology leading to strain optimization, was considered highly relevant to the topic. Likewise, bioinformatic analyses and modeling pertaining to gene network prediction and function were also desirable and therefore invited in the thematic forum. Upon e-book development today, we, at the editorial, strongly believe that all articles presented herein – original research papers, reviews, perspectives and a technology report – significantly contribute to the emerging insights regarding microbial-derived energy production.

Katherine M. Pappas, 2016

**Citation:** Pappas, K. M., Louis, E., Minton, N., Mukhopadhyay, B., Yang, S., eds. (2017). Genetic and Genome-Wide Insights into Microbes Studied for Bioenergy. Lausanne: Frontiers Media. doi: 10.3389/978-2-88945-085-5



# Table of Contents

- 06 Genetic resources for methane production from biomass described with the Gene Ontology**  
Endang Purwantini, Trudy Torto-Alalibo, Jane Lomax, João C. Setubal, Brett M. Tyler and Biswarup Mukhopadhyay
- 24 Genetic resources for advanced biofuel production described with the Gene Ontology**  
Trudy Torto-Alalibo, Endang Purwantini, Jane Lomax, João C. Setubal, Biswarup Mukhopadhyay and Brett M. Tyler
- 41 Aromatic inhibitors derived from ammonia-pretreated lignocellulose hinder bacterial ethanogenesis by activating regulatory circuits controlling inhibitor efflux and detoxification**  
David H. Keating, Yaoping Zhang, Irene M. Ong, Sean McIlwain, Eduardo H. Morales, Jeffrey A. Grass, Mary Tremaine, William Bothfeld, Alan Higbee, Arne Ulbrich, Allison J. Balloon, Michael S. Westphall, Josh Aldrich, Mary S. Lipton, Joonhoon Kim, Oleg V. Moskvina, Yury V. Bukhman, Joshua J. Coon, Patricia J. Kiley, Donna M. Bates and Robert Landick
- 58 Genomic insights into the fungal lignocellulolytic system of *Myceliophthora thermophila***  
Anthi Karnaouri, Evangelos Topakas, Io Antonopoulou and Paul Christakopoulos
- 80 Comparative genomics and evolution of regulons of the *LacI*-family transcription factors**  
Dmitry A. Ravcheev, Matvei S. Khoroshkin, Olga N. Laikova, Olga V. Tsoy, Natalia V. Sernova, Svetlana A. Petrova, Aleksandra B. Rakhmaninova, Pavel S. Novichkov, Mikhail S. Gelfand and Dmitry A. Rodionov
- 96 Connecting lignin-degradation pathway with pre-treatment inhibitor sensitivity of *Cupriavidus necator***  
Wei Wang, Shihui Yang, Glendon B. Hunsinger, Philip T. Pienkos and David K. Johnson
- 106 Elucidation of *Zymomonas mobilis* physiology and stress responses by quantitative proteomics and transcriptomics**  
Shihui Yang, Chongle Pan, Gregory B. Hurst, Lezlee Dice, Brian H. Davison and Steven D. Brown
- 119 Metabolic engineering of yeasts by heterologous enzyme production for degradation of cellulose and hemicellulose from biomass: a perspective**  
William Kricka, James Fitzpatrick and Ursula Bond
- 130 Comparison of transcriptional profiles of *Clostridium thermocellum* grown on cellobiose and pretreated yellow poplar using RNA-Seq**  
Hui Wei, Yan Fu, Lauren Magnusson, John O. Baker, Pin-Ching Maness, Qi Xu, Shihui Yang, Andrew Bowersox, Igor Bogorad, Wei Wang, Melvin P. Tucker, Michael E. Himmel and Shi-You Ding

- 146** *A mathematical model of metabolism and regulation provides a systems-level view of how Escherichia coli responds to oxygen*  
Michael Ederer, Sonja Steinsiek, Stefan Stagge, Matthew D. Rolfe, Alexander Ter Beek, David Knies, M. Joost Teixeira de Mattos, Thomas Sauter, Jeffrey Green, Robert K. Poole, Katja Bettenbrock and Oliver Sawodny
- 158** *Death by a thousand cuts: the challenges and diverse landscape of lignocellulosic hydrolysate inhibitors*  
Jeff S. Piotrowski, Yaoping Zhang, Donna M. Bates, David H. Keating, Trey K. Sato, Irene M. Ong and Robert Landick
- 166** *Modeling of Zymomonas mobilis central metabolism for novel metabolic engineering strategies*  
Uldis Kalnenieks, Agris Pentjuss, Reinis Rutkis, Egils Stalidzans and David A. Fell
- 173** *Comparative genomics and functional analysis of rhamnose catabolic pathways and regulons in bacteria*  
Irina A. Rodionova, Xiaoqing Li, Vera Thiel, Sergey Stolyar, Krista Stanton, James K. Fredrickson, Donald A. Bryant, Andrei L. Osterman, Aaron A. Best and Dmitry A. Rodionov



# Genetic resources for methane production from biomass described with the Gene Ontology

Endang Purwantini<sup>1</sup>, Trudy Torto-Alalibo<sup>1</sup>, Jane Lomax<sup>2</sup>, João C. Setubal<sup>3,4</sup>, Brett M. Tyler<sup>4,5</sup> and Biswarup Mukhopadhyay<sup>1,4,6\*</sup>

<sup>1</sup> Department of Biochemistry, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

<sup>2</sup> European Bioinformatics Institute (EMBL-EBI), European Molecular Biology Laboratory, Hinxton, UK

<sup>3</sup> Department of Biochemistry, Universidade de São Paulo, São Paulo, Brazil

<sup>4</sup> Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

<sup>5</sup> Center for Genome Research and Biocomputing, Oregon State University, Corvallis, OR, USA

<sup>6</sup> Department of Biological Sciences, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

## Edited by:

Katherine M. Pappas, University of Athens, Greece

## Reviewed by:

Jim Spain, Georgia Institute of Technology, USA

Marcus Constantine Chibucos, University of Maryland School of Medicine, USA

## \*Correspondence:

Biswarup Mukhopadhyay,  
Department of Biochemistry,  
Virginia Polytechnic Institute and  
State University, 123 Engel Hall, 340  
West Campus Drive, Blacksburg, VA  
24060, USA  
e-mail: biswarup@vt.edu

Methane (CH<sub>4</sub>) is a valuable fuel, constituting 70–95% of natural gas, and a potent greenhouse gas. Release of CH<sub>4</sub> into the atmosphere contributes to climate change. Biological CH<sub>4</sub> production or methanogenesis is mostly performed by methanogens, a group of strictly anaerobic archaea. The direct substrates for methanogenesis are H<sub>2</sub> plus CO<sub>2</sub>, acetate, formate, methylamines, methanol, methyl sulfides, and ethanol or a secondary alcohol plus CO<sub>2</sub>. In numerous anaerobic niches in nature, methanogenesis facilitates mineralization of complex biopolymers such as carbohydrates, lipids and proteins generated by primary producers. Thus, methanogens are critical players in the global carbon cycle. The same process is used in anaerobic treatment of municipal, industrial and agricultural wastes, reducing the biological pollutants in the wastes and generating methane. It also holds potential for commercial production of natural gas from renewable resources. This process operates in digestive systems of many animals, including cattle, and humans. In contrast, in deep-sea hydrothermal vents methanogenesis is a primary production process, allowing chemosynthesis of biomaterials from H<sub>2</sub> plus CO<sub>2</sub>. In this report we present Gene Ontology (GO) terms that can be used to describe processes, functions and cellular components involved in methanogenic biodegradation and biosynthesis of specialized coenzymes that methanogens use. Some of these GO terms were previously available and the rest were generated in our Microbial Energy Gene Ontology (MENGO) project. A recently discovered non-canonical CH<sub>4</sub> production process is also described. We have performed manual GO annotation of selected methanogenesis genes, based on experimental evidence, providing “gold standards” for machine annotation and automated discovery of methanogenesis genes or systems in diverse genomes. Most of the GO-related information presented in this report is available at the MENGO website (<http://www.mengo.biochem.vt.edu/>).

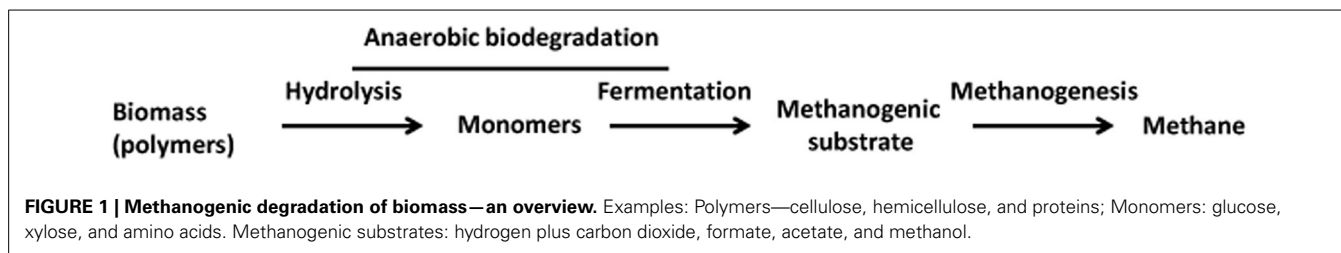
**Keywords: Gene Ontology, biomass, biodegradation, methanogenesis, methanogen, bioenergy, carbon cycle, waste treatment**

## INTRODUCTION

Methane (CH<sub>4</sub>), the simplest aliphatic hydrocarbon, is a valuable fuel. It constitutes 70–95% (volume/volume) of natural gas (Strapoc et al., 2011). The biological production of methane, which occurs under strictly anaerobic conditions, is critical to the operation of the global carbon cycle, nutrient recovery in the digestive systems of numerous animals, and treatment of

municipal and industrial wastes, and it could potentially allow commercial production of methane from renewable resources (Zinder, 1993; Thauer et al., 2008; McInerney et al., 2009). The methane present in geological deposits such as oil and gas reservoirs and coal beds also originated in part from microbial degradation of biomass, and the rest of it was derived from thermal maturation of the remnants from biodegradation (Strapoc et al., 2011). Each of these cases involves anaerobic degradation of biopolymers such as carbohydrates and proteins, as well as lipids, and this process is composed of two broad steps (**Figure 1**): first, generation of substrates for methanogens through a combination of hydrolysis and fermentation; second, methanogenesis or methane production. Methanogenesis is also one of the most ancient respiratory processes on Earth, developing 2.7–3.2 billion

**Abbreviations:** CODH/ACDS, acetyl-CoA decarbonylase/synthase-carbon monoxide dehydrogenase complex; F<sub>420</sub>, coenzyme F<sub>420</sub> or 7,8-didemethyl-8-hydroxy-5-deazaflavin derivative; F<sub>430</sub>, coenzyme F<sub>430</sub> - a tetrapyrrole; GO:MENGO-UR, GO terms generated in the MENGO project, submitted to the GO consortium and awaiting acceptance; HS-CoM, coenzyme M; HS-CoB or HS-HTP, coenzyme B; H<sub>4</sub>MPT, tetrahydromethanopterin; H<sub>4</sub>SPT, tetrahydroscarinapterin; MF, methanofuran.



years ago, and by virtue of the processes described above it continues to be an important process on the present day Earth (Leigh, 2002). Furthermore, biological methanogenesis is a significant contributor to climate change as together with water vapor, carbon dioxide and ozone, methane also contributes to the greenhouse effect (Strapoc et al., 2011). According to United States Environmental Protection Agency (US EPA) “Pound for pound, the comparative impact of CH<sub>4</sub> on climate change is over 20 times greater than CO<sub>2</sub> over a 100-year period” (<http://epa.gov/climatechange/ghgemissions/gases/ch4.html>).

For the ecological, evolutionary, and applied interests discussed in the preceding paragraph, methanogens have been investigated intensely in the past six decades (Wolfe, 1991; Thauer, 1998, 2012). This research has resulted into a detailed understanding of the biochemistry of these archaea, especially their unique energy metabolism, methanogenesis, and the mechanistic details of their interactions with other microorganisms in numerous ecological niches. For the same reasons, genomes of methanogens have been analyzed from the early days of genome sequencing. In fact, *Methanocaldococcus jannaschii*, a methanogen, was the first archaeon and third organism to be targeted for complete genome sequence determination (Bult et al., 1996). Since then the genome sequences of more than 170 methanogens have appeared in public databases. These genomes have not only helped to advance the research on methanogens, but also have catalyzed major shifts in our understanding of the relationships of these organisms with the rest of the biological world. It is now known that many of the biological parts and processes that were once thought to be specific to methanogenic archaea are major contributors to the metabolism of numerous non-methanogenic organisms from all domains of life (Takao et al., 1989; Batschauer, 1993; Purwantini et al., 1997; Graham and White, 2002; Chistoserdova et al., 2004; Krishnakumar et al., 2008). Often such discoveries have been based on the detection of methanogen genes in non-methanogen genomes, followed by biochemical analysis of their molecular functions and knowledge based deductions of their roles in the metabolic pathways in those organisms. In this context a rich set of GO terms fully describing methanogenesis together with manually generated gene annotations based on experimental evidence (gold standards) could bring great strength, as it would provide expanded qualifications of the methanogen genes in a non-methanogen genome, such as predicted functions and cellular locations of the gene products, through automated analysis. This resource will then allow facile mining of useful parts of methanogenesis systems from both methanogens as well as non-methanogenic organisms. The Gene Ontology (GO) consists of three sets of terms for describing

gene products in terms of biological processes (GO:0008150), cellular components (GO:0005575), and molecular functions (GO:0003674) (Ashburner et al., 2000). These terms are related to each other in a semi-hierarchical fashion (a directed graph structure), from very broad terms (at the top of the hierarchy) to specific (at the bottom). GO annotation can thus provide both specific and broader attributes to gene products. This is the primary motivation for the work described in this report.

### GENE ONTOLOGY (GO) DESCRIBING METHANOGENESIS

The promise cited above has inspired the work on the methanogenesis component of our MENO (Microbial Energy production Gene Ontology) project. The goal of the MENO project is to develop a set of GO terms for describing gene products involved in energy-related microbial processes. A major focus is on the microbial biomass degradation for the production of biofuel (fuel from renewable resources) such as methane, alcohols, fatty acid esters, hydrocarbons, and hydrogen. Until now we have generated 667 terms and these are available at the GO website (AMIGO: <http://tinyurl.com/kh7fqne>) as well as at our MENO website (<http://www.mengo.biochem.vt.edu/>). Of these, 563 terms are in the Biological Process ontology, 88 in the Molecular Function ontology and 16 in the Cellular Component ontology. More terms are still under review (GO:MENO-UR) and when the respective GO ID's are assigned, we will post those at the MENO website (<http://www.mengo.biochem.vt.edu/>). We generated these terms in two ways: 1. Our own effort, which involved a review of the relevant literature and creation of terms as the needs were identified. 2. Community input, where MENO terms were generated following suggestions from the members of the bioenergy research community who attended the MENO workshops organized by us at the following locations: Great Lakes Bioenergy Research Center, University of Wisconsin, Madison, WI (2011); Annual User Meeting of the US Department of Energy Joint Genome Institute, Walnut Creek, CA (2011 and 2012); US Department of Energy's Genomic Science Awardee Meeting IX and X (Crystal City, VA, and Bethesda, MD, respectively) (2011 and 2012).

In this report we present GO terms suitable for describing processes, functions and cellular components involved in methanogenic biodegradation of biomass, including methanogenesis, in the context of both natural and engineered processes. We begin this description with a brief review of the relevant systems. More detailed information, especially the mechanistic details of methanogenesis, is available in several reviews including some cited here (Wolfe, 1993; Zinder, 1993; Ferry, 1999; Deppenmeier, 2002; Liu and Whitman, 2008; Thauer et al., 2008;

Thauer, 2012; Costa and Leigh, 2014; Welte and Deppenmeier, 2014). Furthermore, to remain focused on bioconversion or catabolism, the general cellular biosynthesis processes have not been covered in this report; exceptions are the syntheses of coenzymes that were once thought to be unique to methanogens (Wolfe, 1991, 1992; Graham and White, 2002) and afterwards some of these were found to occur in the bacteria (Purwantini et al., 1997; Chistoserdova et al., 2004; Krishnakumar et al., 2008). Recently, a non-canonical route that contributes significantly to global biological production of methane has been described (Metcalf et al., 2012) and we describe this system briefly. In numerous environments, complete anaerobic biodegradation of biomass can occur without the formation of methane and here processes such as sulfate reduction and acetogenesis provide avenues for the disposal of reductants (Isa et al., 1986; Gibson et al., 1988; Widdel, 1988; Zinder, 1993; Breznak, 1994; De Graeve et al., 1994; Raskin et al., 1996; Muller, 2003). Those processes will not be covered here.

The work on the GO for methanogenesis began with a review of the GO database. This showed that, although the GO terms describing many of the biological processes and molecular functions associated with methane biosynthesis were available, the coverage of this area was incomplete. To fill this gap we generated an additional 110 GO terms for methanogenesis. A comprehensive source of this information is on our website (<http://www.mengo.biochem.vt.edu/>) where the data are available under two menus: MENGO (All MENGO Terms; Process Specific; Ontology Specific; New MENGO Terms) and PATHWAYS (Natural Pathways; Synthetic/Engineered Pathways). Under the MENGO menu a form (Submit MENGO Term) is available for the submission of new terms that will help to describe gene products involved in methanogenesis in a comprehensive manner and to validate the resource through research community input.

## METHANOGENIC DEGRADATION OF BIOMASS

As mentioned in the Introduction, this process is composed of two broad steps, anaerobic biodegradation of biomass generating substrates for methanogens, and methanogenesis (Figure 1). The narrative appearing below covers both natural and engineered systems.

### ANAEROBIC BIODEGRADATION OF BIOMASS

#### Natural systems

**Anaerobic biodegradation of biomass in sediments.** Annually, plant and photosynthetic microorganisms fix 70 billion tons of carbon into biomass made up of complex biopolymers, such as cellulose, hemicellulose, lignin, proteins and lipids (Thauer et al., 2008). About 1% of this material is mineralized in various anaerobic niches of nature through a process that yields methane and carbon dioxide as end products (Figure 2). The combination of photosynthesis (GO:0015979) and macromolecule catabolism (GO:0009057) constitutes the biological component of the biogeochemical process of carbon cycling.

Cellulose is a polymer of D-glucose units connected by  $\beta(1\rightarrow4)$  bonds. The anaerobic mineralization of cellulose (synonym of “cellulose catabolic process, anaerobic,” GO:1990488)

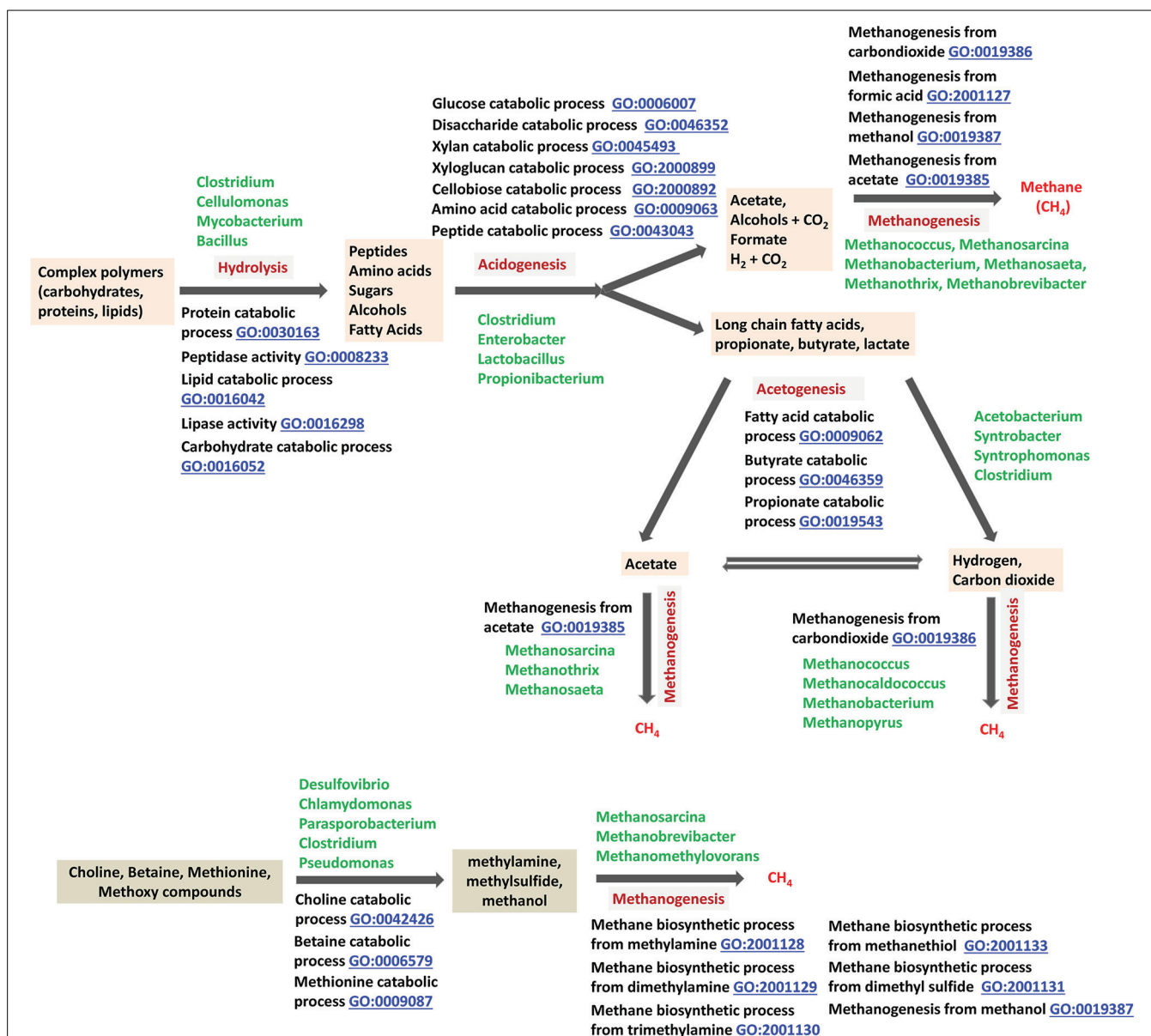
starts with hydrolysis of the  $\beta(1\rightarrow4)$  bonds by cellulases (GO:0008810) produced by anaerobic cellulolytic bacteria and fungi (Adney et al., 1991; Teunissen and Op Den Camp, 1993; Leschine, 1995; Li et al., 1997; Schwarz, 2001; Ljungdahl, 2008; Ransom-Jones et al., 2012). These organisms either secrete the cellulases or carry these enzymes on their cell surfaces (Teunissen and Op Den Camp, 1993; Li et al., 1997). A recent study shows that excreted enzymes with multiple catalytic sites and multiple cellulose-binding modules provide *Caldicellulosiruptor bescii*, an anaerobic thermophile with a high activity of cellulose degradation (Brunecky et al., 2013). The cellulose catabolic process (GO:0030245) involves the actions of endo- $\beta$ -1,4-glucanases (GO:0052859) and exo-1,4- $\beta$ -glucanases or cellobiohydrolases (CBH) (reducing-end-specific, GO:0033945; non-reducing-end-specific, GO:0016162) that generate cellobiose, with intermediate formation of fragments with multiple glucose units (Akin, 1980; Beguin and Aubert, 1994; Bayer et al., 1998; Perez et al., 2002; Hilden and Johansson, 2004), and hydrolysis of cellobiose to glucose (cellobiose catabolic process, GO:2000892) by  $\beta$ -glucosidase (GO:0008422).

The cellulose degrading anaerobic microorganisms and other non-cellulolytic anaerobes with access to the products generated by cellulolytic microbes take up and ferment D-glucose to acetate, alcohols, lactate and fatty acids (e.g., propionate, butyrate) via respective biosynthetic processes (Figure 3) (Zinder, 1993; Schink, 1997; Ahring, 2003; McInerney et al., 2009). The butyrate biosynthetic process (GO:0046358) involves an intermediate formation of acetyl-CoA (acetyl-CoA biosynthetic process, GO:0006085) whereas for propionate biosynthesis (GO:0019542) succinate generated via the tricarboxylic acid metabolic process or TCA cycle (GO:0072350) serves as the direct precursor (Zinder, 1993; Schink, 1997; Ahring, 2003; McInerney et al., 2009).

Butyrate and propionate, which are called short-chain fatty acids (SCFAs), are further fermented to acetate, hydrogen and CO<sub>2</sub> (fatty acid catabolic process: GO:0009062; child term, anaerobic fatty acid catabolic process, GO:1990486) via their respective catabolic processes (butyrate catabolic process, GO:0046359; propionate catabolic process, GO:0019543) (Zinder, 1993; Schink, 1997; Ahring, 2003; McInerney et al., 2009); *Syntrophobacter*, *Syntrophomonas*, *Syntrophus*, *Smithella*, and *Pelotomaculum* species are some of the bacteria that produce these SCFAs. Ethanol and lactate are also fermented to acetate, hydrogen and CO<sub>2</sub> (ethanol catabolic process, GO:0006068; anaerobic lactic acid catabolic, process GO:1990485). The hydrogen biosynthetic process (GO:1902422) is a key element of these fermentation processes and those described in the preceding paragraph for the following reason. Several steps of fermentation lead to the reduction of electron carriers such as NAD<sup>+</sup> and ferredoxin, producing NADH and reduced-ferredoxin. For the fermentation process to continue, NAD<sup>+</sup> and ferredoxin must be regenerated, and often the only available route to meet this requirement is the reduction of protons, yielding molecular hydrogen (H<sub>2</sub>) (GO:1902422) (Zinder, 1993; Schink, 1997; Ahring, 2003; McInerney et al., 2009).

Degradation of hemicellulose follows a path similar to that described for cellulose (summarized in Figure 3). The term hemicellulose includes xylan (polymer of xylose), glucuronoxylan



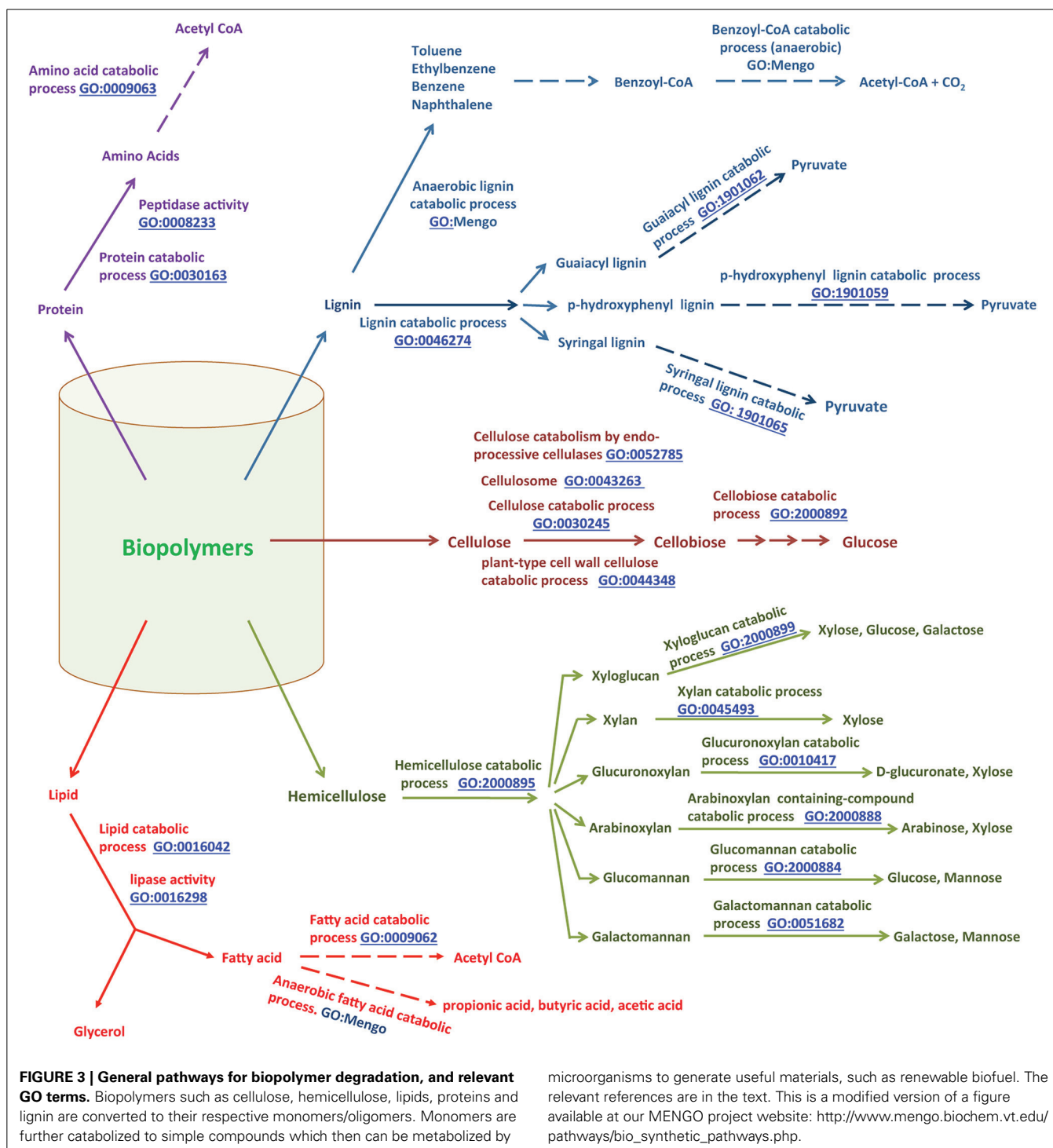


**FIGURE 2 | Anaerobic digestion of complex biopolymers to methane, and relevant GO terms.** The degradation of large polymers found in biomass starts with their hydrolytic fragmentation to monomers by biological hydrolysis (Zinder, 1993); in industry, chemical processes are often used for this deconstruction step (Blanch et al., 2011). Monomers are degraded further to methanogenic substrates first by acidogenic and then by acetogenic

microorganisms (Zinder, 1993). Finally, methanogenic archaea transform these methanogenic substrates such as formate, hydrogen, carbon dioxide, acetate, methylamines and methanol to methane (Zinder, 1993). Methylamines originate from choline and betaine by the actions of choline/betaine degrading microorganisms. Microbial degradation of pectin is a common source of methanol in nature (Schink et al., 1981).

(polymer of D-glucuronate and xylose), arabinoxylan (polymer of arabinose and xylose), glucomannan (polymer of glucose and mannose), galactomannan (polymer of galactose and mannose), and xyloglucan (polymer of xylose, glucose and galactose) (Akin, 1980; Perez et al., 2002). These are degraded via specific hemi-cellulose catabolic processes (GO:2000895) to their respective monomers (Akin, 1980; Perez et al., 2002). The fermentation of monomers yields acetate, hydrogen and CO<sub>2</sub> (Wolin and Miller, 1983; Schink, 1997).

Lignin degradation in anaerobic environments (anaerobic lignin catabolic process, GO:1990487) is not well studied and is considered rare to impossible (Akin, 1980; Harwood and Gibson, 1988; Perez et al., 2002; Fuchs, 2008); the broader lignin catabolic process is generally considered an aerobic process (Perez et al., 2002). However, following the degradation of lignin by aerobic microorganisms such as fungi, a variety of aromatic compounds (catechol, benzoate, p-hydroxybenzoate, vanillate-, ferulate, syringate, p-hydroxybenzoate, p-hydroxycinnamate, and



3-methoxy-4-hydroxyphenylpyruvate) (Kaiser and Hanselmann, 1982a,b) become available in anaerobic environments. Fermentation of these aromatic compounds by anaerobic bacteria leads to acetate, CO<sub>2</sub> and hydrogen (Harwood and Parales, 1996; Fuchs, 2008). Anaerobic degradation of benzoate, one of the lignin monomers, has been studied in detail and this catabolic process (benzoate catabolic process via CoA ligation, [GO:0010128](#)) yields acetate and CO<sub>2</sub> (Figure 3). The metabolism

of several other lignin monomers by anaerobes has also been investigated (Harwood and Parales, 1996; Fuchs, 2008) and some of the relevant information for vanillin, ferulate and catechol is summarized in Figure 3.

Anaerobic lipid catabolic processes also lead to acetate and hydrogen (McInerney, 1988; Zinder, 1993; Schink, 1997). The process begins with the hydrolysis of lipids (lipase activity, [GO:0016298](#)); the broader lipid catabolic process is represented

by GO:0016042. The glycerol released by hydrolysis enters the glycolysis pathway generating acetate and hydrogen (Zinder, 1993) (**Figure 3**). The fatty acid units are degraded via the  $\beta$ -oxidation pathway (fatty acid beta-oxidation, GO:0006635) to acetate and the excess reducing equivalents are released as hydrogen (**Figure 3**). In the case of proteins, the amino acids released by the action of proteases or peptidases (peptidase activity, GO:0008233; endopeptidase activity GO:0004175, exopeptidase activity, GO:0008238) are deaminated oxidatively, releasing ammonia and hydrogen, and then the resulting ketoacids are fermented to acetate and hydrogen (McInerney, 1988; Zinder, 1993; Schink, 1997) (**Figure 3**).

In each of the above cases, as  $H_2$  accumulates the oxidation of reduced electron carriers becomes thermodynamically unfavorable and consequently the fermentation process slows down or even halts (McInerney, 1988; Zinder, 1993; Schink, 1997). Methanogens consume hydrogen and reduce  $CO_2$  to methane, thus relieving the block on fermentation (McInerney, 1988; Zinder, 1993; Schink, 1997). These archaea also convert acetate to methane and  $CO_2$  and this action also improves the thermodynamics of biodegradation (Zinder, 1993). As  $CH_4$  moves to aerobic zones, such as the surface of water overlaying sediments, methanotrophic bacteria oxidize this hydrocarbon to  $CO_2$  (methane catabolic process, GO:0046188) (Kiene, 1991; Zinder, 1993; Conrad, 2007, 2009). More recent work shows that significant amount of methane is oxidized anaerobically and the microbial basis and mechanistic details of this process are beginning to emerge (Conrad, 2009; Thauer, 2011; Milucka et al., 2012; Shima et al., 2012; Haroon et al., 2013; Offre et al., 2013). Hence, by removing the hydrogen-induced thermodynamic block and converting acetate to methane, methanogens facilitate the complete degradation of the biopolymers discussed above.

In marine anaerobic sediments rich in sulfates some of the products of biomass degradation also lead to methane production. In general however, in this environment hydrogen and acetate are not available to the methanogens as, in the presence of sulfate, sulfate-reducing bacteria readily use these materials to reduce sulfate to hydrogen sulfide (dissimilatory sulfate reduction, GO:0019420), and the growth rates and affinities for  $H_2$  of the sulfate-reducing bacteria are much higher than those of the methanogens (Widdel, 1988). However, several hydrogen-consuming methanogens belonging to the class of Methanococci have been isolated from marine environments (Whitman et al., 1986). It has been speculated that these organisms may depend primarily on formate which could arise from the catabolism of oxalate (GO:0033611) derived from plant materials (Allison et al., 1985); most methanococci are capable of consuming both hydrogen and formate (Boone et al., 1993).

A significant amount of methane is also produced from methylamines, methylsulfides and carbon monoxide (Zinder, 1993; Thauer, 1998; Deppenmeier et al., 1999; Ferry, 1999, 2011; Liu and Whitman, 2008; Thauer et al., 2008). The sources of methylamines are betaine and choline, (GO:0006579 and GO:0042426, respectively) while methylsulfides are generated from sulfur-containing compounds such as methionine and dimethylpropiothetin (GO:0009087; and GO:0047869,

respectively; **Figure 3**) (Boone et al., 1993). In certain marine environments, carbon monoxide provided by kelp algae provides both reductant and carbon for methanogenesis (methane biosynthetic process from carbon monoxide, GO:2001134) (Rother and Metcalf, 2004; Lessner et al., 2006).

#### **Anaerobic biodegradation of biomass in animal intestines.**

Foregut fermenting animals such as the ruminants (cattle, sheep, goats) as well as hindgut fermenters such as human, termites, and horse, employ variations of the overall process shown in **Figure 2** for deriving nutrients from feed or food (Wolin, 1981; Wolin and Miller, 1983; Zinder, 1993; Miller and Wolin, 1996; Weimer, 1998; Hook et al., 2010; Sahakian et al., 2010). In cattle and many other foregut fermenters, the rumen serves as the first site for the degradation of forage (Wolin, 1981; Weimer, 1998; Hook et al., 2010). The residence time for the feed in rumen is rather short (5.6 h for the fluid and 35 h from particulates in rumen; compared to about 4.5 months even for nitrate, a soluble compound, in freshwater sediment) (Hristov et al., 2003), which is not conducive for significant growth and activity of slow-growing fatty acid-fermenting bacteria and acetoclastic methanogens (Zinder, 1993). Thus, in this digestive chamber the fatty acids and acetate are not converted to methane, rather are absorbed by the animal for nutrition (Zinder, 1993). The hydrogen and formate produced during the fermentation are converted to methane by methanogenic archaea. All plant material contains pectin, a methylated carbohydrate, and leaves, shoots and fruit are particularly rich in it. Anaerobic degradation of pectin (anaerobic pectin catabolic process, GO:1990489) serves as an important source of methanol in anaerobic environments (Schink et al., 1981). Thus, ruminants could carry methanogens in their rumens capable of utilizing methanol for methanogenesis and in some cases this has been shown to be true (Mukhopadhyay et al., 1992; Zinder, 1993).

In the hindgut of humans, i.e., the large intestine, the undigested material delivered from the small intestine is fermented, generating fatty acids, some hydrogen, and formate, and the latter two are converted to methane (Wolin, 1981; Zinder, 1993; Miller and Wolin, 1996; Sahakian et al., 2010; Flint, 2011). The process is beneficial to the host as it provides the fatty acids as additional nutrients. However, an uncontrolled production of fatty acids in this hindgut activity has been identified as one of the possible causes of obesity (Schmitz and Langmann, 2006; Nakamura et al., 2010; Sahakian et al., 2010; Flint, 2011).

In certain foregut fermenters such as kangaroos and wallabies and hindgut fermenters such as termites, the removal of hydrogen during biodegradation of complex polymers occurs through acetate formation and not methanogenesis (Brune and Friedrich, 2000; Gagen et al., 2010; Klieve et al., 2012).

#### **Anaerobic biomass degradation in engineered systems: waste treatment and methane production from renewable resources**

Aerobic treatment of municipal and industrial wastes via methods such as activated sludge requires energy input for supplying oxygen (Switzenbaum, 1983; Zinder, 1993). The process also generates a significant amount of microbial biomass (Zinder, 1993), which cannot be discharged to waterways (Zinder, 1993;



Paul and Debellemfontaine, 2007). In contrast, anaerobic methods not only require much less energy input and produce very little microbial biomass, but also conserve most of the energy present in the waste materials in the form of methane, which can be recovered as fuel (Zinder, 1993; Gao et al., 2014). Anaerobic waste treatment and the production of methane as bio-fuel from renewable resources follow the basic biological process (macromolecules catabolic process, GO:0009057) that has been described above for methanogenic biodegradation of biomass in sediments (Zinder, 1993). The mixture of methane and carbon dioxide that is produced in all these cases is known as biogas (Ducom et al., 2009). The biogas obtained from waste treatment facilities as well as from bio-digesters processing high sulfur feedstock contains a substantial amount of hydrogen sulfide and nitrogen oxides (Zinder, 1993; Fdz-Polanco et al., 2001; Janssen et al., 2001; Ducom et al., 2009; Diaz and Fdz-Polanco, 2012). Several methods for the removal of these unwanted compounds have been developed and research for developing even better separation methods continues (Fdz-Polanco et al., 2001; Janssen et al., 2001; Ducom et al., 2009; Diaz and Fdz-Polanco, 2012).

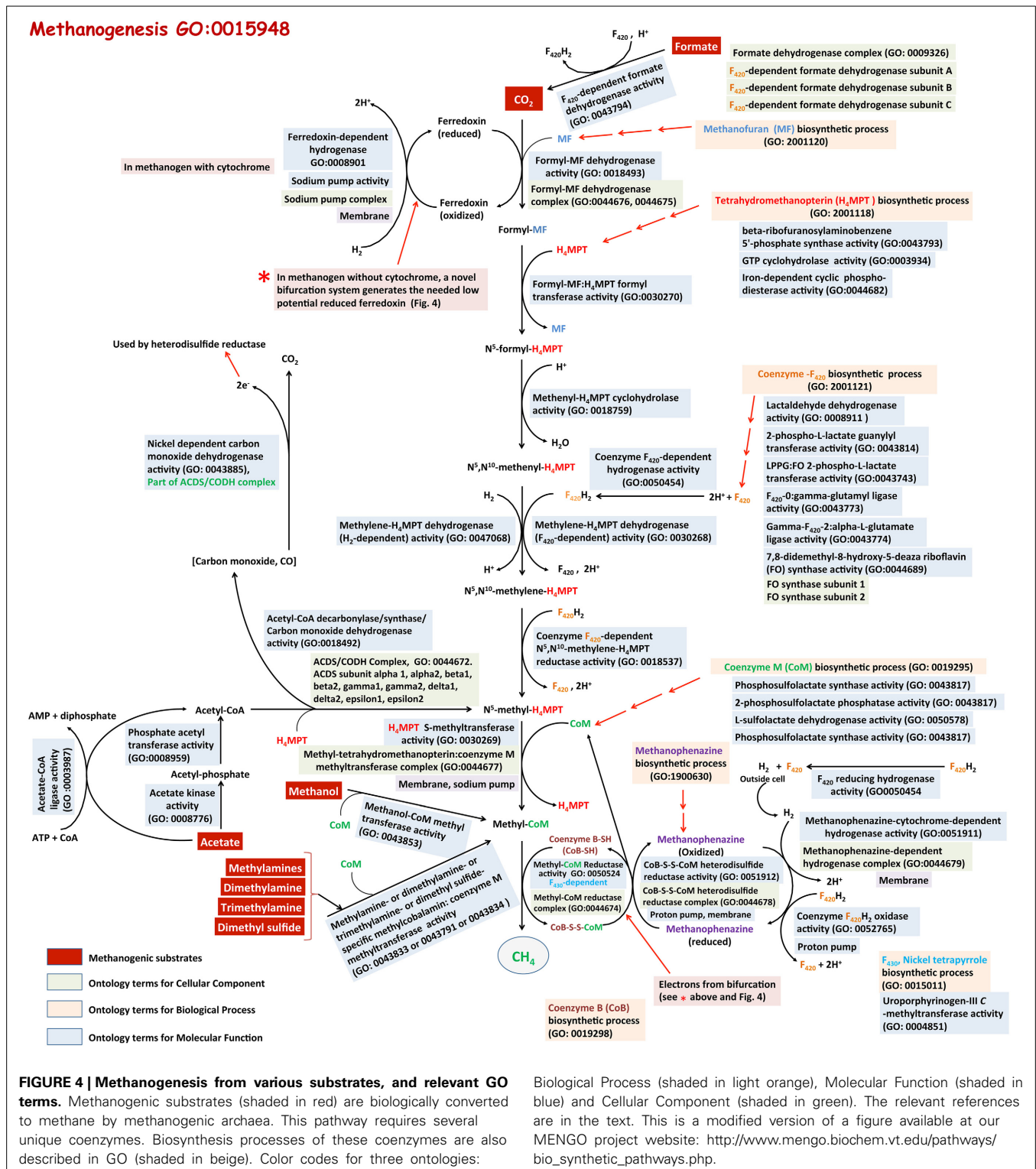
## METHANOGENESIS

The pathways for methanogenesis or methane biosynthetic process (GO: 0015948) from various substrates and the respective molecular functions are shown in **Figure 4**. Here, the steps leading from CO<sub>2</sub> to CH<sub>4</sub> form the core, which is used in part or its entirety with other substrates as well (Wolfe, 1991, 1993; Ferry, 1993, 1999, 2011; Thauer et al., 1993, 2008; Deppenmeier et al., 1999; Deppenmeier, 2002; Deppenmeier and Muller, 2008; Liu and Whitman, 2008). These pathways utilize several unusual coenzymes of which methanofuran (MF), tetrahydromethanopterin (H<sub>4</sub>MPT), tetrahydrosarcinapterin (H<sub>4</sub>SPT), and coenzyme M (or HS-CoM) carry the carbon moiety destined to generate methane, while coenzyme F<sub>420</sub> (a deazaflavin derivative), coenzyme B (HS-CoB or HS-HTP), methanophenazine, and coenzyme F<sub>430</sub> (a tetrapyrrole) transfer electrons that are used in carbon reduction (Wolfe, 1991, 1993; Ferry, 1999; Deppenmeier and Muller, 2008; Thauer et al., 2008). Many unique enzymes and unusual mechanisms are also involved (Wolfe, 1991, 1993; Ferry, 1999; Deppenmeier and Muller, 2008; Thauer et al., 2008). In the following narrative the term H<sub>4</sub>MPT represents both H<sub>4</sub>MPT and H<sub>4</sub>SPT, which serve the same function in different organisms.

### Methanogenesis from H<sub>2</sub> plus CO<sub>2</sub>

This process (GO:0019386) utilizes hydrogen as the primary source of electron or reductant (Ferry, 1999; Deppenmeier and Muller, 2008; Thauer et al., 2008). It can operate with or without the involvement of cytochromes (Ferry, 1999; Deppenmeier and Muller, 2008; Thauer et al., 2008). The latter is utilized by methanogens that lack cytochromes (Ferry, 1999; Deppenmeier and Muller, 2008; Thauer et al., 2008) and is considered one of the most ancient respiratory metabolisms on earth (Leigh, 2002). We describe the process starting with the carbon transfer and reduction steps, followed by the energy production avenues.

**Carbon transfer and reduction.** It is believed that carbon dioxide (CO<sub>2</sub>) is captured by methanofuran (MF) to form an unstable compound called carboxy-MF (Thauer et al., 1993) which is reduced by formyl-MF dehydrogenase (GO:0018493) in an energy-dependent (endergonic) manner to formyl-MF with a low-potential ferredoxin (Fd) serving as electron carrier (Thauer et al., 2008). Formyl-MF dehydrogenase exists in two forms, one of which contains molybdenum (Fmd) and the other tungsten (Fwd) (Thauer et al., 1993); molybdenum and tungsten are found to be bound to a molybdopterin and growth conditions dictate which metal will be incorporated (Hochheimer et al., 1995). At the next step the formyl group is transferred to H<sub>4</sub>MPT by a transferase enzyme (Ftr, GO:0030270) to form formyl-H<sub>4</sub>MPT (Donnelly and Wolfe, 1986; Breitung and Thauer, 1990; Thauer et al., 1993). From this stage H<sub>4</sub>MPT carries four forms of the fixed carbon representing three oxidation states (Wolfe, 1991, 1993; Thauer et al., 1993, 2008; Ferry, 1999; Deppenmeier and Muller, 2008). First formyl-H<sub>4</sub>MPT is dehydrated by methenyl-H<sub>4</sub>MPT cyclohydrolase (Mch, GO:0018759) to form methenyl-H<sub>4</sub>MPT (Donnelly et al., 1985; Dimarco et al., 1986; Mukhopadhyay and Daniels, 1989; Klein et al., 1993), which in turn is reduced to methylene-H<sub>4</sub>MPT by the action of one of the two enzymes, F<sub>420</sub>-dependent methylene-H<sub>4</sub>MPT dehydrogenase (Mtd, GO:0030268) and a Fe-containing hydrogenase (Hmd, GO:0047068) (Hartzell et al., 1985; Mukhopadhyay and Daniels, 1989; Von Bunau et al., 1991; Schworer et al., 1993; Thauer et al., 1993; Mukhopadhyay et al., 1995). Mtd utilizes reduced F<sub>420</sub> (F<sub>420</sub>H<sub>2</sub>) as reductant whereas Hmd retrieves electrons from molecular hydrogen (H<sub>2</sub>) (Hartzell et al., 1985; Mukhopadhyay and Daniels, 1989; Von Bunau et al., 1991; Schworer et al., 1993; Thauer et al., 1993; Mukhopadhyay et al., 1995). Methanogens with Hmd also carry paralogs of this protein (HmdII and HmdIII), but these proteins do not reduce methylene-H<sub>4</sub>MPT (Lie et al., 2013). Two roles of HmdII and HmdIII have been proposed: a. guiding the maturation of Hmd and b. linking energy production and protein synthesis (Oza et al., 2012; Lie et al., 2013). Methylene-H<sub>4</sub>MPT is reduced with F<sub>420</sub>H<sub>2</sub> and by the action of F<sub>420</sub>-dependent methylene-H<sub>4</sub>MPT reductase (Mer, GO:0018537), providing the last H<sub>4</sub>MPT derivative on the pathway, methyl-H<sub>4</sub>MPT (Ma and Thauer, 1990; Te Brommelstroet et al., 1990; Ma et al., 1991; Thauer et al., 2008). The transfer of the methyl group from methyl-H<sub>4</sub>MPT to coenzyme M is catalyzed by a membrane-bound sodium ion (Na<sup>+</sup>)-pumping enzyme complex called methyl-H<sub>4</sub>MPT:coenzyme M methyl transferase (Mtr, GO:0044677) (Becher et al., 1992; Kengen et al., 1992; Gartner et al., 1993). This complex not only yields methyl-coenzyme M (CH<sub>3</sub>-CoM), but also generates a Na<sup>+</sup>-gradient that is used for energy production (see below) (Ferry, 1999; Deppenmeier and Muller, 2008; Thauer et al., 2008). The next step in the sequence yields methane. This last carbon-reduction reaction is catalyzed by CH<sub>3</sub>-CoM reductase (GO:0044674) with coenzyme B (HS-CoB or HS-HTP) serving as an electron source, resulting in a heterodisulfide, CoM-S-S-CoB, as product in addition to methane (Wolfe, 1991, 1992; Ferry, 1999; Deppenmeier and Muller, 2008; Thauer et al., 2008). The heterodisulfide is reduced by a reductase (Hdr, GO:0051912) to regenerate HS-CoM and HS-CoB (Ferry, 1999; Deppenmeier and Muller,



2008; Thauer et al., 2008). Hydrogen-oxidizing methanogens often carry two  $\text{CH}_3\text{-CoM}$  reductase isozymes (McrI and McrII) (Rospert et al., 1990), one of which is effective under high hydrogen availability and the other under low hydrogen conditions (Rospert et al., 1990).

**Energy conservation.** First, we describe the details for methanogens lacking cytochromes. The first site of energy conservation is the Mtr reaction (Ferry, 1999; Deppenmeier and Muller, 2008; Thauer et al., 2008). The  $\text{Na}^+$ -gradient generated at this step is directly used for the production of



ATP by a membrane-bound AoA1-ATP synthase (GO:1990490) (Deppenmeier and Muller, 2008). Under certain conditions this gradient assists two membrane-associated and energy-converting hydrogenase complexes, EhaA-T and EhbA-Q, to generate reduced Fd with the ability to deliver low redox potential electrons (Thauer et al., 2008; Costa et al., 2010; Lie et al., 2012). The reduced Fd molecules generated by EhaA-T are used for the endergonic formyl-MF dehydrogenase reaction that yields formyl-MF, and those provided by EhbA-Q are used for cellular biosynthesis (Porat et al., 2006; Thauer et al., 2008; Costa et al., 2010; Major et al., 2010; Kaster et al., 2011; Lie et al., 2012). The next energy yielding step is the reduction of CH<sub>3</sub>-CoM and it is not clear whether the methanogens conserve this energy or the energy is released to strongly favor the forward reaction toward methane formation (Thauer et al., 2008). The reduction of CoM-S-S-CoB involves rather complex electron transfer mechanisms and also is a major site for energy conservation (Thauer et al., 2008; Costa and Leigh, 2014).

In certain methanogens without cytochromes, the reduction of CoM-S-S-CoB and formyl-MF generation is coupled via a novel mechanism called bifurcation (Thauer et al., 2008; Costa et al., 2010; Kaster et al., 2011; Lie et al., 2012) (**Figure 5**). Here, the Vhu hydrogenase retrieves electrons from hydrogen and transfers those to soluble heterodisulfide reductase (Hdr). Hdr utilizes these electrons for two purposes (**Figure 5**): (i) converting CoM-S-S-CoB to HS-CoM and HS-CoB, which requires a relatively lower investment of energy; and (ii) reducing a low potential ferredoxin, which is energetically suitable for the highly energy intensive reduction of CO<sub>2</sub> and generation of formyl-MF (Thauer et al., 2008; Costa et al., 2010; Kaster et al., 2011; Lie et al., 2012). This novel mechanism, where a single input (electrons of moderately low potential) is used to generate two outputs (two pools of electrons, with potentials that are higher and much lower than the input) is called electron bifurcation (GO:MENGO-UR). It is a major factor in energy conservation in methanogens as it helps to perform a highly endergonic reaction, such as the generation of formyl-MF, without an investment of ATP or an ion gradient (Thauer et al., 2008; Costa et al., 2010; Kaster et al., 2011; Lie et al., 2012). It seems to be an important tool for energy poor anaerobes (Thauer et al., 2008; Kaster et al., 2011). When withdrawal of intermediates from the methanogenesis pathway for biosynthesis causes a drop in CoM-S-S-CoB levels, the bifurcation process is rendered less efficient (Lie et al., 2012); in that case, as described in the preceding paragraph, an ion-driven hydrogenase system (EhaA-T) is employed for the generation of formyl-MF and this could be considered to be a type of anaplerosis (GO:MENGO-UR) (Lie et al., 2012).

Methanogens with cytochromes do not employ the bifurcation mechanism. Instead, a membrane-bound complex composed of a cytochrome-containing heterodisulfide reductase (HdrDE) (GO:0044678) and a hydrogenase (VhoECG) where VhoC is a b-type cytochrome is utilized (Thauer et al., 2008). The electrons derived from hydrogen by VhoECG are utilized by HdrDE for reduction of CoM-S-S-CoB (Thauer et al., 2008). The overall process is exergonic and thus, in addition to reducing CoM-S-S-CoB, the VhoECG-Hdr complex utilizes excess energy to extrude protons out of the cell. The low potential reduced ferredoxin,

which is needed for the generation of formyl-MF, is provided by a proton-gradient-driven membrane-bound energy-converting hydrogenase complex (EchA-F).

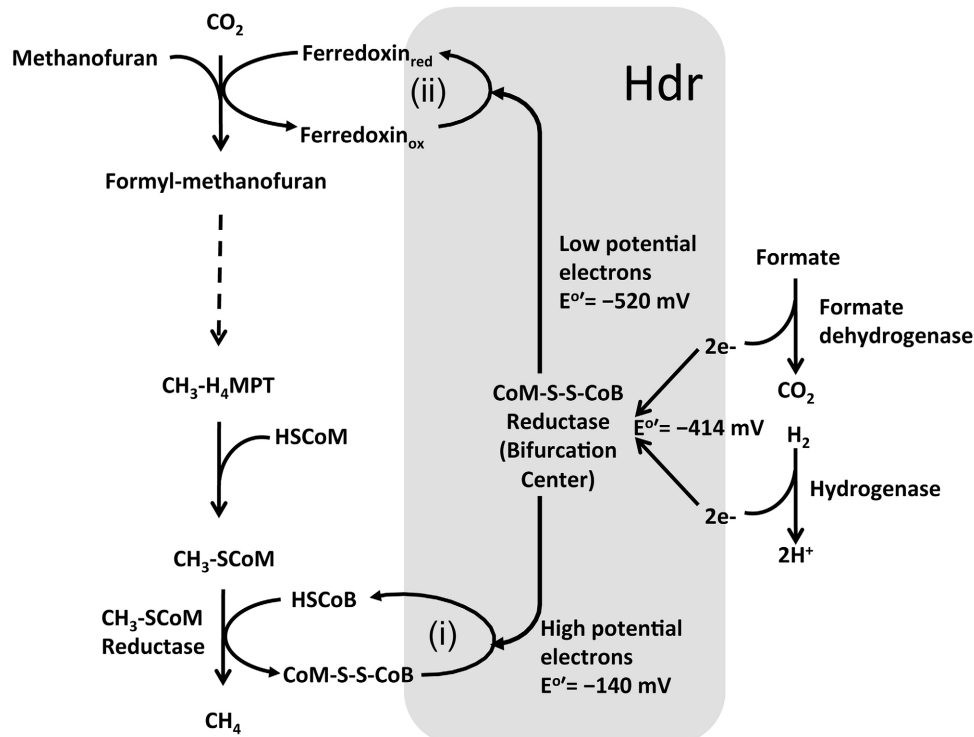
As mentioned above, several methanogenesis enzymes form large protein complexes (GO:0043234) and some these are membrane bound (GO:0019898) and include specialized non-enzyme units such as ion pumps and lipid soluble small compounds. One example is the soluble heterodisulfide reductase complexes of methanogens that lack cytochromes, and can be described using a molecular function term, GO:0051912 ("CoB-CoM heterodisulfide reductase activity") and two cellular component terms GO:0044678 ("CoB-CoM heterodisulfide reductase complex") and GO:0043234 ("protein complex"). In the case of cytochrome-containing methanogens, one further additional cellular component GO term is available for a full description, namely GO:0019898 ("extrinsic component of membrane").

### ***Methanogenesis from formate***

The carbon transfer and reduction steps in this process (GO:2001127) are similar to those described above for methanogenesis from H<sub>2</sub>+CO<sub>2</sub> (GO: 0015948). Both the CO<sub>2</sub> and reducing power are derived from formate by the action of an F<sub>420</sub>-dependent formate dehydrogenase (FdhABC) (GO:0043794) (Schauer and Ferry, 1982; Lie et al., 2012); FdhAB subunits form the enzyme that produces CO<sub>2</sub> and reduced F<sub>420</sub> or F<sub>420</sub>H<sub>2</sub> (HCOO<sup>-</sup> + H<sup>+</sup> + F<sub>420</sub> → CO<sub>2</sub> + F<sub>420</sub>H<sub>2</sub>) and FdhC is thought to import formate into the cell (Wood et al., 2003). CO<sub>2</sub> is converted to methane using the CO<sub>2</sub>-reduction pathway described in **Figure 4**. Some of the reduced F<sub>420</sub> (F<sub>420</sub>H<sub>2</sub>) participates directly in the Mtd and Mer reactions and a part of it is used by a bifurcating complex that provides electrons of appropriate redox potentials to heterodisulfide reductase and formyl-MF dehydrogenase. In the composition and some of the properties this bifurcating complex differ from the one employed for methanogenesis from H<sub>2</sub>+CO<sub>2</sub> (see above). When a methanogen grows on formate, a part of the Fdh pool associates with the Hdr and Vhu/Vhc hydrogenases, and together with a formyl-MF dehydrogenase they form a bifurcating complex (Lie et al., 2012). This Fdh-containing bifurcating complex utilizes electrons from F<sub>420</sub>H<sub>2</sub> (produced by Fdh) and generates high and low potential electrons, either directly or via production of hydrogen as intermediate, that are consumed in the reduction of CoM-S-S-CoB and the generation of formyl-MF, respectively (Lie et al., 2012).

### ***Methanogenesis from ethanol or secondary alcohols plus carbon dioxide***

Only a few methanogens can perform methanogenesis with secondary alcohol as electron source (GO:MENGO-UR; also secondary alcohol catabolic process, GO:MENGO-UR) (Widdel, 1986; Bleicher et al., 1989; Widdel and Wolfe, 1989; Schirmack et al., 2014). These substrates are oxidized to their respective ketones to provide reducing equivalents for the reduction of carbon dioxide to methane via the pathway shown in **Figure 3** (Boone et al., 1993; Zinder, 1993). Ethanol, when used, is converted to acetaldehyde (methanogenesis with ethanol as electron source, GO:MENGO-UR; ethanol catabolic process,



**FIGURE 5 | Electron bifurcation in Methanogenesis.** HS-CoB or HS-HTP, coenzyme B; HS-CoM, coenzyme M; CoM-S-S-CoB, heterodisulfide of coenzyme M and coenzyme B. Heterodisulfide reductase (Hdr) utilizes bifurcated energy electrons for two purposes: (i) converting CoM-S-S-CoB to

HS-CoM and HS-CoB, using high potential electrons; and (ii) reducing a low potential ferredoxin using low potential electrons, which is energetically suitable for the highly endergonic reduction of CO<sub>2</sub> and generation of formyl-MF (Thauer et al., 2008; Thauer, 2012; Costa and Leigh, 2014).

GO:0006068). These conversions are consistent with the general observation that methanogens cannot break carbon-carbon bonds in energy substrates other than that found in acetate (see below). Two types of alcohol dehydrogenase have been found in these organisms: one reduces nicotinamides ( $\text{NAD}^+$  or  $\text{NADP}^+$ ) (GO:0004022 and GO:0008106), and the other transfers electrons to coenzyme  $\text{F}_{420}$  during alcohol oxidation (GO:0052753) (Widdel and Wolfe, 1989). Most of these enzymes have broad specificities allowing the organisms to use ethanol, 2-propanol, 2-butanol, 2-pentanol, cyclopentanol, cyclohexanol, and 2,3-butanediol (Bleicher et al., 1989).

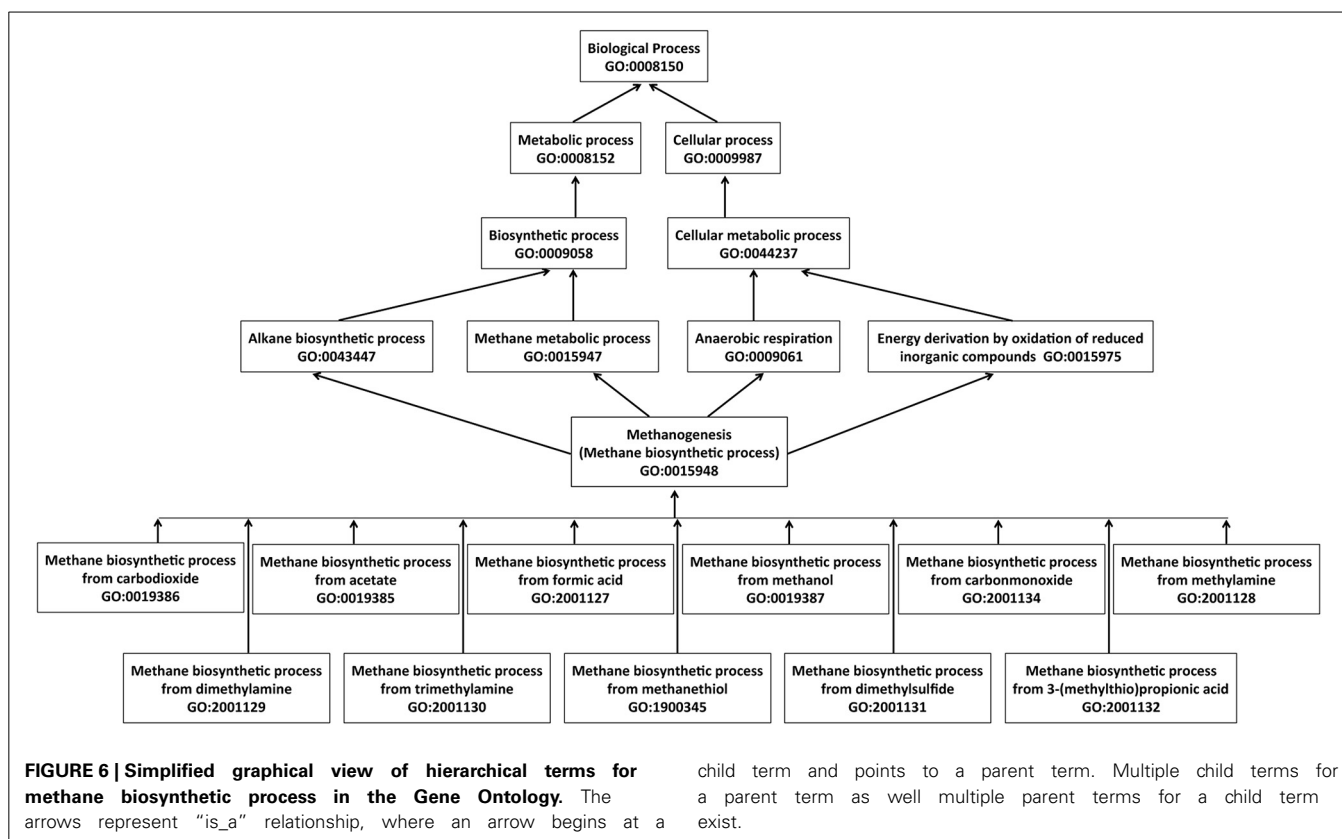
### ***Methanogenesis from carbon monoxide***

Many methanogens can utilize carbon monoxide (CO) although higher levels of this gas inhibit growth of these archaea (Daniels et al., 1977; O'Brien et al., 1984; Rother and Metcalf, 2004; Lessner et al., 2006). Three routes of CO utilization have been found in these organisms. In one, called methanogenesis from carbon monoxide (GO:2001134), CO is simply oxidized to CO<sub>2</sub> by carbon monoxide dehydrogenase (CODH) (GO:0008805), and the resulting two electrons are used for either hydrogen production (GO:1902422) or ferredoxin reduction (Daniels et al., 1977; Ferry, 1999; Vepachedu and Ferry, 2012). Then the hydrogen and/or reduced ferredoxin are used for methanogenesis from CO<sub>2</sub> (GO:0019386). Overall, for every four moles of CO oxidized, one mole of methane and 3 moles of CO<sub>2</sub> are produced. The

second mode of CO utilization has been found in *Methanosarcina acetivorans* where methanogenesis (GO: 0015948) is inhibited by CO but growth is not (Rother and Metcalf, 2004). This organism uses two non-methanogenic routes for energy production (Rother and Metcalf, 2004), the primary one being acetogenic (acetate biosynthetic process, GO:0019413) and the secondary one being formate-forming (formate biosynthetic process, GO:0015943). Even under these conditions methanogenesis operates at low rates, primarily to provide cellular biosynthetic precursors (Rother and Metcalf, 2004). Here methanogenesis from CO<sub>2</sub> involves novel enzymes that transfer the methyl group of CH<sub>3</sub>-H<sub>4</sub>MPT to CH<sub>3</sub>-CoM and serve in the accompanying energy conservation (Lessner et al., 2006); the methyl transfer step could involve a cytoplasmic methyltransferase (CmtA) in addition to a membrane-bound methyl-H<sub>4</sub>MPT:coenzyme M methyl transferase (Mtr) (Vepachedu and Ferry, 2012). In the third route, CO promotes the production of dimethyl sulfide and methanethiol ( $3\text{CO} + \text{H}_2\text{S} + \text{H}_2\text{O} \rightarrow \text{CH}_3\text{SH} + 2\text{CO}_2$ ) and energy is conserved via a yet to be identified system (Moran et al., 2008).

### ***Methanogenesis from methanol, methylamines and methanethiols***

Methanogenesis from all of these substrates involves the formation of methyl-CoM as an intermediate (Ferry, 1999). When methanol serves as the sole substrate for methanogenesis (GO:0019387), it provides both carbon and reductant



for methanogenesis and this process consumes four moles of methanol for every three moles methane generated. Of these, one mole of methanol is oxidized to  $\text{CO}_2$ , generating six-electron equivalents of reductant, which are then used to convert three moles of methanol to three moles of methane (Keltjens and Vogels, 1993). The oxidation of methanol to  $\text{CO}_2$  involves a part of the  $\text{CO}_2$  reduction pathway, but in the reverse direction (Figure 3) (Wassenaar et al., 1998). The methyl groups enter this oxidation process at the methyl-coenzyme M stage by the action of two methyl transferases, MT1 and MT2 or MT2-M (Van Der Meijden et al., 1984a,b; Keltjens and Vogels, 1993; Wassenaar et al., 1998; Ferry, 1999). MT1 is a two-subunit enzyme (MtaBC) and MT2-M has one subunit (MtaA). The first reaction involves transfer of the methyl group of methanol by MT1 to the corrinoid co-factor of its MtaC subunit; this is an automethylation process (Van Der Meijden et al., 1984a,b; Wassenaar et al., 1998). Then MT2-M or MtaA transfers the methyl group from MtaC to HS-CoM, generating methyl-coenzyme M (Van Der Meijden et al., 1984a,b; Wassenaar et al., 1998). Existence of isozymes of MT1 catering to the growth on methanol under various conditions has also been reported (Bose et al., 2006). The methyl groups destined for oxidation are transferred from  $\text{CH}_3\text{-CoM}$  to  $\text{H}_4\text{MPT}$  by the membrane-bound methyl- $\text{H}_4\text{MPT}$ :coenzyme M methyl transferase (Mtr) (Fischer et al., 1992; Sauer et al., 1997; Ferry, 1999). This endergonic reaction is assisted by a  $\text{Na}^+$ -gradient and generates  $\text{CH}_3\text{-H}_4\text{MPT}$  (Deppenmeier et al., 1999; Ferry, 1999; Deppenmeier and Muller, 2008). The steps from  $\text{CH}_3\text{-H}_4\text{MPT}$  to  $\text{CO}_2$  are a reversal of those used for  $\text{CO}_2$ -reduction,

except the organisms performing this process lack Hmd and  $\text{F}_{420}$ -dependent Mtd performs the oxidation of methylene- $\text{H}_4\text{MPT}$  to methenyl- $\text{H}_4\text{MPT}$  (Thauer et al., 1993; Deppenmeier et al., 1999; Ferry, 1999; Deppenmeier and Muller, 2008).

Utilization of mono-, di- and tri-methylamines (MMA, DMA, and TMA) for methanogenesis (GO:2001128, GO:2001129, GO:2001130 respectively) follows the general process that is described above for methanol except that substrate-specific methyl transferases are involved in the transfer of methyl groups to coenzyme M. Using MT1 and MT2 of the methanol systems as the reference the methylamine-specific methyl transferases have been named as follows (Wassenaar et al., 1996, 1998; Ferguson and Krzycki, 1997; Burke et al., 1998; Ferry, 1999; Ferguson et al., 2000; Paul et al., 2000; Bose et al., 2008): for MMA, MMAMT+MMCP (MT1) and MT2-A (MT2); for DMA, DMA-MT (MT1) and MT2-A (MT2); for TMA, TMA-52+TCP (MT1) and MT2-A (MT2). For methanogenesis from TMA, MT2-A could be substituted by MT2-M (Ferry, 1999). Methanogenesis from methylated thiols (methanethiol, dimethylsulfide, or methylmercaptopropionate; GO:2001133, GO:2001131, and GO:2001132) also involves special methyl transferase proteins (Tallant et al., 2001; Bose et al., 2009). For example, dimethylsulfide is converted to methyl-CoM by the actions of MtsB (MT2) and MtsA (MT2) (Tallant et al., 2001).

Energy conservation during methanogenesis from methylated compounds occurs in at least two ways. The  $\text{F}_{420}\text{H}_2$  generated during the oxidation of the methyl group of  $\text{CH}_3\text{-H}_4\text{MPT}$  to  $\text{CO}_2$  is oxidized via the membrane-bound

F<sub>420</sub>H<sub>2</sub>-dehydrogenase complex (reduced coenzyme F<sub>420</sub> dehydrogenase activity, GO:0043738), and in the process a lipid soluble membrane-resident cofactor called methanophenazine is reduced (Deppenmeier and Muller, 2008). These events lead to the extrusion of two protons per F<sub>420</sub>H<sub>2</sub> oxidized (Deppenmeier and Muller, 2008). There is another avenue that produces the same outcome and it begins with the release of molecular hydrogen through the oxidation of F<sub>420</sub>H<sub>2</sub> by a soluble F<sub>420</sub>-dependent hydrogenase (Frh, GO:0050454) (Kulkarni et al., 2009). This hydrogen upon its release from the cell is captured by a membrane-bound hydrogenase complex (Vht/Vtx) (GO:MENGO-UR, GO:MENGO-UR), which transfers electrons generated from the oxidation of hydrogen to methanophenazine and releases two protons outside the cell (Kulkarni et al., 2009). In certain methanogens the latter process is the major route of F<sub>420</sub>H<sub>2</sub> oxidation (Kulkarni et al., 2009). The reduced methanophenazine produced by these reactions is utilized by the membrane-bound heterodisulfide reductase (Hdr)-cytochrome b2 complex (GO:MENGO-UR) for the reduction of CoM-S-S-HTP, and this process provides two more protons outside the cell (Deppenmeier and Muller, 2008; Kulkarni et al., 2009; Welte and Deppenmeier, 2014). All these extruded protons generate proton-motive force, which drives ATP synthesis (GO:0015986) via an ATP synthase (GO:0045259) (Deppenmeier and Muller, 2008; Kulkarni et al., 2009; Welte and Deppenmeier, 2014).

#### **Methanogenesis from H<sub>2</sub> plus methanol**

A GO term for this process has recently been proposed by us (methanogenesis from H<sub>2</sub> and methanol, GO:1990491). Here, the methyl group of methanol is transferred to coenzyme M by two methyl transferases, MT1 and MT2, producing methyl-CoM (Keltjens and Vogels, 1993). The rest of the process, the reduction of methyl-CoM by HS-CoB, the reduction of CoM-S-S-CoB by electrons derived from hydrogen, and the energy conservation, likely follows the system described in the section on methanogenesis from H<sub>2</sub> plus CO<sub>2</sub>; an exception is *Methanosphaera stadtmanae*, which grows only on H<sub>2</sub> plus methanol with a supplement of acetate (Miller and Wolin, 1985), as it would employ the cytochrome-independent system (Fricke et al., 2006).

#### **Methanogenesis from acetate**

About 70% of the biologically produced methane originates from acetate (GO:0019385) (Ferry, 1992, 1993, 1999). The methyl group of acetate is reduced to methane and the carboxyl group is oxidized to CO<sub>2</sub> providing the reductant for methyl reduction (Ferry, 1992, 1993, 1999, 2011; Thauer et al., 2008). The process begins with the activation of acetate by the action of one of two systems, one involving acetate kinase and phosphotransacetylase (GO:0008776 and GO:0008959) and the other catalyzed by acetyl-CoA synthase (synonym of acetate-CoA ligase activity, GO:0003987), both generating acetyl-CoA (Aceti and Ferry, 1988; Jetten et al., 1989; Lundie and Ferry, 1989; Ferry, 1992, 1993, 1999, 2011; Thauer et al., 2008). The first route generates ADP that is converted back to ATP via electron transport phosphorylation at an ATPase (Ferry, 1992, 1993, 1999). In contrast, the second route generates AMP and pyrophosphate, and AMP has to be converted to ADP by adenylate kinase (GO:0004017)

through the consumption of one ATP (AMP + ATP → 2ADP) before it can be used for the regeneration of ATP (ADP + P<sub>i</sub> + energy → ATP) (Jetten et al., 1989; Zinder, 1993; Berger et al., 2012). Thus, organisms utilizing the acetyl-CoA synthase reaction are placed in an energetically unfavorable situation and exhibit slow growth rates (Zinder, 1993). However, by virtue of this investment they are able to utilize acetate even at very low concentrations and consequently are the predominant acetotrophic methanogens in many anaerobic niches of nature (Zinder, 1993). It is not known whether the energy present in pyrophosphate is conserved or is released via hydrolysis for the purpose of making the acetate activation process thermodynamically more favorable (Welte and Deppenmeier, 2014). The methanogens employing acetyl-CoA synthase carry pyrophosphatase (GO:0016462) and whether the enzyme is positioned to harvest or release energy is not known (Berger et al., 2012; Welte and Deppenmeier, 2014). The next step, the breakage of the carbon-carbon bond of the acetate moiety in acetyl-CoA, is catalyzed by an acetyl-CoA decarbonylase/synthase-carbon monoxide dehydrogenase complex (GO:0044672) (Ferry, 1993, 1999; Lu et al., 1994; Grahame, 2003; Li et al., 2006; Wang et al., 2011). The carbonyl group of acetyl-CoA is oxidized to CO<sub>2</sub> by the CODH component (GO:0043885) and the reducing equivalents (two-electrons) generated by this process help to reduce ferredoxin (Ferry, 1993, 1999, 2011; Lu et al., 1994; Grahame, 2003; Li et al., 2006; Wang et al., 2011). The methyl group of the acetyl group is transferred to H<sub>4</sub>MPT via a corrinoid cofactor of the CODH/ACDS complex, producing CH<sub>3</sub>-H<sub>4</sub>MPT (Ferry, 1999; Grahame, 2003). The methyl group of CH<sub>3</sub>-H<sub>4</sub>MPT leads to methane via the actions of methyl-H<sub>4</sub>MPT:coenzyme M methyl transferase (Mtr) and methyl-CoM reductase (Figure 4). The CO<sub>2</sub> produced from acetate is hydrated to bicarbonate by a membrane-bound gamma-type carbonic anhydrase (GO:0004089) and is efficiently exported out of the cell (Ferry, 2011). This process is thought to improve the thermodynamic efficiency of methanogenesis from acetate (Ferry, 2011).

There are two avenues for energy conservation in methanogenesis from acetate (Deppenmeier and Muller, 2008). One is via the use of the sodium potential generated by Mtr and has been described above. The other is through the oxidation of reduced ferredoxin through one of two complex processes. Certain acetotrophic methanogens oxidize reduced ferredoxin by use of Ech hydrogenase, generating molecular hydrogen and proton potential (Meuer et al., 1999, 2002; Kulkarni et al., 2009). The molecular hydrogen is utilized for the extrusion of additional protons and for heterodisulfide reduction via the Vho hydrogenase, methanophenazine and heterodisulfide reductase, as during methylotrophic methanogenesis (Figure 4; see above) (Kulkarni et al., 2009). In methanogens lacking Ech hydrogenase, a complex called Rnf utilizes reduced ferredoxin, producing a sodium gradient and transferring electrons to heterodisulfide reductase via methanophenazine. Thus, Rnf is considered a replacement of the Ech and Vho hydrogenases (Li et al., 2006; Wang et al., 2011). Both the H<sup>+</sup> and Na<sup>+</sup> potentials are utilized by an A<sub>1</sub>A<sub>0</sub> ATP synthase (GO:1990490) for ATP production (Deppenmeier and Muller, 2008); in some cases a Na<sup>+</sup>/H<sup>+</sup> antiporter (GO:0015385) called Mrp adjusts the ratio of the two gradients for optimizing



the thermodynamic efficiency of the ATP synthase (Li et al., 2006; Wang et al., 2011; Jasso-Chavez et al., 2013).

### Biosynthesis of methanogenesis coenzymes

Many of the coenzymes involved in methanogenesis, namely methanofuran, tetrahydromethanopterin, tetrahydrodarsinapterin, coenzyme M, coenzyme F<sub>420</sub>, coenzyme B, methanophenazine, and coenzyme F<sub>430</sub>, have unusual properties. As a result, the respective biosynthesis pathways have attracted attention (Graham and White, 2002). This interest has increased further as some of these coenzymes have been found to perform critical functions in other organisms, such as in actinobacteria (includes mycobacteria and streptomyces groups), methanotrophic and methylotrophic bacteria, cyanobacteria, and plants (Takao et al., 1989; Batschauer, 1993; Purwantini et al., 1997; Chistoserdova et al., 2004; Krishnakumar et al., 2008). Some of the existing knowledge has been summarized at the MENGO website.

### SYNTHETIC BIOLOGY EXPLOITATION OF METHANOGENESIS PATHWAYS IN METHANOGENS

Exploitation of methanogens for the production of methane from unnatural substrates has begun. For example, *Methanosarcina acetivorans* has been made proficient in converting methyl acetate to methane and carbon dioxide, and in converting methyl propionate to methane and propionate. This was achieved by expressing a broad-specificity esterase (hydrolase activity, acting on ester bonds, GO:0016788) from *Pseudomonas veronii* in *M. acetivorans* (Lessner et al., 2010). Wild type *M. acetivorans* exhibits only a minor esterase activity. The heterologous esterase in the engineered strain releases methanol from these two esters, and methanol is used for methanogenesis following the pathways described above. Acetate, the other product from methyl acetate is also converted to methane whereas propionate generated from methyl propionate is excreted (Lessner et al., 2010).

### A NEW ROUTE FOR BIOLOGICAL PRODUCTION OF METHANE

A recent discovery (Metcalf et al., 2012; Yu et al., 2013) shows that some of the abundant marine archaea and bacteria, which are distinct from the well-known methanogenic archaea, are likely major producers of methane in nature. Methane is abundant in the oceans, but the source was unclear (Reeburgh, 2007). Methylphosphonate was suspected as the source as genes encoding carbon-phosphorus lyases are common in marine microbes, but the biosynthetic pathway for methylphosphonate was unknown (Karl et al., 2008). It has recently been shown that the marine archaeon *Nitrosopumilus maritimus* encodes a pathway for methylphosphonate biosynthesis and it produces cell-associated methylphosphonate esters (Metcalf et al., 2012). The production of methylphosphonate seems to be a widespread process in marine microorganisms, and that when facing phosphorus-limitation these organisms would degrade methylphosphonate to obtain phosphorus, thus releasing methane (Metcalf et al., 2012). The GO database lacks description for methylphosphonate biosynthetic and catabolic processes, as well for the following key enzymes: carbon-phosphorus (C-P) lyase, producer of methane

from methylphosphonate; phosphonoacetaldehyde dehydrogenase (Pdh) and methylphosphonate synthase (MPn), two key enzymes for methylphosphonate biosynthesis (Metcalf et al., 2012). However, the terms for the first two enzymes on the methylphosphonate biosynthesis pathway that starts from phosphoenolpyruvate, namely “phosphoenolpyruvate mutase (Ppm) activity” (GO:0050188) and “phosphonopyruvate decarboxylase (Ppd) activity,” do exist (GO:0033980). To cover this new biological process for methane production we have proposed the following new GO terms: phosphonate carbon-phosphorus lyase activity (GO:MENGO-UR); “methane biosynthetic process” (GO:0015948), a parent term; two child terms, “aerobic methane biosynthetic process” (GO:MENGO-UR) and “anaerobic methane biosynthetic process” (GO:MENGO-UR).

### GO ANNOTATION

To begin the application of the GO terms to annotating genomes of methanogenic microbes, we have performed GO annotation of the relevant gene products encoded by these genomes. The annotations we created were based solely on experimental evidence (e.g., results from direct assays or mutant phenotypes), in order to provide “gold standards” for subsequent machine annotations. These annotations are available at the MENGO website under the Gene Annotations menu (Gene Annotations for Natural Biological System; Gene Annotations for Synthetic Biological System). Forms for the submission of new annotations (Submit New Gene Annotation for Natural or Synthetic Biological System) are available under the same menu.

We have annotated 80 gene products with the parent term “methane biosynthetic process” (GO: 0015948) along with appropriate child terms (Figure 6). These genes were categorized into three groups; 51 gene products for methanogenesis pathways, 19 gene products for biosynthesis of coenzymes specifically used in methanogenesis, and 10 genes for coenzyme metabolism (see Table S1, Supplementary material).

### CONCLUDING REMARKS

The goal of the MENGO project is to develop a set of GO terms for describing gene products involved in energy-related microbial processes. GO allows annotations of gene products using terms from three ontologies: molecular function, biological process, and cellular component. The GO embodies structured relationships among the terms and the annotations provide links between gene products and the terms (Figure 6). This combination allows researchers to infer possible functional roles of gene products in diverse organisms. A set of relevant gene products well-annotated with GO terms will assist bioenergy researchers to efficiently design synthetic biological systems for commercially viable biofuel production, as it will allow effective mining for optimal parts from a larger natural inventory. For example, one could mine for amenable parts of a methanogenesis system from all available genomes, including those of organisms that do not produce methane. Thus, the GO terms and associated “gold standard” manual annotations that the MENGO has developed should provide the foundation for a growing resource that is of wide value to the microbial bioenergy community. We encourage members of the research community to participate in our effort toward



the development of additional GO terms and performing manual annotations of gene products with potentials of application in bioenergy production and bioremediation. The MENGO website provides electronic forms for the submission of candidate GO terms and annotations for review and subsequent submission to the GO database.

## ACKNOWLEDGMENTS

This work was supported by grant DE-SC0005011 from the US Department of Energy. João C. Setubal was funded by CNPq and FAPESP, and Jane Lomax was supported by EMBL-EBI core funds. We thank Tirtha Bhattacharjee, Morgan Pixa, and Stephen Slaughter for help with the MENGO website and Morgan Pixa and Sujung Kang for help in annotation. We thank the GO consortium for collaboration and members of the bioenergy and bioremediation research community who attended our MENGO workshops and provided suggestions for GO term generation and annotation of gene products.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fmicb.2014.00634/abstract>

## REFERENCES

- Aceti, D. J., and Ferry, J. G. (1988). Purification and characterization of acetate kinase from acetate-grown *Methanosarcina thermophila*. Evidence for regulation of synthesis. *J. Biol. Chem.* 263, 15444–15448.
- Adney, W. S., Rivard, C. J., Ming, S. A., and Himmel, M. E. (1991). Anaerobic digestion of lignocellulosic biomass and wastes. Cellulases and related enzymes. *Appl. Biochem. Biotechnol.* 30, 165–183. doi: 10.1007/BF02921684
- Ahring, B. K. (2003). Perspectives for anaerobic digestion. *Adv. Biochem. Eng. Biotechnol.* 81, 1–30. doi: 10.1007/3-540-45839-5\_1
- Akin, D. E. (1980). Attack on lignified grass cell walls by a facultatively anaerobic bacterium. *Appl. Environ. Microbiol.* 40, 809–820.
- Allison, M. J., Dawson, K. A., Mayberry, W. R., and Foss, J. G. (1985). *Oxalobacter formigenes* gen. nov., sp. nov.: oxalate-degrading anaerobes that inhabit the gastrointestinal tract. *Arch. Microbiol.* 141, 1–7. doi: 10.1007/BF00446731
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology consortium. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556
- Batschauer, A. (1993). A plant gene for photolyase: an enzyme catalyzing the repair of UV-light-induced DNA damage. *Plant J.* 4, 705–709. doi: 10.1046/j.1365-313X.1993.04040705.x
- Bayer, E. A., Chanzy, H., Lamed, R., and Shoham, Y. (1998). Cellulose, cellulases and cellulosomes. *Curr. Opin. Struct. Biol.* 8, 548–557. doi: 10.1016/S0959-440X(98)80143-7
- Becher, B., Muller, V., and Gottschalk, G. (1992). N<sup>5</sup>-methyltetrahydromethanopterin:coenzyme M methyltransferase of *Methanosarcina* strain Go1 is an Na(+)-translocating membrane protein. *J. Bacteriol.* 174, 7656–7660.
- Begu, P., and Aubert, J. P. (1994). The biological degradation of cellulose. *FEMS Microbiol. Rev.* 13, 25–58. doi: 10.1111/j.1574-6976.1994.tb00033.x
- Berger, S., Welte, C., and Deppenmeier, U. (2012). Acetate activation in *Methanosaeta thermophila*: characterization of the key enzymes pyrophosphatase and acetyl-CoA synthetase. *Archaea* 2012:315153. doi: 10.1155/2012/315153
- Blanch, H. W., Simmons, B. A., and Klein-Marcuschamer, D. (2011). Biomass deconstruction to sugars. *Biotechnol. J.* 6, 1086–1102. doi: 10.1002/biot.201000180
- Bleicher, K., Zellner, G., and Winter, J. (1989). Growth of methanogens on cyclopentanol/CO<sub>2</sub> and specificity of alcohol dehydrogenase. *FEMS Microbiol. Lett.* 59, 307–312. doi: 10.1111/j.1574-6968.1989.tb03130.x
- Boone, D. R., Whitman, W. B., and Rouviere, P. (1993). “Diversity and taxonomy of methanogens,” in *Methanogenesis: Ecology, Physiology, Biochemistry and Genetics*, ed J. G. Ferry. (New York, NY: Chapman and Hall), 35–80.
- Bose, A., Kulkarni, G., and Metcalf, W. W. (2009). Regulation of putative methylsulphide methyltransferases in *Methanosarcina acetivorans* C2A. *Mol. Microbiol.* 74, 227–238. doi: 10.1111/j.1365-2958.2009.06864.x
- Bose, A., Pritchett, M. A., and Metcalf, W. W. (2008). Genetic analysis of the methanol- and methylamine-specific methyltransferase 2 genes of *Methanosarcina acetivorans* C2A. *J. Bacteriol.* 190, 4017–4026. doi: 10.1128/JB.00117-08
- Bose, A., Pritchett, M. A., Rother, M., and Metcalf, W. W. (2006). Differential regulation of the three methanol methyltransferase isozymes in *Methanosarcina acetivorans* C2A. *J. Bacteriol.* 188, 7274–7283. doi: 10.1128/JB.00535-06
- Breitung, J., and Thauer, R. K. (1990). Formylmethanofuran: tetrahydromethanopterin formyltransferase from *Methanosarcina barkeri*. Identification of N<sup>5</sup>-formyltetrahydromethanopterin as the product. *FEBS Lett.* 275, 226–230. doi: 10.1016/0014-5793(90)81477-6
- Breznak, J. A. (1994). “Acetogenesis from carbon dioxide in termite guts,” in *Acetogenesis*, ed H. L. Drake. (New York, NY: Chapman & Hall), 303–330.
- Brune, A., and Friedrich, M. (2000). Microecology of the termite gut: structure and function on a microscale. *Curr. Opin. Microbiol.* 3, 263–269. doi: 10.1016/S1369-5274(00)00087-4
- Brunecky, R., Alahuhta, M., Xu, Q., Donohoe, B. S., Crowley, M. F., Kataeva, I. A., et al. (2013). Revealing nature's cellulase diversity: the digestion mechanism of *Caldicellulosiruptor bescii* CelA. *Science* 342, 1513–1516. doi: 10.1126/science.1244273
- Bult, C. J., White, O., Olsen, G. J., Zhou, L., Fleischmann, R. D., Sutton, G. G., et al. (1996). Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* 273, 1058–1073. doi: 10.1126/science.273.5278.1058
- Burke, S. A., Lo, S. L., and Krzycki, J. A. (1998). Clustered genes encoding the methyltransferases of methanogenesis from monomethylamine. *J. Bacteriol.* 180, 3432–3440.
- Chistoserdova, L., Jenkins, C., Kalyuzhnaya, M. G., Marx, C. J., Lapidus, A., Vorholt, J. A., et al. (2004). The enigmatic planctomycetes may hold a key to the origins of methanogenesis and methylotrophy. *Mol. Biol. Evol.* 21, 1234–1241. doi: 10.1093/molbev/msh113
- Conrad, R. (2007). Microbial ecology of methanogens and methanotrophs. *Adv. Agron.* 96, 1–63. doi: 10.1016/S0065-2113(07)96005-8
- Conrad, R. (2009). The global methane cycle: recent advances in understanding the microbial processes involved. *Environ. Microbiol. Rep.* 1, 285–292. doi: 10.1111/j.1758-2229.2009.00038.x
- Costa, K. C., and Leigh, J. A. (2014). Metabolic versatility in methanogens. *Curr. Opin. Biotechnol.* 29C, 70–75. doi: 10.1016/j.copbio.2014.02.012
- Costa, K. C., Wong, P. M., Wang, T., Lie, T. J., Dodsworth, J. A., Swanson, I., et al. (2010). Protein complexing in a methanogen suggests electron bifurcation and electron delivery from formate to heterodisulfide reductase. *Proc. Natl. Acad. Sci. U.S.A.* 107, 11050–11055. doi: 10.1073/pnas.1003653107
- Daniels, L., Fuchs, G., Thauer, R. K., and Zeikus, J. G. (1977). Carbon monoxide oxidation by methanogenic bacteria. *J. Bacteriol.* 132, 118–126.
- De Graeve, K. G., Grivet, J. P., Durand, M., Beaumatin, P., Cordelet, C., Hannequart, G., et al. (1994). Competition between reductive acetogenesis and methanogenesis in the pig large-intestinal flora. *J. Appl. Bacteriol.* 76, 55–61. doi: 10.1111/j.1365-2672.1994.tb04415.x
- Deppenmeier, U. (2002). The unique biochemistry of methanogenesis. *Prog. Nucleic Acid Res. Mol. Biol.* 71, 223–283. doi: 10.1016/S0079-6603(02)71045-3
- Deppenmeier, U., Lienard, T., and Gottschalk, G. (1999). Novel reactions involved in energy conservation by methanogenic archaea. *FEBS Lett.* 457, 291–297. doi: 10.1016/S0014-5793(99)01026-1
- Deppenmeier, U., and Muller, V. (2008). Life close to the thermodynamic limit: how methanogenic archaea conserve energy. *Results Probl. Cell Differ.* 45, 123–152. doi: 10.1007/400\_2006\_026
- Diaz, I., and Fdz-Polanco, M. (2012). Robustness of the microaerobic removal of hydrogen sulfide from biogas. *Water Sci. Technol.* 65, 1368–1374. doi: 10.2166/wst.2012.013
- Dimarco, A. A., Donnelly, M. I., and Wolfe, R. S. (1986). Purification and properties of the 5,10-methenyltetrahydromethanopterin cyclohydrolase from *Methanobacterium thermoautotrophicum*. *J. Bacteriol.* 168, 1372–1377.

- Donnelly, M. I., Escalante-Semerena, J. C., Rinehart, K. L. Jr., and Wolfe, R. S. (1985). Methenyl-tetrahydromethanopterin cyclohydrolase in cell extracts of *Methanobacterium*. *Arch. Biochem. Biophys.* 242, 430–439. doi: 10.1016/0003-9861(85)90227-9
- Donnelly, M. I., and Wolfe, R. S. (1986). The role of formylmethanofuran: tetrahydromethanopterin formyltransferase in methanogenesis from carbon dioxide. *J. Biol. Chem.* 261, 16653–16659.
- Ducom, G., Radu-Tirnovanu, D., Pascual, C., Benadda, B., and Germain, P. (2009). Biogas–municipal solid waste incinerator bottom ash interactions: sulphur compounds removal. *J. Hazard. Mater.* 166, 1102–1108. doi: 10.1016/j.jhazmat.2008.12.024
- Fdz-Polanco, F., Fdz-Polanco, M., Fernandez, N., Uruena, M. A., Garcia, P. A., and Villaverde, S. (2001). New process for simultaneous removal of nitrogen and sulphur under anaerobic conditions. *Water Res.* 35, 1111–1114. doi: 10.1016/S0043-1354(00)00474-7
- Ferguson, D. J. Jr., Gorlatova, N., Grahame, D. A., and Krzycki, J. A. (2000). Reconstitution of dimethylamine:coenzyme M methyl transfer with a discrete corrinoid protein and two methyltransferases purified from *Methanosarcina barkeri*. *J. Biol. Chem.* 275, 29053–29060. doi: 10.1074/jbc.M910218199
- Ferguson, D. J. Jr., and Krzycki, J. A. (1997). Reconstitution of trimethylamine-dependent coenzyme M methylation with the trimethylamine corrinoid protein and the isozymes of methyltransferase II from *Methanosarcina barkeri*. *J. Bacteriol.* 179, 846–852.
- Ferry, J. G. (1992). Methane from acetate. *J. Bacteriol.* 174, 5489–5495.
- Ferry, J. G. (1993). “Fermentation of acetate,” in *Methanogenesis: Ecology, Physiology, Biochemistry and Genetics*, ed J. G. Ferry (New York, NY: Chapman and Hall), 304–334.
- Ferry, J. G. (1999). Enzymology of one-carbon metabolism in methanogenic pathways. *FEMS Microbiol. Rev.* 23, 13–38. doi: 10.1111/j.1574-6976.1999.tb00390.x
- Ferry, J. G. (2011). Fundamentals of methanogenic pathways that are key to the biometanation of complex biomass. *Curr. Opin. Biotechnol.* 22, 351–357. doi: 10.1016/j.copbio.2011.04.011
- Fischer, R., Gartner, P., Yeliseev, A., and Thauer, R. K. (1992). N<sup>5</sup>-methyltetrahydromethanopterin: coenzyme M methyltransferase in methanogenic archaeobacteria is a membrane protein. *Arch. Microbiol.* 158, 208–217. doi: 10.1007/BF00290817
- Flint, H. J. (2011). Obesity and the gut microbiota. *J. Clin. Gastroenterol.* 45(Suppl.), S128–S132. doi: 10.1097/MCG.0b013e31821f44c4
- Fricke, W. F., Seedorf, H., Henne, A., Kruer, M., Liesegang, H., Hedderich, R., et al. (2006). The genome sequence of *Methanospaera stadtmanae* reveals why this human intestinal archaeon is restricted to methanol and H<sub>2</sub> for methane formation and ATP synthesis. *J. Bacteriol.* 188, 642–658. doi: 10.1128/JB.188.2.642-658.2006
- Fuchs, G. (2008). Anaerobic metabolism of aromatic compounds. *Ann. N.Y. Acad. Sci.* 1125, 82–99. doi: 10.1196/annals.1419.010
- Gagen, E. J., Denman, S. E., Padmanabha, J., Zadbucke, S., Al Jassim, R., Morrison, M., et al. (2010). Functional gene analysis suggests different acetogen populations in the bovine rumen and tammar wallaby forestomach. *Appl. Environ. Microbiol.* 76, 7785–7795. doi: 10.1128/AEM.01679-10
- Gao, H., Scherson, Y. D., and Wells, G. F. (2014). Towards energy neutral wastewater treatment: methodology and state of the art. *Environ. Sci. Process. Impacts* 16, 1223–1246. doi: 10.1039/c4em00069b
- Gartner, P., Ecker, A., Fischer, R., Linder, D., Fuchs, G., and Thauer, R. K. (1993). Purification and properties of N<sup>5</sup>-methyltetrahydromethanopterin:coenzyme M methyltransferase from *Methanobacterium thermoautotrophicum*. *Eur. J. Biochem.* 213, 537–545. doi: 10.1111/j.1432-1033.1993.tb17792.x
- Gibson, G. R., Macfarlane, G. T., and Cummings, J. H. (1988). Occurrence of sulphate-reducing bacteria in human faeces and the relationship of dissimilatory sulphate reduction to methanogenesis in the large gut. *J. Appl. Bacteriol.* 65, 103–111. doi: 10.1111/j.1365-2672.1988.tb01498.x
- Graham, D. E., and White, R. H. (2002). Elucidation of methanogenic coenzyme biosyntheses: from spectroscopy to genomics. *Nat. Prod. Rep.* 19, 133–147. doi: 10.1039/b103714p
- Grahame, D. A. (2003). Acetate C-C bond formation and decomposition in the anaerobic world: the structure of a central enzyme and its key active-site metal cluster. *Trends Biochem. Sci.* 28, 221–224. doi: 10.1016/S0968-0004(03)00063-X
- Haroon, M. F., Hu, S., Shi, Y., Imelfort, M., Keller, J., Hugenholtz, P., et al. (2013). Anaerobic oxidation of methane coupled to nitrate reduction in a novel archaeal lineage. *Nature* 500, 567–570. doi: 10.1038/nature12375
- Hartzell, P. L., Zvilius, G., Escalante-Semerena, J. C., and Donnelly, M. I. (1985). Coenzyme F<sub>420</sub> dependence of the methylenetetrahydromethanopterin dehydrogenase of *Methanobacterium thermoautotrophicum*. *Biochem. Biophys. Res. Commun.* 133, 884–890. doi: 10.1016/0006-291X(85)91218-5
- Harwood, C. S., and Gibson, J. (1988). Anaerobic and aerobic metabolism of diverse aromatic compounds by the photosynthetic bacterium *Rhodospseudomonas palustris*. *Appl. Environ. Microbiol.* 54, 712–717.
- Harwood, C. S., and Parales, R. E. (1996). The beta-ketoadipate pathway and the biology of self-identity. *Annu. Rev. Microbiol.* 50, 553–590. doi: 10.1146/Annurev.Micro.50.1.553
- Hilden, L., and Johansson, G. (2004). Recent developments on cellulases and carbohydrate-binding modules with cellulose affinity. *Biotechnol. Lett.* 26, 1683–1693. doi: 10.1007/s10529-004-4579-8
- Hochheimer, A., Schmitz, R. A., Thauer, R. K., and Hedderich, R. (1995). The tungsten formylmethanofuran dehydrogenase from *Methanobacterium thermoautotrophicum* contains sequence motifs characteristic for enzymes containing molybdopterin dinucleotide. *Eur. J. Biochem.* 234, 910–920. doi: 10.1111/j.1432-1033.1995.910.a.x
- Hook, S. E., Wright, A. D., and McBride, B. W. (2010). Methanogens: methane producers of the rumen and mitigation strategies. *Archaea* 2010:945785. doi: 10.1155/2010/945785
- Hristov, A. N., Ahvenjarvi, S., McAllister, T. A., and Huhtanen, P. (2003). Composition and digestive tract retention time of ruminal particles with functional specific gravity greater or less than 1.02. *J. Anim. Sci.* 81, 2639–2648.
- Isa, Z., Grusenmeyer, S., and Verstraete, W. (1986). Sulfate reduction relative to methane production in high-rate anaerobic digestion: microbiological aspects. *Appl. Environ. Microbiol.* 51, 580–587.
- Janssen, A. J., Ruitenberg, R., and Buisman, C. J. (2001). Industrial applications of new sulphur biotechnology. *Water Sci. Technol.* 44, 85–90.
- Jasso-Chavez, R., Apolinario, E. E., Sowers, K. R., and Ferry, J. G. (2013). MrpA functions in energy conversion during acetate-dependent growth of *Methanosarcina acetivorans*. *J. Bacteriol.* 195, 3987–3994. doi: 10.1128/JB.00581-13
- Jetten, M. S., Stams, A. J., and Zehnder, A. J. (1989). Isolation and characterization of acetyl-coenzyme A synthetase from *Methanoxanthus soehngenii*. *J. Bacteriol.* 171, 5430–5435.
- Kaiser, J. P., and Hanselmann, K. W. (1982a). Anaerobic degradation and conversion of methoxylated and hydroxylated benzoic-acid derivatives by bacterial communities. *Experientia* 38, 1379–1379.
- Kaiser, J. P., and Hanselmann, K. W. (1982b). Aromatic chemicals through anaerobic microbial conversion of lignin monomers. *Experientia* 38, 167–176. doi: 10.1007/BF01945070
- Karl, D. M., Beversdorf, L., Bjorkman, K. M., Church, M. J., Martinez, A., and Delong, E. F. (2008). Aerobic production of methane in the sea. *Nat. Geosci.* 1, 473–478. doi: 10.1038/Ngeo234
- Kaster, A. K., Moll, J., Parey, K., and Thauer, R. K. (2011). Coupling of ferredoxin and heterodisulfide reduction via electron bifurcation in hydrogenotrophic methanogenic archaea. *Proc. Natl. Acad. Sci. U.S.A.* 108, 2981–2986. doi: 10.1073/pnas.1016761108
- Keltjens, J. T., and Vogels, G. D. (1993). “Conversion of methanol and methylamines,” in *Methanogenesis: Ecology, Physiology, Biochemistry and Genetics*, ed J. G. Ferry (New York, NY: Chapman and Hall), 253–303.
- Kengen, S. W., Daas, P. J., Duits, E. F., Keltjens, J. T., Van Der Drift, C., and Vogels, G. D. (1992). Isolation of a 5-hydroxybenzimidazolyl cobamide-containing enzyme involved in the methyltetrahydromethanopterin: coenzyme M methyltransferase reaction in *Methanobacterium thermoautotrophicum*. *Biochim. Biophys. Acta* 1118, 249–260. doi: 10.1016/0167-4838(92)90282-1
- Kiene, R. P. (1991). “Production and consumption of methane in aquatic systems,” in *Microbial Production and Consumption of Greenhouse Gases: Methane, Nitrogen Oxides, and Halo-Methanes*, eds J. E. Rogers and W. B. Whitman (Washington, DC: American Society for Microbiology), 111–146.
- Klein, A. R., Breitung, J., Linder, D., Stetter, K. O., and Thauer, R. K. (1993). N<sub>5</sub>,N<sub>10</sub>-methenyltetrahydromethanopterin cyclohydrolase from the extremely thermophilic sulfate reducing *Archaeoglobus fulgidus*: comparison of its properties with those of the cyclohydrolase from the extremely

- thermophilic *Methanopyrus kandleri*. *Arch. Microbiol.* 159, 213–219. doi: 10.1007/BF00248474
- Klieve, A. V., Ouwerkerk, D., and Maguire, A. J. (2012). Archaea in the foregut of macropod marsupials: PCR and amplicon sequence-based observations. *J. Appl. Microbiol.* 113, 1065–1075. doi: 10.1111/j.1365-2672.2012.05428.x
- Krishnakumar, A. M., Sliwa, D., Endrizzi, J. A., Boyd, E. S., Ensign, S. A., and Peters, J. W. (2008). Getting a handle on the role of coenzyme M in alkene metabolism. *Microbiol. Mol. Biol. Rev.* 72, 445–456. doi: 10.1128/MMBR.00005-08
- Kulkarni, G., Kridelbaugh, D. M., Guss, A. M., and Metcalf, W. W. (2009). Hydrogen is a preferred intermediate in the energy-conserving electron transport chain of *Methanosarcina barkeri*. *Proc. Natl. Acad. Sci. U.S.A.* 106, 15915–15920. doi: 10.1073/pnas.0905914106
- Leigh, J. A. (2002). “Evolution of energy metabolism,” in *Biodiversity of Microbial Life: Foundation of Earth Biosphere*, eds J. T. Staley and A. L. Reysenbach (New York, NY: John Wiley & Sons), 103–120.
- Leschine, S. B. (1995). Cellulose degradation in anaerobic environments. *Annu. Rev. Microbiol.* 49, 399–426. doi: 10.1146/annurev.mi.49.100195.002151
- Lessner, D. J., Lhu, L., Wahal, C. S., and Ferry, J. G. (2010). An engineered methanogenic pathway derived from the domains Bacteria and Archaea. *MBiol.* doi: 10.1128/mBio.00243-10
- Lessner, D. J., Li, L., Li, Q., Rejtar, T., Andreev, V. P., Reichlen, M., et al. (2006). An unconventional pathway for reduction of CO<sub>2</sub> to methane in CO-grown *Methanosarcina acetivorans* revealed by proteomics. *Proc. Natl. Acad. Sci. U.S.A.* 103, 17921–17926. doi: 10.1073/pnas.0608833103
- Li, Q., Li, L., Rejtar, T., Lessner, D. J., Karger, B. L., and Ferry, J. G. (2006). Electron transport in the pathway of acetate conversion to methane in the marine archaeon *Methanosarcina acetivorans*. *J. Bacteriol.* 188, 702–710. doi: 10.1128/JB.188.2.702-710.2006
- Li, X. L., Chen, H., and Ljungdahl, L. G. (1997). Monocentric and polycentric anaerobic fungi produce structurally related cellulases and xylanases. *Appl. Environ. Microbiol.* 63, 628–635.
- Lie, T. J., Costa, K. C., Lupa, B., Korpole, S., Whitman, W. B., and Leigh, J. A. (2012). Essential anaerobic role for the energy-converting hydrogenase Eha in hydrogenotrophic methanogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 109, 15473–15478. doi: 10.1073/pnas.1208779109
- Lie, T. J., Costa, K. C., Pak, D., Sakesan, V., and Leigh, J. A. (2013). Phenotypic evidence that the function of the [Fe]-hydrogenase Hmd in *Methanococcus maripaludis* requires seven hcg (*hmd* co-occurring genes) but not hmdII. *FEMS Microbiol. Lett.* 343, 156–160. doi: 10.1111/1574-6968.12141
- Liu, Y., and Whitman, W. B. (2008). Metabolic, phylogenetic, and ecological diversity of the methanogenic archaea. *Ann. N.Y. Acad. Sci.* 1125, 171–189. doi: 10.1196/annals.1419.019
- Ljungdahl, L. G. (2008). The cellulase/hemicellulase system of the anaerobic fungus *Orpinomyces* PC-2 and aspects of its applied use. *Ann. N.Y. Acad. Sci.* 1125, 308–321. doi: 10.1196/annals.1419.030
- Lu, W. P., Jablonski, P. E., Rasche, M., Ferry, J. G., and Ragsdale, S. W. (1994). Characterization of the metal centers of the Ni/Fe-S component of the carbon-monoxide dehydrogenase enzyme complex from *Methanosarcina thermophila*. *J. Biol. Chem.* 269, 9736–9742.
- Lundie, L. L. Jr., and Ferry, J. G. (1989). Activation of acetate by *Methanosarcina thermophila*. Purification and characterization of phosphotransacetylase. *J. Biol. Chem.* 264, 18392–18396.
- Ma, K., Linder, D., Stetter, K. O., and Thauer, R. K. (1991). Purification and properties of N<sup>5</sup>,N<sup>10</sup>-methylene-tetrahydromethanopterin reductase (coenzyme F<sub>420</sub>-dependent) from the extreme thermophile *Methanopyrus kandleri*. *Arch. Microbiol.* 155, 593–600. doi: 10.1007/BF00245355
- Ma, K., and Thauer, R. K. (1990). Purification and properties of N<sup>5</sup>, N<sup>10</sup>-methylene-tetrahydromethanopterin reductase from *Methanobacterium thermoautotrophicum* (strain Marburg). *Eur. J. Biochem.* 191, 187–193. doi: 10.1111/j.1432-1033.1990.tb19109.x
- Major, T. A., Liu, Y., and Whitman, W. B. (2010). Characterization of energy-conserving hydrogenase B in *Methanococcus maripaludis*. *J. Bacteriol.* 192, 4022–4030. doi: 10.1128/JB.01446-09
- McInerney, M. J. (1988). “Anaerobic degradation of proteins and lipids,” in *Biology of Anaerobic Microorganisms*, ed A. J. B. Zehnder (New York, NY: John Wiley & Sons, Inc.), 373–415.
- McInerney, M. J., Sieber, J. R., and Gunsalus, R. P. (2009). Syntrophy in anaerobic global carbon cycles. *Curr. Opin. Biotechnol.* 20, 623–632. doi: 10.1016/j.copbio.2009.10.001
- Metcalf, W. W., Griffin, B. M., Cicchillo, R. M., Gao, J., Janga, S. C., Cooke, H. A., et al. (2012). Synthesis of methylphosphonic acid by marine microbes: a source for methane in the aerobic ocean. *Science* 337, 1104–1107. doi: 10.1126/science.1219875
- Meurer, J., Bartoschek, S., Koch, J., Kunkel, A., and Hedderich, R. (1999). Purification and catalytic properties of Ech hydrogenase from *Methanosarcina barkeri*. *Eur. J. Biochem.* 265, 325–335. doi: 10.1046/j.1432-1327.1999.00738.x
- Meurer, J., Kuettner, H. C., Zhang, J. K., Hedderich, R., and Metcalf, W. W. (2002). Genetic analysis of the archaeon *Methanosarcina barkeri* Fusaro reveals a central role for Ech hydrogenase and ferredoxin in methanogenesis and carbon fixation. *Proc. Natl. Acad. Sci. U.S.A.* 99, 5632–5637. doi: 10.1073/pnas.072615499
- Miller, T. L., and Wolin, M. J. (1985). *Methanosphaera stadtmaniae* gen. nov., sp. nov.: a species that forms methane by reducing methanol with hydrogen. *Arch. Microbiol.* 141, 116–122. doi: 10.1007/BF00423270
- Miller, T. L., and Wolin, M. J. (1996). Pathways of acetate, propionate, and butyrate formation by the human fecal microbial flora. *Appl. Environ. Microbiol.* 62, 1589–1592.
- Milucka, J., Ferdelman, T. G., Polerecky, L., Franzke, D., Wegener, G., Schmid, M., et al. (2012). Zero-valent sulphur is a key intermediate in marine methane oxidation. *Nature* 491, 541–546. doi: 10.1038/nature11656
- Moran, J. J., House, C. H., Vrentas, J. M., and Freeman, K. H. (2008). Methyl sulfide production by a novel carbon monoxide metabolism in *Methanosarcina acetivorans*. *Appl. Environ. Microbiol.* 74, 540–542. doi: 10.1128/AEM.01750-07
- Mukhopadhyay, B., and Daniels, L. (1989). Aerobic purification of N<sup>5</sup>,N<sup>10</sup>-methylene-tetrahydromethanopterin dehydrogenase, separated from N<sup>5</sup>,N<sup>10</sup>-methylene-tetrahydromethanopterin cyclohydrolase, from *Methanobacterium thermoautotrophicum* strain Marburg. *Can. J. Microbiol.* 35, 499–507. doi: 10.1139/m89-077
- Mukhopadhyay, B., Purwantini, E., De Macario, E. C., and Daniels, L. (1992). Characterization of *Methanosarcina* strain isolated from goat feces, and that grows on H<sub>2</sub>-CO<sub>2</sub> only after adaptation. *Curr. Microbiol.* 23, 165–173. doi: 10.1007/BF02091977
- Mukhopadhyay, B., Purwantini, E., Pihl, T. D., Reeve, J. N., and Daniels, L. (1995). Cloning, sequencing, and transcriptional analysis of the coenzyme F<sub>420</sub>-dependent methylene-5,6,7,8-tetrahydromethanopterin dehydrogenase gene from *Methanobacterium thermoautotrophicum* strain Marburg and functional expression in *Escherichia coli*. *J. Biol. Chem.* 270, 2827–2832. doi: 10.1074/jbc.270.6.2827
- Muller, V. (2003). Energy conservation in acetogenic bacteria. *Appl. Environ. Microbiol.* 69, 6345–6353. doi: 10.1128/AEM.69.11.6345-6353.2003
- Nakamura, N., Lin, H. C., McSweeney, C. S., Mackie, R. I., and Gaskins, H. R. (2010). Mechanisms of microbial hydrogen disposal in the human colon and implications for health and disease. *Annu. Rev. Food Sci. Technol.* 1, 363–395. doi: 10.1146/annurev.food.102308.124101
- O'Brien, J. M., Wolk, R. H., Moench, T. T., Morgan, J. B., and Zeikus, J. G. (1984). Association of hydrogen metabolism with unitrophic or mixotrophic growth of *Methanosarcina barkeri* on carbon monoxide. *J. Bacteriol.* 158, 373–375.
- Offe, P., Spang, A., and Schleper, C. (2013). Archaea in biogeochemical cycles. *Annu. Rev. Microbiol.* 67, 437–457. doi: 10.1146/annurev-micro-092412-155614
- Oza, J. P., Sowers, K. R., and Perona, J. J. (2012). Linking energy production and protein synthesis in hydrogenotrophic methanogens. *Biochemistry* 51, 2378–2389. doi: 10.1021/bi300106p
- Paul, E., and DeBellefontaine, H. (2007). Reduction of excess sludge produced by biological treatment processes: effect of ozonation on biomass and on sludge. *Ozone Sci. Eng.* 29, 415–427. doi: 10.1080/01919510701593762
- Paul, L., Ferguson, D. J. Jr., and Krzycki, J. A. (2000). The trimethylamine methyltransferase gene and multiple dimethylamine methyltransferase genes of *Methanosarcina barkeri* contain in-frame and read-through amber codons. *J. Bacteriol.* 182, 2520–2529. doi: 10.1128/JB.182.9.2520-2529.2000
- Perez, J., Munoz-Dorado, J., De La Rubia, T., and Martinez, J. (2002). Biodegradation and biological treatments of cellulose, hemicellulose and lignin: an overview. *Int. Microbiol.* 5, 53–63. doi: 10.1007/s10123-002-0062-3
- Porat, I., Kim, W., Hendrickson, E. L., Xia, Q., Zhang, Y., Wang, T., et al. (2006). Disruption of the operon encoding Ehb hydrogenase limits anabolic CO<sub>2</sub> assimilation in the archaeon *Methanococcus maripaludis*. *J. Bacteriol.* 188, 1373–1380. doi: 10.1128/JB.188.4.1373-1380.2006
- Purwantini, E., Gillis, T. P., and Daniels, L. (1997). Presence of F<sub>420</sub>-dependent glucose-6-phosphate dehydrogenase in *Mycobacterium* and *Nocardia* species, but absence from *Streptomyces* and *Corynebacterium* species

- and methanogenic Archaea. *FEMS Microbiol. Lett.* 146, 129–134. doi: 10.1111/j.1574-6968.1997.tb10182.x
- Ransom-Jones, E., Jones, D. L., McCarthy, A. J., and McDonald, J. E. (2012). The Fibrobacteres: an important phylum of cellulose-degrading bacteria. *Microb. Ecol.* 63, 267–281. doi: 10.1007/s00248-011-9998-1
- Raskin, L., Rittmann, B. E., and Stahl, D. A. (1996). Competition and coexistence of sulfate-reducing and methanogenic populations in anaerobic biofilms. *Appl. Environ. Microbiol.* 62, 3847–3857.
- Reeburgh, W. S. (2007). Oceanic methane biogeochemistry. *Chem. Rev.* 107, 486–513. doi: 10.1021/cr050362v
- Rospert, S., Linder, D., Ellermann, J., and Thauer, R. K. (1990). Two genetically distinct methyl-coenzyme M reductases in *Methanobacterium thermoautotrophicum* strain Marburg and delta H. *Eur. J. Biochem.* 194, 871–877. doi: 10.1111/j.1432-1033.1990.tb19481.x
- Rother, M., and Metcalf, W. W. (2004). Anaerobic growth of *Methanosarcina acetivorans* C2A on carbon monoxide: an unusual way of life for a methanogenic archaeon. *Proc. Natl. Acad. Sci. U.S.A.* 101, 16929–16934. doi: 10.1073/pnas.0407486101
- Sahakian, A. B., Jee, S. R., and Pimentel, M. (2010). Methane and the gastrointestinal tract. *Dig. Dis. Sci.* 55, 2135–2143. doi: 10.1007/s10620-009-1012-0
- Sauer, K., Harms, U., and Thauer, R. K. (1997). Methanol:coenzyme M methyltransferase from *Methanosarcina barkeri*. Purification, properties and encoding genes of the corrinoid protein MT1. *Eur. J. Biochem.* 243, 670–677. doi: 10.1111/j.1432-1033.1997.t01-1-00670.x
- Schauer, N. L., and Ferry, J. G. (1982). Properties of formate dehydrogenase in *Methanobacterium formicum*. *J. Bacteriol.* 150, 1–7.
- Schink, B. (1997). Energetics of syntrophic cooperation in methanogenic degradation. *Microbiol. Mol. Biol. Rev.* 61, 262–280.
- Schink, B., Ward, J. C., and Zeikus, J. G. (1981). Microbiology of wetwood: importance of pectin degradation and clostridium species in living trees. *Appl. Environ. Microbiol.* 42, 526–532.
- Schirmack, J., Mangelsdorf, K., Ganzert, L., Sand, W., Hillebrand-Voiculescu, A., and Wagner, D. (2014). *Methanobacterium movilense* sp. nov., a hydrogenotrophic, secondary-alcohol-utilizing methanogen from the anoxic sediment of a subsurface lake. *Int. J. Syst. Evol. Microbiol.* 64, 522–527. doi: 10.1099/ij.s.0.057224-0
- Schmitz, G., and Langmann, T. (2006). Metabolic learning in the intestine: adaptation to nutrition and luminal factors. *Horm. Metab. Res.* 38, 452–454. doi: 10.1055/s-2006-947841
- Schwarz, W. H. (2001). The cellulosome and cellulose degradation by anaerobic bacteria. *Appl. Microbiol. Biotechnol.* 56, 634–649. doi: 10.1007/s002530100710
- Schworer, B., Fernandez, V. M., Zirngibl, C., and Thauer, R. K. (1993). H<sub>2</sub>-forming N<sup>5</sup>,N<sup>10</sup>-methylene-tetrahydromethanopterin dehydrogenase from *Methanobacterium thermoautotrophicum*. Studies of the catalytic mechanism of H<sub>2</sub> formation using hydrogen isotopes. *Eur. J. Biochem.* 212, 255–261. doi: 10.1111/j.1432-1033.1993.tb17657.x
- Shima, S., Krueger, M., Weinert, T., Demmer, U., Kahnt, J., Thauer, R. K., et al. (2012). Structure of a methyl-coenzyme M reductase from Black Sea mats that oxidize methane anaerobically. *Nature* 481, 98–101. doi: 10.1038/nature10663
- Strapoc, D., Mastalerz, M., Dawson, K., Macalady, J., Callaghan, A. V., Wawrik, B., et al. (2011). Biogeochemistry of microbial coal-bed methane. *Annu. Rev. Earth Planet. Sci.* 39, 617–656. doi: 10.1146/Annurev-Earth-040610-133343
- Switzenbaum, M. S. (1983). Anaerobic treatment of wastewater recent developments. *ASM News* 49, 532–536.
- Takao, M., Oikawa, A., Eker, A. P., and Yasui, A. (1989). Expression of an Anacystis nidulans photolase gene in *Escherichia coli*; functional complementation and modified action spectrum of photoreactivation. *Photochem. Photobiol.* 50, 633–637. doi: 10.1111/j.1751-1097.1989.tb04319.x
- Tallant, T. C., Paul, L., and Krzycki, J. A. (2001). The MtsA subunit of the methylthiol:coenzyme M methyltransferase of *Methanosarcina barkeri* catalyzes both half-reactions of corrinoid-dependent dimethylsulfide: coenzyme M methyl transfer. *J. Biol. Chem.* 276, 4485–4493. doi: 10.1074/jbc.M007514200
- Te Bommelstroet, B. W., Hensgens, C. M., Keltjens, J. T., Van Der Drift, C., and Vogels, G. D. (1990). Purification and properties of 5,10-methylenetetrahydromethanopterin reductase, a coenzyme F<sub>420</sub>-dependent enzyme, from *Methanobacterium thermoautotrophicum* strain delta H. *J. Biol. Chem.* 265, 1852–1857.
- Teunissen, M. J., and Op Den Camp, H. J. (1993). Anaerobic fungi and their cellulolytic and xylanolytic enzymes. *Antonie Van Leeuwenhoek* 63, 63–76. doi: 10.1007/BF00871733
- Thauer, R. K. (1998). Biochemistry of methanogenesis: a tribute to marjory stephenson. 1998 marjory stephenson prize lecture. *Microbiology* 144(Pt 9), 2377–2406. doi: 10.1099/00221287-144-9-2377
- Thauer, R. K. (2011). Anaerobic oxidation of methane with sulfate: on the reversibility of the reactions that are catalyzed by enzymes also involved in methanogenesis from CO<sub>2</sub>. *Curr. Opin. Microbiol.* 14, 292–299. doi: 10.1016/j.mib.2011.03.003
- Thauer, R. K. (2012). The Wolfe cycle comes full circle. *Proc. Natl. Acad. Sci. U.S.A.* 109, 15084–15085. doi: 10.1073/pnas.1213193109
- Thauer, R. K., Hedderich, R., and Fischer, R. (1993). “Reactions and enzymes involved in methanogenesis from CO<sub>2</sub> and H<sub>2</sub>,” in *Methanogenesis: Ecology, Physiology, Biochemistry, and Genetics*, ed J.G. Ferry (New York, NY: Chapman & Hall), 209–252.
- Thauer, R. K., Kaster, A. K., Seedorf, H., Buckel, W., and Hedderich, R. (2008). Methanogenic archaea: ecologically relevant differences in energy conservation. *Nat. Rev. Microbiol.* 6, 579–591. doi: 10.1038/nrmicro1931
- Van Der Meijden, P., Te Bommelstroet, B. W., Poirot, C. M., Van Der Drift, C., and Vogels, G. D. (1984a). Purification and properties of methanol:5-hydroxybenzimidazolylcobamide methyltransferase from *Methanosarcina barkeri*. *J. Bacteriol.* 160, 629–635.
- Van Der Meijden, P., Van Der Lest, C., Van Der Drift, C., and Vogels, G. D. (1984b). Reductive activation of methanol: 5-hydroxybenzimidazolylcobamide methyltransferase of *Methanosarcina barkeri*. *Biochem. Biophys. Res. Commun.* 118, 760–766. doi: 10.1016/0006-291X(84)91460-8
- Vepachedu, V. R., and Ferry, J. G. (2012). Role of the fused corrinoid/methyl transfer protein CmtA during CO-dependent growth of *Methanosarcina acetivorans*. *J. Bacteriol.* 194, 4161–4168. doi: 10.1128/JB.00593-12
- Von Bunau, R., Zirngibl, C., Thauer, R. K., and Klein, A. (1991). Hydrogen-forming and coenzyme-F<sub>420</sub>-reducing methylene tetrahydromethanopterin dehydrogenase are genetically distinct enzymes in *Methanobacterium thermoautotrophicum* (Marburg). *Eur. J. Biochem.* 202, 1205–1208. doi: 10.1111/j.1432-1033.1991.tb16491.x
- Wang, M., Tomb, J. F., and Ferry, J. G. (2011). Electron transport in acetate-grown *Methanosarcina acetivorans*. *BMC Microbiol.* 11:165. doi: 10.1186/1471-2180-11-165
- Wassenaar, R. W., Daas, P. J., Geerts, W. J., Keltjens, J. T., and Van Der Drift, C. (1996). Involvement of methyltransferase-activating protein and methyltransferase 2 isoenzyme II in methylamine:coenzyme M methyltransferase reactions in *Methanosarcina barkeri* Fusaro. *J. Bacteriol.* 178, 6937–6944.
- Wassenaar, R. W., Keltjens, J. T., Van Der Drift, C., and Vogels, G. D. (1998). Purification and characterization of dimethylamine:5-hydroxybenzimidazolyl-cobamide methyltransferase from *Methanosarcina barkeri* Fusaro. *Eur. J. Biochem.* 253, 692–697. doi: 10.1046/j.1432-1327.1998.2530692.x
- Weimer, P. J. (1998). Manipulating ruminal fermentation: a microbial ecological perspective. *J. Anim. Sci.* 76, 3114–3122.
- Welte, C., and Deppenmeier, U. (2014). Bioenergetics and anaerobic respiratory chains of aceticlastic methanogens. *Biochim. Biophys. Acta* 1837, 1130–1147. doi: 10.1016/j.bbabi.2013.12.002
- Whitman, W. B., Shieh, J., Sohn, S., Cars, D. S., and Premachandran, U. (1986). Isolation and characterization of 22 mesophilic methanococci. *Sys. Appl. Microbiol.* 7, 235–240. doi: 10.1016/S0723-2020(86)80012-1
- Widdel, F. (1986). Growth of methanogenic bacteria in pure culture with 2-propanol and other alcohols as hydrogen donors. *Appl. Environ. Microbiol.* 51, 1056–1062.
- Widdel, F. (1988). “Microbiology and ecology of sulfate- and sulfur-reducing bacteria,” in *Biology of Anaerobic Microorganisms*, ed A. Zehnder (New York, NY: John Wiley & Sons), 469–585.
- Widdel, F., and Wolfe, R. S. (1989). Expression of secondary alcohol dehydrogenase in methanogenic bacteria and purification of the F<sub>420</sub>-specific enzyme from *Methanogenium thermophilum* strain TCI. *Arch. Microbiol.* 152, 322–328. doi: 10.1007/BF00425168
- Wolfe, R. S. (1991). My kind of biology. *Annu. Rev. Microbiol.* 45, 1–35. doi: 10.1146/annurev.mi.45.100191.000245

- Wolfe, R. S. (1992). Biochemistry of methanogenesis. *Biochem. Soc. Symp.* 58, 41–49.
- Wolfe, R. S. (1993). “An historical overview of methanogenesis,” in *Methanogenesis: Ecology, Physiology, Biochemistry, and Genetics*, ed J. G. Ferry (New York, NY: Chapman & Hall), 1–32.
- Wolin, M. J. (1981). Fermentation in the rumen and human large intestine. *Science* 213, 1463–1468. doi: 10.1126/science.7280665
- Wolin, M. J., and Miller, T. L. (1983). Interactions of microbial populations in cellulose fermentation. *Fed. Proc.* 42, 109–113.
- Wood, G. E., Haydock, A. K., and Leigh, J. A. (2003). Function and regulation of the formate dehydrogenase genes of the methanogenic archaeon *Methanococcus maripaludis*. *J. Bacteriol.* 185, 2548–2554. doi: 10.1128/JB.185.8.2548-2554.2003
- Yu, X., Doroghazi, J. R., Janga, S. C., Zhang, J. K., Circello, B., Griffin, B. M., et al. (2013). Diversity and abundance of phosphonate biosynthetic genes in nature. *Proc. Natl. Acad. Sci. U.S.A.* 110, 20759–20764. doi: 10.1073/pnas.1315107110
- Zinder, S. H. (1993). “Physiological ecology of methanogens,” in *Methanogenesis: Ecology, Physiology, Biochemistry, and Genetics*, ed J. G. Ferry (New York, NY: Chapman and Hall), 128–206.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 09 July 2014; accepted: 05 November 2014; published online: 03 December 2014.

Citation: Purwantini E, Torto-Alalibo T, Lomax J, Setubal JC, Tyler BM and Mukhopadhyay B (2014) Genetic resources for methane production from biomass described with the Gene Ontology. *Front. Microbiol.* 5:634. doi: 10.3389/fmicb.2014.00634

This article was submitted to *Microbial Physiology and Metabolism*, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Purwantini, Torto-Alalibo, Lomax, Setubal, Tyler and Mukhopadhyay. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Genetic resources for advanced biofuel production described with the Gene Ontology

Trudy Torto-Alalibo<sup>1,2</sup>, Endang Purwantini<sup>1,2</sup>, Jane Lomax<sup>3</sup>, João C. Setubal<sup>2,4</sup>, Biswarup Mukhopadhyay<sup>1,2,5</sup> and Brett M. Tyler<sup>2,6</sup> \*

<sup>1</sup> Department of Biochemistry, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

<sup>2</sup> Virginia Bioinformatics Institute, Virginia Polytechnic Institute and State University, Blacksburg, VA, USA

<sup>3</sup> European Molecular Biology Laboratory, European Bioinformatics Institute (EMBL-EBI), Wellcome Trust Genome Campus, Cambridge, UK

<sup>4</sup> Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil

<sup>5</sup> Department of Biological Sciences, Oregon State University, Corvallis, OR, USA

<sup>6</sup> Center for Genome Research and Biocomputing, Oregon State University, Corvallis, OR, USA

## Edited by:

Katherine M. Pappas, University of Athens, Greece

## Reviewed by:

Kesen Ma, University of Waterloo, Canada

Christopher J. Brigham, University of Massachusetts at Dartmouth, USA

David E. Graham, Oak Ridge National Laboratory, USA

## \*Correspondence:

Brett M. Tyler, Center for Genome Research and Biocomputing, Oregon State University, 4750 Campus Way ALS3021, Corvallis, OR 97331, USA  
e-mail: brett.tyler@oregonstate.edu

Dramatic increases in research in the area of microbial biofuel production coupled with high-throughput data generation on bioenergy-related microbes has led to a deluge of information in the scientific literature and in databases. Consolidating this information and making it easily accessible requires a unified vocabulary. The Gene Ontology (GO) fulfills that requirement, as it is a well-developed structured vocabulary that describes the activities and locations of gene products in a consistent manner across all kingdoms of life. The Microbial ENergy processes Gene Ontology (<http://www.mengo.biochem.vt.edu>) project is extending the GO to include new terms to describe microbial processes of interest to bioenergy production. Our effort has added over 600 bioenergy related terms to the Gene Ontology. These terms will aid in the comprehensive annotation of gene products from diverse energy-related microbial genomes. An area of microbial energy research that has received a lot of attention is microbial production of advanced biofuels. These include alcohols such as butanol, isopropanol, isobutanol, and fuels derived from fatty acids, isoprenoids, and polyhydroxyalkanoates. These fuels are superior to first generation biofuels (ethanol and biodiesel esterified from vegetable oil or animal fat), can be generated from non-food feedstock sources, can be used as supplements or substitutes for gasoline, diesel and jet fuels, and can be stored and distributed using existing infrastructure. Here we review the roles of genes associated with synthesis of advanced biofuels, and at the same time introduce the use of the GO to describe the functions of these genes in a standardized way.

**Keywords: Gene Ontology, advanced biofuels, synthetic biology, cellulosome, advanced alcohols, fatty acid-derived fuel, isoprenoid-derived fuel**

## INTRODUCTION

Depletion of the world's fossil fuel resources and environmental concerns associated with the emission of greenhouse gases has fueled interest in renewable and environmentally friendly alternatives (Köne and Büke, 2010; Hansen et al., 2013; Suranovic, 2013). In this context, advanced biofuels have been of growing interest as these compounds can be generated from non-food cellulosic biomass, can be added directly to gasoline or diesel or sometimes used as stand-alone fuel, and can be stored and distributed using existing infrastructure (Dürre, 2007; Mehta et al., 2010; Weber et al., 2010). Advanced biofuels include alcohols such as butanol, isopropanol, and isobutanol, and fuels derived from fatty acids, isoprenoids, and polyhydroxyalkanoates (PHAs). Advanced biofuels from lignocellulose feedstock begins with biomass deconstruction. The cellulosic component of the biomass is degraded into pentoses and hexoses. The multi-enzyme complexes involved in the degradation of cellulosic biomass are discussed in this review. The native and engineered pathways leading to the production of advanced biofuels have been

studied extensively. As such the primary published literature is rich in information derived from research on these fuels but this information has yet to be aggregated in a manner that is easily accessible and amenable to computational analysis. The well-established Gene Ontology (GO; Ashburner et al., 2000) provides a basis for harnessing this information. The GO provides sets of standardized terms to describe molecular functions, biological processes and cellular components across all kingdoms of life (Ashburner et al., 2000; Harris et al., 2004; Torto-Alalibo et al., 2009; Dimmer et al., 2012). The Microbial ENergy processes Gene Ontology (MENGO) consortium<sup>1</sup>, an associate of the GO, initiated an effort in 2011 to extend the GO by developing missing terms associated with bioenergy processes. This effort has generated over 600 bioenergy-related GO terms, most of which are described in this review and the MENGO website<sup>1</sup>. These new terms, together with existing ones in the GO, were used to annotate gene products involved in bioenergy processes. Our

<sup>1</sup><http://www.mengo.biochem.vt.edu>

emphasis is on selected native pathways leading to the generation of advanced biofuels. The scope of the review is summarized in Figure 1.

## THE GENE ONTOLOGY

The GO is a structured, species-neutral ontology that describes the attributes of gene products. The GO consists of three distinct ontologies: “GO:0003674 molecular function,” “GO:0008150 biological process” and “GO:0005575 cellular component,” which are made up of terms (otherwise known as classes) arranged in a graph structure. Terms in the graph are related to one another via relationships, which are of a given type. Examples of relations used in GO include “is\_a,” “part\_of,” “regulates,” “has\_part” and “occurs\_in.” Some relationships, namely “is\_a” (is a type of) and “part\_of,” are taxonomic such that “child terms” are more specific than the more general “parent” terms (Ashburner et al., 2000; Harris et al., 2004; Figure 2). Terms can have one or more “parent terms” and gene products annotated with specific child terms are automatically associated with the corresponding “parent terms” in the graph. Currently, GO uses the CHEBI database<sup>2</sup> as the source of all primary chemical names (Hill et al., 2013). Alternative names are added as synonyms. We made several additions to the CHEBI as it lacked most of the chemical entities used by the MENGO consortium. The GO also actively interacts

with and maintains mappings to EC, MetaCyc, Rhea, Reactome, and several other systems<sup>3</sup>. GO is working with Rhea and Reactome on a system of automatic import for enzymatic reactions to avoid duplication of curation effort and to further improve interoperability.

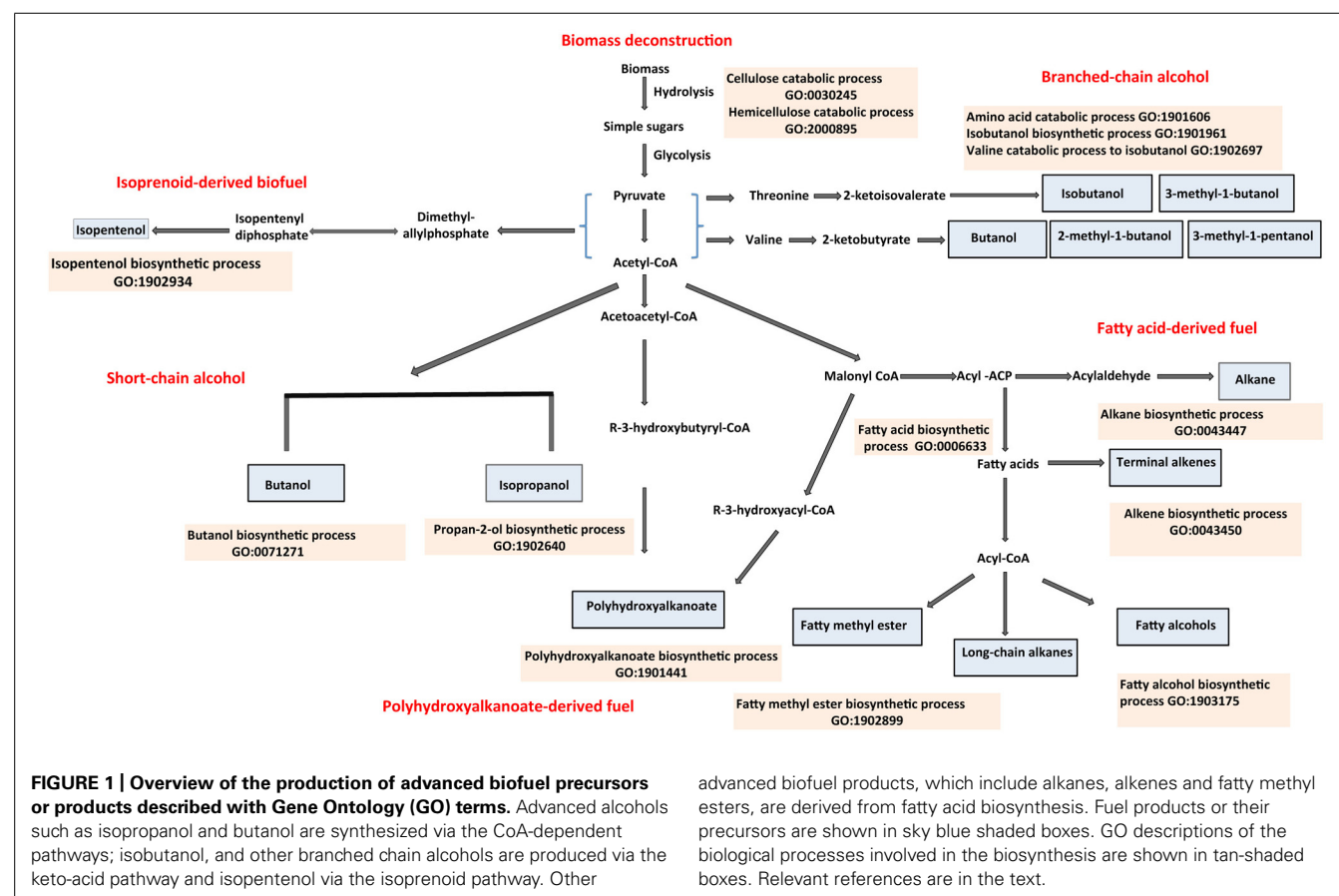
The GO is widely recognized as a powerful tool for the annotation of gene products of all organisms including important biofuel producers like *Saccharomyces cerevisiae* and cyanobacteria (Christie et al., 2009; Beck et al., 2012). The ongoing maintenance and expansion of the GO results from international collaborative efforts that are managed by the Gene Ontology consortium (GOC). The GOC encourages and works with associated groups, such as MENGO, on term development and annotation in focused subject areas.

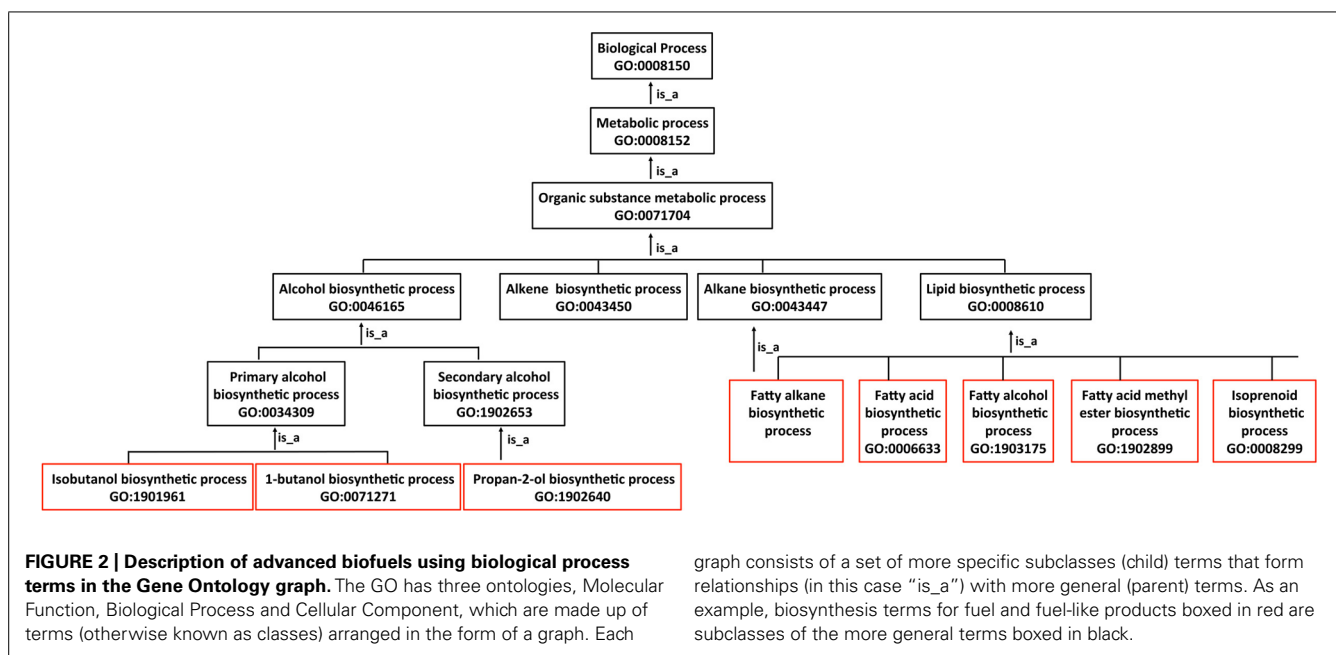
The MENGO team has produced over 600 bioenergy-related terms<sup>4,5</sup>. Included in this set are some key terms relevant to the description of advanced biofuel production. For example, terms for the microbial production of isobutanol “GO:1901961 isobutanol biosynthetic process,” isopropanol “GO:1902640 propan-2-ol biosynthetic process” and isopentenol “GO:1902934 isopentenol biosynthetic process” were additions made by the MENGO group. When specific gene products are annotated using GO terms, the

<sup>3</sup>[http://wiki.geneontology.org/index.php/Enzymes\\_and\\_EC\\_mappings](http://wiki.geneontology.org/index.php/Enzymes_and_EC_mappings)

<sup>4</sup><http://www.mengo.biochem.vt.edu/>

<sup>5</sup><http://tinyurl.com/o2l7nsy>





information is stored in tabular form as “gene association” statements<sup>6</sup>, which are consolidated into a central GO database to enable public access to the annotations (Christie et al., 2009; Beck et al., 2012). Information stored in a gene association statement includes the gene product ID, the GO terms specifying the function of the gene product, its biological context and/or the location of activity. Experimental or bioinformatics evidence is represented by terms from the Evidence Code Ontology<sup>7</sup> (ECO). Also included is the taxonomic source of the gene product and the reference(s) (often a PubMed identity number) indicating the publication(s) from which information was obtained. Selected gene products, annotated with GO bioenergy related terms are provided in **Table 1** and Supplementary Table 1. All other bioenergy related annotations made by MENGO can be found at the MENGO website<sup>8</sup> or the GO website<sup>9</sup> (AMIGO).

The GO has also been used extensively in the analysis of high-throughput data including genomic, transcriptomic, proteomic, and metagenomic data (Marinković et al., 2012; Molnar et al., 2012; Benso et al., 2013; Plewniak et al., 2013; Schmidt et al., 2014). Currently, the GO is limited to terms for describing natural processes (Blake and Harris, 2008). It does not provide terms for non-natural processes such as disease states, nor does it provide terms for novel biological functions that have been produced by genetic manipulations or synthetic biology. This is a current limitation of the GO in the context of bioenergy research. The MENGO team is currently discussing with the community the creation of a complementary set of GO-like terms suitable for synthetic biological functions (which we provisionally call SYNGO). To

underscore the anticipated broad scope of SYNGO, at the end of this review we provide brief descriptions of some synthetic processes and cell parts that are driving the development of this resource.

## MULTIENZYME COMPLEXES FOR DECONSTRUCTION OF BIOMASS: CELLULOSOMES AND XYLANOSOMES

Complete hydrolysis of cellulose to glucose requires the synergistic action of three general types of glycoside hydrolases (Doi, 2008; Gomez del Pulgar and Saadeddin, 2014): (i) Cellulases or endo-1,4- $\beta$ -glucanases randomly hydrolyzes internal bonds in a cellulose chain releasing products of varying chain lengths. Non-processive cellulases are truly random and processive cellulases, upon binding a cellulose chain continue cutting through the bound substrate. (ii) Exo-1,4- $\beta$ -glucanases “GO:0031217 glucan 1,4-beta-glucosidase activity” works from either the reducing or non-reducing end of the cellulose polymer to release cellobiose. They are sometimes called exocellulases and their processive forms are known as cellobiohydrolases “GO:0016162 cellulose 1,4-beta-cellobiosidase activity.” A variation of this group, glucohydrolase “GO:0080079 cellobiose glucosidase activity,” acts on the non-reducing end and releases glucose. (iii)  $\beta$ -glucosidases “GO:0008422  $\beta$ -glucosidase activity” hydrolyze cellobiose released by other enzymes to glucose and are also known as cellobiases (Wilson, 2011, 2012). Activities of these enzymes on oligosaccharides have also been reported and their actions from the non-reducing end release glucose, endoglucanases, exoglucanases, and cellobiases (Chang et al., 2013).

The major component of hemicellulose is xylan. Two key enzymes collectively called xylanases “GO:0097599 xylanase activity” are responsible for hydrolysis of xylan. Endo-xylanase (endo-1,4- $\beta$ -xylanase) “GO:0031176 endo-1,4- $\beta$ -xylanase activity” acts on the homopolymeric backbone of 1,4-linked  $\beta$ -D-xylopyranose

<sup>6</sup><http://www.geneontology.org/GO.format.annotation.shtml>

<sup>7</sup><http://www.evidenceontology.org/>

<sup>8</sup><http://www.mengo.biochem.vt.edu/>

<sup>9</sup><http://tinyurl.com/o2l7nys>

**Table 1 | Abbreviated Gene Ontology table showing annotations of selected experimentally characterized gene products associated with advanced biofuel production.**

Gene product	Ontology*	GO term	Evidence†	Taxon	Reference
<b>Butanol biosynthetic process</b>					
GO:0071271					
Thl	F	GO:0003985: acetyl-CoA acetyltransferase activity	IDA	1488	PMID:16347774
Hbd	F	GO:0008691: 3-hydroxybutyryl-CoA dehydrogenase activity	IDA	1520	PMID:1444364
Crt	F	GO:0003859: 3-hydroxybutyryl-CoA dehydratase activity	IDA	1488	PMID:16346566
Bcd	F	GO:0004085: butyryl-CoA dehydrogenase	IDA	1488	PMID:8655474
BYDH	F	GO:0004029: aldehyde dehydrogenase (NAD) activity	IDA	1488	PMID:8300540
BDH	F	GO:0004022: alcohol dehydrogenase activity	IDA	272562	PMID:11790753
<b>Isobutanol biosynthetic process</b>					
GO:1901961					
Bat1	F	GO:0004084: branched-chain aminotransferase activity	IDA	4932	PMID:8798704
Bat1	C	GO:0005739: mitochondrion	IDA	4932	PMID:8798704
Bat2	F	GO:0004084: branched-chain aminotransferase	IDA	4932	PMID:8798704
Bat2	C	GO:0005829: cytosol	IDA	4932	PMID:8798704
PDC	F	GO:0004737: pyruvate decarboxylase	IMP	4932	PMID:9546164
ADH2	F	GO:0004022: alcohol dehydrogenase activity	IMP	4932	PMID:3546317
<b>Isopropanol biosynthetic process</b>					
GO:1902639					
THL	F	GO:0003985: acetyl-CoA acetyltransferase activity	IDA	1488	PMID:16347774
CTF	F	GO:0008410: CoA-transferase activity	IDA	1488	PMID:2719476
ADC	F	GO:0047602: Acetoacetate decarboxylase activity	IDA	272562	PMID:2268159
SADH	F	GO:0004022: alcohol dehydrogenase activity	IDA	1520	PMID:8349550
<b>Tolerance</b>					
ARI1	P	GO:1990370: process resulting in tolerance to aldehyde			
Spo0A	P	GO:1990336: process resulting in tolerance to butanol	IDA	1488	PMID:15028679
Cfa	P	GO:1990336: process resulting in tolerance response to butanol	IDA	1350	PMID:24014527
Cfa	P	GO:1990337: process resulting in tolerance to isobutanol	IDA	1350	PMID:24014527
SII0690	P	GO:1990336: process resulting in tolerance to butanol	IMP	1148	PMID:23883549
Slr0947	P	GO:1990336: process resulting in tolerance to butanol	IMP	1148	PMID:23883549
slr1295	P	GO:1990336: process resulting in tolerance to butanol	IMP	1148	PMID:23883549

\* Molecular function (F), biological process (P), cellular component (C).

† Inferred from direct assay (IDA), inferred from mutant phenotype (IMP).

producing xylooligomers, and  $\beta$ -xylosidase (xylan-1,4- $\beta$ -xylosidase) “GO:0009044 xylan-1,4- $\beta$ -xylosidase activity” act on the xylooligomers releasing xylose (Biely et al., 1985; Ahmed et al., 2009). Additionally, accessory enzymes such as acetyl xylan esterases “GO:0046555 acetylxylan esterase activity” act to remove the side chain substitution along the xylan backbone (Juturu and Wu, 2013). The GO uses the CHEBI database<sup>10</sup> as the source of all primary chemical names and all alternate names are listed as synonyms. These hydrolytic enzymes exist as free independent enzymes or as multienzyme complexes called cellulosomes “GO:0043263 cellulosome” and xylanosomes “GO:1990358 xylanosome” (Jiang et al., 2004, 2006). This section will focus on using the GO to describe these multienzyme

complexes that mediate degradation of polysaccharides. Other aspects of deconstruction of biomass for biofuel can be found in several other reviews (Dodd and Cann, 2009; Blanch et al., 2011; Chundawat et al., 2011; Blumer-Schuette et al., 2014).

The cellulosome of *Clostridium thermocellum* serve as a paradigm for this enzymatic nanomachine, and thus a model for studies of the structure and assembly process. The central component of the cellulosome is a non-catalytic “scaffoldin” subunit, which mediates a highly specific interaction between the enzyme-bearing type I dockerin modules and the resident type I cohesin modules (Bras et al., 2012; Smith and Bayer, 2013; Hong et al., 2014). The cellulose-binding domain (CBD) of the scaffoldin subunit aids in binding the complex to the cellulosic substrate. The type II dockerin module, which forms

<sup>10</sup><http://www.ebi.ac.uk/chebi/>

part of the primary scaffoldin in turn, binds to the type II cohesin of anchoring scaffoldins, which connect to the bacteria cell surface (**Figure 3**; Bayer et al., 1985; Bayer and Lamed, 1986). Recently, a third type of dockerin-cohesin interaction (type III) has been characterized in *Ruminococcus flavefaciens*. The type III dockerins and cohesins show a high degree of sequence divergence compared to their type I and II counterparts (Rincon et al., 2005; Karpol et al., 2013). Prior to our work, GO terms related to cellulosomes were “GO:0043263 cellulosome,” its regulation terms, and the different hydrolytic enzyme terms. However, the latter terms were not necessarily linked to the cellulosome term in a way that clearly indicated their associations. We used the “domain binding” aspect of the GO to make the multiple functions of the scaffoldin more obvious. To start with, the ability of the non-enzymatic scaffoldin to bind to several cellulose degradative enzymes via the cohesin domain makes it a complex. Therefore we introduced the term “GO:1990296 scaffoldin complex” which has a “part of” relationship to “GO:0043263 cellulosome.” In addition, domain-specific terms were also developed, namely: “GO:1990311 type I cohesin domain binding,” “GO:1990308 type I dockerin domain binding,” “GO:1990312 type II cohesin domain binding,” “GO:1990309 type II dockerin domain binding,” “GO:1990313 type III cohesin domain binding” and “GO:1990310 type III dockerin domain binding.” Gene products binding to each of

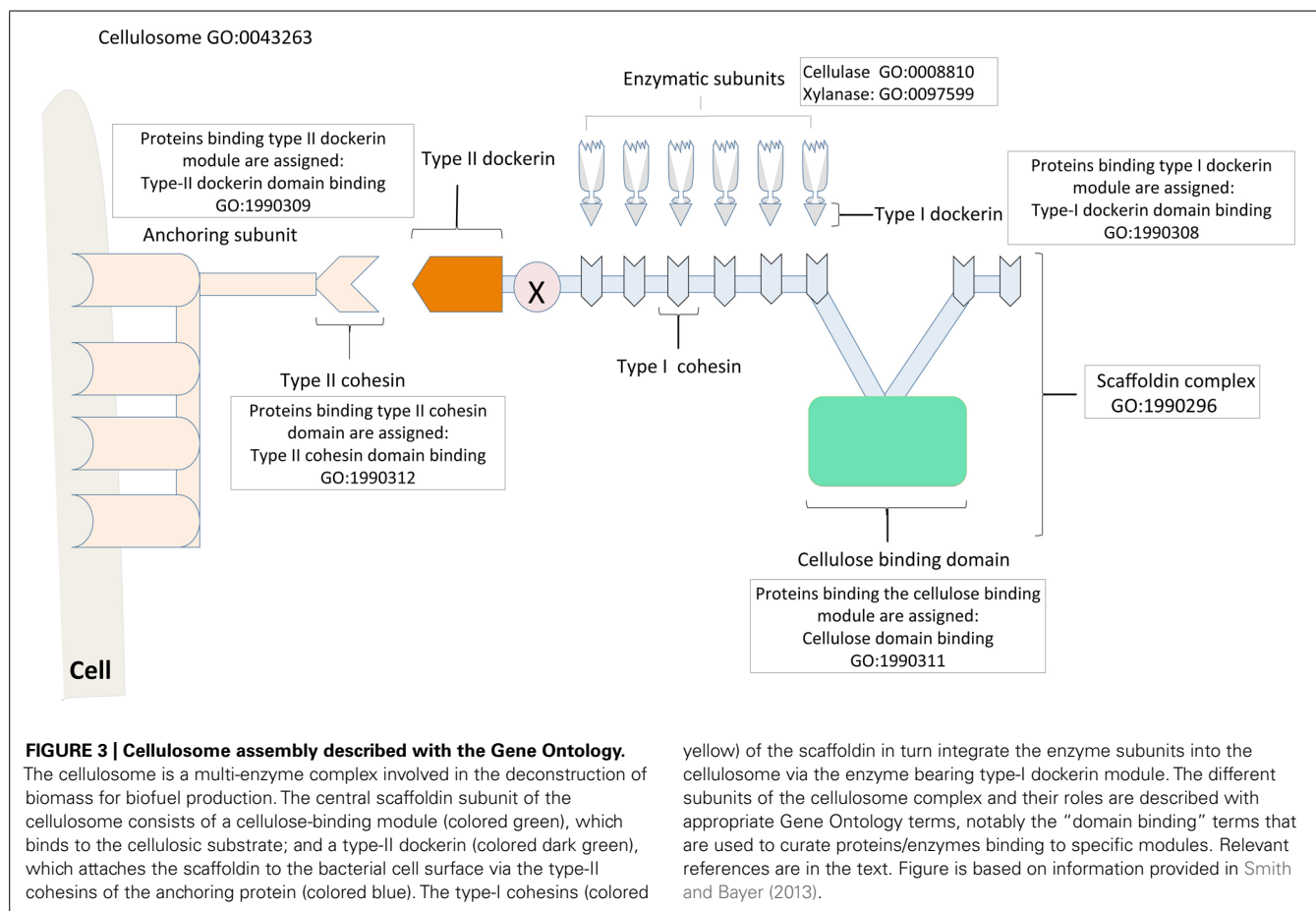
these modules were assigned the appropriate domain binding term.

The xylanosome occurs naturally in anaerobic fungi and bacteria. However, besides containing several hemicellulases, not much is known about its assembly. We introduced the term “GO:1990358 xylanosome” into the GO. A modular nature, if any, has not been described for the xylanosome. Therefore for now, the only way to identify components of the xylanosome will be at the annotation level where the hemicellulases are annotated with their unique enzymatic activity terms using the Molecular Function ontology and also with the cellular component term “GO:1990358 xylanosome” where there is experimental evidence of an association with the xylanosome.

The synergy and combined action of hemicellulases and cellulases within these multi-enzyme systems makes them efficient in the degradation of biomass (Morais et al., 2010). Systematic annotation of the ever-increasing repertoire of dockerin-cohesin pairs of the cellulosome, using the GO terms described above, will serve to inform their use in engineering more efficient cellulosomes, xylanosomes, and cellulosome-like complexes.

## PRODUCTION OF ISOPROPANOL AND BUTANOL

Clostridia can produce isopropanol and butanol from sugars via pyruvate and acetyl-CoA [for a review of early work in this area see Stephenson (1949) and references therein]. Short chain

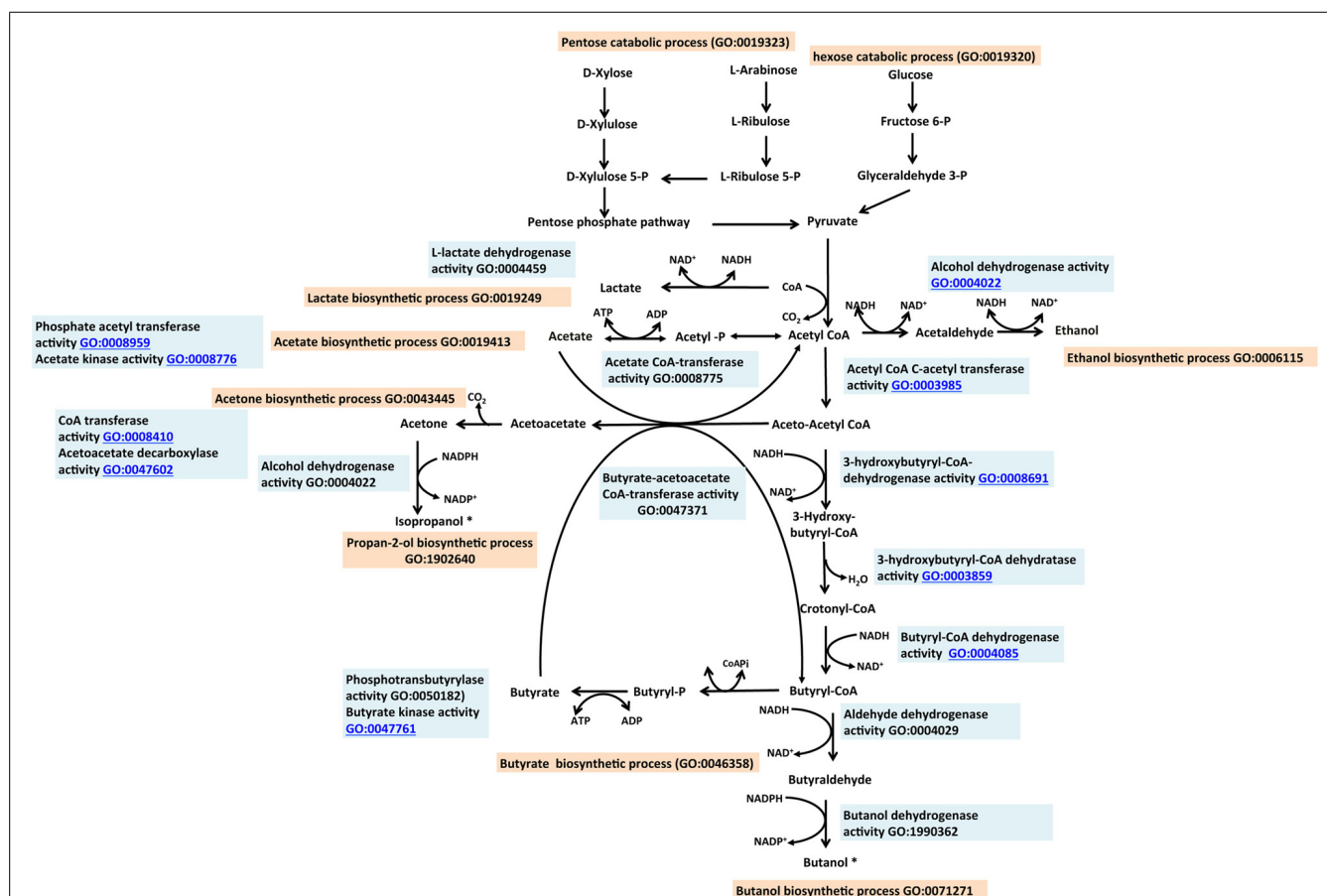




alcohols proposed as advanced alcohols include those with C3–C4 units such as isopropanol and butanol (Gronenberg et al., 2013; Xue et al., 2013). These alcohols have superior properties relative to ethanol in terms of energy density or energy content, and ease of storage and distribution. Similar to ethanol, these can also be produced from renewable non-food feedstocks. Traditionally, butanol and isopropanol are produced as a mixture by some clostridial species such as *Clostridium acetobutylicum* possessing the acetone/isopropanol butanol–ethanol (A/IBE) pathway (Figure 4). Following breakdown of the pentose sugars via the pentose phosphate pathway “GO:0019323 pentose catabolic process” and hexose sugars via glycolysis, “GO:0006096 glycolysis” to pyruvate, the A/IBE process can be divided into two distinct metabolic phases: one phase involves the formation of the organic acids, butyrate and acetate (called “acidogenesis”). In the second phase, these organic acids may act as substrates for the biosynthesis of acetone and alcohols (butanol and ethanol). The latter phase is referred to as “solventogenesis.” The production of butyrate and acetate are represented in the GO as “GO:0046358 butyrate biosynthetic process” and “GO:0019413 acetate biosynthetic process,” respectively. These two child terms fall under

the parent term “GO:0016053 organic acid biosynthetic process.” Solventogenic enzymes and gene products associated with the formation of acetone, butanol, ethanol, and isopropanol are annotated with appropriate GO molecular function terms and also with the biological process terms “GO:0043445 acetone biosynthetic process,” “GO:0071271 1-butanol biosynthetic process,” “GO:0006115 ethanol biosynthetic process” and “GO:1902640 isopropanol biosynthetic process,” respectively. The processes represented by these terms fall under “GO:0042181 ketone biosynthetic process” and “GO:0046165 alcohol biosynthetic process.” The terms for organic acid and solvent productions have synonyms, “acidogenesis” and “solventogenesis,” respectively. The addition of synonyms is a feature of the GO that facilitates easy access to concepts represented by diverse descriptions in the literature.

In *Clostridium*, the physiological transition from acidogenesis to solventogenesis is associated with the stationary phase of the organism’s growth cycle. The *Clostridium* Spo0A gene product has been shown to regulate solventogenesis (Ravagnani et al., 2000; Harris et al., 2002; Dürre and Hollergschwandner, 2004; Hu et al., 2011) and activate the downstream genes of the acidogenic process,



**FIGURE 4 | The isopropanol, butanol ethanol (IBE) pathway described with the Gene Ontology.** Butanol and propan-2-ol (asterisk) designated as advanced biofuels are naturally produced by the IBE pathway. Role of enzymes are labeled with appropriate GO Molecular

Function term (sky blue). Processes leading to the production of each alcohol are designated with appropriate GO Biological Process term (tan). Relevant references are in the text. Figure is based on information provided in Lee et al. (2008b).

in addition to controlling sporulation “GO:0043934 sporulation.” We annotated Spo0A with the term “GO:1902930 regulation of alcohol biosynthetic process” which has “regulation of solventogenesis” as a synonym, as well as with the term “GO:0043937 regulation of sporulation.” Another gene product, SpoIIE was shown to control sporulation in *C. acetobutylicum* but was not associated with solvent synthesis (Scotcher and Bennett, 2005; Bi et al., 2011). The multiple roles of Spo0A are thus documented as separate entries in the GO association table with its role in sporulation overlapping that of SpoIIE. Biosynthesis of butanol and isopropanol share the same metabolic pathway from pyruvate to acetyl-CoA and thereafter follow respective branches (Figure 4).

Microbial production of butanol occurs naturally in certain *Clostridium* species as part of the acetone butanol ethanol (ABE) pathway (Lee et al., 2008a; Qureshi et al., 2010). Some of these bacteria can use both pentose and hexose sugars as substrates. Typically, the breakdown of pentose and hexose sugars to pyruvate follows different paths but subsequent steps from pyruvate to butanol production may comprise the same reactions. The GO terms “GO:1990284 hexose catabolic process to 1-butanol” and “GO:1990290 pentose catabolic process to 1-butanol” are child terms of “GO:0071271 1-butanol biosynthetic process.” In the so-called CoA-dependent pathway for butanol production, the generation of butanol is initiated with two molecules of acetyl-CoA. Six enzymes encoded by seven genes mediate reactions leading to the production of butanol. Two molecules of acetyl-CoA are condensed, reduced, and dehydrated to form crotonyl-CoA, which is then reduced to butanol. The enzymes responsible for these reactions are shown in Figure 4 and Table 1, with appropriate GO annotations. The key enzymes leading to butanol production are each described with a distinct molecular function term. The transition from acidogenesis to solventogenesis in some *Clostridium* species occurs at low pH as a result of acid accumulation. Butyrate, produced during acidogenesis, becomes a substrate for butanol production in the solventogenesis phase. This process is mediated by the enzymes butyrate acetoacetate CoA-transferase “GO:0047371 butyrate acetoacetate CoA-transferase activity,” butyraldehyde dehydrogenase and butanol dehydrogenase “GO:1990362 butanol dehydrogenase activity.”

Isopropanol is naturally produced by *Clostridium beijerinckii* and *Clostridium aurantibutyricum* (George et al., 1983), although the natural yield is very low (Chen and Hiu, 1986). It can be used directly as an additive to gasoline or as a feedstock for biodiesel production (Lee et al., 1995); in the latter case esterification with isopropanol is used to reduce the chances of crystallization at low temperatures. The native pathway involves the condensation of two molecules of acetyl-CoA into a molecule of acetoacetyl-CoA; the coenzyme A is then transferred to acetate or butyrate, catalyzed by CoA transferase “GO:0008410 CoA-transferase activity.” The two child terms of CoA transferase are “GO:0008775 acetate CoA transferase” and “GO:0047371 butyrate-acetoacetate CoA-transferase.” A reaction catalyzed by acetoacetate decarboxylase “GO:0047602 acetoacetate decarboxylase activity” converts acetoacetate to acetone. Subsequently acetone is converted to isopropanol in a

NADPH-dependent reaction catalyzed by a secondary alcohol dehydrogenase (ADH) “GO:0050009 isopropanol dehydrogenase activity.”

## ISOBUTANOL PRODUCTION FROM VALINE AND GLYCINE

The branched-chain alcohol isobutanol, (2-methylpropan-1-ol) exhibits superior physiochemical properties similar to butanol. Isobutanol is naturally produced during fermentation by *S. cerevisiae* albeit in low amounts (Dickinson et al., 1998; Branduardi et al., 2013). Dickinson and coworkers examined the metabolic pathways used in the degradation of valine to isobutanol. Catalytic breakdown of valine to isobutanol is mediated by the Ehrlich pathway, which starts with transamination of valine to  $\alpha$ -ketovalerate via the branched chain amino acid aminotransferase, Bat2, which is also known as Twt2p and Eca40p. The alternate names are curated in the GO association table under “database object symbol synonyms.” The location of activity of gene products is described with the GO Cellular Component ontology. Bat2 has been shown to be localized in the cytosol “GO:0005829 cytosol” and a homolog Bat1 is localized to the mitochondrial matrix “GO:0005759 mitochondrial matrix,” both having the same molecular function “GO:0004084 branched chain amino acid aminotransferase” (Table 1). The subsequent decarboxylation step resulting in the formation of isobutyraldehyde is mediated by ketoacid decarboxylase. All of the three isozymes of pyruvate decarboxylase (Pdc1p, Pdc5p and Pdc6p) have been shown to be capable of decarboxylating  $\alpha$ -ketovalerate (Dickinson et al., 1998). Finally a reduction step mediated by ADH converts the aldehyde to isobutanol (Dickinson et al., 1998). The PDCs are each assigned the same GO Molecular Function terms “GO:0004737 pyruvate decarboxylase activity” in separate entries in the GO association table and the aldehyde dehydrogenase gene is assigned “GO:0004022 ADH activity.” All these gene products are also associated with the biological process term “GO:1901961 isobutanol biosynthetic process.”

Another pathway was recently discovered in *S. cerevisiae* for *de novo* isobutanol biosynthesis. A study by Villas-Bôas et al. (2005) suggested that glycine deamination led to the generation of glyoxylate and subsequently the formation of  $\alpha$ -ketovalerate and  $\alpha$ -isoketovalerate. Based on the above work, Branduardi et al. (2013) conducted a study to decipher the components of the pathways leading to the formation of butanol and isobutanol with glycine as substrate. Briefly, glycine is converted into serine through serine hydroxymethyltransferase (Shm2; McNeil et al., 1994), which is deaminated by serine deaminase CHA1 to form pyruvate. Pyruvate is then converted to isoketovalerate and then to isobutanol following the reactions as described above. These findings emphasize that it is important to define formation of a product based on the substrate utilized, as sometimes the same end is reached from multiple starting points. The terms “GO:1902697 valine catabolic process to isobutanol” and “glycine catabolic process to isobutanol” can appropriately capture this difference and these two terms are linked to the parent term “GO:1901961 isobutanol biosynthetic process.” The pathway also produces butanol, but since a high level of isobutanol production was observed, McNeil et al. (1994) hypothesized that carbon flux was being shunted away

from butanol production toward isobutanol production by the isomerization of  $\alpha$ -ketovalerate into  $\alpha$ -isoketovalerate. Analogous to these natural processes in *S. cerevisiae*, Atsumi et al. (2008) employed a metabolic engineering approach, which involved diverting 2-keto acid intermediates from the amino acid biosynthetic pathway of *Escherichia coli* to produce other branched-chain alcohols such as 2-methyl-1-butanol, 3-methyl-1-butanol and 2-phenylethanol.

## FATTY ACID BIOSYNTHESIS AND FATTY ACID-DERIVED BIOFUELS

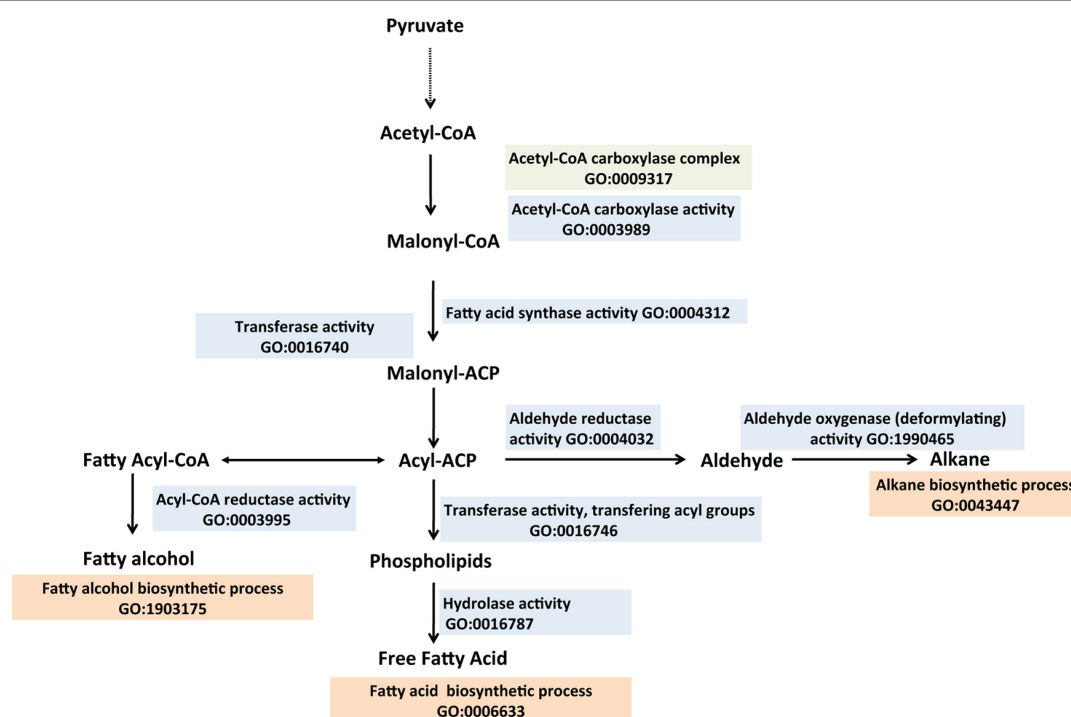
Fatty acid metabolism has attracted the most attention as a biological route to convert sugars to liquid transportation fuels (Zhou and Zhao, 2011; Lennen and Pfleger, 2013; Janssen and Steinbuchel, 2014). It is particularly suited to provide precursors for advanced biofuel because of its high efficiency and the high-energy content of the end product (Atsumi and Liao, 2008; Rude and Schirmer, 2009; Peralta-Yahya and Keasling, 2010; Wen et al., 2013). Additionally, natural metabolic pathways have been identified that convert these precursors into the biofuel product. Generally, the first committed step in fatty acid biosynthesis in *E. coli* is the conversion of acetyl-CoA to malonyl-CoA catalyzed by acetyl-CoA carboxylase (AccABCD) “GO:0003989 acetyl-CoA carboxylase activity,” which is a protein complex (GO:0009317 acetyl CoA carboxylase complex) comprising four subunits (Li and Cronan, 1992). Malonyl-CoA is then transferred to acyl carrier protein (ACP) via a malonyl-CoA:ACP transacylase (FabD, GO:0004314 malonyl-CoA:ACP transacylase activity; Campbell and Cronan, 2001; Oefner et al., 2006). Subsequently cycles of fatty acid elongation are initiated by malonyl-ACP and acetyl CoA catalyzed by  $\beta$ -ketoacyl-ACP synthase III (FabH, GO:0033818 beta-ketoacyl-ACP synthase III activity; Lai and Cronan, 2003). FabB and FabF use acyl-ACP as substrates to initiate successive chain elongation reactions (Feng and Cronan, 2009). A  $\beta$ -hydroxyacyl-ACP is the product of the second step in elongation, generated by  $\beta$ -ketoacyl ACP reductase (FabG, GO:0004316 beta-ketoacyl ACP reductase activity) while expending one molecule of NADPH. FabA and FabZ catalyze formation of an enoyl-ACP (Mohan et al., 1994). Following enoyl-ACP formation, the last intermediate in the fatty acid elongation cycle, acyl-ACP is formed by FabI with the consumption of NADPH (Heath and Rock, 1995). Two thioesterases (TEs) in *E. coli*, TesA and TesB release the fatty acid chains from the ACP to produce free fatty acids (Lee et al., 2006). In the biofuel field there are preferences for fatty acids of various chain lengths (Lennen and Pfleger, 2012; Torella et al., 2013). For example, longer chain products ( $C_{12}$ – $C_{20}$ ) fall within the diesel range, which have high energy densities (Wen et al., 2013). The GO has terms describing the synthesis of fatty acids of different chain lengths: for short-chain fatty acids, “GO:0051790 short-chain fatty acid biosynthetic process”; for medium-chain fatty acids, “GO:0051792 medium-chain fatty acid biosynthetic process”; for long-chain fatty acids “GO:0042759 long-chain fatty acid biosynthetic process”; and for very long-chain fatty acids, “GO:0042761 very long-chain fatty acid biosynthetic process.” While free fatty acids are valuable, they cannot be used directly as fuels and must first be converted either to fatty acid alkyl esters (for biodiesel), or to fatty acid-derived alkanes, alkenes

or fatty alcohols (Zhang et al., 2011; Peralta-Yahya et al., 2012; Wen et al., 2013).

## FATTY ACID-DERIVED ALKANES AND ALKENES

Alkanes, an integral part of fossil fuels (gasoline, diesel and jet fuel), are naturally found in diverse organisms including plants, insects and microbial species, but the genetic and biochemical bases behind the production of alkanes have been elusive. Those with C4–C23 carbon chain length possess higher energy densities, hydrophobic properties and compatibilities with existing liquid fuel infrastructure. Most evidence supporting the decarbonylation of aldehydes (a fatty acid metabolite) as the primary mechanism for alkane production have been obtained in eukaryotic systems (Cheesbrough and Kolattukudy, 1984; Dennis and Kolattukudy, 1991, 1992). This in part informed the identification of the *Arabidopsis cer* gene as encoding a protein with decarbonylase activity (GO:0071771 decarbonylase activity) involved in alkane biosynthesis (Aarts et al., 1995). It is only recently that the pathway of alkane biosynthesis was elucidated in cyanobacteria (Figure 5; Schirmer et al., 2010; Li et al., 2011, 2012) making this class of hydrocarbons eligible to be categorized as a next generation biofuel (advanced). In this pathway, fatty acids are reduced by a two-step reaction to alkanes. Fatty acyl-ACP (fatty acid metabolite) is reduced to a fatty aldehyde via a fatty acyl-ACP reductase. This is then followed by a deformylation step catalyzed by an aldehyde deformylating oxidase “GO:1990465 aldehyde oxygenase (deformylating) activity,” resulting in alkane production (Schirmer et al., 2010; Li et al., 2011, 2012; Warui et al., 2011; Coates et al., 2014). Commonly found hydrocarbons in cyanobacteria are heptadecane (GO:1900636 heptadecane biosynthetic process) and methyl heptadecane (Schirmer et al., 2010; Zhang et al., 2011), which have cetane numbers of 105 and 66 respectively (requirement for US diesel or ASTM standard, 47 minimum (Rashid et al., 2008) making the hydrocarbon products from cyanobacteria ideal candidates for diesel fuel applications.

Another group of hydrocarbons produced from fatty acid derivatives are olefins (alkenes, GO:0043450 alkene biosynthetic process). In *Jeotgalicoccus* sp., the terminal alkenes, 8-methyl-1-nonadecene and 17-methyl-1-nonadecene have been identified. A terminal olefin-forming fatty acid decarboxylase belonging to the cytochrome P450 family of enzymes (OleT) has been identified as the key enzyme involved in the production of these olefins in *Jeotgalicoccus* sp. (Rude et al., 2011). Other studies proposed a head-to-head condensation of fatty acids, which involves the formation of a carbon-to-carbon bond between the carboxyl carbon of one fatty acid and the  $\alpha$ -carbon of another fatty acid as another mechanism to generate long chain ( $C_{23}$ – $C_{33}$ ) olefins. OleA, a homolog of the condensing enzyme in fatty acid biosynthesis (FabH-3-oxo-acyl-ACP ketosynthase (KS) III), was identified as a key enzyme in long chain olefin production (Beller et al., 2010; Sukovich et al., 2010). In cyanobacteria, the olefin-producing pathway (OLS) involves a polyketide synthase (GO:0034081 polyketide synthase complex) that first elongates fatty acyl-CoA with two carbons from malonyl-CoA via KS and acyl transferase (AT) domains. This is followed by reduction to the hydroxyacid by a ketoreductase (KR, GO:0045703 ketoreductase activity; Mendez-Perez et al.,



**FIGURE 5 | Biosynthesis of fatty acids, fatty alkanes and fatty alcohols described with the Gene Ontology.** Fatty acids are generated from acetyl-CoA and malonyl-CoA precursors through fatty acid synthases. Acyl-ACPs, which are intermediates of the fatty acid biosynthetic process, are reduced to fatty alcohols or deformylated to form alkanes. Enzymes

involved in the biosynthesis of the fatty acid derivatives are described with appropriate GO Molecular Function terms (colored sky blue), Biological Process terms (tan) and the Cellular Component terms (green). Relevant references are in the text. Figure is based on information provided in Lennen and Pfleger (2013).

2011). In the final step, sulfotransferases (ST, GO:0008146 sulfotransferase activity) activate the  $\beta$ -hydroxy group via sulfonation, and then a TE acts on this substrate to catalyze decarboxylation and loss of sulfate to form the terminal alkene (McCarthy et al., 2012).

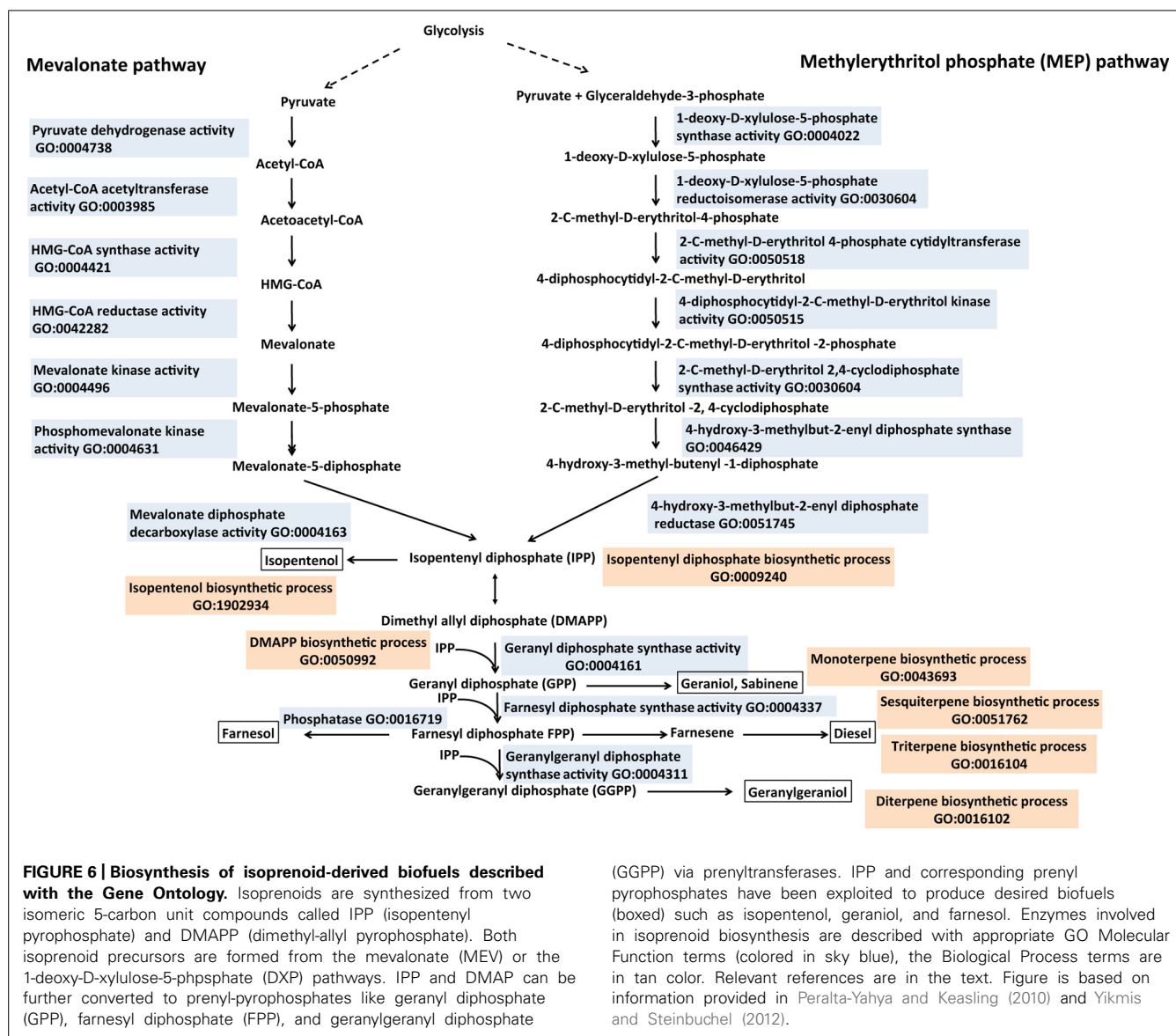
#### FATTY ACID ALKYL ESTERS (BIODIESEL)

Biodiesel is a substitute for petroleum-based diesel fuel. Like its counterparts described above, biodiesel has properties similar to those of diesel, and therefore, can be used in the diesel engines and stored and distributed using the existing infrastructure. Other advantages of biodiesel include reduced fuel toxicity and increased lubricity. Additionally the use of biodiesel leads to lower carbon monoxide and soot emissions than conventional diesel fuels. Biodiesel is traditionally produced by the trans-esterification of mostly plant-derived triacylglycerols yielding glycerol and fatty acid alkyl esters (FAAE), particularly fatty acid methyl esters (FAMEs; “GO:1902899 FAME biosynthetic process”). However, biodiesel production is limited by the availability of inexpensive vegetable oil feedstocks. This has prompted a search for sustainable alternatives. Direct microbial production of FAAEs is an area of intense research as it bypasses the transesterification step, reducing cost and energy and also avoids the use of methanol, (Kalscheuer et al., 2006; Nawabi et al., 2011) which is an expensive and toxic feedstock.

#### ISOPRENOID-DERIVED BIOFUEL

Isoprenoids, also called terpenes, have been evaluated for pharmaceutical, nutritional and fuel products (Bohlmann and Keeling, 2008; Peralta-Yahya and Keasling, 2010). Two independent biosynthetic pathways for isoprenoid production “GO:0008299 isoprenoid biosynthetic process” are found in nature (Figure 6). The methylerythritol phosphate (MEP) pathway “GO:1902768 isoprenoid biosynthetic process via 1-deoxy-D-xylulose 5-phosphate” and the mevalonate pathway “GO:1902767 isoprenoid biosynthetic process via mevalonate” have evolved for the production of the key five-carbon isoprenoid intermediates, isopentenyl diphosphate (IPP), and dimethylallyl diphosphate (DMAPP; Beytia and Porter, 1976; Edwards et al., 1992; Figure 6). The MEP pathway consists of seven steps resulting in the conversion of glyceraldehyde-3-phosphate and pyruvate to IPP and DMAPP and it is found in most bacteria, chloroplast, unicellular eukaryotes, and certain parasites (Zhao et al., 2013; Jarchow-Choy et al., 2014; Figure 6). The mevalonate pathway is responsible for all the isoprenoid production in archaea, some bacteria and most eukaryotes (Miziorko, 2011; Figure 6). It converts acetyl CoA in six steps to IPP via the key intermediate mevalonate. Specifically, acetyl-CoA and acetoacetyl-CoA are condensed into 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA), which is reduced to mevalonate via HMG-CoA reductase (HMGR, GO:0042282 HMG-CoA reductase activity). Following this are two phosphorylation steps, which convert mevalonate





**FIGURE 6 | Biosynthesis of isoprenoid-derived biofuels described with the Gene Ontology.** Isoprenoids are synthesized from two isomeric 5-carbon unit compounds called IPP (isopentenyl pyrophosphate) and DMAPP (dimethyl-allyl pyrophosphate). Both isoprenoid precursors are formed from the mevalonate (MEV) or the 1-deoxy-D-xylulose-5-phosphate (DXP) pathways. IPP and DMAPP can be further converted to prenyl-pyrophosphates like geranyl diphosphate (GPP), farnesyl diphosphate (FPP), and geranylgeranyl diphosphate (GGPP) via prenyltransferases. IPP and corresponding prenyl pyrophosphates have been exploited to produce desired biofuels (boxed) such as isopentenol, geraniol, and farnesol. Enzymes involved in isoprenoid biosynthesis are described with appropriate GO Molecular Function terms (colored in sky blue), the Biological Process terms are in tan color. Relevant references are in the text. Figure is based on information provided in Peralta-Yahya and Keasling (2010) and Yikmis and Steinbuchel (2012).

pyrophosphatase “GO:0016462 pyrophosphatase activity,” nudF from *B. subtilis*, was shown to produce isopentenol in *E. coli*. Curating the biosynthesis of the intermediates and the vast diversity of terpene synthases “GO:0010333 terpene synthase activity” therefore will provide a useful resource, which will inform decisions in the synthesis of terpene-based fuels. A new class of terpenes ( $C_{35}$ ), the sesquiterpenes has recently been classified (Sato, 2013) and it will be interesting to know if these compounds could provide precursors for terpene-based biofuels.

## POLYHYDROXYALKANOATE-DERIVED BIOFUEL

Microbially produced PHAs have attracted attention as biodegradable polyesters and quite recently as biofuels (Zhang et al., 2009). PHAs are synthesized inside cells during oxygen, phosphorous or nitrogen starvation in the presence of excess carbon and found as insoluble cytoplasmic inclusions called polyhydroxyalkanoate



granules (GO:0070088 PHA granule; Pfeiffer and Jendrossek, 2012; Jendrossek and Pfeiffer, 2013). Synthesis of a PHA could occur through several metabolic pathways and from a variety of carbon precursors including glucose, glycerol, and fatty acids (Park et al., 2012; Shahid et al., 2013). PHAs comprise over 150 both homo- and hetero-polymers of which poly-3-hydroxybutyrate (poly (3HB), P (3HB) or PHB) being the most abundant naturally produced and most studied polyhydroxyalkanoate “GO:0042618 polyhydroxybutyrate biosynthetic process” (Laycock et al., 2014). On the basis of the carbon chain lengths of the monomers, PHAs are divided into two main groups: short chain length PHAs (scl-PHA) containing monomer units with 3–5 carbon atoms, and medium chain length PHAs (mcl-PHA) composed of monomer units with 6–18 carbon atoms. Based on the bacterial strains, PHA synthases and carbon sources involved, products of PHA biosynthesis can be a homopolymer, copolymer, block polymer or even blends (Lu et al., 2009). Functional groups such as unsaturated bonds, benzene, halogens and cyclic chemicals and epoxides can also modify PHA structures, opening up the potential for the production of a vast number of diverse PHAs. The biosynthesis of PHA “GO:190144 polyhydroxyalkanoate biosynthetic process” is quite complex and involves several enzymes that are directly or indirectly involved in PHA synthesis (Reddy et al., 2003; Park et al., 2012). PHA synthesis is well studied in the model organism *Ralstonia eutropha* (Riedel et al., 2014). In this organism, the carbon source is converted into coenzyme A thioesters of (R)-hydroxyalkanoic acids. Following this step,  $\beta$  ketothiolase “GO:0003988 acetyl-CoA C-acyltransferase activity” catalyzes the condensation of two coenzyme-A thioester monomers such as acetyl-CoA and propionyl-CoA. An (R)-specific reduction step involving acetoacetyl-CoA reductase “GO:0018454 acetoacetyl-CoA reductase activity” produces (R)-3 hydroxybutyryl-CoA or (R)-3 hydroxyvaleryl-CoA which is then converted into PHA by the action of PHA synthase. In regards to biofuels, 3-hydroxyalkanoates (3HAs; mcl PHAs) are linked by ester bonds formed with the hydroxyl group (–OH) of one monomer and the carboxyl (–COOH) group of the other monomer through the catalysis by various PHA synthases (Reddy et al., 2003). These hydroxyalkanoate esters (3HA esters) have been found to be similar to methyl esters of long chain fatty acids (in biodiesel). As such they can be used as fuel additives. Hydroxybutyrate and hydroxyalkanoate methyl esters (3HBME and 3HAME), generated from esterification of scl PHB and mcl PHA with methanol, respectively, are considered equivalent to ethanol (Gao et al., 2011). These esters can also be used as gasoline and biodiesel additive (Gao et al., 2011).

## TOLERANCE TO BIOFUELS AND BY-PRODUCTS

Accumulation of biofuels as well as by-products often affect the integrity of the fuel producing organism's cell membrane and also impairs metabolic pathways associated with cell growth, thus compromising product titers (Heipieper et al., 2007; Segura et al., 2012). Most native producers of biofuels are sensitive to the solvents they produce. Advances are being made toward removing this block through investigations on the mechanisms of tolerance

to product accumulation (McEvoy et al., 2004; Fujita et al., 2006; Liu and Qureshi, 2009a; Ghiaci et al., 2013; Zingaro and Terry Papoutsakis, 2013) and several organisms are being investigated as better choices for biofuel production from this standpoint (Dunlop, 2011). These efforts will be substantially aided by curating all gene products associated with native or improved tolerance to biofuels and their by-products. Tolerance is considered a phenotype, not a process and is therefore out of the scope for GO Biological Process ontology. To circumvent this problem, we created the terms “process resulting in tolerance to x” instead of “tolerance to x” (x is any compound). Here we present some examples.

The common inhibitors to microbial growth that are found in lignocellulose hydrolysates include aldehydes, ketones, phenols and organic acids (Klinke et al., 2004; Jonsson et al., 2013). A *S. cerevisiae* gene, *ARI1*, encoding a novel NADPH-dependent aldehyde reductase has been shown to be involved in providing tolerance to inhibitors such as furfural, vanillin and cinnamaldehyde that arise from lignocellulose hydrolysis (Liu and Moon, 2009b). We annotated the gene product ARI1 with GO terms for its molecular role “GO:0018455 aldehyde reductase (NADPH/NADH) activity” and a biological process term “GO:1990370 process resulting in tolerance to aldehyde” to describe its ability to confer tolerance to several aldehydes. Other notable *S. cerevisiae* enzymes with efficient aldehyde reduction activities include ADHs (ADH1, ADH6 and ADH7; GO:0004022 ADH activity), aldehyde dehydrogenase (ALD4; GO:0004030 aldehyde dehydrogenase activity) and methylglyoxal reductases (GRE2 and GRE3; GO:0043892 methylglyoxal reductase activity; Petersson et al., 2006; Laadan et al., 2008).

Tolerance to alcohols such as butanol in *C. acetobutylicum* is influenced by the multifunctional Spo0A protein, which, as discussed in an earlier section, controls sporulation and solventogenesis (Alsaker et al., 2004). Kanno et al. (2013) obtained several aerobic and anaerobic bacterial isolates from soil, spanning diverse genera that tolerate butanol and isobutanol levels greater than 2% (vol/vol). The *cfa* gene, which encodes cyclopropane fatty acid (CFA) synthase in one of these bacteria (belonging to the *Firmicutes* phylum), confers solvent tolerance in recombinant *E. coli* (Kanno et al., 2013). CFA is annotated with the terms “GO:1990336 process resulting in tolerance to butanol” and “GO:1990337 process resulting in tolerance to isobutanol.” A quantitative transcriptomic analysis revealed that in the cyanobacterium *Synechocystis* sp. PCC 6803, which uses solar energy and carbon dioxide as sole energy and carbon sources, over 250 genes are induced upon exposure to butanol (Zhu et al., 2013). Of these, three, *slr0690*, *slr0947*, and *slr1295* were further characterized using knock-out mutants. The results indicated butanol sensitivity in strains that lacked any of these genes, indicating their involvement in resistance to butanol. Using a genome-scale analysis in *S. cerevisiae*, González-Ramos et al. (2013) demonstrated the role of protein degradation in tolerance to C3 and C4 alcohols (butanol, 2-butanol, isobutanol, and isopropanol). Specifically, the YLR224W gene was found to be associated with increased butanol tolerance and it encodes a subunit of the Skp-Cullin\_F-box (SCF) ubiquitin ligase that recognizes damaged proteins. The

efflux pumps involved in the extrusion of toxins from cells have been considered appropriate candidates for studies on solvent tolerance (Segura et al., 2012). Such a pump consists of multiple proteins forming a multicomponent complex “GO:1990281 efflux pump complex,” which span the inner to the outer membrane of bacteria cells (Andersen, 2003; Du et al., 2014). An exposure to solvents induces the expression of *srpABC* genes in *P. putida* (Kieboom et al., 1998b). The SrpABC share considerable sequence similarities to multi-drug efflux pump proteins (Kieboom et al., 1998a). Specifically, SrpA functions as a periplasmic linker protein, SrpB as an inner membrane transporter, and SrpC, as an outer membrane channel. For these reasons, besides annotations for their roles in efflux pump and cellular locations (“GO:0015562 efflux transmembrane transporter activity” from Molecular Function ontology; “GO:1990281 efflux pump complex,” “GO:0042597 periplasmic space,” “GO:0009276 Gram-negative-bacterium type cell wall,” and “GO:0019867 outer membrane from the Cellular Component ontology”), we have qualified these proteins with a biological process term “GO:1990367 process resulting in tolerance to organic substance.” Rojas et al. (2001) have identified three different solvent efflux pumps in *P. putida*, TtgABC, TtgDEF, and TtgGHI, which extrude toluene. TtgABC also could extrude antibiotics (Ramos et al., 1998; Mosqueda and Ramos, 2000; Rojas et al., 2001). All these multi-protein efflux pumps are assigned the GO term “GO:1990281 efflux pump complex.” Hydrocarbons including nonane, decane, and undecane were shown to be toxic to *S. cerevisiae* when these solvents accumulated inside the cell. A transcriptomic analysis identified modified cell membrane and efflux pumps as contributing to alkane export and tolerance. Specifically efflux pumps Snp2p and Pdr5 were shown to reduce intracellular levels of decane and undecane thereby enhancing tolerance to the alkanes “GO:1990373 process resulting in tolerance to alkane” (Ling et al., 2013).

## SYNTHETIC BIOLOGY

The yields of most fuels or fuel derivatives in the native biological systems discussed above are insufficient to be cost-competitive with petroleum-derived fuels, and therefore the technologies of metabolic engineering and synthetic biology are being employed to increase production and/or generate entirely new fuels with superior qualities. As mentioned in the introduction, currently the GO is limited to terms for describing natural processes. The MENGO team is currently discussing with the community the creation of a complementary set of GO-like terms suitable for the annotation of synthetic biological systems, which we provisionally call SYNGO. To underscore the anticipated broad scope of SYNGO, we provide here brief descriptions of selected synthetic processes and parts that are driving the development of commercially viable routes for the production of advanced biofuels.

Based on the modular nature of cellulosomes and the availability of a variety of dockerin-cohesin pairs, efforts are underway to construct designer cellulosomes, xylanosomes, and cellulosome-like structures. In such a precision-engineered multienzyme complex, the molecular architecture and enzyme content are well controlled, and the result is enhanced synergistic deconstruction of biomass (Mitsuzawa et al., 2009; Nordon et al., 2009; Smith and Bayer, 2013; Srikrishnan et al., 2013; Vazana

et al., 2013). For example, Mitsuzawa et al. (2009) modified a thermostable group II chaperonin (18-subunit self-assembling protein complex called rosettasome), from the archaeon *Sulfolobus shibatae*, for use as a scaffold to assemble selected hydrolytic enzymes. A cohesin module was fused to each of the eighteen subunits which in turn was combined with dockerin-containing cellulases from *C. thermocellum* to build an 18 enzyme cellulosome-like structure they termed rosettazyme. Truncated cellulosomes also called minicellulosomes have been constructed in an effort to fully understand the relationship between cellulosome structure and enzymatic activity (Arai et al., 2007; Cha et al., 2007; Mingardon et al., 2007). Examples are the mini-CipA and mini-CipC1 cellulosomes from *Clostridium cellulovorans* and *Clostridium cellulolyticum*, respectively (Murashima et al., 2002; Perret et al., 2004). The engineered minicellulosome platforms can be extended to build complex designer cellulosomes. Designer cellulosomes are not described in the GO and a complementary term “synthetic multi-cellular complex” would be useful to capture designer multienzyme complexes such as the rosettazyme, mini-CipA and mini-CipC1 and their components.

Generally, yield is one of the focal points for improving alcohol production by means of synthetic biology technology (Peralta-Yahya et al., 2012; Gronenberg et al., 2013). For example, low butanol yield in the native *Clostridium* species is not economically competitive, and thus the *Clostridium* CoA-dependent pathway has been heterologously expressed in genetically tractable industrial microbes hosts such as *E. coli* and *S. cerevisiae* (Inui et al., 2008; Steen et al., 2008) for facile manipulation toward better yield and productivity. For example, *E. coli* strain bearing the CoA-dependent pathway was further modified by substituting the reversible, flavin-dependent butyryl-CoA dehydrogenase (Bcd) with an irreversible trans-enoyl-CoA reductase (Ter) for the reduction of crotonyl-CoA (Bond-Watts et al., 2011). The increased flux to butanol improved product yield (Bond-Watts et al., 2011). While Bcd is associated with “butanol biosynthetic process” in the GO, Ter is not, as it is part of a non-natural (synthetic) pathway. This limitation in the GO could potentially be addressed with complementary SYNGO terms, for example, Ter could be assigned “enoyl-CoA reductase involved in increased butanol biosynthesis.”

The natural ability of *E. coli* to produce fatty acids and creation of new biochemical reactions through synthetic biology have provided the means to divert fatty acid metabolism toward the production of fuels (Clomburg and Gonzalez, 2010; Wen et al., 2013). This approach is a more sustainable alternative than, for example, the production of biodiesel from plant oils. Direct microbial production of fatty acid esters (biodiesel) eliminates the need for a subsequent chemical transesterification step. A pathway which includes an acyl-CoA ligase and the broad specificity acyltransferase (WS/DGAT; AtfA), together with the enzymes that provide ethanol for esterification [pyruvate decarboxylase (pdc) and alcohol dehydrogenase B (adhB)] were incorporated into a fatty acid overproducing *E. coli* strain for the production of fatty acid ethyl esters (Stoveken et al., 2005; Steen et al., 2010). Curating the components of the engineered pathway with complementary SYNGO terms,

associating them with fatty acid ethyl ester biosynthesis, should be a powerful tool for researchers working to improve these processes.

## CONCLUSION

The MENGO group has created and continues to create GO terms relevant to microbial bioenergy research. In addition, the team has used the terms to produce high quality “gold standard” annotations of gene products based on the experimental scientific literature, which are useful for engineering of microbial strains to optimize bioenergy production. These resources provide easy access to otherwise dispersed information in the scientific literature and also aids in the computational analysis of large datasets. The GO annotations can also be used to uncover metabolic pathways, and to understand bioenergy-relevant microbes at the system biology level. Community involvement in MENGO term development and annotations is welcome and will create a more comprehensive resource for use by all. Since the GO currently is restricted to natural processes, a need has arisen for a complementary set of GO-like terms to describe engineered processes and the gene products that comprise them. Community involvement in this other development is also essential.

## ACKNOWLEDGMENTS

We thank the editors at the GOC for reviewing the MENGO terms. This work was supported by grant DE-SC0005011 from the US Department of Energy. Jane Lomax is funded by European Molecular Biology Laboratories (EMBL) core funds. João C. Setubal is funded by CNPq and FAPESP. The authors wish to thank all participants in the MENGO workshops held at various venues in 2011 and 2012.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fmicb.2014.00528/abstract>

## REFERENCES

- Aarts, M. G., Keijzer, C. J., Stiekema, W. J., and Pereira, A. (1995). Molecular characterization of the CER1 gene of arabidopsis involved in epicuticular wax biosynthesis and pollen fertility. *Plant Cell* 7, 2115–2127. doi: 10.1105/tpc.7.12.2115
- Ahmed, S., Riaz, S., and Jamil, A. (2009). Molecular cloning of fungal xylanases: an overview. *Appl. Microbiol. Biotechnol.* 84, 19–35. doi: 10.1007/s00253-009-2079-4
- Alsaker, K. V., Spitzer, T. R., and Papoutsakis, E. T. (2004). Transcriptional analysis of spo0A overexpression in *Clostridium acetobutylicum* and its effect on the cell's response to butanol stress. *J. Bacteriol.* 186, 1959–1971. doi: 10.1128/JB.186.7.1959-1971.2004
- Andersen, C. (2003). Channel-tunnels: outer membrane components of type I secretion systems and multidrug efflux pumps of Gram-negative bacteria. *Rev. Physiol. Biochem. Pharmacol.* 147, 122–165. doi: 10.1007/s10254-003-0008-y
- Arai, T., Matsuoka, S., Cho, H. Y., Yukawa, H., Inui, M., Wong, S. L., et al. (2007). Synthesis of *Clostridium cellulovorans* minicellulosomes by intercellular complementation. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1456–1460. doi: 10.1073/pnas.0610740104
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., et al. (2000). Gene Ontology: tool for the unification of biology. The Gene Ontology consortium. *Nat. Genet.* 25, 25–29. doi: 10.1038/75556
- Atsumi, S., Hanai, T., and Liao, J. C. (2008). Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* 451, 86–89. doi: 10.1038/nature06450
- Atsumi, S., and Liao, J. C. (2008). Metabolic engineering for advanced biofuels production from *Escherichia coli*. *Curr. Opin. Biotechnol.* 19, 414–419. doi: 10.1016/j.copbio.2008.08.008
- Bayer, E. A., and Lamed, R. (1986). Ultrastructure of the cell surface cellulosome of *Clostridium thermocellum* and its interaction with cellulose. *J. Bacteriol.* 167, 828–836.
- Bayer, E. A., Setter, E., and Lamed, R. (1985). Organization and distribution of the cellulosome in *Clostridium thermocellum*. *J. Bacteriol.* 163, 552–559.
- Beck, C., Knoop, H., Axmann, I. M., and Steuer, R. (2012). The diversity of cyanobacterial metabolism: genome analysis of multiple phototrophic microorganisms. *BMC Genomics* 13:56. doi: 10.1186/1471-2164-13-56
- Beller, H. R., Goh, E. B., and Keasling, J. D. (2010). Genes involved in long-chain alkene biosynthesis in *Micrococcus luteus*. *Appl. Environ. Microbiol.* 76, 1212–1223. doi: 10.1128/AEM.02312-09
- Benso, A., Di Carlo, S., Ur Rehman, H., Politano, G., Savino, A., and Suravajhala, P. (2013). A combined approach for genome wide protein function annotation/prediction. *Proteome Sci.* 11, S1. doi: 10.1186/1477-5956-11-S1-S1
- Beytia, E. D., and Porter, J. W. (1976). Biochemistry of polyisoprenoid biosynthesis. *Annu. Rev. Biochem.* 45, 113–142. doi: 10.1146/annurev.bi.45.070176.000553
- Bi, C., Jones, S. W., Hess, D. R., Tracy, B. P., and Papoutsakis, E. T. (2011). SpoIIE is necessary for asymmetric division, sporulation, and expression of sigmaF, sigmaE, and sigmaG but does not control solvent production in *Clostridium acetobutylicum* ATCC 824. *J. Bacteriol.* 193, 5130–5137. doi: 10.1128/JB.05474-11
- Biely, P., Mislovicova, D., and Toman, R. (1985). Soluble chromogenic substrates for the assay of endo-1,4-beta-xylanases and endo-1,4-beta-glucanases. *Anal. Biochem.* 144, 142–146. doi: 10.1016/0003-2697(85)90095-8
- Blake, J. A., and Harris, M. A. (2008). The Gene Ontology (GO) project: structured vocabularies for molecular biology and their application to genome and expression analysis. *Curr. Protoc. Bioinformatics* Chap. 7, Unit 7.2. doi: 10.1002/0471250953.bi0702s23
- Blanch, H. W., Simmons, B. A., and Klein-Marcuschamer, D. (2011). Biomass deconstruction to sugars. *Biotechnol. J.* 6, 1086–1102. doi: 10.1002/biot.201000180
- Blumer-Schuette, S. E., Brown, S. D., Sander, K. B., Bayer, E. A., Kataeva, I., Zurawski, J. V., et al. (2014). Thermophilic lignocellulose deconstruction. *FEMS Microbiol. Rev.* 38, 393–448. doi: 10.1111/1574-6976.12044
- Bohlmann, J., and Keeling, C. I. (2008). Terpenoid biomaterials. *Plant J.* 54, 656–669. doi: 10.1111/j.1365-313X.2008.03449.x
- Bond-Watts, B. B., Bellerose, R. J., and Chang, M. C. (2011). Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways. *Nat. Chem. Biol.* 7, 222–227. doi: 10.1038/nchembio.537
- Branduardi, P., Longo, V., Berterame, N. M., Rossi, G., and Porro, D. (2013). A novel pathway to produce butanol and isobutanol in *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* 6, 68. doi: 10.1186/1754-6834-6-68
- Bras, J. L., Alves, V. D., Carvalho, A. L., Najmudin, S., Prates, J. A., Ferreira, L. M., et al. (2012). Novel *Clostridium thermocellum* type I cohesin-dockerin complexes reveal a single binding mode. *J. Biol. Chem.* 287, 44394–44405. doi: 10.1074/jbc.M112.407700
- Campbell, J. W., and Cronan, J. E. Jr. (2001). Bacterial fatty acid biosynthesis: targets for antibacterial drug discovery. *Annu. Rev. Microbiol.* 55, 305–332. doi: 10.1146/annurev.micro.55.1.305
- Cha, J., Matsuoka, S., Chan, H., Yukawa, H., Inui, M., and Doi, R. H. (2007). Effect of multiple copies of cohesins on cellulase and hemicellulase activities of *Clostridium cellulovorans* mini-cellulosomes. *J. Microbiol. Biotechnol.* 17, 1782–1788.
- Chang, J. J., Ho, F. J., Ho, C. Y., Wu, Y. C., Hou, Y. H., Huang, C. C., et al. (2013). Assembling a cellulase cocktail and a cellobextrin transporter into a yeast host for CBP ethanol production. *Biotechnol. Biofuels* 6:19. doi: 10.1186/1754-6834-6-19
- Cheesbrough, T. M., and Kolattukudy, P. E. (1984). Alkane biosynthesis by decarbonylation of aldehydes catalyzed by a particulate preparation from *Pisum sativum*. *Proc. Natl. Acad. Sci. U.S.A.* 81, 6613–6617. doi: 10.1073/pnas.81.2.6613
- Chen, J.-S., and Hiu, S. F. (1986). Acetone-butanol-isopropanol production by *Clostridium beijerinckii* (Synonym. *Clostridium butylicum*). *Biotechnol. Lett.* 8, 371–376. doi: 10.1007/BF01040869
- Christie, K. R., Hong, E. L., and Cherry, J. M. (2009). Functional annotations for the *Saccharomyces cerevisiae* genome: the knowns and the

- known unknowns. *Trends Microbiol.* 17, 286–294. doi: 10.1016/j.tim.2009.04.005
- Chundawat, S. P., Beckham, G. T., Himmel, M. E., and Dale, B. E. (2011). Deconstruction of lignocellulosic biomass to fuels and chemicals. *Annu. Rev. Chem. Biomol. Eng.* 2, 121–145. doi: 10.1146/annurev-chembioeng-061010-114205
- Clomburg, J. M., and Gonzalez, R. (2010). Biofuel production in *Escherichia coli*: the role of metabolic engineering and synthetic biology. *Appl. Microbiol. Biotechnol.* 86, 419–434. doi: 10.1007/s00253-010-2446-1
- Coates, R. C., Podell, S., Korobeynikov, A., Lapidus, A., Pevzner, P., Sherman, D. H., et al. (2014). Characterization of cyanobacterial hydrocarbon composition and distribution of biosynthetic pathways. *PLoS ONE* 9:e85140. doi: 10.1371/journal.pone.0085140
- Dennis, M., and Kolattukudy, P. E. (1992). A cobalt-porphyrin enzyme converts a fatty aldehyde to a hydrocarbon and CO. *Proc. Natl. Acad. Sci. U.S.A.* 89, 5306–5310. doi: 10.1073/pnas.89.12.5306
- Dennis, M. W., and Kolattukudy, P. E. (1991). Alkane biosynthesis by decarbonylation of aldehyde catalyzed by a microsomal preparation from *Botryococcus braunii*. *Arch. Biochem. Biophys.* 287, 268–275. doi: 10.1016/0003-9861(91)90478-2
- Dickinson, J. R., Harrison, S. J., and Hewlins, M. J. (1998). An investigation of the metabolism of valine to isobutyl alcohol in *Saccharomyces cerevisiae*. *J. Biol. Chem.* 273, 25751–25756. doi: 10.1074/jbc.273.40.25751
- Dimmer, E. C., Huntley, R. P., Alam-Faruque, Y., Sawford, T., O'donovan, C., Martin, M. J., et al. (2012). The UniProt-GO annotation database in 2011. *Nucleic Acids Res.* 40, D565–D570. doi: 10.1093/nar/gkr1048
- Dodd, D., and Cann, I. K. (2009). Enzymatic deconstruction of xylan for biofuel production. *Glob. Change Biol. Bioenergy* 1, 2–17. doi: 10.1111/j.1757-1707.2009.01004.x
- Doi, R. H. (2008). Cellulases of mesophilic microorganisms: cellulosome and noncellulosome producers. *Ann. N. Y. Acad. Sci.* 1125, 267–279. doi: 10.1196/annals.1419.002
- Du, D., Wang, Z., James, N. R., Voss, J. E., Klimont, E., Ohene-Agyei, T., et al. (2014). Structure of the AcrAB-TolC multidrug efflux pump. *Nature* 509, 512–515. doi: 10.1038/nature13205
- Dunlop, M. J. (2011). Engineering microbes for tolerance to next-generation biofuels. *Biotechnol. Biofuels* 4:32. doi: 10.1186/1754-6834-4-32
- Dürre, P. (2007). Biobutanol: an attractive biofuel. *Biotechnol. J.* 2, 1525–1534. doi: 10.1002/biot.200700168
- Dürre, P., and Hollerschwandner, C. (2004). Initiation of endospore formation in *Clostridium acetobutylicum*. *Anaerobe* 10, 69–74. doi: 10.1016/j.anaerobe.2003.11.001
- Edwards, P. A., Ashby, M. N., Spear, D. H., Marrero, P. F., Joly, A., and Popjak, G. (1992). Polyisoprenoid synthesis and metabolism. *Biochem. Soc. Trans.* 20, 475–479.
- Feng, Y., and Cronan, J. E. (2009). *Escherichia coli* unsaturated fatty acid synthesis: complex transcription of the fabA gene and in vivo identification of the essential reaction catalyzed by FabB. *J. Biol. Chem.* 284, 29526–29535. doi: 10.1074/jbc.M109.023440
- Fujita, K., Matsuyama, A., Kobayashi, Y., and Iwashita, H. (2006). The genome-wide screening of yeast deletion mutants to identify the genes required for tolerance to ethanol and other alcohols. *FEMS Yeast Res.* 6, 744–750. doi: 10.1111/j.1567-1364.2006.00040.x
- Gao, X., Chen, J. C., Wu, Q., and Chen, G. Q. (2011). Polyhydroxyalkanoates as a source of chemicals, polymers, and biofuels. *Curr. Opin. Biotechnol.* 22, 768–774. doi: 10.1016/j.copbio.2011.06.005
- George, H. A., Johnson, J. L., Moore, W. E., Holdeman, L. V., and Chen, J. S. (1983). Acetone, Isopropanol, and Butanol production by *Clostridium beijerinckii* (syn. *Clostridium butylicum*) and *Clostridium aurantibutyricum*. *Appl. Environ. Microbiol.* 45, 1160–1163.
- Ghiaci, P., Norbeck, J., and Larsson, C. (2013). Physiological adaptations of *Saccharomyces cerevisiae* evolved for improved butanol tolerance. *Biotechnol. Biofuels* 6:101. doi: 10.1186/1754-6834-6-101
- Gomez del Pulgar, E. M., and Saadeddin, A. (2014). The cellulolytic system of *Thermobifida fusca*. *Crit. Rev. Microbiol.* 40, 236–247. doi: 10.3109/1040841X.2013.776512
- González-Ramos, D., Van Den Broek, M., Van Maris, A. J. A., Pronk, J. T., and Daran, J.-M. G. (2013). Genome-scale analyses of butanol tolerance in *Saccharomyces cerevisiae* reveal an essential role of protein degradation. *Biotechnol. Biofuels* 6:48. doi: 10.1186/1754-6834-6-48
- Gronenberg, L. S., Marcheschi, R. J., and Liao, J. C. (2013). Next generation biofuel engineering in prokaryotes. *Curr. Opin. Chem. Biol.* 17, 462–471. doi: 10.1016/j.cbpa.2013.03.037
- Hansen, J., Kharecha, P., Sato, M., Masson-Delmotte, V., Ackerman, F., Beerling, D. J., et al. (2013). Assessing “dangerous climate change”: required reduction of carbon emissions to protect young people, future generations and nature. *PLoS ONE* 8:e81648. doi: 10.1371/journal.pone.0081648
- Harris, L. M., Welker, N. E., and Papoutsakis, E. T. (2002). Northern, morphological, and fermentation analysis of spo0A inactivation and overexpression in *Clostridium acetobutylicum* ATCC 824. *J. Bacteriol.* 184, 3586–3597. doi: 10.1128/JB.184.13.3586-3597.2002
- Harris, M. A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., et al. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* 32, D258–D261. doi: 10.1093/nar/gkh036
- Heath, R. J., and Rock, C. O. (1995). Enoyl-acyl carrier protein reductase (fabI) plays a determinant role in completing cycles of fatty acid elongation in *Escherichia coli*. *J. Biol. Chem.* 270, 26538–26542. doi: 10.1074/jbc.270.44.26538
- Heipieper, H. J., Neumann, G., Cornelissen, S., and Meinhardt, F. (2007). Solvent-tolerant bacteria for biotransformations in two-phase fermentation systems. *Appl. Microbiol. Biotechnol.* 74, 961–973. doi: 10.1007/s00253-006-0833-4
- Hill, D. P., Adams, N., Bada, M., Batchelor, C., Berardini, T. Z., Dietze, H., et al. (2013). Doretailing biology and chemistry: integrating the Gene Ontology with the ChEBI chemical ontology. *BMC Genomics* 14:513. doi: 10.1186/1471-2164-14-513
- Hong, W., Zhang, J., Feng, Y., Mohr, G., Lambowitz, A. M., Cui, G. Z., et al. (2014). The contribution of cellulosomal scaffoldins to cellulose hydrolysis by *Clostridium thermocellum* analyzed by using thermotargetrons. *Biotechnol. Biofuels* 7, 80. doi: 10.1186/1754-6834-7-80
- Hu, S., Zheng, H., Gu, Y., Zhao, J., Zhang, W., Yang, Y., et al. (2011). Comparative genomic and transcriptomic analysis revealed genetic characteristics related to solvent formation and xylose utilization in *Clostridium acetobutylicum* EA 1018. *BMC Genomics* 12:93. doi: 10.1186/1471-2164-12-93
- Inui, M., Suda, M., Kimura, S., Yasuda, K., Suzuki, H., Toda, H., et al. (2008). Expression of *Clostridium acetobutylicum* butanol synthetic genes in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 77, 1305–1316. doi: 10.1007/s00253-007-1257-5
- Janssen, H. J., and Steinbüchel, A. (2014). Fatty acid synthesis in *Escherichia coli* and its applications towards the production of fatty acid based biofuels. *Biotechnol. Biofuels* 7, 7. doi: 10.1186/1754-6834-7-7
- Jarchow-Choy, S. K., Koppisch, A. T., and Fox, D. T. (2014). Synthetic routes to methylerythritol phosphate pathway intermediates and downstream isoprenoids. *Curr. Org. Chem.* 18, 1050–1072. doi: 10.2174/1385272819666140501001101
- Jendrossek, D., and Pfeiffer, D. (2013). New insights in the formation of polyhydroxyalkanoate (PHA) granules (carbonosomes) and novel functions of poly(3-hydroxybutyrate; PHB). *Environ. Microbiol.* 16, 2357–2373. doi: 10.1111/1462-2920.12356
- Jiang, Z., Dang, W., Yan, Q., Zhai, Q., Li, L., and Kusakabe, I. (2006). Subunit composition of a large xylanolytic complex (xylanosome) from *Streptomyces olivaceoviridis* E-86. *J. Biotechnol.* 126, 304–312. doi: 10.1016/j.jbiotec.2006.05.006
- Jiang, Z. Q., Deng, W., Li, L. T., Ding, C. H., Kusakabe, I., and Tan, S. S. (2004). A novel, ultra-large xylanolytic complex (xylanosome) secreted by *Streptomyces olivaceoviridis*. *Biotechnol. Lett.* 26, 431–436. doi: 10.1023/B:BILE.0000018266.75248.03
- Jonsson, L. J., Alriksson, B., and Nilvebrant, N. O. (2013). Bioconversion of lignocellulose: inhibitors and detoxification. *Biotechnol. Biofuels* 6:16. doi: 10.1186/1754-6834-6-16
- Juturu, V., and Wu, J. C. (2013). Insight into microbial hemicellulases other than xylanases: a review. *J. Chem. Technol. Biotechnol.* 88, 353–363. doi: 10.1002/jctb.3969
- Kalscheuer, R., Stoltz, T., and Steinbüchel, A. (2006). Microdiesel: *Escherichia coli* engineered for fuel production. *Microbiology* 152, 2529–2536. doi: 10.1099/mic.0.29028-0
- Kanno, M., Katayama, T., Tamaki, H., Mitani, Y., Meng, X. Y., Hori, T., et al. (2013). Isolation of butanol- and isobutanol-tolerant bacteria and physiological characterization of their butanol tolerance. *Appl. Environ. Microbiol.* 79, 6998–7005. doi: 10.1128/AEM.02900-13



- Karpol, A., Jobby, M. K., Slutzki, M., Noach, I., Chitayat, S., Smith, S. P., et al. (2013). Structural and functional characterization of a novel type-III dockerin from *Ruminococcus flavefaciens*. *FEBS Lett.* 587, 30–36. doi: 10.1016/j.febslet.2012.11.012
- Kieboom, J., Dennis, J. J., De Bont, J. A., and Zylstra, G. J. (1998a). Identification and molecular characterization of an efflux pump involved in *Pseudomonas putida* S12 solvent tolerance. *J. Biol. Chem.* 273, 85–91. doi: 10.1074/jbc.273.1.85
- Kieboom, J., Dennis, J. J., Zylstra, G. J., and De Bont, J. A. (1998b). Active efflux of organic solvents by *Pseudomonas putida* S12 is induced by solvents. *J. Bacteriol.* 180, 6769–6772.
- Klinke, H. B., Thomsen, A. B., and Ahring, B. K. (2004). Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pretreatment of biomass. *Appl. Microbiol. Biotechnol.* 66, 10–26. doi: 10.1007/s00253-004-1642-2
- Köne, A. C., and Büke, T. (2010). Forecasting of CO<sub>2</sub> emissions from fuel combustion using trend analysis. *Renew. Sust. Energ. Rev.* 14, 2906–2915. doi: 10.1016/j.rser.2010.06.006
- Laadan, B., Almeida, J. R., Radstrom, P., Hahn-Hagerdal, B., and Gorwa-Grauslund, M. (2008). Identification of an NADH-dependent 5-hydroxymethylfurfural-reducing alcohol dehydrogenase in *Saccharomyces cerevisiae*. *Yeast* 25, 191–198. doi: 10.1002/yea.1578
- Lai, C. Y., and Cronan, J. E. (2003). Beta-ketoacyl-acyl carrier protein synthase III (FabH) is essential for bacterial fatty acid synthesis. *J. Biol. Chem.* 278, 51494–51503. doi: 10.1074/jbc.M308638200
- Laycock, B., Halley, P., Pratt, S., Werker, A., and Lant, P. (2014). The chemomechanical properties of microbial polyhydroxyalkanoates. *Prog. Polym. Sci.* 39, 397–442. doi: 10.1016/j.progpolymsci.2013.06.008
- Lee, I., Johnson, L. A., and Hammond, E. G. (1995). Use of branched-chain esters to reduce the crystallization temperature of biodiesel. *J. Am. Oil Chem. Soc.* 72, 1155–1160. doi: 10.1007/BF02540982
- Lee, L. C., Lee, Y. L., Leu, R. J., and Shaw, J. F. (2006). Functional role of catalytic triad and oxyanion hole-forming residues on enzyme activity of *Escherichia coli* thioesterase I/protease I/phospholipase L1. *Biochem. J.* 397, 69–76. doi: 10.1042/BJ20051645
- Lee, S. M., Cho, M. K., Park, C. H., Chung, Y. C., Kim, J. H., Sang, B. I., et al. (2008a). Continuous butanol production using suspended and immobilized *Clostridium beijerinckii* NCIMB 8052 with supplementary butyrate. *Energ. Fuel.* 22, 3459–3464. doi: 10.1021/ef800076j
- Lee, S. Y., Park, J. H., Jang, S. H., Nielsen, L. K., Kim, J., and Jung, K. S. (2008b). Fermentative butanol production by *Clostridia*. *Biotechnol. Bioeng.* 101, 209–228. doi: 10.1002/bit.22003
- Lennen, R. M., and Pfleger, B. F. (2012). Engineering *Escherichia coli* to synthesize free fatty acids. *Trends Biotechnol.* 30, 659–667. doi: 10.1016/j.tibtech.2012.09.006
- Lennen, R. M., and Pfleger, B. F. (2013). Microbial production of fatty acid-derived fuels and chemicals. *Curr. Opin. Biotechnol.* 24, 1044–1053. doi: 10.1016/j.copbio.2013.02.028
- Li, N., Chang, W. C., Warui, D. M., Booker, S. J., Krebs, C., and Bollinger, J. M. Jr. (2012). Evidence for only oxygenative cleavage of aldehydes to alk(a)enes and formate by cyanobacterial aldehyde decarboxylases. *Biochemistry* 51, 7908–7916. doi: 10.1021/bi300912n
- Li, N., Norgaard, H., Warui, D. M., Booker, S. J., Krebs, C., and Bollinger, J. M. Jr. (2011). Conversion of fatty aldehydes to alk(a)enes and formate by a cyanobacterial aldehyde decarboxylase: cryptic redox by an unusual dimetal oxygenase. *J. Am. Chem. Soc.* 133, 6158–6161. doi: 10.1021/ja2013517
- Li, S. J., and Cronan, J. E. Jr. (1992). The gene encoding the biotin carboxylase subunit of *Escherichia coli* acetyl-CoA carboxylase. *J. Biol. Chem.* 267, 855–863.
- Ling, H., Chen, B., Kang, A., Lee, J. M., and Chang, M. W. (2013). Transcriptome response to alkane biofuels in *Saccharomyces cerevisiae*: identification of efflux pumps involved in alkane tolerance. *Biotechnol. Biofuels* 6:95. doi: 10.1186/1754-6834-6-95
- Liu, S., and Qureshi, N. (2009a). How microbes tolerate ethanol and butanol. *N. Biotechnol.* 26, 117–121. doi: 10.1016/j.nbt.2009.06.984
- Liu, Z. L., and Moon, J. (2009b). A novel NADPH-dependent aldehyde reductase gene from *Saccharomyces cerevisiae* NRRL Y-12632 involved in the detoxification of aldehyde inhibitors derived from lignocellulosic biomass conversion. *Gene* 446, 1–10. doi: 10.1016/j.gene.2009.06.018
- Lu, J. N., Tappel, R. C., and Nomura, C. T. (2009). Mini-review: biosynthesis of poly(hydroxyalkanoates). *J. Macromol. Sci.* 49, 226–248.
- Marinković, M., De Leeuw, W. C., De Jong, M., Kraak, M. H., Admiraal, W., Breit, T. M., et al. (2012). Combining next-generation sequencing and microarray technology into a transcriptomics approach for the non-model organism *Chironomus riparius*. *PLoS ONE* 7:e48096. doi: 10.1371/journal.pone.0048096
- McCarthy, J. G., Eisman, E. B., Kulkarni, S., Gerwick, L., Gerwick, W. H., Wipf, P., et al. (2012). Structural basis of functional group activation by sulfotransferases in complex metabolic pathways. *ACS Chem. Biol.* 7, 1994–2003. doi: 10.1021/cb300385m
- McEvoy, E., Wright, P. C., and Bustard, M. T. (2004). The effect of high concentration isopropanol on the growth of a solvent-tolerant strain of *Chlorella vulgaris*. *Enzyme Microb. Tech.* 35, 140–146. doi: 10.1016/j.enzymitec.2004.01.012
- McNeil, J. B., McIntosh, E. M., Taylor, B. V., Zhang, F. R., Tang, S., and Bognar, A. L. (1994). Cloning and molecular characterization of three genes, including two genes encoding serine hydroxymethyltransferases, whose inactivation is required to render yeast auxotrophic for glycine. *J. Biol. Chem.* 269, 9155–9165.
- Mehta, R. N., Chakraborty, M., Mahanta, P., and Parikh, P. A. (2010). Evaluation of fuel properties of butanol-biodiesel-diesel blends and their impact on engine performance and emissions. *Ind. Eng. Chem. Res.* 49, 7660–7665. doi: 10.1021/ie1006257
- Mendez-Perez, D., Begemann, M. B., and Pfleger, B. F. (2011). Modular synthase-encoding gene involved in alpha-olefin biosynthesis in *Synechococcus* sp. strain PCC 7002. *Appl. Environ. Microbiol.* 77, 4264–4267. doi: 10.1128/AEM.00467-11
- Mingardon, F., Chanal, A., Lopez-Contreras, A. M., Dray, C., Bayer, E. A., and Fierobe, H. P. (2007). Incorporation of fungal cellulases in bacterial minicellulosomes yields viable, synergistically acting cellulolytic complexes. *Appl. Environ. Microbiol.* 73, 3822–3832. doi: 10.1128/AEM.00398-07
- Mitsuzawa, S., Kagawa, H., Li, Y., Chan, S. L., Paavola, C. D., and Trent, J. D. (2009). The rosettazyme: a synthetic cellulosome. *J. Biotechnol.* 143, 139–144. doi: 10.1016/j.jbiotec.2009.06.019
- Miziorko, H. M. (2011). Enzymes of the mevalonate pathway of isoprenoid biosynthesis. *Arch. Biochem. Biophys.* 505, 131–143. doi: 10.1016/j.abb.2010.09.028
- Mohan, S., Kelly, T. M., Eveland, S. S., Raetz, C. R., and Anderson, M. S. (1994). An *Escherichia coli* gene (FabZ) encoding (3R)-hydroxymyristoyl acyl carrier protein dehydrase. Relation to fabA and suppression of mutations in lipid biosynthesis. *J. Biol. Chem.* 269, 32896–32903.
- Molnar, I., Lopez, D., Wisecaver, J. H., Devarenne, T. P., Weiss, T. L., Pellegrini, M., et al. (2012). Bio-crude transcriptomics: gene discovery and metabolic network reconstruction for the biosynthesis of the terpenome of the hydrocarbon oil-producing green alga, *Botryococcus braunii* race B (Showa). *BMC Genomics* 13:576. doi: 10.1186/1471-2164-13-576
- Morais, S., Barak, Y., Caspi, J., Hadar, Y., Lamed, R., Shoham, Y., et al. (2010). Cellulase-xylanase synergy in designer cellulosomes for enhanced degradation of a complex cellulosic substrate. *MBio* 1, e00285-10. doi: 10.1128/mBio.00285-10
- Mosqueda, G., and Ramos, J. L. (2000). A set of genes encoding a second toluene efflux system in *Pseudomonas putida* DOT-T1E is linked to the tod genes for toluene metabolism. *J. Bacteriol.* 182, 937–943. doi: 10.1128/JB.182.4.937-943.2000
- Murashima, K., Chen, C. L., Kosugi, A., Tamaru, Y., Doi, R. H., and Wong, S. L. (2002). Heterologous production of *Clostridium cellulovorans* engB, using protease-deficient *Bacillus subtilis*, and preparation of active recombinant cellulosomes. *J. Bacteriol.* 184, 76–81. doi: 10.1128/JB.184.1.76-81.2002
- Nawabi, P., Bauer, S., Kyrpides, N., and Lykidis, A. (2011). Engineering *Escherichia coli* for biodiesel production utilizing a bacterial fatty acid methyltransferase. *Appl. Environ. Microbiol.* 77, 8052–8061. doi: 10.1128/AEM.05046-11
- Nordon, R. E., Craig, S. J., and Foong, F. C. (2009). Molecular engineering of the cellulosome complex for affinity and bioenergy applications. *Biotechnol. Lett.* 31, 465–476. doi: 10.1007/s10529-008-9899-7
- Oefner, C., Schulz, H., D'arcy, A., and Dale, G. E. (2006). Mapping the active site of *Escherichia coli* malonyl-CoA-acyl carrier protein transacylase (FabD) by protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* 62, 613–618. doi: 10.1107/S0907444906009474
- Park, S. J., Kim, T. W., Kim, M. K., Lee, S. Y., and Lim, S. C. (2012). Advanced bacterial polyhydroxyalkanoates: towards a versatile and sustainable platform for unnatural tailor-made polyesters. *Biotechnol. Adv.* 30, 1196–1206. doi: 10.1016/j.biotechadv.2011.11.007



- Peralta-Yahya, P. P., and Keasling, J. D. (2010). Advanced biofuel production in microbes. *Biotechnol. J.* 5, 147–162. doi: 10.1002/biot.200900220
- Peralta-Yahya, P. P., Zhang, F., Del Cardayre, S. B., and Keasling, J. D. (2012). Microbial engineering for the production of advanced biofuels. *Nature* 488, 320–328. doi: 10.1038/nature11478
- Perret, S., Belaich, A., Fierobe, H. P., Belaich, J. P., and Tardif, C. (2004). Towards designer cellulosomes in *Clostridia*: mannanase enrichment of the cellulosomes produced by *Clostridium cellulolyticum*. *J. Bacteriol.* 186, 6544–6552. doi: 10.1128/JB.186.19.6544-6552.2004
- Petersson, A., Almeida, J. R., Modig, T., Karhumaa, K., Hahn-Hagerdal, B., Gorwa-Grauslund, M. F., et al. (2006). A 5-hydroxymethyl furfural reducing enzyme encoded by the *Saccharomyces cerevisiae* ADH6 gene conveys HMF tolerance. *Yeast* 23, 455–464. doi: 10.1002/yea.1370
- Pfeiffer, D., and Jendrossek, D. (2012). Localization of poly(3-hydroxybutyrate) (PHB) granule-associated proteins during PHB granule formation and identification of two new phasins, PhaP6 and PhaP7, in *Ralstonia eutropha* H16. *J. Bacteriol.* 194, 5909–5921. doi: 10.1128/JB.00779-12
- Plewniak, F., Koehler, S., Navet, B., Dugat-Bony, E., Bouchez, O., Peyret, P., et al. (2013). Metagenomic insights into microbial metabolism affecting arsenic dispersion in Mediterranean marine sediments. *Mol. Ecol.* 22, 4870–4883. doi: 10.1111/mec.12432
- Qureshi, N., Sahal, B. C., Hector, R. E., and Cotta, M. A. (2010). Production of butanol (a biofuel) from agricultural residues: Part 1-Use of barley straw hydrolysate. *Biomass Bioenerg.* 34, 559–565. doi: 10.1016/j.biombioe.2009.12.024
- Ramos, J. L., Duque, E., Godoy, P., and Segura, A. (1998). Efflux pumps involved in toluene tolerance in *Pseudomonas putida* DOT-T1E. *J. Bacteriol.* 180, 3323–3329.
- Rashid, U., Anwar, F., Moser, B. R., and Knothe, G. (2008). Moringa oleifera oil: a possible source of biodiesel. *Bioresour. Technol.* 99, 8175–8179. doi: 10.1016/j.biortech.2008.03.066
- Ravagnani, A., Jennert, K. C., Steiner, E., Grunberg, R., Jefferies, J. R., Wilkinson, S. R., et al. (2000). Spo0A directly controls the switch from acid to solvent production in solvent-forming *Clostridia*. *Mol. Microbiol.* 37, 1172–1185. doi: 10.1046/j.1365-2958.2000.02071.x
- Reddy, C. S., Ghai, R., Rashmi, and Kalia, V. C. (2003). Polyhydroxyalkanoates: an overview. *Bioresour. Technol.* 87, 137–146. doi: 10.1016/S0960-8524(02)00212-2
- Riedel, S. L., Lu, J., Stahl, U., and Brigham, C. J. (2014). Lipid and fatty acid metabolism in *Ralstonia eutropha*: relevance for the biotechnological production of value-added products. *Appl. Microbiol. Biotechnol.* 98, 1469–1483. doi: 10.1007/s00253-013-5430-8
- Rincon, M. T., Cepeljnik, T., Martin, J. C., Lamed, R., Barak, Y., Bayer, E. A., et al. (2005). Unconventional mode of attachment of the *Ruminococcus flavefaciens* cellulosome to the cell surface. *J. Bacteriol.* 187, 7569–7578. doi: 10.1128/JB.187.22.7569-7578.2005
- Rojas, A., Duque, E., Mosqueda, G., Golden, G., Hurtado, A., Ramos, J. L., et al. (2001). Three efflux pumps are required to provide efficient tolerance to toluene in *Pseudomonas putida* DOT-T1E. *J. Bacteriol.* 183, 3967–3973. doi: 10.1128/JB.183.13.3967-3973.2001
- Rude, M. A., Baron, T. S., Brubaker, S., Alibhai, M., Del Cardayre, S. B., and Schirmer, A. (2011). Terminal olefin (1-alkene) biosynthesis by a novel p450 fatty acid decarboxylase from *Jeotgalicoccus* species. *Appl. Environ. Microbiol.* 77, 1718–1727. doi: 10.1128/AEM.02580-10
- Rude, M. A., and Schirmer, A. (2009). New microbial fuels: a biotech perspective. *Curr. Opin. Microbiol.* 12, 274–281. doi: 10.1016/j.mib.2009.04.004
- Sato, T. (2013). Unique biosynthesis of sesquiterpenes (C<sub>15</sub> terpenes). *Biosci. Biotechnol. Biochem.* 77, 1155–1159. doi: 10.1271/bbb.130180
- Schirmer, A., Rude, M. A., Li, X., Popova, E., and Del Cardayre, S. B. (2010). Microbial biosynthesis of alkanes. *Science* 329, 559–562. doi: 10.1126/science.1187936
- Schmidt, A., Forne, I., and Imhof, A. (2014). Bioinformatic analysis of proteomics data. *BMC Syst. Biol.* 8 Suppl. 2:S3. doi: 10.1186/1752-0509-8-S2-S3
- Scotcher, M. C., and Bennett, G. N. (2005). SpoIIE regulates sporulation but does not directly affect solventogenesis in *Clostridium acetobutylicum* ATCC 824. *J. Bacteriol.* 187, 1930–1936. doi: 10.1128/JB.187.6.1930-1936.2005
- Segura, A., Molina, L., Fillet, S., Krell, T., Bernal, P., Munoz-Rojas, J., et al. (2012). Solvent tolerance in Gram-negative bacteria. *Curr. Opin. Biotechnol.* 23, 415–421. doi: 10.1016/j.copbio.2011.11.015
- Shahid, S., Mosrati, R., Ledauphin, J., Amiel, C., Fontaine, P., Gaillard, J. L., et al. (2013). Impact of carbon source and variable nitrogen conditions on bacterial biosynthesis of polyhydroxyalkanoates: evidence of an atypical metabolism in *Bacillus megaterium* DSM 509. *J. Biosci. Bioeng.* 116, 302–308. doi: 10.1016/j.jbiosc.2013.02.017
- Smith, S. P., and Bayer, E. A. (2013). Insights into cellulosome assembly and dynamics: from dissection to reconstruction of the supramolecular enzyme complex. *Curr. Opin. Struct. Biol.* 23, 686–694. doi: 10.1016/j.sbi.2013.09.002
- Srikrishnan, S., Chen, W., and Da Silva, N. A. (2013). Functional assembly and characterization of a modular xylanosome for hemicellulose hydrolysis in yeast. *Biotechnol. Bioeng.* 110, 275–285. doi: 10.1002/bit.24609
- Steen, E. J., Chan, R., Prasad, N., Myers, S., Petzold, C. J., Redding, A., et al. (2008). Metabolic engineering of *Saccharomyces cerevisiae* for the production of n-butanol. *Microb. Cell Fact.* 7:36. doi: 10.1186/1475-2859-7-36
- Steen, E. J., Kang, Y., Bokinsky, G., Hu, Z., Schirmer, A., McClure, A., et al. (2010). Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* 463, 559–562. doi: 10.1038/nature08721
- Stephenson, M. (1949). *Microbial Metabolism*. Cambridge: MIT Press.
- Stoveken, T., Kalscheuer, R., Malkus, U., Reichelt, R., and Steinbuchel, A. (2005). The wax ester synthase/acyl coenzyme A:diacylglycerol acyltransferase from *Acinetobacter* sp. strain ADP1: characterization of a novel type of acyltransferase. *J. Bacteriol.* 187, 1369–1376. doi: 10.1128/JB.187.4.1369-1376.2005
- Sukovich, D. J., Seffernick, J. L., Richman, J. E., Gralnick, J. A., and Wackett, L. P. (2010). Widespread head-to-head hydrocarbon biosynthesis in bacteria and role of OleA. *Appl. Environ. Microbiol.* 76, 3850–3862. doi: 10.1128/AEM.00436-10
- Suranovic, S. (2013). Fossil fuel addiction and the implication for climate change policy. *Global Environ. Chang.* 23, 598–608. doi: 10.1016/j.gloenvcha.2013.02.006
- Torella, J. P., Ford, T. J., Kim, S. N., Chen, A. M., Way, J. C., and Silver, P. A. (2013). Tailored fatty acid synthesis via dynamic control of fatty acid elongation. *Proc. Natl. Acad. Sci. U.S.A.* 110, 11290–11295. doi: 10.1073/pnas.1307129110
- Torto-Alalibo, T., Collmer, C. W., and Gwinn-Giglio, M. (2009). The Plant-Associated Microbe Gene Ontology (PAMGO) consortium: community development of new Gene Ontology terms describing biological processes involved in microbe-host interactions. *BMC Microbiol.* 9(Suppl. 1):S1. doi: 10.1186/1471-2180-9-S1-S1
- Vazana, Y., Barak, Y., Unger, T., Peleg, Y., Shamshoum, M., Ben-Yehzekel, T., et al. (2013). A synthetic biology approach for evaluating the functional contribution of designer cellulosome components to deconstruction of cellulosic substrates. *Biotechnol. Biofuels* 6:182. doi: 10.1186/1754-6834-6-182
- Villas-Bôas, S. G., Kesson, M., and Nielsen, J. (2005). Biosynthesis of glyoxylate from glycine in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* 5, 703–709. doi: 10.1016/j.femsyr.2005.03.001
- Warui, D. M., Li, N., Norgaard, H., Krebs, C., Bollinger, J. M. Jr., and Booker, S. J. (2011). Detection of formate, rather than carbon monoxide, as the stoichiometric coproduct in conversion of fatty aldehydes to alkanes by a cyanobacterial aldehyde decarboxylase. *J. Am. Chem. Soc.* 133, 3316–3319. doi: 10.1021/ja111607x
- Weber, C., Farwick, A., Benisch, F., Brat, D., Dietz, H., Subtil, T., et al. (2010). Trends and challenges in the microbial production of lignocellulosic bioalcohol fuels. *Appl. Microbiol. Biotechnol.* 87, 1303–1315. doi: 10.1007/s00253-010-2707-z
- Wen, M., Bond-Watts, B. B., and Chang, M. C. (2013). Production of advanced biofuels in engineered *E. coli*. *Curr. Opin. Chem. Biol.* 17, 472–479. doi: 10.1016/j.cbpa.2013.03.034
- Wilson, D. B. (2011). Microbial diversity of cellulose hydrolysis. *Curr. Opin. Microbiol.* 14, 259–263. doi: 10.1016/j.mib.2011.04.004
- Wilson, D. B. (2012). Processive and nonprocessive cellulases for biofuel production—lessons from bacterial genomes and structural analysis. *Appl. Microbiol. Biotechnol.* 93, 497–502. doi: 10.1007/s00253-011-3701-9
- Xue, C., Zhao, X. Q., Liu, C. G., Chen, L. J., and Bai, F. W. (2013). Prospective and development of butanol as an advanced biofuel. *Biotechnol. Adv.* 31, 1575–1584. doi: 10.1016/j.biotechadv.2013.08.004
- Yikmis, M., and Steinbuchel, A. (2012). Historical and recent achievements in the field of microbial degradation of natural and synthetic rubber. *Appl. Environ. Microbiol.* 78, 4543–4551. doi: 10.1128/AEM.00001-12
- Zhang, F., Rodriguez, S., and Keasling, J. D. (2011). Metabolic engineering of microbial pathways for advanced biofuels production. *Curr. Opin. Biotechnol.* 22, 775–783. doi: 10.1016/j.copbio.2011.04.024

- Zhang, X., Luo, R., Wang, Z., Deng, Y., and Chen, G. Q. (2009). Application of (R)-3-hydroxyalkanoate methyl esters derived from microbial polyhydroxyalkanoates as novel biofuels. *Biomacromolecules* 10, 707–711. doi: 10.1021/bm801424e
- Zhao, L., Chang, W. C., Xiao, Y., Liu, H. W., and Liu, P. (2013). Methylerythritol phosphate pathway of isoprenoid biosynthesis. *Annu. Rev. Biochem.* 82, 497–530. doi: 10.1146/annurev-biochem-052010-100934
- Zhou, Y. J., and Zhao, Z. K. (2011). Advanced biofuel-oriented engineering of fatty acid pathway: a review. *Sheng Wu Gong Cheng Xue Bao* 27, 1261–1267.
- Zhu, H., Ren, X., Wang, J., Song, Z., Shi, M., Qiao, J., et al. (2013). Integrated OMICS guided engineering of biofuel butanol-tolerance in photosynthetic *Synechocystis* sp. PCC 6803. *Biotechnol. Biofuels* 6:106. doi: 10.1186/1754-6834-6-106
- Zingaro, K. A., and Terry Papoutsakis, E. (2013). GroESL overexpression imparts *Escherichia coli* tolerance to i-, n-, and 2-butanol, 1,2,4-butanetriol and ethanol with complex and unpredictable patterns. *Metab. Eng.* 15, 196–205. doi: 10.1016/j.ymben.2012.07.009
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 28 July 2014; accepted: 22 September 2014; published online: 10 October 2014.

Citation: Torto-Alalibo T, Purwantini E, Lomax J, Setubal JC, Mukhopadhyay B and Tyler BM (2014) Genetic resources for advanced biofuel production described with the Gene Ontology. *Front. Microbiol.* 5:528. doi: 10.3389/fmicb.2014.00528

This article was submitted to *Microbial Physiology and Metabolism*, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Torto-Alalibo, Purwantini, Lomax, Setubal, Mukhopadhyay and Tyler. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Aromatic inhibitors derived from ammonia-pretreated lignocellulose hinder bacterial ethanologenesis by activating regulatory circuits controlling inhibitor efflux and detoxification

David H. Keating<sup>1†</sup>, Yaoping Zhang<sup>1†</sup>, Irene M. Ong<sup>1†</sup>, Sean McIlwain<sup>1†</sup>, Eduardo H. Morales<sup>1,2†</sup>, Jeffrey A. Grass<sup>1,3</sup>, Mary Tremaine<sup>1</sup>, William Bothfeld<sup>1</sup>, Alan Higbee<sup>1</sup>, Arne Ulbrich<sup>4</sup>, Allison J. Balloon<sup>4</sup>, Michael S. Westphall<sup>2,4</sup>, Josh Aldrich<sup>5</sup>, Mary S. Lipton<sup>5</sup>, Joonhoon Kim<sup>1,6</sup>, Oleg V. Moskvina<sup>1</sup>, Yury V. Bukhman<sup>1</sup>, Joshua J. Coon<sup>1,2,4</sup>, Patricia J. Kiley<sup>1,2</sup>, Donna M. Bates<sup>1\*</sup> and Robert Landick<sup>1,3,7\*</sup>

<sup>1</sup> Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI, USA

<sup>2</sup> Department of Biomolecular Chemistry, University of Wisconsin-Madison, Madison, WI, USA

<sup>3</sup> Department of Biochemistry, University of Wisconsin-Madison, Madison, WI, USA

<sup>4</sup> Department of Chemistry, University of Wisconsin-Madison, Madison, WI, USA

<sup>5</sup> Pacific Northwest National Laboratory, Richland, WA, USA

<sup>6</sup> Department of Chemical and Biological Engineering, University of Wisconsin-Madison, Madison, WI, USA

<sup>7</sup> Department of Bacteriology, University of Wisconsin-Madison, Madison, WI, USA

## Edited by:

Katherine M. Pappas, University of Athens, Greece

## Reviewed by:

Carl James Yeoman, Montana State University, USA

Helen Zgurskaya, University of Oklahoma, USA

## \*Correspondence:

Donna M. Bates, Great Lakes Bioenergy Research Center, 4119 Wisconsin Energy Institute, University of Wisconsin - Madison, 1552 University Ave., Madison, WI 53726, USA

e-mail: dbates@glbrc.wisc.edu;  
Robert Landick, Department of Biochemistry, 5441 Microbial Sciences, 1550 Linden Dr., University of Wisconsin, Madison, WI 53706, USA

e-mail: landick@biochem.wisc.edu

<sup>†</sup> These authors have contributed equally to this work.

Efficient microbial conversion of lignocellulosic hydrolysates to biofuels is a key barrier to the economically viable deployment of lignocellulosic biofuels. A chief contributor to this barrier is the impact on microbial processes and energy metabolism of lignocellulose-derived inhibitors, including phenolic carboxylates, phenolic amides (for ammonia-pretreated biomass), phenolic aldehydes, and furfurals. To understand the bacterial pathways induced by inhibitors present in ammonia-pretreated biomass hydrolysates, which are less well studied than acid-pretreated biomass hydrolysates, we developed and exploited synthetic mimics of ammonia-pretreated corn stover hydrolysate (ACSH). To determine regulatory responses to the inhibitors normally present in ACSH, we measured transcript and protein levels in an *Escherichia coli* ethanologen using RNA-seq and quantitative proteomics during fermentation to ethanol of synthetic hydrolysates containing or lacking the inhibitors. Our study identified four major regulators mediating these responses, the MarA/SoxS/Rob network, AaeR, FrmR, and YqhC. Induction of these regulons was correlated with a reduced rate of ethanol production, buildup of pyruvate, depletion of ATP and NAD(P)H, and an inhibition of xylose conversion. The aromatic aldehyde inhibitor 5-hydroxymethylfurfural appeared to be reduced to its alcohol form by the ethanologen during fermentation, whereas phenolic acid and amide inhibitors were not metabolized. Together, our findings establish that the major regulatory responses to lignocellulose-derived inhibitors are mediated by transcriptional rather than translational regulators, suggest that energy consumed for inhibitor efflux and detoxification may limit biofuel production, and identify a network of regulators for future synthetic biology efforts.

**Keywords:** *Escherichia coli*, lignocellulosic hydrolysate, aromatic inhibitors, transcriptomics, RNAseq, proteomics, ethanol, biofuels

## INTRODUCTION

Elucidation of metabolic and regulatory barriers in microbial conversion of lignocellulosic sugars to ethanol is crucial for both the immediate goal of economical cellulosic ethanol and for the long-term development of next-generation biofuels and sustainable chemicals from renewable biomass. Efficient conversion of lignocellulose (LC) hydrolysates is limited by multiple factors (Mills et al., 2009; Lau and Dale, 2010), including high osmolarity (Underwood et al., 2004; Purvis et al., 2005; Miller and Ingram,

2007), toxicity of the conversion products (Ingram and Buttke, 1984), and inhibitors of microbial metabolism and growth generated during the deconstruction of LC (Zaldivar et al., 1999; Wang et al., 2011a; Tang et al., submitted). Understanding and overcoming the barriers created by LC-derived inhibitors presents significant challenges as their composition can vary depending on the biomass source of LC, the methods used to deconstruct the LC, and the diverse metabolic and regulatory responses of microbes to inhibitors (Klinke et al., 2004; Liu, 2011). Synergy among the

inhibitors, the high osmolarity inherent to hydrolysates, and toxicity of conversion products (e.g., ethanol) are additional factors that contribute to the complex molecular landscape of lignocellulosic hydrolysates (Klinke et al., 2004; Liu, 2011; Piotrowski et al., 2014).

Release of sugars from LC typically requires either acidic or alkaline treatment of biomass prior to or coupled with chemical or enzymatic hydrolysis (Chundawat et al., 2011). Acidic treatments generate significant microbial inhibitors by condensation reactions of sugars (e.g., furfural and 5-hydroxymethylfurfural). Microbes typically detoxify these aldehydes by reduction or oxidation to less toxic alcohols or acids (Booth et al., 2003; Herring and Blattner, 2004; Marx et al., 2004; Jarboe, 2011), but these conversions also directly or indirectly consume energy that otherwise would be available for biofuel synthesis (Miller et al., 2009a,b). The impact of these inhibitors is especially significant for C5 sugars like xylose whose catabolism provide slightly less cellular energy (Lawford and Rousseau, 1995), and can be partially ameliorated by replacing NADPH-consuming enzymes with NADH-consuming enzymes (Wang et al., 2013).

Alkaline treatments, for instance with ammonia, are potentially advantageous in generating fewer toxic aldehydes, but the spectrum of inhibitors generated by alkaline treatments is less well characterized and their effects on microbial metabolism are less well understood. We have developed an approach to elucidate the metabolic and regulatory barriers to microbial conversion in LC hydrolysates using ammonia fiber expansion (AFEX) of corn stover, enzymatic hydrolysis, and a model ethanologen (GLBRCE1) engineered from the well-studied bacterium *E. coli* K-12 (Schwalbach et al., 2012). Our strategy is to compare anaerobic metabolic and regulatory responses of the ethanologen in authentic AFEX-pretreated corn stover hydrolysate (ACSH) to responses to synthetic hydrolysates (SynHs) designed to mimic ACSH with a chemically defined medium. GLBRCE1 metabolizes ACSH in exponential, transition, and stationary phases but, unlike growth in traditional rich media (Sezonov et al., 2007), GLBRCE1 enters stationary phase (ceases growth) long before depletion of available glucose but coincident with exhaustion of amino acid sources of organic nitrogen (Schwalbach et al., 2012). The growth-arrested cells remain metabolically active and convert the remaining glucose, but not xylose, into ethanol (Schwalbach et al., 2012).

Our first version of SynH (SynH1) matched ACSH for levels of glucose, xylose, amino acids, and some inorganics, overall osmolality, and the amino-acid-dependent growth arrest of GLBRCE1 (Schwalbach et al., 2012). However, gene expression profiling revealed that SynH1 cells experienced significant osmotic stress relative to ACSH cells, whereas ACSH cells exhibited elevated expression of efflux pumps, notably of *aaeAB* that acts on aromatic carboxylates (Van Dyk et al., 2004), relative to SynH1 cells (Schwalbach et al., 2012). Osmolytes found in ACSH (betaine, choline, and carnitine) likely explained the lower osmotic stress, whereas phenolic carboxylates derived from LC (e.g., coumarate and ferulate) likely explained efflux pump induction possibly via the AaeR and MarA/SoxS/Rob regulons known to be induced by phenolic carboxylates (Sulavik et al., 1995; Dalrymple and Swadling, 1997). We also observed elevated expression of *psp*,

*ibp*, and *srl* genes associated with ethanol stress at ethanol concentrations three-fold lower than previously reported to induce expression (Yomano et al., 1998; Goodarzi et al., 2010) and thus consistent with a synergistic stress response with the LC-derived inhibitors. These findings led us to hypothesize that the collective effects of osmotic, ethanol, and LC-derived inhibitor stresses created an increased need for ATP and reducing equivalents that was partially offset in early growth phase by catabolism of amino acids, as N and possibly S sources. However, as these amino acids are depleted, cells transition to stationary phase where they continue to catabolize glucose for maintenance ATP and NAD(P)H but are unable to generate sufficient energy for cell growth or efficient xylose catabolism.

To test this hypothesis, we developed a new SynH formulation (SynH2) that faithfully replicates the physiological responses in ACSH and the effects of LC-derived inhibitors. Using SynH2 with and without the LC-derived inhibitors, we generated and analyzed metabolomic, gene expression, and proteomic data to define the effects of inhibitors on bacterial gene expression and physiology. The analysis allowed identification of key regulators that may provoke stress responses in the presence of LC-derived inhibitors and suggest that coping mechanisms employed by *E. coli* to deal with lignocellulosic stress drains cellular energy, thus limiting xylose conversion.

## MATERIALS AND METHODS

### REAGENTS

Reagents and chemicals were obtained from Thermo Fisher Scientific (Pittsburgh, Pennsylvania, USA) or Sigma Aldrich Co. (Saint Louis, Missouri, USA) with the following exceptions. 5-hydroxymethyl-2-furancarboxylic acid and 5-(hydroxymethyl)furfuryl alcohol were obtained from Toronto Research Chemicals Inc. (Toronto, Ontario, Canada). Deuterated compounds for HS-SPME-GC/IDMS were obtained from C/D/N Isotopes (Pointe-Claire, Quebec, Canada). D4-acetaldehyde and U<sup>13</sup>C<sub>6</sub>-fructose were obtained from Cambridge Isotope Labs (Andover, Massachusetts, USA).

### SYNTHESIS OF FERULOYL AND COUMAROYL AMIDES

Twenty grams of ferulic or coumaric acid were dissolved in 200 ml of 100% ethanol in a 3-neck, 250 ml round-bottom flask equipped with a magnetic stir bar and a drying tube on one of the outside arms. Ten milliliters of acetyl chloride was added and incubated with stirring at room temperature overnight. Ethanol was removed in a rotary evaporator at 40°C under modest vacuum; the syrup re-dissolved in 250 ml 100% ethanol and re-evaporated twice. When the final syrup was reduced to <25 ml, ~6 ml portions were transferred to heavy-wall 25 × 150 mm tubes containing ~30 ml concentrated ammonium hydroxide and sealed with a Teflon-lined cap. The sealed tubes were incubated at 95°C in a heating block covered with a safety shield overnight. The tubes were cooled and then left open in a hood for 4–8 h to allow evaporation of ammonium hydroxide, during which the feruloyl or coumaroyl amide precipitated. The crystallized products were collected under vacuum on a glass filter and washed with 250 ml ice-cold 150 mM ammonium hydroxide. The product was allowed to air dry in a plastic weigh boat in the

hood at room temperature for 2–3 days. Purity of the products was analyzed by silica gel TLC developed with 5% methanol in chloroform. Only preparations exceeding 90% purity were used for experiments.

### PREPARATION OF ACSH

ACSH was prepared by one of two methods that differed in whether or not CS was autoclaved prior to enzymatic hydrolysis. Non-autoclaved CS hydrolysate more closely replicates an industrial process, was used by Tang et al. (submitted) for compositional analysis, and was used for some of our fermentation experiments. Autoclaved CS hydrolysate ensures sterility for bacterial fermentations and was used for our compositional analysis and for experiments to generate RNA-seq data. We did not observe a significant difference in GLBRCE1 behavior in non-autoclaved vs. autoclaved CS hydrolysates, although HMF was detectable in the former, but not the latter (Table 2). We observed minor variations in growth with CS harvested in different years. For autoclaved CS hydrolysate, AFEX-pretreated CS was mixed with water to 6–10 L final volume at 60 g glucan/L loading (18–22% solids, adjusted for moisture content) and autoclaved for 30–120 min in a 15 L Applikon bioreactor vessel (Schwalbach et al., 2012). For non-autoclaved CS hydrolysate, AFEX pretreated-corn stover was added to the vessel after the water was autoclaved for 30 min. For both, the sample was cooled to ~70°C, adjusted to 10 L volume with water, and pH adjusted with ~30 ml concentrated HCl. Hydrolysis was initiated by adding Novozymes CTec2 to 24 mg/g glucan and HTec2 to 6 mg/g glucan, followed by incubation for 5 days at 50°C with stir speed at 700 rpm. Some older batches of hydrolysate were prepared using Genencor Accellerase, Genencor Accellerase XY, and Multifect pectinase A in place of Novozyme enzymes (Schwalbach et al., 2012). Solids were then removed by centrifugation (8200 × g, 4°C, 10–12 h) and the supernatant was filter-sterilized through 0.5 µm and then 0.2 µm filters. Prior to fermentation, the hydrolysate was adjusted to pH 7.0 using NaOH pellets and filtered again through a 0.2 µm filter to remove precipitates and to ensure sterility.

### PREPARATION OF SYNTHETIC HYDROLYSATE (SYNH2)

SynH2 (Table 1) was prepared by combining per L final volume of SynH2 the following ingredients. Water (~700 ml) was mixed with 6.25 ml of 1.6 M KPO<sub>4</sub> buffer, pH 7.2, 20 ml of 1.5 M ammonium sulfate, 20 ml of 2.25 M KCl, 1.25 M NaCl, 20 ml of a 50X amino acid stock giving the final concentrations shown in Table 1 (except tyrosine), 20 ml of 8.75 mM tyrosine dissolved in 50 mM HCl, 50 ml of 1 mM each adenine, guanine, cytosine and uracil dissolved in 10 mM KOH, 10 ml of vitamin stock (1 mM each thiamine, calcium pantothenate, *p*-aminobenzoic acid, *p*-hydroxybenzoic acid, and 2,3-dihydroxybenzoic acid), 1 ml of a 1000X stock of micronutrients (ZnCl<sub>2</sub>, MnCl<sub>2</sub>, CuCl<sub>2</sub>, CoCl<sub>2</sub>, H<sub>3</sub>BO<sub>3</sub>, (NH<sub>4</sub>)<sub>6</sub>Mo<sub>7</sub>O<sub>24</sub>, and FeCl<sub>3</sub>) giving the final concentrations shown in Table 1, 1 ml of 1 M magnesium chloride, 1 ml of 90 mM CaCl<sub>2</sub>, 10 ml of 1 M sodium formate, 10 mM sodium nitrate, and 50 mM sodium succinate, 1 ml of 3 M glycerol, 1 ml of 500 mM lactic acid, 1 ml of 700 mM glycine betaine, 700 mM choline chloride, 200 mM DL-carnitine (osmolytes), 5.61 g acetamide, 2.71 g sodium acetate, 3.3 g sodium pyruvate, 2.94 g sodium citrate, 1.34 g DL-malic acid, 60 g D-glucose,

30 g D-xylose, 5.1 g D-arabinose, 1.48 g D-fructose, 1.15 g D-galactose, and 468 mg D-mannose. After adjusting to pH 7 with 10 N NaOH, the final volume was adjusted to 1 L. This base recipe corresponds to SynH2<sup>−</sup>. To create SynH2, the aromatic inhibitors were added as solids to the base recipe in the following quantities per L SynH2 and stirred until fully dissolved before filter sterilization; 531 mg feruloyl amide, 448 mg coumaroyl amide, 173 mg *p*-coumaric acid, 69 mg ferulic acid, 69 mg hydroxymethylfurfural, 59 mg benzoic acid, 15 mg syringic acid, 14 mg cinnamic acid, 15 mg vanillic acid, 2 mg caffeic acid, 20 mg vanillin, 30 mg syringaldehyde, 24 mg 4-hydroxybenzaldehyde, 3.4 mg 4-hydroxybenzophenone. For some experiments (Figures S3, S4), feruloyl amide, coumaroyl amide, *p*-coumaric acid, ferulic acid, and hydroxymethylfurfural were added at up to twice these concentrations. The medium was filter-sterilized through a 0.2 µm filter.

### CHEMICAL ANALYSIS OF ACSH

Carbohydrates, ethanol, and short chain acids in ACSH and fermentation media were quantified using HPLC-RID, NMR, and GC-MS as previously described (Schwalbach et al., 2012). ACSH osmolality was measured using a Vapro osmometer 5520 (Wescor Inc., Logan, Utah, USA). The synthetic hydrolysate medium used in these studies (SynH2) was based on a previously described synthetic hydrolysate medium (Schwalbach et al., 2012) that was modified to more closely approximate the composition of ACSH media, particularly with regard to the presence of alternative carbon sources and protective osmolytes. Concentrations of components in the modified SynH2 are described in Table S1.

### FERMENTATIVE GROWTH CONDITIONS

Cell culture was conducted as described previously (Schwalbach et al., 2012), except fermentations were carried out in 3 L bioreactors (Applikon Biotechnology) containing 2.45 L of ACSH or SynH media, and cultures were diluted into ACSH or SynH with initial OD<sub>600</sub> at 0.2, grown anaerobically overnight, and then inoculated into bioreactors to a starting OD<sub>600</sub> of 0.2. For fermentation experiments to determine the effect of osmolytes, it was carried out in 0.5 L Sartorius BIostat Qplus bioreactors (Sartorius Stedium Biotech, Bohemia, NY) containing 0.35 L of SynH2 media in the absence or presence of osmolytes or aromatic inhibitors. Culture density was measured using a Beckman Coulter DU720 in a 1 ml cuvette. Due to the high absorbance of ACSH at 600 nm, cells were diluted 1:10 in water prior to OD<sub>600</sub> measurement, with diluted ACSH (1:10) as a blank. For SynH, diluted SynH (1:10) was used as a blank.

### RNA-seq GENE EXPRESSION ANALYSES

Samples for RNA-seq were captured and RNA extracted as described previously (Schwalbach et al., 2012). FASTQ formatted sequence files from strand-specific Illumina RNA-Seq reads were aligned to the GLBRCE1 reference genome using Bowtie version 0.12.7 (Langmead et al., 2009) with “*—nofw*” strand-specific parameter and maximal distance between the paired reads of 1000 bp. Nucleotide-level read quality information was used to weight each alignment at subsequent probabilistic expression counting step using the RNA-Seq by Expectation-Maximization



**Table 1 | Composition of ACSH, SynH1, SynH2<sup>-</sup>, and SynH2.**

Media component	ACSH	SynH1 <sup>a</sup>	SynH2 <sup>-</sup>	SynH2
<b>CARBOHYDRATES (mM)<sup>a</sup></b>				
D-Glucose	343	333	333	333
D-Xylose	205	200	200	200
L-Arabinose <sup>b</sup>	31	–	–	–
D-Arabinose <sup>b</sup>	– <sup>c</sup>	–	34	34
D-Galactose	76	–	6.4	6.4
D-Mannose	3.5	–	2.6	2.6
L-Rhamnose	0.9	–	–	–
L-Fucose	0.2	–	–	–
D-Fructose	8.2	–	8.2	8.2
<b>MISC. COMPOUNDS (mM)<sup>a</sup></b>				
Lactate	0.5	–	0.5	0.5
Pyruvate	–	–	30	30
Citrate	–	–	10	10
Nitrate	–	–	0.1	0.1
Formate	11.2	–	10	10
Malate	9.3	–	10	10
Succinate	0.8	–	0.5	0.5
Acetate	36	–	33	33
Acetamide	76	–	95	95
Glycerol	5.5	–	3	3
Glycine betaine	0.7	–	0.7	0.7
Choline	0.7	–	0.7	0.7
Carnitine	0.2	–	0.2	0.2
<b>SALTS (mM)</b>				
KH <sub>2</sub> PO <sub>4</sub>	–	22	3.4	3.4
K <sub>2</sub> HPO <sub>4</sub>	–	42	6.6	6.6
KCl	–	45	45	45
NaCl	–	25	25	25
(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	–	30	30	30
MgCl <sub>2</sub>	–	1	1	1
CaCl <sub>2</sub>	–	0.09	0.09	0.09
<b>AMINO ACIDS (μM)<sup>a</sup></b>				
Alanine	760	700	700	700
Arginine	370	400	400	400
Asparagine	163	200	200	200
Aspartate	379	350	350	350
Cysteine	n.d. <sup>d</sup>	50	50	50
Glutamine	102	100	100	100
Glutamate	459	450	450	450
Glycine	542	400	400	400
Histidine	56	80	80	80
Isoleucine	0	250	250	250
Leucine	460	360	360	360
Lysine	187	200	200	200
Methionine <sup>d</sup>	n.d.	100	100	100
Phenylalanine	189	200	200	200
Proline	214	225	225	225
Serine	272	275	275	275
Threonine <sup>d</sup>	216	225	225	225
Tryptophan	n.d.	50	50	50
Tyrosine	171	175	175	175
Valine	270	225	225	225
<b>NUCLEOBASES (μM)</b>				
Adenine	–	100	100	50
Cytosine	–	100	100	50
Uracil	–	100	100	50
Guanine	–	100	100	50
<b>TRACE COMPONENTS (μM)</b>				
Thiamine-HCl	–	10	10	10
Pantothenate	–	10	10	10
p-Aminobenzoic acid	–	10	10	10

(Continued)

**Table 1 | Continued**

Media component	ACSH	SynH1 <sup>a</sup>	SynH2 <sup>-</sup>	SynH2
p-Hydroxybenzoic acid	–	10	10	10
2,3-di-Hydroxybenzoic acid	–	10	10	10
CuCl <sub>2</sub>	–	0.010	0.010	0.010
CoCl <sub>2</sub> ·6H <sub>2</sub> O	–	0.025	0.030	0.030
H <sub>3</sub> BO <sub>4</sub>	–	0.400	10	10
(NH <sub>4</sub> ) <sub>6</sub> Mo <sub>7</sub> O <sub>2</sub> ·4H <sub>2</sub> O	–	0.003	0.003	0.003
FeCl <sub>3</sub>	–	16.6	17	17
ZnCl <sub>2</sub>	–	12	12	12
MnCl <sub>2</sub> ·4H <sub>2</sub> O	–	100	100	100
LC-derived inhibitors (mM) <sup>e,f</sup>	w/o auto-claving <sup>g</sup>	w/auto-claving		
Feruloyl amide	5.5	3.5 ± 0.6	–	2.75
Coumaroyl amide	5.5	7.1 ± 1.3	–	2.75
Hydroxymethylfurfural	1.1	<0.03	–	0.55
p-Coumaric acid	2.1	1.4 ± 0.3	–	1.05
Ferulic acid	0.71	0.091 ± 0.003	–	0.355
Benzoic acid	0.48	0.32 ± 0.01	–	0.48
Syringic acid	0.08	0.036 ± 0.004	–	0.08
Cinnamic acid	0.09	–	–	0.09
Vanillic acid	0.09	0.15 ± 0.02	–	0.09
Caffeic acid	0.01	0.006 ± 0.001	–	0.01
Vanillin	0.132	0.24 ± 0.04	–	0.132
Syringaldehyde	0.162	0.017 ± 0.002	–	0.162
4-Hydroxybenzaldehyde	0.197	0.15 ± 0.02	–	0.197
4-Hydroxyacetophenone	0.025	0.017 ± 0.002	–	0.025
Osmolality (mol/kg)	1.16 ± 0.03	0.97	1.17 ± 0.01	1.19 ± 0.01

<sup>a</sup>ACSH data are from Schwalbach et al. (2012). Sugar concentrations are averages of HPLC-MS and NMR determinations.

<sup>b</sup>In the SynH2 recipe, D-Arabinose was substituted for the L-Arabinose present in ACSH to avoid AraC-mediated repression of xylose-utilization genes (Desai and Rao, 2010). In other contexts, use of L-Arabinose in SynH2 would be appropriate.

<sup>c</sup>–, not determined in ACSH or not added in SynH.

<sup>d</sup>n.d., not detectable by methods used.

<sup>e</sup>Aromatic compounds detected at less than 20 μM in ACSH are not reported here.

<sup>f</sup>The sets of acids, amides, and aldehydes used for supplemental studies in formulating SynH2 consisted of p-Coumaric acid, Ferulic acid, Benzoic acid, Syringic acid, Cinnamic acid, Vanillic acid, and Caffeic acid (acids), Feruloyl amide and Coumaroyl amide (amides), and HMF, Vanillin, Syringaldehyde, 4-Hydroxybenzaldehyde, and 4-Hydroxyacetophenone (aldehydes) at the concentrations listed for non-autoclaved ACSH or fractions thereof as described in the Supplemental Results.

<sup>g</sup>ACSH Inhibitor concentrations for non-autoclaved CS hydrolysate are from (Tang et al., submitted). Hydrolysate preparations are described in Materials and Methods.

(RSEM) version 1.2.4 (Li and Dewey, 2011). Posterior mean estimates of counts and FPKM values were used in the downstream analysis.

The program edgeR v.3.0.2 (Robinson et al., 2010) was used to compute differential expression by using the procedures and steps described in the package documentation in all function calls with median normalization rather than the default TMM procedure. We found that median normalization better adjusted for the particular biases within the dataset. Adjusted *p*-values for multiple hypothesis corrections were used as calculated by edgeR. Pairwise

**Table 2 | Growth, sugar uptake, and ethanol production by GLBRCE1 grown in ACSH and SynH2<sup>−</sup>, and SynH2<sup>a</sup>.**

	Media		
	SynH2 <sup>−</sup>	SynH2	ACSH
Growth (Exponential) (hr <sup>−1</sup> ) <sup>b</sup>	0.13 ± 0.01	0.09 ± 0.02	0.12 ± 0.01
Glucose Rate (Exponential) <sup>b</sup>	4.7 ± 0.5	5.9 ± 1.3	5.6 ± 1.3
Glucose Rate (Transition) <sup>c</sup>	3.2 ± 0.1	2.6 ± 0.4	2.7 ± 0.1
Xylose Rate (Transition) <sup>c</sup>	0.6 ± 0.1	0.5 ± 0.1	0.2 ± 0.1
Glucose Rate (Glu-Stationary) <sup>d</sup>	N/A	1.6 ± 0.2	1.4 ± 0.2
Xylose Rate (Glu-Stationary) <sup>d</sup>	N/A	0.11 ± 0.05	0.11 ± 0.04
Xylose Rate (Xyl-Stationary) <sup>e</sup>	0.19 ± 0.03	0.01 ± 0.01	0.04 ± 0.03
Total Glucose Consumed (mM)	330 ± 20	310 ± 20	300 ± 20
Total Xylose Consumed (mM)	65 ± 30	25 ± 1	25 ± 10
Total Ethanol produced (mM)	540 ± 30	460 ± 60	470 ± 60
Ethanol Yield (%) <sup>f</sup>	70 ± 3	70 ± 6	73 ± 3

<sup>a</sup> Each value is from at least three biological replicates in different bioreactors.

<sup>b</sup> Exponential phase is between 4 and 12 h in all media. Unit for glucose uptake rate is mM·OD<sub>600</sub><sup>−1</sup>·h<sup>−1</sup>.

<sup>c</sup> Transition phase is between 12 and 30 h for SynH2<sup>−</sup>, and between 12 and 23 h for SynH2 and ACSH. Units for glucose and xylose uptake rate are mM·OD<sub>600</sub><sup>−1</sup>·h<sup>−1</sup>.

<sup>d</sup> Stationary phase when glucose is present (Glu-Stationary) is between 23 and 100 h for SynH2 and ACSH. However, there was no Glu-stationary phase for SynH2<sup>−</sup> because it remained in transition phase until the glucose was gone.

<sup>e</sup> Stationary phase when glucose is gone (Xyl-Stationary) is between 47 and 78 h for SynH2<sup>−</sup>. The Xyl-Stationary rates for SynH2 and ACSH were measured in follow-up experiments carried out long enough to exhaust glucose in stationary phase.

<sup>f</sup> Calculated from the total ethanol produced and the total glucose and xylose consumed, assuming 2 ethanol per glucose and 1.67 ethanol per xylose.

fold-changes and adjusted *p*-values are calculated between media types and within each phase and between phases within each media type.

To catalog the most significant effects, we examined the ratios using several different strategies. In addition to identifying the largest changes in expression of individual genes in SynH2 and ACSH relative to SynH2<sup>−</sup> (Table S2), we also used gene set enrichment analyses as described by Subramanian et al. (2005) and Varemo et al. (2013). We compiled gene sets for these analyses from pathways, transporters, and regulons documented in Ecocyc (Keseler et al., 2013) and KEGG.

## PROTEOMIC MEASUREMENTS

Thirty-four *Escherichia coli* samples were processed for analysis by mass spectrometry at PNNL. Each sample was typically digested using a global urea digestion (Pasa-Tolic et al., 2004; Smyth, 2004) prior to isobaric labeling with an iTRAQ 4-plex labeling kit, following the manufacturer's directions (ABSciex, Redwood City, CA) (Ross et al., 2004; Bantscheff et al., 2008). Prior to high pH reverse phase fractionation with concatenated pooling (Wang et al., 2011b), the samples were desalted using C18 solid-phase extraction (SPE) (SUPELCO, Bellefonte, PA). All samples were processed with a custom LC system using reversed-phase C18 columns (unpublished variation of Maiolica et al., 2005) and the

samples were then analyzed with a Velos Orbitrap mass spectrometer (Thermo Scientific, San Jose, CA) that was equipped with an electrospray ionization (ESI) interface (Kelly et al., 2006).

Raw files were searched against a concatenated *Escherichia coli* K-12 database and contaminant database using MS-GF+ (v9018) with oxidation as a dynamic modification on methionine and 4-plex iTRAQ label as a static modification (Kim et al., 2008). The parent ion mass tolerance was set to 50 ppm. The resulting sequence identifications were filtered down to a 1% false discovery rate using target-decoy approach and MS-GF derived *q*-values. Reporter ion intensities were quantified using the tool MASIC (Monroe et al., 2008). Results were then processed with the MAC (Multiple Analysis Chain) pipeline, an internal tool which aggregates and filters data. Missing reporter ion channel results were retained. Degenerate peptides, i.e., peptides occurring in more than one protein, were filtered out. Proteins with one peptide detected were removed if they were not repeatable across at least two replicates. Redundant peptide identification reporter ions were summed across fractions and median central tendency normalization was applied to account for channel bias. Each 4-plex sample group was normalized using a pooled sample for comparison between groups. The final protein values were obtained by averaging their associated peptide intensity values and varied from ~5000 to 350000. Finally, the protein values were then log<sub>2</sub> transformed.

All proteins that had missing values in their replicates were removed and the pair-wise protein expression level changes and significance *p*-values between the SynH2 and SynH2<sup>−</sup> cells at each growth phase were estimated using limma (Smyth, 2004; Smith, 2005), which fits a linear model across the replicates to calculate the fold changes, smooths the standard errors for significance and adjusts the *p*-values via the Benjamini-Hochberg method.

## COMPARISON OF PROTEOMIC DATA TO TRANSCRIPTOMIC DATA

Pair-wise RNA expression level changes and significance *p*-values were estimated using the edgeR package as previously discussed. The log<sub>2</sub>-fold-changes for the Protein and RNA were z-score scaled separately to correct for the difference in dynamic ranges between the protein and RNA measurements.

Significant discrepant Protein/RNA ratios between SynH2 and SynH2<sup>−</sup> cells were estimated using a two-sample z-test and the corresponding *p*-values are adjusted for multiple comparisons using the Benjamini-Hochberg method. All Protein/RNA ratios that are either significant in the RNA or protein ratio (*p* < 0.05) and that significantly disagree (*p* < 0.05) are tabulated in Table S7.

## MEASUREMENT OF INTERNAL METABOLITE ABUNDANCES PREPARATION OF INTRACELLULAR EXTRACTS

Two ml of cell culture was rapidly removed from bioreactors with a 10 ml sterile syringe and cells captured on Whatman 0.45 μm nylon syringe filters (GE Healthcare Bio-Sciences, Pittsburgh, Pennsylvania, USA) as described previously (Schwalbach et al., 2012). To reduce the background associated with metabolites present in ACSH and SynH the cells on the filter were then rapidly washed with 5 ml of M9 medium (Neidhardt et al., 1974) lacking a

carbon source. Acetonitrile-methanol-water (40:40:20; 2 ml) containing 0.1% formic acid was then applied to the filters, and the eluate captured in a 15 ml conical tube. The eluate was passed through the cells a second time to ensure complete cell lysis and then flash frozen in a dry ice/ethanol bath.

### DETECTION/QUANTIFICATION OF METABOLITES

The concentration of internal glycolytic and TCA cycle intermediates were determined using high performance anion exchange chromatography electrospray ionization tandem mass spectrometry (HPAEC-ESI-MS/MS). Reagents and non-labeled reference compounds were from Sigma Aldrich Co.

HPAEC was adapted from a previously reported method (Buescher et al., 2010), and was used for determination of pyruvate, citrate,  $\alpha$ -ketoglutarate, glucose-6-phosphate, fructose-6-phosphate, fructose-1,6-bis phosphate, phospho(enol)pyruvate, and ATP. Chromatography was carried out on an Agilent 1200 series HPLC comprised of a vacuum degasser, binary pump, and a heated column compartment, and a thermostated autosampler set to maintain 6°C. Mobile Phase A was 0.5 mM NaOH and mobile phase B was 100 mM NaOH. Compounds were separated by a gradient elution of 0.35 mL per minute starting at 10% B, increased to 15% B over 5 min and held at 15% B for 10 min, then increased to 100% B over 12 min and held for 10 min before returning to 10% B to be re-equilibrated for 5 min prior to the next injection. The column temperature was 40°C. The injection volume was 20  $\mu$ L of intracellular extract or calibrant standard mixture.

### MEASUREMENT OF AROMATIC INHIBITORS IN ACSH AND SynH

Samples of ACSH and SynH cultures were prepared by centrifugation as described previously (Schwalbach et al., 2012), and then were subjected to reverse phase HPLC high resolution/accurate mass spectrometry (RP-HPLC-HRAM MS) and headspace solid-phase microextraction gas chromatography-isotope dilution mass spectrometry (HS-SPME/IDMS) analysis.

The majority of phenolic compounds were determined by RP-HPLC-HRAM MS, which was carried out with a MicroAS autosampler (Thermo Scientific) equipped with a chilled sample tray and a Surveyor HPLC pump (Thermo Scientific) coupled to a Q-Exactive hybrid quadrupole/orbitrap mass spectrometer by electrospray ionization. The analytical column was an Ascentis Express column (150  $\times$  2.1 mm  $\times$  2.7  $\mu$ m core-shell particles, Supelco, Bellefonte, PA) protected by a 5 mm C18 precolumn (Phenomenex, Torrance, CA). Mobile phase A was 10 mM formic acid adjusted to pH 3 with ammonium hydroxide and mobile phase B was methanol with 10 mM formic acid and the same volume of ammonium hydroxide as was added to mobile phase A. Compounds were separated by gradient elution. The initial composition was 95% A, which was held for 2 min after injection, then decreased to 40% A over the next 8 min, changed immediately to 5% A and held for 5 min, then changed back to 95% A for a column re-equilibration period of 7 min prior to the next injection. The flow rate was 0.3 mL/min.

The HPLC separation was coupled to the mass spectrometer via a heated electrospray (HESI) source (HESI II Probe, Thermo

Scientific). The operating parameters of the source were: spray voltages: +3000, −2500; capillary temperature: 300°C; sheath gas flow: 20 units; auxiliary gas flow: 5 units; HESI probe heater: 300°C. Spectra were acquired with fast polarity switching to obtain positive and negative mode ionization chromatograms in a single analysis. In each mode, a full MS<sup>1</sup> scan was performed by the Orbitrap analyzer followed by a data dependent MS<sup>2</sup> scan of the most abundant ion in the MS<sup>1</sup> scan. The Q-Exactive parameters (both positive and negative modes) were: MS<sup>1</sup> range 85–500 Th, resolution: 17,500 (FWHM at 400 *m/z*), AGC target: 1e6, maximum ion accumulation time 100ms, S-lens level: 50. Settings for data dependent MS<sup>2</sup> scans were: isolation width: 1.8 Th, normalized collision energy: 50 units, resolution: 17,500, AGC target: 2e5, maximum ion accumulation time: 50 ms, under-fill ratio: 1%, apex trigger: 5–12 s, isotope exclusion enabled, dynamic exclusion: 10 s.

HS-SPME/IDMS was used to quantify acetaldehyde, acetamide, furfural, furfuryl alcohol, HMF, 5-(hydroxymethyl)furfural (HMF), and Bis(hydroxymethyl) furan (“HMF alcohol”/BHMF). Samples were thawed and briefly vortex mixed prior to measuring 500 microliters of sample, 500 microliters of stable isotope labeled internal standard mixture, and ~300 mg of NaCl into a 20 mL screw top headspace and quickly capped with magnetic screwtop cap with 4 mm PTFE backed silicone rubber septum for SPME. Automated SPME sample processing and analysis was carried out using a Pegasus 4D GCxGC-TOF MS (Leco Corp. Saint Joseph, Michigan) with an Agilent 6890A gas chromatograph coupled to the ToF mass analyzer via a heated capillary transfer line, and a Gerster-LEAP combi PAL autosampler and sample preparation system with Twister heated sample agitator fitted with an automated SPME holder containing a gray hub 50/30, 23 ga. Stabiliflex DVB/Carboxen/PDMS SPME fiber (Supelco, Inc.). Chromatof software (Leco, Corp.) V. 4.50.8.0 was used for system control during acquisition and for data processing, calibration and calculation of final concentrations. Sample incubation temperature 95°C, agitation speed 100 rpm, during extraction time, 100 rpm, agitation on 4 s/off 15 s, sample extraction time (SPME fiber exposed to the sample headspace in heated agitator) 20 min, desorb time (SPME fiber inserted in hot GC inlet) 60 min. GC cycle time 40 min. Critical injector positions were determined empirically through trial, error, and careful measurement: vial penetration 11 mm, Injector penetration 54 mm, Injector penetration—needle 40 mm. GC was carried out using a StabilWAX-DA column (Restek Corp, Bellefonte, Pennsylvania, USA) 0.25 mm ID  $\times$  30 m, *df* = 0.25  $\mu$ m; carrier gas He, 1 mL/min; split 5:1; purge flow 3 mL/min; inlet temp 250°C; inlet liner type straight split/splitless deactivated glass 0.75 mm ID; equilibration time 1 min; Oven temperature program: initial temperature 30°C, hold 2 min. Increase to 10°C/min to 250°C, hold 10 min; MS transfer line 250°C. ToF mass spectrometer (unit mass resolution) Acquisition delay 85 s; start mass 10 end mass 500; acquisition 10 spectra/s; electron multiplier delta V 1475 (dependent on QC procedure) source temperature 200°C.

Quantification of organic acids in ACSH was carried out by HPAEC-MS/MS in a similar manner to that described for intracellular metabolites.

**Table 2 | Growth, sugar uptake, and ethanol production by GLBRCE1 grown in ACSH and SynH2<sup>−</sup>, and SynH2<sup>a</sup>.**

	Media		
	SynH2 <sup>−</sup>	SynH2	ACSH
Growth (Exponential) (hr <sup>−1</sup> ) <sup>b</sup>	0.13 ± 0.01	0.09 ± 0.02	0.12 ± 0.01
Glucose Rate (Exponential) <sup>b</sup>	4.7 ± 0.5	5.9 ± 1.3	5.6 ± 1.3
Glucose Rate (Transition) <sup>c</sup>	3.2 ± 0.1	2.6 ± 0.4	2.7 ± 0.1
Xylose Rate (Transition) <sup>c</sup>	0.6 ± 0.1	0.5 ± 0.1	0.2 ± 0.1
Glucose Rate (Glu-Stationary) <sup>d</sup>	N/A	1.6 ± 0.2	1.4 ± 0.2
Xylose Rate (Glu-Stationary) <sup>d</sup>	N/A	0.11 ± 0.05	0.11 ± 0.04
Xylose Rate (Xyl-Stationary) <sup>e</sup>	0.19 ± 0.03	0.01 ± 0.01	0.04 ± 0.03
Total Glucose Consumed (mM)	330 ± 20	310 ± 20	300 ± 20
Total Xylose Consumed (mM)	65 ± 30	25 ± 1	25 ± 10
Total Ethanol produced (mM)	540 ± 30	460 ± 60	470 ± 60
Ethanol Yield (%) <sup>f</sup>	70 ± 3	70 ± 6	73 ± 3

<sup>a</sup> Each value is from at least three biological replicates in different bioreactors.

<sup>b</sup> Exponential phase is between 4 and 12 h in all media. Unit for glucose uptake rate is mM·OD<sub>600</sub><sup>−1</sup>·h<sup>−1</sup>.

<sup>c</sup> Transition phase is between 12 and 30 h for SynH2<sup>−</sup>, and between 12 and 23 h for SynH2 and ACSH. Units for glucose and xylose uptake rate are mM·OD<sub>600</sub><sup>−1</sup>·h<sup>−1</sup>.

<sup>d</sup> Stationary phase when glucose is present (Glu-Stationary) is between 23 and 100 h for SynH2 and ACSH. However, there was no Glu-stationary phase for SynH2<sup>−</sup> because it remained in transition phase until the glucose was gone.

<sup>e</sup> Stationary phase when glucose is gone (Xyl-Stationary) is between 47 and 78 h for SynH2<sup>−</sup>. The Xyl-Stationary rates for SynH2 and ACSH were measured in follow-up experiments carried out long enough to exhaust glucose in stationary phase.

<sup>f</sup> Calculated from the total ethanol produced and the total glucose and xylose consumed, assuming 2 ethanol per glucose and 1.67 ethanol per xylose.

fold-changes and adjusted *p*-values are calculated between media types and within each phase and between phases within each media type.

To catalog the most significant effects, we examined the ratios using several different strategies. In addition to identifying the largest changes in expression of individual genes in SynH2 and ACSH relative to SynH2<sup>−</sup> (Table S2), we also used gene set enrichment analyses as described by Subramanian et al. (2005) and Varemo et al. (2013). We compiled gene sets for these analyses from pathways, transporters, and regulons documented in Ecocyc (Keseler et al., 2013) and KEGG.

## PROTEOMIC MEASUREMENTS

Thirty-four *Escherichia coli* samples were processed for analysis by mass spectrometry at PNNL. Each sample was typically digested using a global urea digestion (Pasa-Tolic et al., 2004; Smyth, 2004) prior to isobaric labeling with an iTRAQ 4-plex labeling kit, following the manufacturer's directions (ABSciex, Redwood City, CA) (Ross et al., 2004; Bantscheff et al., 2008). Prior to high pH reverse phase fractionation with concatenated pooling (Wang et al., 2011b), the samples were desalted using C18 solid-phase extraction (SPE) (SUPELCO, Bellefonte, PA). All samples were processed with a custom LC system using reversed-phase C18 columns (unpublished variation of Maiolica et al., 2005) and the

samples were then analyzed with a Velos Orbitrap mass spectrometer (Thermo Scientific, San Jose, CA) that was equipped with an electrospray ionization (ESI) interface (Kelly et al., 2006).

Raw files were searched against a concatenated *Escherichia coli* K-12 database and contaminant database using MS-GF+ (v9018) with oxidation as a dynamic modification on methionine and 4-plex iTRAQ label as a static modification (Kim et al., 2008). The parent ion mass tolerance was set to 50 ppm. The resulting sequence identifications were filtered down to a 1% false discovery rate using target-decoy approach and MS-GF derived *q*-values. Reporter ion intensities were quantified using the tool MASIC (Monroe et al., 2008). Results were then processed with the MAC (Multiple Analysis Chain) pipeline, an internal tool which aggregates and filters data. Missing reporter ion channel results were retained. Degenerate peptides, i.e., peptides occurring in more than one protein, were filtered out. Proteins with one peptide detected were removed if they were not repeatable across at least two replicates. Redundant peptide identification reporter ions were summed across fractions and median central tendency normalization was applied to account for channel bias. Each 4-plex sample group was normalized using a pooled sample for comparison between groups. The final protein values were obtained by averaging their associated peptide intensity values and varied from ~5000 to 350000. Finally, the protein values were then log<sub>2</sub> transformed.

All proteins that had missing values in their replicates were removed and the pair-wise protein expression level changes and significance *p*-values between the SynH2 and SynH2<sup>−</sup> cells at each growth phase were estimated using limma (Smyth, 2004; Smith, 2005), which fits a linear model across the replicates to calculate the fold changes, smooths the standard errors for significance and adjusts the *p*-values via the Benjamini-Hochberg method.

## COMPARISON OF PROTEOMIC DATA TO TRANSCRIPTOMIC DATA

Pair-wise RNA expression level changes and significance *p*-values were estimated using the edgeR package as previously discussed. The log<sub>2</sub>-fold-changes for the Protein and RNA were z-score scaled separately to correct for the difference in dynamic ranges between the protein and RNA measurements.

Significant discrepant Protein/RNA ratios between SynH2 and SynH2<sup>−</sup> cells were estimated using a two-sample z-test and the corresponding *p*-values are adjusted for multiple comparisons using the Benjamini-Hochberg method. All Protein/RNA ratios that are either significant in the RNA or protein ratio (*p* < 0.05) and that significantly disagree (*p* < 0.05) are tabulated in Table S7.

## MEASUREMENT OF INTERNAL METABOLITE ABUNDANCES PREPARATION OF INTRACELLULAR EXTRACTS

Two ml of cell culture was rapidly removed from bioreactors with a 10 ml sterile syringe and cells captured on Whatman 0.45 μm nylon syringe filters (GE Healthcare Bio-Sciences, Pittsburgh, Pennsylvania, USA) as described previously (Schwalbach et al., 2012). To reduce the background associated with metabolites present in ACSH and SynH the cells on the filter were then rapidly washed with 5 ml of M9 medium (Neidhardt et al., 1974) lacking a



cells to cease growth before glucose was consumed, decreased the rate of ethanol production, and to lesser extent decreased final amounts of ethanol produced.

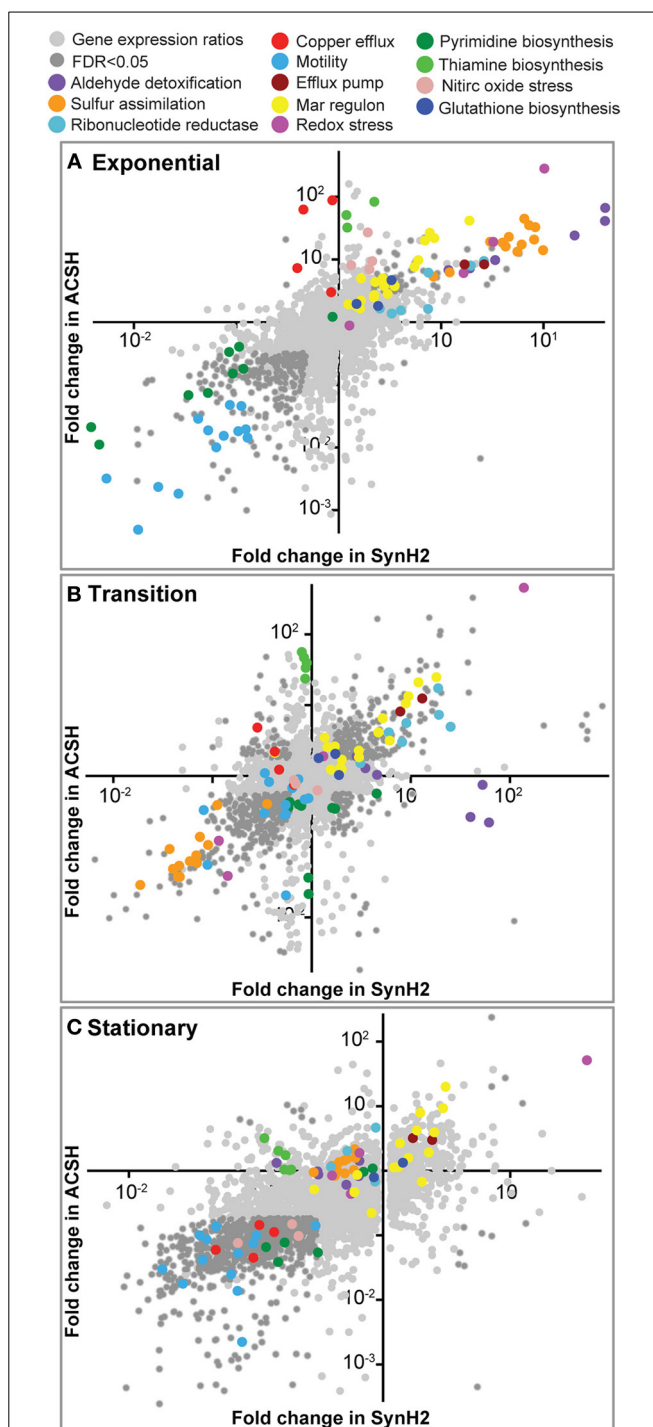
### GLBRCE1 GENE EXPRESSION PATTERNS ARE SIMILAR IN SynH2 AND ACSH

To test the similarity of SynH2 to ACSH and the extent to which LC-derived inhibitors impact ethanologenesis, we next used RNA-seq to compare gene expression patterns of GLBRCE1 grown in the two media relative to cells grown in SynH2<sup>-</sup> (Materials and Methods; Table 1). We computed normalized gene expression ratios of ACSH cells vs. SynH2<sup>-</sup> cells and SynH2 cells vs. SynH2<sup>-</sup> cells, and then plotted these ratios against each other using log<sub>10</sub> scales for exponential phase (Figure 2A), transition phase (Figure 2B), and stationary phase (Figure 2C). For simplicity, we refer to these comparisons as the SynH2 and ACSH ratios. The SynH2 and ACSH ratios were highly correlated in all three phases of growth, although were lower in transition and stationary phases (Pearson's *r* of 0.84, 0.66, and 0.44 in exponential, transition, and stationary, respectively, for genes whose SynH2 and ACSH expression ratios both had corrected *p* < 0.05; *n* = 390, 832, and 1030, respectively). Thus, SynH2 is a reasonable mimic of ACSH.

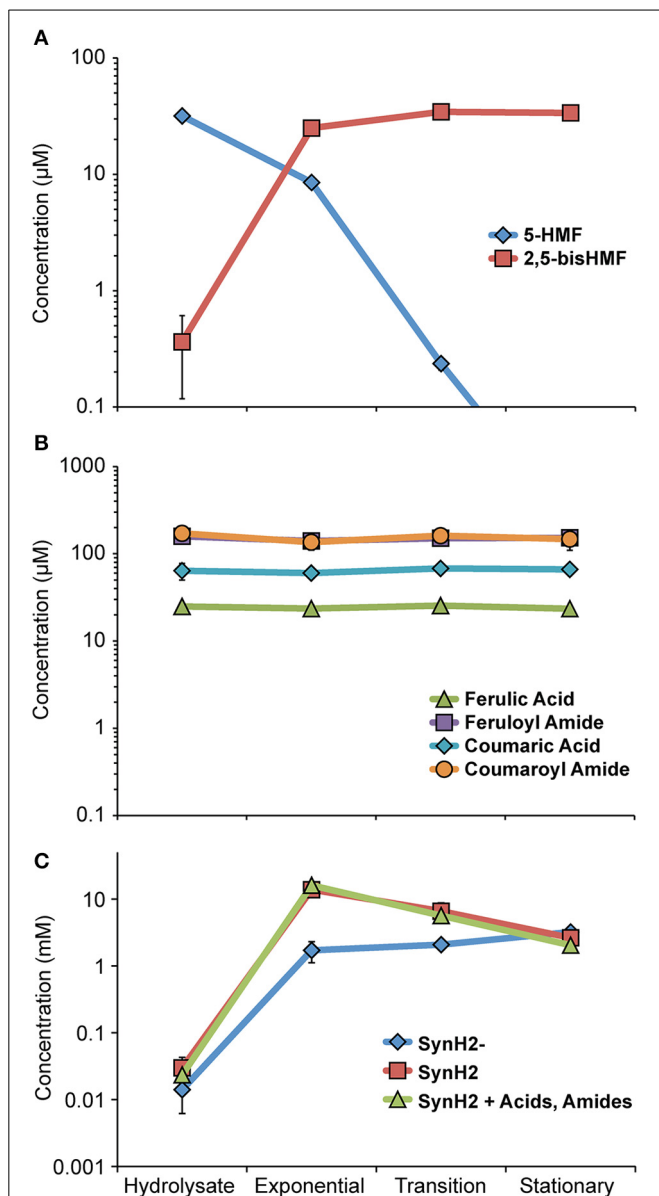
We used these data to investigate the gene expression differences between SynH2 and ACSH (Table S3). Several differences likely reflected the absence of some trace carbon sources in SynH2 (e.g., sorbitol, mannitol), their presence in SynH2 at higher concentrations than found in ACSH (e.g., citrate and malate), and the intentional substitution of D-arabinose for L-arabinose. Elevated expression of genes for biosynthesis or transport of some amino acids and cofactors confirmed or suggested that SynH2 contained somewhat higher levels of Trp, Asn, thiamine and possibly lower levels of biotin and Cu<sup>2+</sup> (Table S3). Although these discrepancies point to minor or intentional differences that can be used to refine the SynH recipe further, overall we conclude that SynH2 can be used to investigate physiology, regulation, and biofuel synthesis in microbes in a chemically defined, and thus reproducible, media to accurately predict behaviors of cells in real hydrolysates like ACSH that are derived from ammonia-pretreated biomass.

### AROMATIC ALDEHYDES IN SynH2 ARE CONVERTED TO ALCOHOLS, BUT PHENOLIC CARBOXYLATES AND AMIDES ARE NOT METABOLIZED

Before evaluating how patterns of gene expression informed the physiology of GLBRCE1 in SynH2, we first determined the profiles of inhibitors, end-products, and intracellular metabolites during ethanologenesis. The most abundant aldehyde inhibitor, HMF, quickly disappeared below the limit of detection as the cells entered transition phase with concomitant and approximately stoichiometric appearance of the product of HMF reduction, 2,5-bis-HMF (hydroxymethylfurfuryl alcohol; Figure 3A, Table S8). Hydroxymethylfuroic acid did not appear during the fermentation, suggesting that HMF is principally reduced by aldehyde reductases such as YqhD and DkgA, as previously reported for HMF and furfural generated from acid-pretreated biomass (Miller et al., 2009a, 2010; Wang et al., 2013). In contrast, the concentrations of ferulic acid, coumaric acid, feruloyl amide, and coumaroyl amide did not change appreciably over the course



**FIGURE 2 | Relative gene expression patterns in SynH2 and ACSH cells relative to SynH2<sup>-</sup> cells.** Scatter plots were prepared with the ACSH/SynH2<sup>-</sup> gene expression ratios plotted on the y-axis and the SynH2/SynH2<sup>-</sup> ratios on the x-axis (both on a log<sub>10</sub> scale). GLBRCE1 was cultured in a bioreactor anaerobically (Figure 1 and Figure S5); RNAs were prepared from exponential (A), transition (B), or stationary (C) phase cells and subjected to RNA-seq analysis (Materials and Methods). Dark gray dots represent genes for which *p* = 0.05 for each expression ratio. Sets of genes with related functions that exhibited significant discrepant or parallel changes are color-coded and described in the legend at the top (see also Tables S3, S4, respectively).



**FIGURE 3 | Growth phase-dependent changes in SynH2 aromatic inhibitor levels.** GLBRCE1 was cultured under anaerobic conditions in SynH2 in bioreactors. Levels of the major LC-derived inhibitors in the culture medium were determined as described in Materials and Methods. “Hydrolysate” refers to medium immediately prior to inoculation, “Exp,” “Trans,” and “Stat” refers to samples collected during exponential, transition, and stationary phase growth, respectively. **(A)** Metabolic fate of hydroxymethylfurfural (HMF). Concentrations of HMF and 2,5-bis-HMF (2,5-bis-hydroxymethylfurfuryl alcohol) are represented. **(B)** Metabolic fates of the major aromatic acids and amides. Concentrations of ferulic acid, feruloyl amide, coumaric acid, and coumaroyl amide are shown. **(C)** Concentration of acetaldehyde in the culture medium when GLBRCE1 was grown in SynH2, SynH2<sup>-</sup>, or SynH2 with aromatic aldehydes only omitted.

of the experiment (Figure 3B, Table S8), suggesting that *E. coli* either does not encode activities for detoxification of phenolic carboxylates and amides, or that expression of such activities is not induced in SynH2.

Although HMF disappeared early in fermentation, acetaldehyde accumulated to >10 mM during exponential and transition phase in both SynH2 and ACSH (Figure 3C, Table S8). Elevated acetaldehyde relative to SynH2<sup>-</sup> was also observed upon omission of aromatic aldehydes from SynH2, demonstrating that LC-derived phenolic acids and amides alone can cause accumulation of acetaldehyde (Figure 3C). Thus, acetaldehyde accumulation was not simply a consequence of diverting reducing equivalents to detoxification of the aromatic aldehydes like HMF but likely resulted from a broader impact of LC-derived inhibitors on cellular energetics that decreased the pools of NADH available for conversion of acetaldehyde to ethanol.

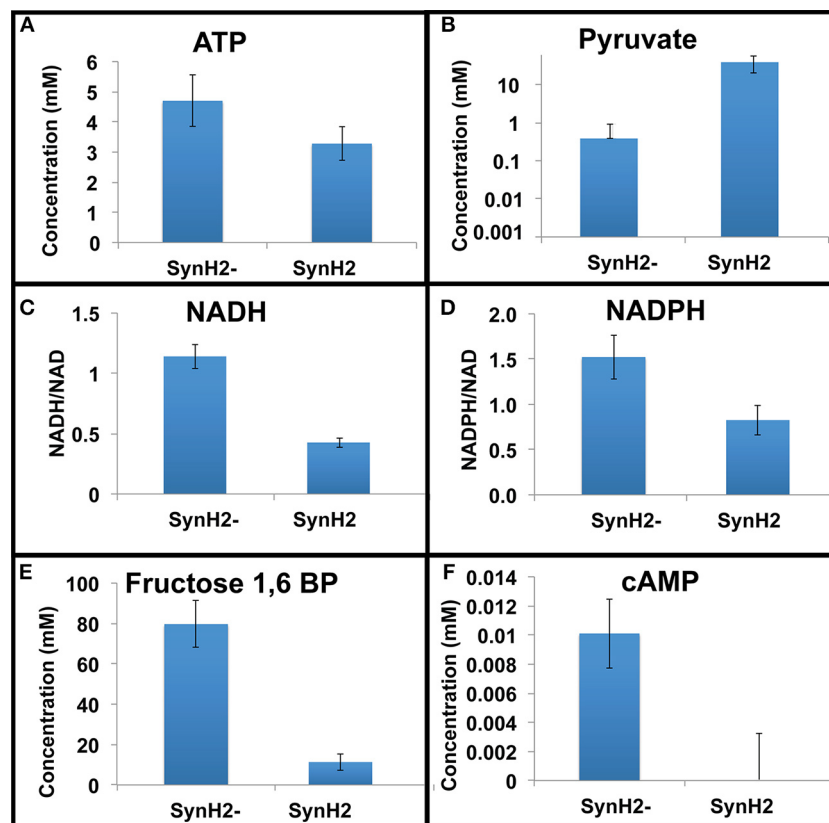
#### LIGNOCELLULOSE-DERIVED INHIBITORS NEGATIVELY IMPACT CARBON AND ENERGY METABOLISM, RESULTING IN ACCUMULATION OF PYRUVATE AND ACETALDEHYDE

Examination of intracellular metabolites revealed that aromatic inhibitors decreased the levels of metabolites associated with glycolysis and the TCA cycle (Figures 4B,E; Table S1). Strikingly, metabolites associated with cellular energetics and redox state were also decreased in SynH2 cells relative to SynH2<sup>-</sup> cells (Figures 4A,C,D,F; Table S1). ATP was reduced 30%; the NADH/NAD<sup>+</sup> ratio decreased by 63%; and the NADPH/NADP<sup>+</sup> ratio decreased 56%. Together, these data indicate that the aromatic inhibitors dramatically decreased cellular energy pools and available reducing equivalents in SynH2 cells. The consequences of energetic depletion were readily apparent with an approximate 100-fold increase in the intracellular levels of pyruvate in SynH2 cells (to ~14 mM), despite the disappearance of pyruvate from the growth medium (Table S1, Figure 4B, and data not shown). The increase in pyruvate and correspondingly in acetaldehyde (Figures 3C, 4B) suggest that the reduced rate of glucose-to-ethanol conversion caused by aromatic inhibitors results from inadequate supplies of NADH to convert acetaldehyde to ethanol.

Transition-phase SynH2 vs. SynH2<sup>-</sup> cells exhibited similar trends in aromatic-inhibitor-dependent depletion of some glycolytic intermediates, some TCA intermediates, and ATP, along with elevation of pyruvate and acetaldehyde (Table S1; Figure 3C). Stationary phase cells displayed several differences, however. Glycolytic intermediates (glucose 6-phosphate, fructose 6-phosphate, fructose 1,6 diphosphate, and 2-, 3-phosphoglycerate) were approximately equivalent in SynH2 and SynH2<sup>-</sup> cells, whereas pyruvate concentrations dropped significantly (Table S1). The impact of the inhibitors was largely attributable to the phenolic carboxylate and amides alone, as removal of the aldehydes from SynH2 changed neither the depletion of glycolytic and TCA intermediates nor the elevation of pyruvate and acetaldehyde (data not shown). We conclude that phenolic carboxylates and amides in SynH2 and ACSH have major negative impacts on the rate at which cells grow and consequently can convert glucose to ethanol.

#### AROMATIC INHIBITORS INDUCE GENE EXPRESSION CHANGES REFLECTING ENERGY STRESS

Given the major impacts of aromatic inhibitors on ethanologenesis, we next sought to address how these inhibitors impacted gene expression and regulation in *E. coli* growing in SynH2.



**FIGURE 4 | Relative metabolite levels in SynH2 and SynH2<sup>-</sup> cells.**

GLBRCE1 was cultured anaerobically in bioreactors in SynH2 and SynH2<sup>-</sup>. Metabolites were prepared from exponential phase cells and analyzed as

described in the Material and Methods. Shown are intracellular concentrations of ATP (A), pyruvate (B), fructose-1,6-bisphosphate (E), and cAMP (F). (C,D) show the ratios of NADH/NAD<sup>+</sup> and NADPH/NADP<sup>+</sup>, respectively.

To that end, we first identified pathways, transporters, and regulons with similar relative expression patterns in SynH2 and ACSH using both conventional gene set enrichment analysis and custom comparisons of aggregated gene expression ratios (Materials and Methods). These comparisons yielded a curated set of regulons, pathways, and transporters whose expression changed significantly in SynH2 or ACSH relative to SynH2<sup>-</sup> (aggregate  $p < 0.05$ ; Table S4).

For many key pathways, transporters, and regulons, similar trends were seen in both SynH2 and ACSH vs. SynH2<sup>-</sup> (Figure 2 and Table S4). The most upregulated gene sets reflected key impacts of aromatic inhibitors on cellular energetics. Anabolic processes requiring a high NADPH/NADP<sup>+</sup> potential were significantly upregulated (e.g., sulfur assimilation and cysteine biosynthesis, glutathione biosynthesis, and ribonucleotide reduction). Additionally, genes encoding efflux of drugs and aromatic carboxylates (e.g., *aaeA*) and regulons encoding efflux functions (e.g., the *rob* regulon), were elevated. Curiously, both transport and metabolism of xylose were downregulated in all three growth phases in both media, suggesting that even prior to glucose depletion aromatic inhibitors reduce expression of xylose genes and thus the potential for xylose conversion. Currently the mechanism of this repression is unclear, but it presumably reflects either an indirect impact of altered energy metabolism or an interaction

of one or more of the aromatic inhibitors with a regulator that decreases xylose gene expression.

During transition phase, a different set of genes involved in nitrogen assimilation were upregulated in SynH2 cells and ACSH cells relative to SynH2<sup>-</sup> cells (Table S5). Previously, we found that transition phase corresponded to depletion of amino acid nitrogen sources (e.g., Glu and Gln; Schwalbach et al., 2012). Thus, this pattern of aromatic-inhibitor-induced increase in the expression of nitrogen assimilation genes during transition phase suggests that the reduced energy supply caused by the inhibitors increased difficulty of ATP-dependent assimilation of ammonia. Interestingly, the impact on gene expression appeared to occur earlier in ACSH than in SynH2, which may suggest that availability of organic nitrogen is even more growth limiting in ACSH.

Of particular interest were the patterns of changes in gene expression related to the detoxification pathways for the aromatic inhibitors. Our gene expression analysis revealed inhibitor induction of genes encoding aldehyde detoxification pathways (*frmA*, *frmB*, *dkgA*, and *yqhD*) that presumably target LC-derived aromatic aldehydes (e.g., HMF and vanillin) and acetaldehyde that accumulates when NADH-dependent reduction to ethanol becomes inefficient (Herring and Blattner, 2004; Gonzalez et al., 2006; Miller et al., 2009b, 2010; Wang et al., 2013) as well as efflux

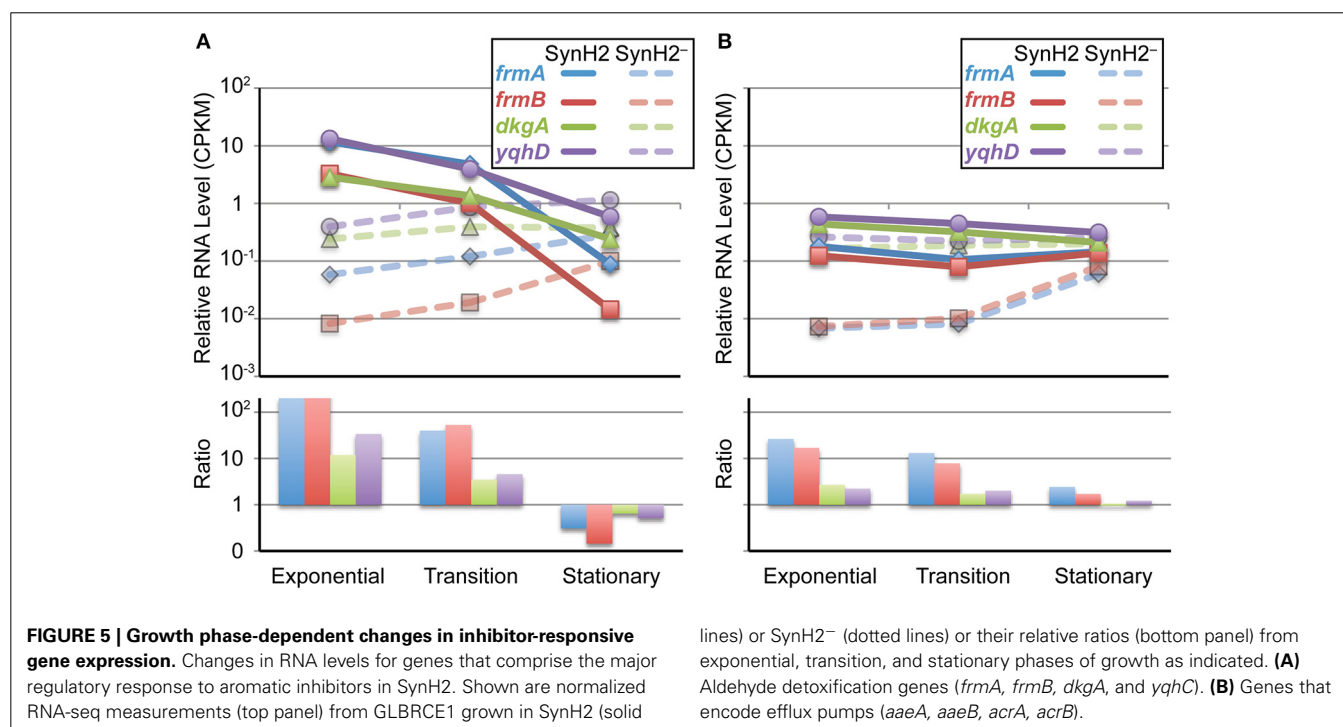
pumps controlled by MarA/SoxS/Rob (e.g., *acrA* and *acrB*) and the separate system for aromatic carboxylates (*aaeA* and *aaeB*) (Van Dyk et al., 2004). Interestingly, we observed that expression of the aldehyde detoxification genes *frmA*, *frmB*, *dkgA*, and *yqhD* paralleled the levels of LC-derived aromatic aldehydes and acetaldehyde detected in the media (Figure 3). Initially high-level expression was observed in SynH2 cells, which decreased as the aldehydes were inactivated (Figure 5A). Conversely, expression of these genes increased in SynH2<sup>-</sup> cells, surpassing the levels in SynH2 cells in stationary phase when the level of acetaldehyde in the SynH2<sup>-</sup> culture spiked past that in the SynH2 culture. The elevation of *frmA* and *frmB* is particularly noteworthy as the only reported substrate for FrmAB is formaldehyde. We speculate that this system, which has not been extensively studied in *E. coli*, may also act on acetaldehyde. Alternatively, formaldehyde, which we did not assay, may have accumulated in parallel to acetaldehyde.

In contrast to the decrease in *frmA*, *frmB*, *dkgA*, and *yqhD* expression as SynH2 cells entered stationary phase, expression of *aaeA*, *aaeB*, *acrA*, and *acrB* remained high (Figure 5B). This continued high-level expression is consistent with the persistence of phenolic carboxylates and amides in the SynH2 culture (Figure 3), and presumably reflect the futile cycle of antiporter excretion of these inhibitors to compete with constant leakage back into cells.

#### POST-TRANSCRIPTIONAL EFFECTS OF AROMATIC INHIBITORS WERE LIMITED PRIMARILY TO STATIONARY PHASE

We next investigated the extent to which the aromatic inhibitors could exert effects on cellular regulation post-transcriptionally rather than *via* transcriptional regulators by comparing inhibitor-induced changes in protein levels to changes in RNA levels. For this purpose, we used iTRAQ quantitative proteomics to assess

changes in protein levels (Material and Methods). We then normalized the log<sub>2</sub>-fold-changes in protein levels in each of the three growth phases to changes in RNA levels determined by RNA-seq and plotted the normalized values against each other (Figures 6A–C; Tables S6, S7). Most proteome and transcriptome fold-changes fall within a factor of 2 of the diagonal, consistent with concordant changes in mRNA and protein and thus limited post-transcriptional effects of aromatic inhibitors. A small number of RNA-protein pairs exhibited an >2-fold change with  $p < 0.05$ . During exponential phase, four proteins were present at elevated levels relative to changes in RNA levels, which actually decreased (RpoS, TnaA, MalE, and GlnH; red circles, Figure 6A; Table S7A), whereas 26 RNAs increased or decreased significantly with little difference in proteins levels (blue circles, Figure 6A; Table S7A). These disparate increases in RNA levels included some of the major transcriptional responses to the inhibitors (S assimilation and the FrmA aldehyde detoxification pathway), and these proteins were present at high levels both with and without inhibitors (Table S7D). Several observations led us to conclude that these discrepancies in protein and RNA levels between SynH2<sup>-</sup> and SynH2 cells reflect induction of expression in SynH2 cells but carryover of elevated protein levels in the inoculum of SynH2<sup>-</sup> cells not yet diluted in exponential phase. First, we sampled exponential phase between one and two cell doublings so that proteins elevated in stationary phase in the inoculum might still be present. Second, FrmRAB and S assimilation genes are elevated in stationary SynH2<sup>-</sup> cells relative to SynH2 cells (Table S7C), likely reflecting the greater accumulation of acetaldehyde in SynH2<sup>-</sup> cells in stationary phase (Figure 3C). Finally, RpoS and TnaA are markers of stationary phase (Lacour and Landini, 2004) and may reflect elevation of these proteins in SynH stationary cells carried over from the inoculum. In a similar





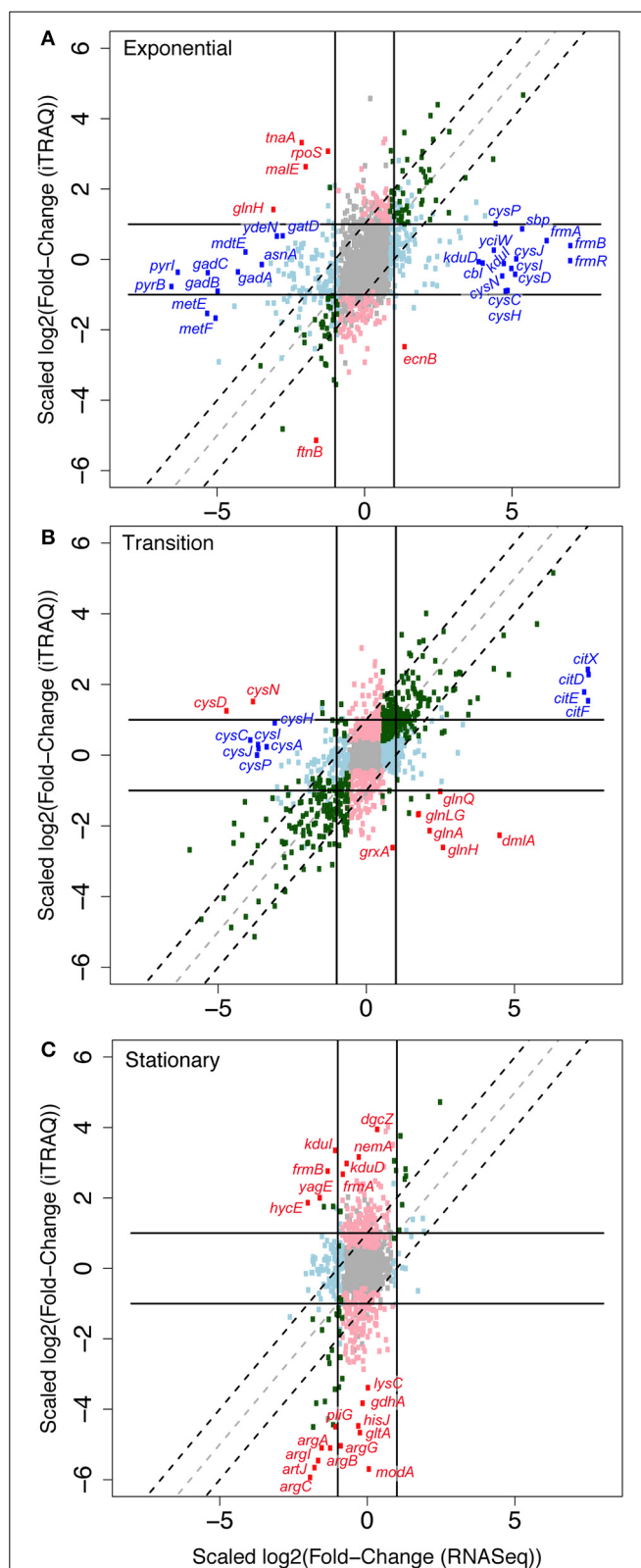
vein, the apparent overrepresentation of PyrBI, GadABC, and MetEF proteins in SynH2 cells could reflect their greater abundance in stationary phase SynH2 cells that were carried over to early exponential phase.

Supporting this view, transition phase cells in which the inoculum was diluted >5-fold exhibited a higher correlation between protein and RNA levels and only limited evidence of post-transcriptional regulation caused by the aromatic inhibitors (**Figure 6B**). Three clusters of outliers reflected (i) reduced transcript levels for S assimilation genes in SynH2<sup>−</sup> without a corresponding drop in protein level (*cys* genes), (ii) higher levels of *glnAGHLQ* transcripts in SynH2 cells than SynH2<sup>−</sup> cells with high protein levels in both, and (iii) high induction of transcripts for the citrate assimilation system (*citDEFX*) in SynH2 with lesser induction of protein levels. These effects likely reflect adjustment of S assimilation gene expression during transition phase, a greater induction of N assimilation in the more rapidly growing SynH2 cells, and induction of citrate assimilation by the aromatic inhibitors.

The clearest evidence for post-transcriptional regulation caused by the aromatic inhibitors appeared in stationary phase (**Figure 6C**). A set of proteins involved in arginine, glutamate, lysine and citrate biosynthesis (ArgABCGI, GdhA, LysC, GltA) and periplasmic proteins arginine high-affinity import (ArtJ), histidine high-affinity import (HisJ), molybdate import (ModA), and lysozyme inhibition (PliG) decreased dramatically in SynH2 cells relative to SynH2<sup>−</sup> cells without corresponding reductions of their transcripts. GdhA, other biosynthetic enzymes, and other periplasmic binding proteins are degraded by the ClpP protease during C or N starvation (Maurizi and Rasuloova, 2002; Weichert et al., 2003); Lon protease also has been implicated in proteolysis upon C starvation (Luo et al., 2008). Thus, we suggest that aromatic inhibitors may enhance degradation of proteins involved in N and C metabolism in stationary phase cells. The periplasmic proteins must be degraded as precursors or mediated by an additional effect involving periplasmic proteases.

## DISCUSSION

Results of our investigation into the effects of LC-derived inhibitors on *E. coli* ethanologenesis support several key conclusions that will guide future work. First, a chemically defined mimic of ACSH (SynH2) that contained the major inhibitors found by chemical analysis of ACSH adequately replicated both growth and the rates of glucose and xylose conversion to ethanol by *E. coli*. SynH2-replication of ACSH required inclusion of osmolytes found in ACSH and established that, at the ratios present in ACSH, phenolic carboxylates and amides, which are not metabolized by *E. coli*, had a greater overall impact on cell growth than phenolic aldehydes and furfurals, which were metabolized. In both SynH2 and ACSH, *E. coli* entered a metabolically active stationary phase as cells exhausted organic sources of N and S (e.g., amino acids) and during which the inhibitors greatly reduced xylose conversion. The impact of inhibitors on cellular energetics reduced levels of ATP, NADH, and NADPH and was seen most dramatically for energetically challenging processes requiring NADPH (like  $\text{SO}_4^{2-}$  assimilation and deoxyribonucleotide production), during transition to the stationary phase



**FIGURE 6 | Effects of aromatic inhibitors on protein levels compared to effects on cognate RNA levels.** Scatter plot comparing log<sub>2</sub>-fold RNA ratios (x-axis) to log<sub>2</sub>-fold protein ratios (y-axis) of GLBRCE1 genes and gene (Continued)

**FIGURE 6 | Continued**

products for cells for grown in SynH2 compared to the reference medium, SynH2<sup>-</sup>. Cells were collected and proteomic samples prepared from exponential (A), transition (B), and stationary (C) growth phases. The lines indicate boundaries beyond which changes exceed 2-fold. The dotted lines demarcate the area expected for parallel changes in protein and RNA levels. *Red*, genes for which changes in protein levels were not paralleled by changes in the corresponding RNA and for which the discrepancy had a  $p \leq 0.05$  (see Table S7). *Blue*, genes for which changes in RNA levels were not paralleled by changes in the corresponding protein and for which the discrepancy had a  $p \leq 0.05$ . *Gray*,  $p > 0.05$  for both RNA and protein ratios. *Light blue*,  $p \leq 0.05$  for RNA ratio but not for protein ratio. *Light pink*,  $p \leq 0.05$  for protein ratio but not for RNA ratio. *Green*,  $p \leq 0.05$  for both RNA and protein ratios and effects are parallel.

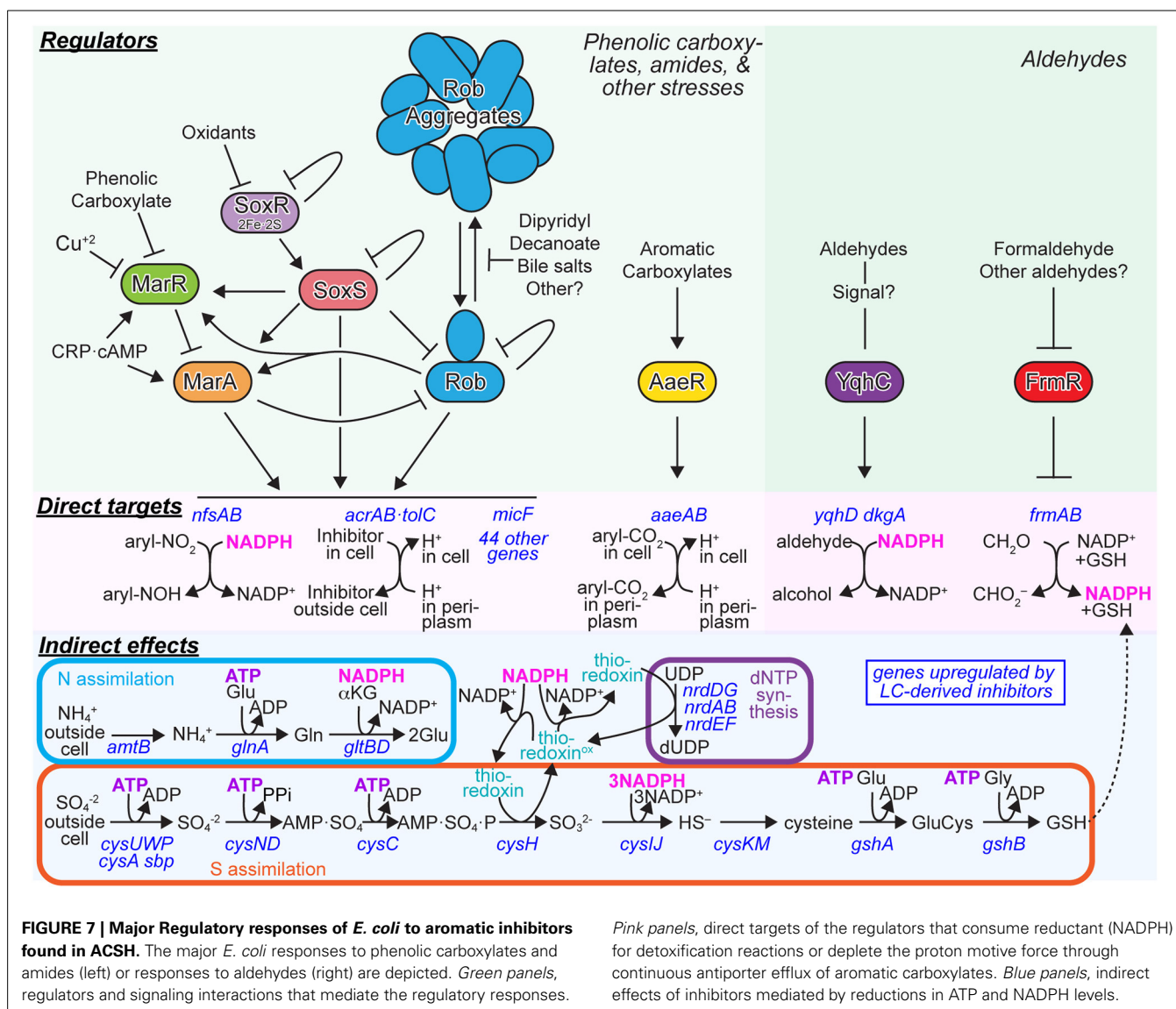
on ATP-dependent  $\text{NH}_3$  assimilation, and in elevated pyruvate levels presumably reflecting reduced NADH-dependent flux of pyruvate to ethanol (Figure 7). The direct effects of the inhibitors on cells appear to be principally mediated by transcriptional rather than translational regulators, with the MarA/SoxS/Rob network, AaeR, FrmR, and YqhC being the most prominent players. Although the effect of the inhibitors on transcriptional regulation of the efflux pumps was striking, increased efflux activity itself may perturb cellular metabolism. For example, Dhamdhare and Zgurskaya (2010) have shown that deletion of the AcrAB-TolC complex results in metabolic shutdown and high NADH/NAD<sup>+</sup> ratios. By analogy, overexpression of efflux pumps may have the opposite effect (e.g., lowering of NADH/NAD<sup>+</sup> ratios), which is consistent with observations in this study. In addition, recent work suggests that the *acrAB* promoter is upregulated in response to certain cellular metabolites (including those related to cysteine and purine biosynthesis), which are normally effluxed by this pump (Ruiz and Levy, 2014). Therefore, upregulation of AcrAB-TolC may impact homeostatic mechanisms of cellular biosynthetic pathways, resulting in continuous upregulation of pathways that require large amounts of reducing power in the form of NADPH. It is also possible that LC-derived inhibitors perturb metabolism directly in ways that generate additional AcrAB-TolC substrates, potentially increasing energy-consuming efflux further. Given these intricacies, further studies to unravel the mechanistic details of the effects of efflux pump activity on cellular metabolism, as a result of exposure to LC-derived inhibitors, are warranted.

The inability of cells to convert xylose in the presence of inhibitors appears to result from a combination of both effects on gene expression and some additional effect on transport or metabolism. The inhibitors lowered xylose gene expression (*XylR* regulon; *xylABFGH*) by a factor of 3–5 during all three growth phases (Table S4). This effect was not caused by the previously documented AraC repression (Desai and Rao, 2010), since it persisted in SynH2 when we replaced the AraC effector L-arabinose with D-arabinose, but might reflect lower levels of cAMP caused by the inhibitors (Figure 4); both the *xylAB* and *xylFGH* operons are also regulated by CRP-cAMP. Nonetheless, significant levels of XylA, B, and F were detected even in the presence of inhibitors (Table S7D), even though xylose conversion remained inhibited even after glucose depletion (Table 2). Thus, the inability to convert xylose may also reflect either the

overall impact of inhibitors on cellular energetics somehow making xylose conversion unfavorable or an effect of xylose transport or metabolism that remains to be discovered. Further studies of the impact of inhibitors on xylose transport and metabolism are warranted. It would be particularly interesting to test SynH formulations designed to compare the conversion efficiencies of xylose, arabinose, and C6 sugars other than glucose.

The central focus of this study was to understand the impact of inhibitors of gene expression regulatory networks. The apparent lack of involvement of post-transcriptional regulation suggests that *E. coli* mounts a defense against LC-derived inhibitors principally by controlling gene transcription, probably reflecting evolution of specific bacterial responses to LC-derived inhibitors. Although enteric bacteria do not ordinarily encounter industrial lignocellulosic hydrolysates, they likely encounter the same suite of compounds from digested plant material in the mammalian gut. Thus, evolution of specific responses is reasonable. A key question for future studies is whether phenolic amides, not ordinarily present in digested biomass, will also invoke these responses in the absence of carboxylates or aldehydes. We note that the apparent absence of a translational regulatory response in the cellular defense against LC-derived inhibitors does not preclude involvement of either direct or indirect post-transcriptional regulation in fine-tuning the response. Our proteomic measurements would likely not have detected fine-tuning. Additionally, we did detect an apparently indirect induction by inhibitors of protein degradation in stationary phase, possibly in response to C starvation (Figure 6C). Finally, we note that the sRNA *micF*, a known post-transcriptional regulator, is a constituent of the MarA/SoxS/Rob regulon and was upregulated by inhibitors. Although confidence was insignificant due to poor detection of sRNAs in RNAseq data, the induction of *micF* was confirmed in a separate study of sRNAs (Ong and Landick, in preparation). Thus, a more focused study of the involvement of sRNAs in responses to LC inhibitors would likely be informative.

MarA/SoxS/Rob is a complex regulon consisting of the three inter-connected primary AraC-class regulators that bind as monomers to 20-bp sites in promoters with highly overlapping specificity and synergistically regulate ~50 genes implicated in resistance to multiple antibiotics and xenobiotics, solvent tolerance, outer membrane permeability, DNA repair, and other functions (Chubiz et al., 2012; Duval and Lister, 2013; Garcia-Bernardo and Dunlop, 2013) (Figure 7). Twenty-three genes, including those encoding the AcrAB-TolC efflux pump, the NfsAB nitroreductases, the *micF* sRNA, superoxide dismutase, some metabolic enzymes (e.g., *Zwf*, *AcnA*, and *FumC*) and incompletely characterized stress proteins are controlled by all three regulators, whereas other genes are annotated as being controlled by only a subset of the regulators (Duval and Lister, 2013), [www.ecocyc.org](http://www.ecocyc.org); (Keseler et al., 2013). MarA and SoxS lack the C-terminal dimerization domain of AraC; this domain is present on Rob and appears to mediate regulation by aggregation that can be reversed by effectors (Griffith et al., 2009). Inputs capable of inducing these genes, either through the MarR and SoxR repressors that control MarA and SoxS, respectively, or by direct effects on Rob include phenolic carboxylates,  $\text{Cu}^{2+}$ , a variety of organic oxidants, dipyrrolyl, decanoate, bile salts, Fis, and Crp-cAMP



(Martin and Rosner, 1997; Rosner et al., 2002; Rosenberg et al., 2003; Chubiz and Rao, 2010; Duval and Lister, 2013; Hao et al., 2014) (Figure 7). Given these diverse inputs, it seems highly likely that ferulate and coumarate in ACSH induce the MarA/SoxS/Rob regulon via MarR. Indeed, LC-hydrolysate and ferulate induction of MarA has been reported (Lee et al., 2012). Interestingly,  $\text{Cu}^{2+}$  recently was shown to induce MarR by oxidation to create MarR disulfide dimer (Hao et al., 2014). Given the elevated levels of  $\text{Cu}^{2+}$  in ACSH reflected by induction of  $\text{Cu}^{2+}$  efflux (Figure 2; Table S4), induction of MarA/SoxS/Rob in ACSH may result from synergistic effects of  $\text{Cu}^{2+}$  and phenolic carboxylates, oxidants that affect SoxR, and yet-to-be-determined compounds that affect Rob. A second response in LC-derived inhibitors appears to be mounted by the LysR-type regulator AaeR, which controls the AaeAB aromatic carboxylate efflux system (Van Dyk et al., 2004) (Figure 7). Both phenolic and aryl carboxylates induce AaeAB through AaeR, but little is known about its substrate specificity or mechanism of activation.

Two distinct regulators, YqhC and FrmR, control synthesis of the YqhD/DkgA NADPH-dependent aldehyde reductases and the FrmAB formaldehyde oxidase, respectively (Herring and Blattner, 2004; Turner et al., 2011). Even less is known about these regulators, although the DNA-binding properties of YqhC have been determined. In particular, it is unclear how aldehydes cause induction, although the current evidence suggests effects on YqhC are likely to be indirect. Given the central role of the regulators AaeR, YqhC, and FrmR in the cellular response to LC-derived inhibitors, further study of their properties and mechanisms is likely to be profitable. With sufficient understanding and engineering, they could be used as response regulators to engineer cells that respond to LC-inhibitors in ways that maximize microbial conversion of sugars to biofuels.

What types of responses would optimize biofuel synthesis? It appears the naturally evolved responses, namely induction of efflux systems and NADPH-dependent detoxification pathways, may not be optimal for efficient synthesis of biofuels. We infer



this conclusion for several reasons. First, our gene expression results reveal that crucial pathways for cellular biosynthesis that are among the most energetically challenging processes in cells, S assimilation, N assimilation, and ribonucleotide reduction, are highly induced by LC-derived inhibitors (Figures 2, 7; Table S4). A reasonable conjecture is that the diversion of energy pools, including NADPH and ATP, to detoxification makes S assimilation, N assimilation, and ribonucleotide reduction difficult, increasing expression of genes for these pathways indirectly. The continued presence of the phenolic carboxylates and amides (Figure 3) likely causes futile cycles of efflux. As both the AcrAB and AaeAB efflux pumps function as proton antiporters (Figure 7), continuous efflux is expected to decrease ATP synthesis by depleting the proton-motive force. Although this response makes sense evolutionarily because it protects DNA from damage by xenobiotics, it does not necessarily aid conversion of sugars to biofuels. Disabling these efflux and detoxification systems, especially during stationary phase when cell growth is no longer necessary, could improve rates of ethanologenesis. Indeed, Ingram and colleagues have shown that disabling the NADPH-dependent YqhD/DkgA enzymes or better yet replacing them with NADH-dependent aldehyde reductases (e.g., FucO) can improve ethanologenesis in furfural-containing hydrolysates of acid-pretreated biomass (Wang et al., 2011a, 2013). That simply deleting *yqhD* improves ethanologenesis argues that, in at least some cases, it is better to expose cells to LC-derived inhibitors than to spend energy detoxifying the inhibitors.

Some previous efforts to engineer cells for improved biofuel synthesis have focused on overexpression of selected efflux pumps to reduce the toxic effects of biofuel products (Dunlop et al., 2011). Although this strategy may help cells cope with the effects of biofuel products, our results suggest an added potential issue when dealing with real hydrolysates, namely that efflux pumps may also reduce the rates of biofuel yields by futile cycling of LC-derived inhibitors. Thus, effective use of efflux pumps will require careful control of their synthesis (Harrison and Dunlop, 2012). An alternative strategy to cope with LC-derived inhibitors may be to devise metabolic routes to assimilate them into cellular metabolism.

In conclusion, our findings illustrate the utility of using chemically defined mimics of biomass hydrolysates for genome-scale study of microbial biofuel synthesis as a strategy to identify barriers to biofuel synthesis. By identifying the main inhibitors present in ammonia-pretreated biomass hydrolysate and using these inhibitors in a synthetic hydrolysate, we were able to identify the key regulators responsible for the cellular responses that reduced the rate of ethanol production and limited xylose conversion to ethanol. Knowledge of these regulators will enable design of new control circuits to improve microbial biofuel production.

## ACKNOWLEDGMENTS

The authors thank Trey Sato and Jeff Piotrowski for critical reading of the manuscript, Fachuang Lu and John Ralph for advice on synthesis of feruloyl and coumaroyl amide, and Christa Pennacchio and colleagues at the Joint Genome Institute for cDNA library preparation and sequencing. This work was funded by the DOE Great Lakes Bioenergy Research Center (DOE BER

Office of Science DE-FC02-07ER64494). Portions of this research were enabled by the DOE GSP under the Pan-omics project. Work was performed in the Environmental Molecular Science Laboratory, a U.S. Department of Energy (DOE) national scientific user facility at Pacific Northwest National Laboratory (PNNL) in Richland, WA. Battelle operates PNNL for the DOE under contract DE-AC05-76RLO01830.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fmicb.2014.00402/abstract>

## REFERENCES

- Bantscheff, M., Boesche, M., Eberhard, D., Matthieson, T., Sweetman, G., and Kuster, B. (2008). Robust and sensitive iTRAQ quantification on an LTQ Orbitrap mass spectrometer. *Mol. Cell. Proteomics* 7, 1702–1713. doi: 10.1074/mcp.M800029-MCP200
- Booth, I. R., Ferguson, G. P., Miller, S., Li, C., Gunasekera, B., and Kinghorn, S. (2003). Bacterial production of methylglyoxal: a survival strategy or death by misadventure? *Biochem. Soc. Trans.* 31, 1406–1408. doi: 10.1042/BST0311406
- Buescher, J. M., Moco, S., Sauer, U., and Zamboni, N. (2010). Ultrahigh performance liquid chromatography-tandem mass spectrometry method for fast and robust quantification of anionic and aromatic metabolites. *Anal. Chem.* 82, 4403–4412. doi: 10.1021/ac100101d
- Chubiz, L. M., Glekas, G. D., and Rao, C. V. (2012). Transcriptional cross talk within the mar-sox-rob regulon in *Escherichia coli* is limited to the rob and marRAB operons. *J. Bacteriol.* 194, 4867–4875. doi: 10.1128/JB.00680-12
- Chubiz, L. M., and Rao, C. V. (2010). Aromatic acid metabolites of *Escherichia coli* K-12 can induce the marRAB operon. *J. Bacteriol.* 192, 4786–4789. doi: 10.1128/JB.00371-10
- Chundawat, S. P., Beckham, G. T., Himmel, M. E., and Dale, B. E. (2011). Deconstruction of lignocellulosic biomass to fuels and chemicals. *Annu. Rev. Chem. Biomol. Eng.* 2, 121–145. doi: 10.1146/annurev-chembioeng-061010-114205
- Dalrymple, B. P., and Swadling, Y. (1997). Expression of a *Butyrivibrio fibrisolvens* E14 gene (*cinB*) encoding an enzyme with cinnamoyl ester hydrolase activity is negatively regulated by the product of an adjacent gene (*cinR*). *Microbiology* 143, 1203–1210. doi: 10.1099/00221287-143-4-1203
- Desai, T. A., and Rao, C. V. (2010). Regulation of arabinose and xylose metabolism in *Escherichia coli*. *Appl. Environ. Microbiol.* 76, 1524–1532. doi: 10.1128/AEM.01970-09
- Dhamdhare, G., and Zgurskaya, H. I. (2010). Metabolic shutdown in *Escherichia coli* cells lacking the outer membrane channel TolC. *Mol. Microbiol.* 77, 743–754. doi: 10.1111/j.1365-2958.2010.07245.x
- Dunlop, M. J., Dossani, Z. Y., Szmidi, H. L., Chu, H. C., Lee, T. S., Keasling, J. D., et al. (2011). Engineering microbial biofuel tolerance and export using efflux pumps. *Mol. Syst. Biol.* 7, 487. doi: 10.1038/msb.2011.21
- Duval, V., and Lister, I. M. (2013). MarA, SoxS and Rob of *Escherichia coli*—Global regulators of multidrug resistance, virulence, and stress response. *Int. J. Biotechnol. Wellness Ind.* 2, 101–124. doi: 10.6000/1927-3037.2013.02.03.2
- Garcia-Bernardo, J., and Dunlop, M. J. (2013). Tunable stochastic pulsing in the *Escherichia coli* multiple antibiotic resistance network from interlinked positive and negative feedback loops. *PLoS Comput. Biol.* 9:e1003229. doi: 10.1371/journal.pcbi.1003229
- Gonzalez, C. F., Proudfoot, M., Brown, G., Korniyenko, Y., Mori, H., Savchenko, A. V., et al. (2006). Molecular basis of formaldehyde detoxification. Characterization of two S-formylglutathione hydrolases from *Escherichia coli*, FrmB and YeiG. *J. Biol. Chem.* 281, 14514–14522. doi: 10.1074/jbc.M600996200
- Goodarzi, H., Bennett, B. D., Amini, S., Reaves, M. L., Hottes, A. K., Rabinowitz, J. D., et al. (2010). Regulatory and metabolic rewiring during laboratory evolution of ethanol tolerance in *E. coli*. *Mol. Syst. Biol.* 6, 378. doi: 10.1038/msb.2010.33
- Griffith, K. L., Fitzpatrick, M. M., Keen, E. F., and Wolf, R. E. (2009). Two functions of the C-terminal domain of *Escherichia coli* Rob: mediating “sequestration–dispersal” as a novel off–on switch for regulating Rob’s activity as a transcription



- activator and preventing degradation of Rob by Lon protease. *J. Mol. Biol.* 388, 415–430. doi: 10.1016/j.jmb.2009.03.023
- Hao, Z., Lou, H., Zhu, R., Zhu, J., Zhang, D., Zhao, B. S., et al. (2014). The multiple antibiotic resistance regulator MarR is a copper sensor in *Escherichia coli*. *Nat. Chem. Biol.* 10, 21–28. doi: 10.1038/nchembio.1380
- Harrison, M. E., and Dunlop, M. J. (2012). Synthetic feedback loop model for increasing microbial biofuel production using a biosensor. *Front. Microbiol.* 3:360. doi: 10.3389/fmicb.2012.00360
- Herring, C. D., and Blattner, F. R. (2004). Global transcriptional effects of a suppressor tRNA and the inactivation of the regulator frnR. *J. Bacteriol.* 186, 6714–6720. doi: 10.1128/JB.186.20.6714-6720.2004
- Ingram, L. O., and Buttke, T. M. (1984). Effects of alcohols on micro-organisms. *Adv. Microb. Physiol.* 25, 253–300.
- Jarboe, L. R. (2011). YqhD: a broad-substrate range aldehyde reductase with various applications in production of biorenewable fuels and chemicals. *Appl. Microbiol. Biotechnol.* 89, 249–257. doi: 10.1007/s00253-010-2912-9
- Kelly, R. T., Page, J. S., Luo, Q., Moore, R. J., Orton, D. J., Tang, K., et al. (2006). Chemically etched open tubular and monolithic emitters for nano-electrospray ionization mass spectrometry. *Anal. Chem.* 78, 7796–7801. doi: 10.1021/ac061133r
- Keseler, I. M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonavides-Martinez, C., et al. (2013). EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res.* 41, D605–D612. doi: 10.1093/nar/gks1027
- Kim, S., Gupta, N., and Pevzner, P. A. (2008). Spectral probabilities and generating functions of tandem mass spectra: a strike against decoy databases. *J. Proteome Res.* 7, 3354–3363. doi: 10.1021/pr8001244
- Klinke, H. B., Thomsen, A. B., and Ahiring, B. K. (2004). Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass. *Appl. Microbiol. Biotechnol.* 66, 10–26. doi: 10.1007/s00253-004-1642-2
- Lacour, S., and Landini, P. (2004). SigmaS-dependent gene expression at the onset of stationary phase in *Escherichia coli*: function of sigmaS-dependent genes and identification of their promoter sequences. *J. Bacteriol.* 186, 7186–7195. doi: 10.1128/JB.186.21.7186-7195.2004
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25. doi: 10.1186/gb-2009-10-3-r25
- Lau, M. W., and Dale, B. E. (2010). Effect of primary degradation-reaction products from Ammonia Fiber Expansion (AFEX)-treated corn stover on the growth and fermentation of *Escherichia coli* KO11. *Bioresour. Technol.* 101, 7849–7855. doi: 10.1016/j.biortech.2010.04.076
- Lawford, H. G., and Rousseau, J. D. (1995). Comparative energetics of glucose and xylose metabolism in ethanologenic recombinant *Escherichia coli* B. *Appl. Biochem. Biotechnol.* 51–52, 179–195.
- Lee, S., Nam, D., Jung, J. Y., Oh, M. K., Sang, B. I., and Mitchell, R. J. (2012). Identification of *Escherichia coli* biomarkers responsive to various lignin-hydrolysate compounds. *Bioresour. Technol.* 114, 450–456. doi: 10.1016/j.biortech.2012.02.085
- Li, B., and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12:323. doi: 10.1186/1471-2105-12-323
- Liu, Z. L. (2011). Molecular mechanisms of yeast tolerance and *in situ* detoxification of lignocellulose hydrolysates. *Appl. Microbiol. Biotechnol.* 90, 809–825. doi: 10.1007/s00253-011-3167-9
- Luo, S., McNeill, M., Myers, T. G., Hohman, R. J., and Levine, R. L. (2008). Lon protease promotes survival of *Escherichia coli* during anaerobic glucose starvation. *Arch. Microbiol.* 189, 181–185. doi: 10.1007/s00203-007-0304-z
- Maiolica, A., Borsotti, D., and Rappsilber, J. (2005). Self-made frits for nanoscale columns in proteomics. *Proteomics* 5, 3847–3850. doi: 10.1002/pmic.200402010
- Martin, R. G., and Rosner, J. L. (1997). Fis, an accessory factor for transcriptional activation of the mar (multiple antibiotic resistance) promoter of *Escherichia coli* in the presence of the activator MarA, SoxS, or Rob. *J. Bacteriol.* 179, 7410–7419.
- Marx, C. J., Miller, J. A., Chistoserdova, L., and Lidstrom, M. E. (2004). Multiple formaldehyde oxidation/detoxification pathways in *Burkholderia fungorum* LB400. *J. Bacteriol.* 186, 2173–2178. doi: 10.1128/JB.186.7.2173-2178.2004
- Maurizi, M. R., and Rasulo, F. (2002). Degradation of L-glutamate dehydrogenase from *Escherichia coli*: allosteric regulation of enzyme stability. *Arch. Biochem. Biophys.* 397, 206–216. doi: 10.1006/abbi.2001.2703
- Miller, E. N., and Ingram, L. O. (2007). Combined effect of betaine and trehalose on osmotic tolerance of *Escherichia coli* in mineral salts medium. *Biotechnol. Lett.* 29, 213–217. doi: 10.1007/s10529-006-9226-0
- Miller, E. N., Jarboe, L. R., Turner, P. C., Pharkya, P., Yomano, L. P., York, S. W., et al. (2009a). Furfural inhibits growth by limiting sulfur assimilation in ethanologenic *Escherichia coli* strain LY180. *Appl. Environ. Microbiol.* 75, 6132–6141. doi: 10.1128/AEM.01187-09
- Miller, E. N., Jarboe, L. R., Yomano, L. P., York, S. W., Shanmugam, K. T., and Ingram, L. O. (2009b). Silencing of NADPH-dependent oxidoreductase genes (yqhD and dkgA) in furfural-resistant ethanologenic *Escherichia coli*. *Appl. Environ. Microbiol.* 75, 4315–4323. doi: 10.1128/AEM.00567-09
- Miller, E. N., Turner, P. C., Jarboe, L. R., and Ingram, L. O. (2010). Genetic changes that increase 5-hydroxymethyl furfural resistance in ethanol-producing *Escherichia coli* LY180. *Biotechnol. Lett.* 32, 661–667. doi: 10.1007/s10529-010-0209-9
- Mills, T. Y., Sandoval, N. R., and Gill, R. T. (2009). Cellulosic hydrolysate toxicity and tolerance mechanisms in *Escherichia coli*. *Biotechnol. Biofuels* 2:26. doi: 10.1186/1754-6834-2-26
- Monroe, M. E., Shaw, J. L., Daly, D. S., Adkins, J. N., and Smith, R. D. (2008). MASiC: a software program for fast quantitation and flexible visualization of chromatographic profiles from detected LC-MS/MS features. *Comput. Biol. Chem.* 32, 215–217. doi: 10.1016/j.compbiolchem.2008.02.006
- Neidhardt, F. C., Bloch, P. L., and Smith, D. F. (1974). Culture medium for enterobacteria. *J. Bacteriol.* 119, 736–747.
- Pasa-Tolic, L., Masselon, C., Barry, R. C., Shen, Y., and Smith, R. D. (2004). Proteomic analyses using an accurate mass and time tag strategy. *Biotechniques* 37, 621–624, 626–633, 636 passim.
- Piotrowski, J. S., Zhang, Y., Bates, D. M., Keating, D. H., Sato, T. K., Ong, I. M., et al. (2014). Death by a thousand cuts: the challenges and diverse landscape of lignocellulosic hydrolysate inhibitors. *Front. Microbiol.* 5:90. doi: 10.3389/fmicb.2014.00090
- Purvis, J. E., Yomano, L. P., and Ingram, L. O. (2005). Enhanced trehalose production improves growth of *Escherichia coli* under osmotic stress. *Appl. Environ. Microbiol.* 71, 3761–3769. doi: 10.1128/AEM.71.7.3761-3769.2005
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616
- Rosenberg, E. Y., Bertenthal, D., Nilles, M. L., Bertrand, K. P., and Nikaido, H. (2003). Bile salts and fatty acids induce the expression of *Escherichia coli* AcrAB multidrug efflux pump through their interaction with Rob regulatory protein. *Mol. Microbiol.* 48, 1609–1619. doi: 10.1046/j.1365-2958.2003.03531.x
- Rosner, J. L., Dangi, B., Gronenborn, A. M., and Martin, R. G. (2002). Posttranscriptional activation of the transcriptional activator Rob by dipyrrolyl in *Escherichia coli*. *J. Bacteriol.* 184, 1407–1416. doi: 10.1128/JB.184.5.1407-1416.2002
- Ross, P. L., Huang, Y. N., Marchese, J. N., Williamson, B., Parker, K., Hattan, S., et al. (2004). Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteomics* 3, 1154–1169. doi: 10.1074/mcp.M400129-MCP200
- Ruiz, C., and Levy, S. B. (2014). Regulation of acrAB expression by cellular metabolites in *Escherichia coli*. *J. Antimicrob. Chemother.* 69, 390–399. doi: 10.1093/jac/dkt352
- Schwalbach, M. S., Keating, D. H., Tremaine, M., Marner, W. D., Zhang, Y., Bothfeld, W., et al. (2012). Complex physiology and compound stress responses during fermentation of alkali-pretreated corn stover hydrolysate by an *Escherichia coli* ethanologen. *Appl. Environ. Microbiol.* 78, 3442–3457. doi: 10.1128/AEM.07329-11
- Sezonov, G., Joseleau-Petit, D., and D'Ari, R. (2007). *Escherichia coli* physiology in Luria-Bertani broth. *J. Bacteriol.* 189, 8746–8749. doi: 10.1128/JB.01368-07
- Smith, G. K. (2005). “Limma: linear models for microarray data,” in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, eds R. Gentleman, V. Carey, S. Dudoit, R. Irizarry, and W. Huber (New York, NY: Springer), 397–420.

- Smyth, G. K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* 3, Article3. doi: 10.2202/1544-6115.1027
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* 102, 15545–15550. doi: 10.1073/pnas.0506580102
- Sulavik, M. C., Gambino, L. F., and Miller, P. F. (1995). The MarR repressor of the multiple antibiotic resistance (*mar*) operon in *Escherichia coli*: prototypic member of a family of bacterial regulatory proteins involved in sensing phenolic compounds. *Mol. Med.* 1, 436–446.
- Turner, P. C., Miller, E. N., Jarboe, L. R., Baggett, C. L., Shanmugam, K. T., and Ingram, L. O. (2011). YqhC regulates transcription of the adjacent *Escherichia coli* genes yqhD and dkgA that are involved in furfural tolerance. *J. Ind. Microbiol. Biotechnol.* 38, 431–439. doi: 10.1007/s10295-010-0787-5
- Underwood, S. A., Buszko, M. L., Shanmugam, K. T., and Ingram, L. O. (2004). Lack of protective osmolytes limits final cell density and volumetric productivity of ethanologenic *Escherichia coli* KO11 during xylose fermentation. *Appl. Environ. Microbiol.* 70, 2734–2740. doi: 10.1128/AEM.70.5.2734-2740.2004
- Van Dyk, T. K., Templeton, L. J., Cantera, K. A., Sharpe, P. L., and Sariaslani, F. S. (2004). Characterization of the *Escherichia coli* AaeAB efflux pump: a metabolic relief valve? *J. Bacteriol.* 186, 7196–7204. doi: 10.1128/JB.186.21.7196-7204.2004
- Varemo, L., Nielsen, J., and Nookaew, I. (2013). Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. *Nucleic Acids Res.* 41, 4378–4391. doi: 10.1093/nar/gkt111
- Wang, X., Miller, E. N., Yomano, L. P., Zhang, X., Shanmugam, K. T., and Ingram, L. O. (2011a). Increased furfural tolerance due to overexpression of NADH-dependent oxidoreductase FucO in *Escherichia coli* strains engineered for the production of ethanol and lactate. *Appl. Environ. Microbiol.* 77, 5132–5140. doi: 10.1128/AEM.05008-11
- Wang, X., Yomano, L. P., Lee, J. Y., York, S. W., Zheng, H., Mullinnix, M. T., et al. (2013). Engineering furfural tolerance in *Escherichia coli* improves the fermentation of lignocellulosic sugars into renewable chemicals. *Proc. Natl. Acad. Sci. U.S.A.* 110, 4021–4026. doi: 10.1073/pnas.1217958110
- Wang, Y., Yang, F., Gritsenko, M. A., Wang, Y., Clauss, T., Liu, T., et al. (2011b). Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *Proteomics* 11, 2019–2026. doi: 10.1002/pmic.201000722
- Weichart, D., Querfurth, N., Dreger, M., and Hengge-Aronis, R. (2003). Global role for ClpP-containing proteases in stationary-phase adaptation of *Escherichia coli*. *J. Bacteriol.* 185, 115–125. doi: 10.1128/JB.185.1.115-125.2003
- Yomano, L. P., York, S. W., and Ingram, L. O. (1998). Isolation and characterization of ethanol-tolerant mutants of *Escherichia coli* KO11 for fuel ethanol production. *J. Ind. Microbiol. Biotechnol.* 20, 132–138.
- Zaldivar, J., Martinez, A., and Ingram, L. O. (1999). Effect of selected aldehydes on the growth and fermentation of ethanologenic *Escherichia coli*. *Biotechnol. Bioeng.* 65, 24–33. doi: 10.1002/(SICI)1097-0290(19991005)65:1<24::AID-BIT4>3.0.CO;2-2

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 April 2014; accepted: 17 July 2014; published online: 13 August 2014.

Citation: Keating DH, Zhang Y, Ong IM, McIlwain S, Morales EH, Grass JA, Tremaine M, Bothfeld W, Higbee A, Ulbrich A, Balloon AJ, Westphall MS, Aldrich J, Lipton MS, Kim J, Moskvina OV, Bukhman YV, Coon JJ, Kiley PJ, Bates DM and Landick R (2014) Aromatic inhibitors derived from ammonia-pretreated lignocellulose hinder bacterial ethanologenesis by activating regulatory circuits controlling inhibitor efflux and detoxification. *Front. Microbiol.* 5:402. doi: 10.3389/fmicb.2014.00402

This article was submitted to Microbial Physiology and Metabolism, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Keating, Zhang, Ong, McIlwain, Morales, Grass, Tremaine, Bothfeld, Higbee, Ulbrich, Balloon, Westphall, Aldrich, Lipton, Kim, Moskvina, Bukhman, Coon, Kiley, Bates and Landick. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Genomic insights into the fungal lignocellulolytic system of *Myceliophthora thermophila*

Anthi Kamaouri<sup>1,2†</sup>, Evangelos Topakas<sup>1†</sup>, Io Antonopoulou<sup>2</sup> and Paul Christakopoulos<sup>2\*</sup>

<sup>1</sup> Biotechnology Laboratory, Department of Synthesis and Development of Industrial Processes, School of Chemical Engineering, National Technical University of Athens, Athens, Greece

<sup>2</sup> Biochemical Process Engineering, Chemical Engineering, Department of Civil, Environmental and Natural Resources Engineering, Luleå University of Technology, Luleå, Sweden

## Edited by:

Katherine M. Pappas, University of Athens, Greece

## Reviewed by:

Gianni Panagiotou, The University of Hong Kong, Hong Kong  
Ulrika Rova, Luleå University of Technology, Sweden

## \*Correspondence:

Paul Christakopoulos, Biochemical Process Engineering, Chemical Engineering, Department of Civil, Environmental and Natural Resources Engineering, Luleå University of Technology, C-Building, Universitetsområdet, Porsön, SE-97187 Luleå, Sweden  
e-mail: paul.christakopoulos@ltu.se

<sup>†</sup> These authors have contributed equally to this work.

The microbial conversion of solid cellulosic biomass to liquid biofuels may provide a renewable energy source for transportation fuels. Cellulolytic fungi represent a promising group of organisms, as they have evolved complex systems for adaptation to their natural habitat. The filamentous fungus *Myceliophthora thermophila* constitutes an exceptionally powerful cellulolytic microorganism that synthesizes a complete set of enzymes necessary for the breakdown of plant cell wall. The genome of this fungus has been recently sequenced and annotated, allowing systematic examination and identification of enzymes required for the degradation of lignocellulosic biomass. The genomic analysis revealed the existence of an expanded enzymatic repertoire including numerous cellulases, hemicellulases, and enzymes with auxiliary activities, covering the most of the recognized CAZy families. Most of them were predicted to possess a secretion signal and undergo through post-translational glycosylation modifications. These data offer a better understanding of activities embedded in fungal lignocellulose decomposition mechanisms and suggest that *M. thermophila* could be made usable as an industrial production host for cellulolytic and hemicellulolytic enzymes.

**Keywords:** *Myceliophthora thermophila*, plant biomass, lignocellulolytic enzymes, CAZy, biofuels

## INTRODUCTION

Ethanol production from lignocellulosic biomass, comprised primarily of cellulose and hemicellulose, appears to evolve as one of the most important technologies for sustainable development. Given its renewable nature, biomass is a potential raw material not only for the production of biofuels, but also chemicals, energy and other materials of main industrial interest (Zhang et al., 2006). The monosaccharides contained in the cellulosic (glucose) and hemicellulosic fractions (xylose, arabinose, mannose, and galactose) represent substrates that can be used for ethanol production via fermentation. To initiate the degradation of these fractions, it is necessary to overcome the physical and chemical barriers presented by the cohesive combination of the main biomass components, which hinders the hydrolysis of cellulose and hemicellulose into fermentable sugars. The above include high substrate viscosity, poor mass transfer conditions and long reaction times, during which hydrolysis reactors are susceptible to contamination. Fungi are the main decomposers of lignocellulosic biomass in terrestrial ecosystems and the enzymes they secrete to break down lignocellulose may be useful in industrial processes. Thermophilic fungi provide a potential source of plant cell wall degrading enzymes with higher levels of specific activity and better stability at higher temperatures, thus making it feasible to minimize the hydrolysis time, reduce substrate viscosity and contamination levels (Margaritis and Merchant, 1986).

*Myceliophthora thermophila* (synonym *Sporotrichum thermophile*) is a thermophilic filamentous fungus, classified as an

ascomycete, which was isolated from soil in eastern Russia and constitutes an exceptionally powerful cellulolytic organism, which synthesizes a complete set of enzymes necessary for the breakdown of cellulose. The growth rate and cell density of this microorganism appear to be similar in media containing cellulose or glucose (Bhat and Maheshwari, 1987). The 38.7 Mbp genome of *M. thermophila*, comprising about 9500 genes, organized in 7 chromosomes, has been sequenced and annotated (Joint Genome Institute, University of California, <http://genome.jgi-psf.org>; Berka et al., 2011). It revealed a large number of genes putatively encoding industrially important enzymes, such as carbohydrate-active enzymes (CAZy), proteases, oxidoreductases, and lipases, while more than 200 sequences have been identified exclusively for plant cell-wall-degrading enzymes. These sequences encode a large number of glycoside hydrolases (GH) and polysaccharide lyases, covering the most of the recognized families (Table 1). In addition, *M. thermophila* was developed into a proprietary mature enzyme production system with easy scaling (C1 strain; Visser et al., 2011). The main features of C1 are the high production levels (up to 100 g/L protein), as well as the maintenance of low viscosity levels of the culture medium, thus enabling fermentation process to reach very high densities.

*M. thermophila* exhibits an impressive number of accessory enzymes belonging to AA9 (previously described as GH61) and family 1 carbohydrate binding modules (CBM), which are the highest found in fungi (Berka et al., 2011). Family 1 CBM presents a cellulose-binding function and is almost exclusively found in

**Table 1 | Number of predicted CAZymes encoded in the genome of *M. thermophila*.**

	Specific activity	CAZy module(s)	No. id. seq.
Cellulases	Endoglucanases	GH 5, 7, 12, 45	8
	Cellobiohydrolases	GH 6, 7	7
	$\beta$ -glucosidases	GH 1, 3	8
	LPMOs	AA9	25
Xylanases	Xylanases	GH 10, 11	12
	Xylosidases	GH 3, 43	4
Arabinases	Endoarabinases	GH 43	3
	Exo-arabinases/ arabinofuranosidases	GH 43, 51, 62	11
Mannanases	Endomannanases	GH 5, 26	3
	Mannosidases	GH 2	2
Pectinases	Polygalacturonases	GH28	2
	Rhamnosidases	GH78	1
	Pectin lyases	PL 1, 3, 4, 20	8
	Pectin esterases	CE 8, 12	4
Esterases	Feruloyl esterases	CE 1	4
	Acetyl esterases	CE 3, 5, 16	8
	Acetylmannan esterases	CE 12	2
	Glycuronoyl esterases	CE 15	2

GHs, Glycoside hydrolases; CEs, carbohydrate esterases; and PLs, polysaccharide lyases are included, covering the most of the recognized families.

enzymes of fungal origin (<http://www.cazy.org>; Guillén et al., 2009). In addition, *M. thermophila* distinguishes itself from other cellulolytic fungi, such as *Aspergillus niger* and *Trichoderma reesei* by the presence of a relatively high number of (glucurono) arabinoxylan degrading enzymes (Hinz et al., 2009). Eleven putative xylanases were found that belong into GH 10 and 11 families compared to five in both *A. niger* (Broad Institute of Harvard and MIT, <http://www.broadinstitute.org>) and *T. reesei* (Joint Genome Institute, University of California, <http://genome.jgi-psf.org>), while 14 arabinofuranosidases belonging to GH 43, 51, and 62 families were found compared to 13 in *A. niger* and three in *T. reesei*, rendering *M. thermophila* a promising source of hemicellulolytic enzymes. Studying the secretome of *M. thermophila* after 30 h of growth in barley and alfalfa straws, it was found to comprise of 683 predicted proteins, 230 of which are proteins with unknown function (Berka et al., 2011). Based on transcriptome analysis, many secreted enzymes including accessory enzymes, hypothetical proteins and proteins with unknown function were upregulated, when the fungus is grown in more complex substrates, such as agricultural straws, compared to glucose, indicating their crucial role in lignocellulose degradation (Berka et al., 2011).

*M. thermophila* grows in temperatures between 25 and 55°C, while a relative growth performance study on mycobroth agar plates indicated that the optimum condition is at 45°C (Morgenstern et al., 2012). The temperature optima for

several enzymes with the same specific activity, characterized from *M. thermophila*, range from 50 to 70°C. For example, *StEG5* endoglucanase, expressed in *A. niger*, exhibits a  $T_{opt}$  of 70°C (Tambor et al., 2012), while recombinant *MtEG7* expressed in *Pichia pastoris* exhibited an optimal temperature of 60°C (Karnaouri et al., 2014). The same characteristic is also observed for *M. thermophila* xylanases expressed in *A. niger*, showing optimal activity at temperatures between 50 and 70°C (Berka et al., 2011), underpinning the enzymatic potential that is not only diverse in catalytic activities, but also in properties increasing its efficiency in various temperatures.

Individual cellulolytic enzymes exhibit comparable activities on cellulose; however, synthetically composed multienzyme mixtures display a much higher performance than those from other lignocellulolytic thermostable fungi (Szijártó et al., 2011; Zhang et al., 2013). This can be attributed to synergistic mode of action between the enzymes. For example, synergism between GH 11 xylanase and type C feruloyl esterase has been proved (Moukoulis et al., 2010), as well as between cellobiohydrolases acting on the reducing and the non-reducing end of cellulose molecules (Gusakov et al., 2007).

In this review, an overview will be given of the cellulolytic and hemicellulolytic potential of *M. thermophila* regarding the degradation of plant cell wall material. The genomic potential of this thermophilic fungus demonstrates a strong enzymatic toolbox including hydrolytic, oxidative and accessory activities that may enhance its ability to decompose plant biomass. Many of these enzymes have been isolated from culture supernatant or selectively overexpressed in *M. thermophila* (C1 strain) or in other heterologous hosts and have been characterized. All sequences used in this study were extracted from Genome Portal database (<http://genome.jgi-psf.org>) and the continually updated CAZy database (<http://www.cazy.org>; Lombard et al., 2014). The conserved domains were found with Pfam/InterProscan (<http://pfam.sanger.ac.uk/>; Punta et al., 2012), while the theoretical molecular mass and isoelectric point for each protein were calculated using the ProtParam tool of ExPASy (<http://web.expasy.org/protparam/>). Post-translational glycosylation sites were predicted with NetNGlyc 1.0 server (<http://www.cbs.dtu.dk/services/NetNGlyc/>) and NetOGlyc 3.1 server (<http://www.cbs.dtu.dk/services/NetOGlyc/>). Predicted secretome was extracted using SignalP v4.0 (<http://www.cbs.dtu.dk/services/SignalP/>).

## CELLULOLYTIC SYSTEM

Cellulose is composed of  $\beta$ -D-anhydroglucopyranose units linked by (1,4)-glycosidic bonds. Polymorphism or allotropy refers to the existence of more than one crystalline forms differing in physical and chemical properties. Cellulose degradation is attributed to the synergistic action of three complementary enzyme activities: (1) endoglucanases (EGs, EC 3.2.1.4); (2) exoglucanases, including cellobiohydrolases (EC 3.2.1.74) and cellobiohydrolases (CBHs, EC 3.2.1.91 for the non-reducing end acting CBHs and EC 3.2.1.176 for the reducing end acting ones) and (3)  $\beta$ -glucosidases (BGs, EC 3.2.1.21) (Lynd et al., 2002). Amorphous regions of the polysaccharide chain are cleaved randomly by EGs, while CBHs remove processively cellobiosaccharides from chain ends. The latter are the most abundant enzymes in the secretome



of cellulolytic fungi (Jun et al., 2011; Ribeiro et al., 2012). Their main representatives are GH family 7 (CBH I) that attack the reducing end of a cellulose chain and GH family 6 (CBH II) that are specific toward the non-reducing end of the chain. Until very recently, CBHs were considered as the main degraders of the crystalline part of cellulose (Sweeney and Xu, 2012).

EGs are widespread among GH families, with examples described for families 5–9, 12, 44, 45, 48, 51, and 74 on the continually updated CAZy database (<http://www.cazy.org/>; Lombard et al., 2014). Most of them show optimal activity at neutral or acidic pH and at temperatures below 50°C (Maheshwari et al., 2000). Exo-glucanases (or CBHs) act in a processive manner (Davies and Henrissat, 1995) and are classified only to two families, as referred previously. One of the important features of all CBHs is that they can act on microcrystalline cellulose (Terri, 1997). BGs include enzymes of GH1 and GH3 families that hydrolyze cellobiose and short (soluble) celooligosaccharides to glucose that could subsequently fermented to ethanol; e.g., the hydrolysis reaction is performed in the liquid phase, rather than on the surface of the insoluble cellulose particles, such as EGs and CBHs. The removal of cellobiose is an important step of the enzymatic hydrolysis process, as it assists in reduction of the inhibitory effect of cellobiose on EG and CBH. BG activity has often been found to be rate-limiting during enzymatic hydrolysis of cellulose (Duff and Murray, 1996; Tolan and Foody, 1999), and due to that the commercial cellulase enzyme preparations are often supplemented with BG activity.

Until recently, only hydrolytic enzymes were thought to play a role in the degradation of recalcitrant cellulose and hemicelluloses to fermentable sugars. Recent studies demonstrate that enzymes from the GH family 61 show lytic polysaccharide monooxygenase activity (LPMO) and have an enhancing cellulolytic effect when combined with common cellulases (Horn et al., 2012).

Together with cellobiose dehydrogenase (CDH; EC 1.1.99.18), an enzymatic system capable of oxidative cellulose cleavage is formed, which increases the efficiency of cellulases and boosts the enzymatic conversion of lignocellulose. It has long been thought that the proteins of GH family 61 are accessory proteins enhancing cellulose decomposition. They were thus frequently referred to as the “cellulose enhancing factors” (Harris et al., 2010) and previously thought to have no or only weak endoglucanase activity (Karlsson et al., 2001). Now, these enzymes are now reclassified to AA9 family of CAZy database and their mode of action provide a new dimension to the classical concept of cellulose degradation, as recently reviewed by Dimarogona et al. (2013). These copper-dependent enzymes were shown to cleave cellulose by an oxidative mechanism provided that reduction equivalents from CDH or low molecular weight reducing agents (e.g., ascorbate) are available (Langston et al., 2011). In some genomes, AA9 genes even outnumber cellulose genes. It remains to be elucidated whether all of these encoded enzymes have PMO activity, but their large number emphasizes the importance of oxidative cellulose cleavage. *M. thermophila*'s genome has 25 AA9 genes, encoding putative proteins acting as accessory LPMOs enzymes (Berka et al., 2011). This number is outstanding in comparison to common lignocellulolytic organisms, as *A. niger* (seven sequences) and *T. reesei* (nine sequences). This difference can explain the high efficiency of hydrolysis of *Myceliophthora* in nature substrates and reveals the crucial role of these enzymes in the whole procedure.

Throughout the genome of *M. thermophila*, there are eight sequences encoding EG activity, seven sequences of CBH activity and nine sequences of BG activity (Figure 1; Table 2). The theoretical average molecular weight of the translated proteins is calculated at  $51.05 \pm 16.2$  kDa and the theoretical pI at  $5.58 \pm 0.3$ . EGs are distributed to families GH5, 7, 12, and 45, all predicted to possess a secretion signal and several N- and O-glycosylation

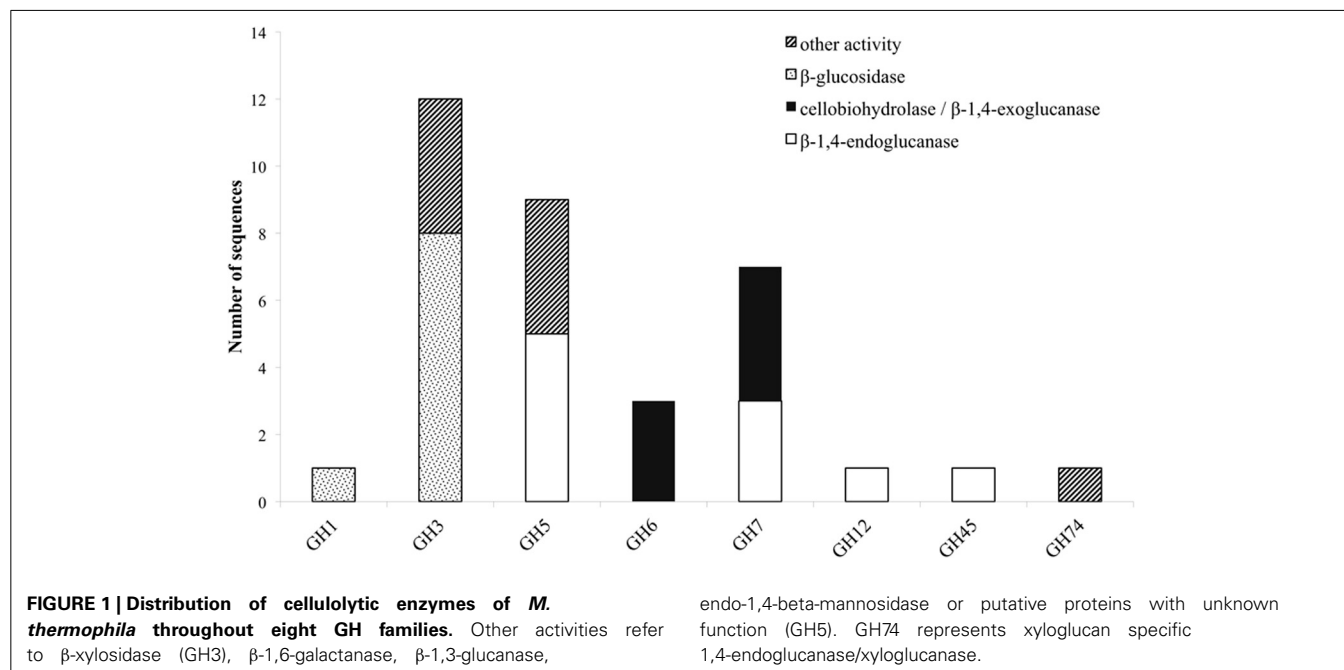


Table 2 | Number of predicted sequences encoding enzymes with cellulolytic activity (EGs, CBHs, and BGLs).

Protein_ ID	InterPro_ ID	InterPro_ description	CAZy module	GenBank ID	Location	Secretion signal	Exons	pI/MW (kDa)	Length	N-gly positions	O-gly positions	CBM1	BLAST parameters
EGs													
MYCTH_ 52068	IPR017853	Glycoside hydrolase, catalytic core	GH5	AE058455.1	4:188041–189524	18aa	7	4.66/35.2	319aa	3	2	–	
MYCTH_ 86753	IPR017853	Glycoside hydrolase, catalytic core	GH5	AE053769.1	1:2823610– 2825549	16aa	3	5.07/40.85	389aa	3	17	+	65% identity with endoglucanase GH5 from <i>Thermoascus aurantiacus</i> [PDB ID: 1GZJ] and 79% with β-1,4-endoglucanase from <i>Penicillium brasilianum</i> [GenBank: ACB06750]
MYCTH_ 94336	IPR017853	Glycoside hydrolase, catalytic core	GH5	AE057401.1	3:908597–909880	21aa	1	4.92/44.6	406aa	2	4	–	61% identity with cellulase family protein from <i>Ophiostoma piceae</i> UAMH 11346 [GenBank: EPE03278.1]
MYCTH_ 43356	IPR017853	Glycoside hydrolase, catalytic core	GH5	AE053175.1	1:580471–581733	26aa	1	7.86/44.18	394aa	1	2	–	71% identity with EG from endoglucanase <i>Verticillium dahliae</i> [GenBank: EGY21718.1]
MYCTH_ 116157	IPR017853	Glycoside hydrolase, catalytic core	GH7	AE059361.1	4:4081701– 4083768	20aa	2	4.75/47.20	436aa	2	2	–	68% identity with endo-1,4-beta-glucanase from <i>Humicola grisea</i> [UniProtKB/Swiss- Prot:Q12622.1]
MYCTH_ 111372	IPR017853	Glycoside hydrolase, catalytic core	GH7	AE058196.1	3:4135959– 4137568	22aa	1	4.61/46.632	442aa	2	16	+	64% identity with endoglucanase I from <i>Hypocrea orientalis</i> [GenBank: AFD50194.1]
MYCTH_ 109444	IPR008985	Concanavalin A-like lectin/glucanases superfamily	GH12	AE060532.1	6:212893–214302	15aa	3	5.46/25.48	247aa	–	–	–	56% identity with Cel12A from <i>Humicola grisea</i> [PDB ID: 1OLR] and 63% with endoglucanase from <i>Fusarium oxysporum</i> [GenBank: ENH72093.1]
MYCTH_ 76901	IPR014733	Barwin-like endoglu- canase	GH45	AE054078.1	1:4001048– 4002222	18aa	4	4.66/21.8	207aa	–	3	–	78% identity with endoglucanase from <i>Humicola insolens</i> [PDB ID: 1HD5] and 82% with endoglucanase from <i>Humicola grisea</i> [GenBank: BAA74957.1]

(Continued)

Table 2 | Continued

Protein_ ID	InterPro_ ID	InterPro_ description	CAZy module	GenBank ID	Location	Secretion signal	Exons	pI/MW (kDa)	Length	N-gly positions	O-gly positions	CBM1	BLAST parameters
<b>CBHs</b>													
MYCTH_ 51545	IPR016288	1,4-beta cellobiohy- drolase	GH6	AE059280.1	4:3726723– 3728113	17aa	3	4.66/40.64	378aa	–	4	–	49% identity with CBHII from <i>Trichoderma reesei</i> [GenBank: ADC83999.1] and 51% with CBHII from <i>Trichoderma reesei</i> [PDB ID: 3CBH]
MYCTH_ 66729	IPR016288	1,4-beta cellobiohy- drolase	GH6	AE055787.1	2:46305–48489	17aa	4	5.28/49.41	465aa	1	35	+	79% identity with CBH II from <i>Humicola insolens</i> [PDB ID: 1BVW] and 64% with CBHII from <i>Trichoderma viride</i> [GenBank: AAQ76094.1]
MYCTH_ 2303045	IPR016288	1,4-beta cellobiohy- drolase	GH6	AE057190.1	2:5389544– 5391132	18aa	2	5.94/39.41	363aa	4	5	–	56% identity with endoglucanase Cel6b from <i>Humicola insolens</i> [PDB ID: 1DYS] and 39% with Cellobiohydrolase II (Cel6a) from <i>Humicola insolens</i> [1BVW]
MYCTH_ 109566	IPR008985	Concanavalin A-like lectin/glucanases superfamily	GH7	AE055544.1	1:9753507– 9755507	17aa	2	4.77/54	509aa	1	20	+	61% identity with CBH I from <i>Humicola grisea</i> [GenBank: BAA09785.1] and 67% with cellobiohydrolase Cel7d (Cbh 58) from <i>Phanerochaete chrysosporium</i> [PDB: 1GPI]
MYCTH_ 42937	IPR008985	Concanavalin A-like lectin/glucanases superfamily	GH7	AE053522.1	1:1922814– 1924693	20aa	5	4.34/47.08	436aa	2	2	–	66% identity with β- 1,4-cellobiohydrolase from <i>Phanerochaete chrysosporium</i> [UniProtKB: P13860.1]
MYCTH_ 97137	IPR008985	Concanavalin A-like lectin/glucanases superfamily	GH7	AE061262.1	7:24661–26448	20aa	6	4.48/46.5	430aa	–	–	–	79% identity with β- 1,4-beta-cellobiosidase from <i>Melanocarpus albomyces</i> [GenBank: CAD56667.1]
MYCTH_ 95095 frag- ment	IPR008985	Concanavalin A-like lectin/glucanases superfamily	GH7	AE058824.1	4:1967106– 1967560	19aa	2	6.87/11.82	109aa	1	4	C- term, no pre- dic- tion	77% identity with β- 1,4-beta-cellobiosidase from <i>Acremonium thermophilum</i> [GenBank: CAM984446.1]
<b>BGLs</b>													
MYCTH_ 115968	IPR017853	Glycoside hydrolase, catalytic core	GH1	AE057459.1	3:1092854– 1095039	–	2	5.53/54.1	476aa	1	–	–	92% identity with β-glucosidase from <i>Humicola grisea</i> [GenBank: BAA74958.1]

(Continued)

Table 2 | Continued

Protein_ID	InterPro_ID	InterPro_description	CAZy module	GenBank ID	Location	Secretion signal	Exons	pI/MW (kDa)	Length	N-gly positions	O-gly positions	CBM1	BLAST parameters
MYCTH_38200	IPR017853/ IPR026891	Glycoside hydrolase, catalytic core	GH3	AE061246.1	6:4096305–4099311	–	2	5.54/105.3	968aa	7	3	–	78% identity with $\beta$ -glucosidase from <i>Chaetomium thermophilum</i> var. <i>thermophilum</i> [GenBank: EGS22574.1]
MYCTH_2059579	IPR017853/ IPR026891	Glycoside hydrolase, catalytic core/fibronectin type III-like domain	GH3	AE056238.1	2:1843191–1846185	23aa	4	5.94/94.8	880aa	6	5	–	57% identity with $\beta$ -1,4-glucosidase from <i>Marssonina brunnea</i> [GenBank: EKD11918.1]
MYCTH_66804	IPR017853/ IPR026891	Glycoside hydrolase, catalytic core/fibronectin type III-like domain	GH3	AE058343.1	3:4861135–4863642	17aa	2	4.99/75.83	716aa	2	2	–	72% identity with $\beta$ -glucosidase from <i>Trichoderma reesei</i> [PRF: 227874]
MYCTH_80304	IPR017853/ IPR026891	Glycoside hydrolase, catalytic core/fibronectin type III-like domain	GH3	AE058175.1	3:3949561–3952737	19aa	4	5.05/93.3	851aa	11	–	–	77% identity with $\beta$ -glucosidase from <i>Chaetomium thermophilum</i> [GenBank: ABR57325.2]
MYCTH_62925	IPR017853	Glycoside hydrolase, catalytic core	GH3	AE053892.1	1:3330354–3333832	–	7	5.39/9705	884aa	3	4	–	76% identity with $\beta$ -glucosidase from <i>Fusarium oxysporum</i> [GenBank: EMT71863.1]
MYCTH_2302509	IPR026891	Fibronectin type III-like domain	GH3	AE056946.1	2:4612453–4614052	–	1	5.15/47.75	440aa	6	2	–	60% identity with $\beta$ -glucosidase from <i>Grosmanina clavigera</i> [GenBank: EFX01027.1]
MYCTH_58882	IPR017853/ IPR026891	Glycoside hydrolase, catalytic core/fibronectin type III-like domain	GH3	AE060477.1	6:51326–53769	16aa	2	5.21/82.46	761aa	5	–	–	–
MYCTH_2129052 (fragment)	IPR001764	Glycoside hydrolase, family 3, N terminal	GH3	AE059952.1	5:2414101–2414414	–	2	4.70/92.48	83aa	1	–	–	46% identity with $\beta$ -glucosidase from <i>Parastagonospora avenae</i> [GenBank: CAB82861.1]

Most of them predicted to possess a secretion signal and several N- and O-glycosylation sites, thus undergoing through post-translational modifications.



sites. Only two of them exhibit a CBM that belong in family 1. CBHs represent three non-reducing acting enzymes of GH6 family and four reducing-end acting enzymes of GH7 family. All of these enzymes seem to be targeted to secretion pathway and modified with glycans during post-translational modifications. BGs are classified to GH3 family, except one GH1 sequence, while none of them exhibit a CBM, as expected. Four are secreted and have potential *N*- and *O*-glycosylation sites, showing the highest molecular weight compared to the other cellulases, with a theoretical average value of  $85.18 \pm 3.2$  kDa.

Totally, 12 cellulases have been isolated and characterized (Table 3). The group of Bukhtojarov et al. (2004) investigated the properties of individual cellulases from the multienzyme complex produced by a mutant strain of *M. thermophila* C1 (Visser et al., 2011). Among EGs, the highest saccharification activity was displayed by EG60 and EG51, representing enzymes of 60 and 51 kDa, respectively, which exhibited pI values of 3.6 and 5.0,

respectively. It has been shown later that the EG51 and EG60 represent the GH5 and GH7 EGs from *M. thermophila*, respectively, (Gusakov et al., 2011). A different EG (StCel5A) displays a typical GH5 domain, exhibiting optimal activity at pH 6.0 and 70°C and retained greater than 50% of its activity following 2 h of incubation at 55°C, diluted in 10 mM citrate buffer pH 4.5 (Tambor et al., 2012). A GH7 EG gene was functionally expressed in methylotrophic yeast *P. pastoris* and subsequently characterized (Karnaouri et al., 2014). Substrate specificity analysis revealed that the enzyme is one of the most thermostable fungal enzymes reported up to now and exhibits high activity on substrates containing  $\beta$ -1,4-glycosidic bonds as well as activity on xylan-containing substrates. Moreover, MtEG7a was proved to liquefy rapidly and efficiently pretreated wheat straw, indicating EGs' key role to the initial step of hydrolysis of high-solids lignocellulose substrates (Karnaouri et al., 2014). This change in viscosity of these substrates is probably due to the gradual reduction of

**Table 3 | Description of the characterized cellulolytic enzymes either isolated from the culture broth of a *M. thermophila* C1 mutant strain or expressed in a heterologous host.**

Enzyme	Type of action	GH	MW-monomer (kDa)	pH <sub>opt</sub>	T <sub>opt</sub> (°C)	pI	Gene	Source	References
StCel5A	Endoglucanase	5	46	6	70	ND		cDNA library-EST analysis, expressed in <i>A. niger</i>	Tambor et al., 2012
EG51	Endoglucanase	5	51	4.7	70	4.8		Isolated from the culture broth of a C1 mutant strain	Bukhtojarov et al., 2004; Gusakov et al., 2005
MtEG7a	Endoglucanase	7	65	5	60	Multiple bands in 3.8–4.5	<i>eg7a</i>	Expressed in <i>P. pastoris</i>	Karnaouri et al., 2014
EG60	Endoglucanase	7	60	4.7	60	3.7		Isolated from the culture broth of a C1 mutant strain	Bukhtojarov et al., 2004
EG28	Endoglucanase	12	28	5.35	60	5.7		Isolated from the culture broth of a C1 mutant strain	Bukhtojarov et al., 2004
EG25	Endoglucanase	45	25	5.5	65	4		Isolated from the culture broth of a C1 mutant strain	Bukhtojarov et al., 2004; Gusakov et al., 2005
CBH Ia	1,4-beta cellobiohydrolase	7	65	5	ND	4.5		Isolated from the culture broth of a C1 mutant strain	Bukhtojarov et al., 2004; Gusakov et al., 2005
CBH IIa	1,4-beta cellobiohydrolase	6	43	5.4	65	4.2		Isolated from the culture broth of a C1 mutant strain	Bukhtojarov et al., 2004; Gusakov et al., 2005
CBH Ib	1,4-beta cellobiohydrolase	7	60	ND	ND	ND		Isolated from the culture broth of a C1 mutant strain	Gusakov et al., 2007
CBH IIb	1,4-beta cellobiohydrolase	6	70	ND	ND	5.6		Isolated from the culture broth of a C1 mutant strain	Gusakov et al., 2007
MtBgl3a	beta-glucosidase	3	90	5	70	4.0	<i>bgl3a</i>	Expressed in <i>P. pastoris</i>	Karnaouri et al., 2013
Bxl5	beta-glucosidase	3	120 ± 5	4.6	75	5.2	<i>bxl5</i>	Homologously expressed in C1	Dotsenko et al., 2012

Most of them exhibit optimum temperatures above 60°C and pH around 5.0.

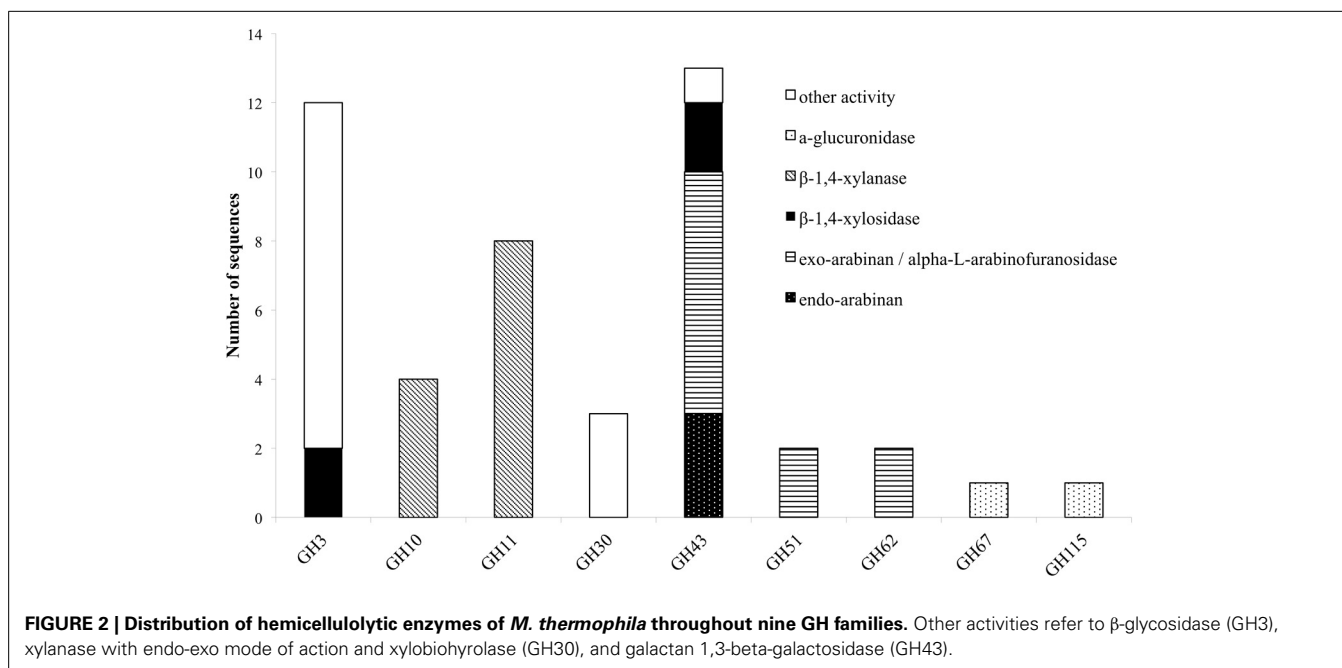
the average chain length of cellulose polysaccharides by endo-acting enzymes, such as endoglucanases. Totally, four CBHs and two BGs have been isolated from *M. thermophila* crude supernatant and studied. CBH IIB is the product of MYCTH\_66729 gene that represents an enzyme of GH6 family, which is attached to polysaccharide substrate through a CBM and exhibits high levels of activity in comparison to other CBHs (Gusakov et al., 2007). In the same study, the isolation of CBH Ib, a GH7 family enzyme (MYCTH\_2140736) is reported, which acts mainly against microcrystalline cellulose and CMC. Bukhtojarov et al. (2004) studied the properties of CBH Ia and CBH IIA, which are classified to GH7 and GH6 family, respectively. CBH Ia is the product of MYCTH\_109566 gene, and seems to be expressed in two isoforms with distinct molecular weights, one exhibiting the catalytic domain owing a CMB and the other only the catalytic domain and part of the linker, after proteolysis. This enzyme is produced as a major protein of fungi's secretome (20–25% of the total extracellular protein) and adsorbed strongly on microcrystalline cellulose. It has been shown that there is a significant synergism between CBH IIB and CBH Ia enzymes during substrate hydrolysis (Gusakov et al., 2007).

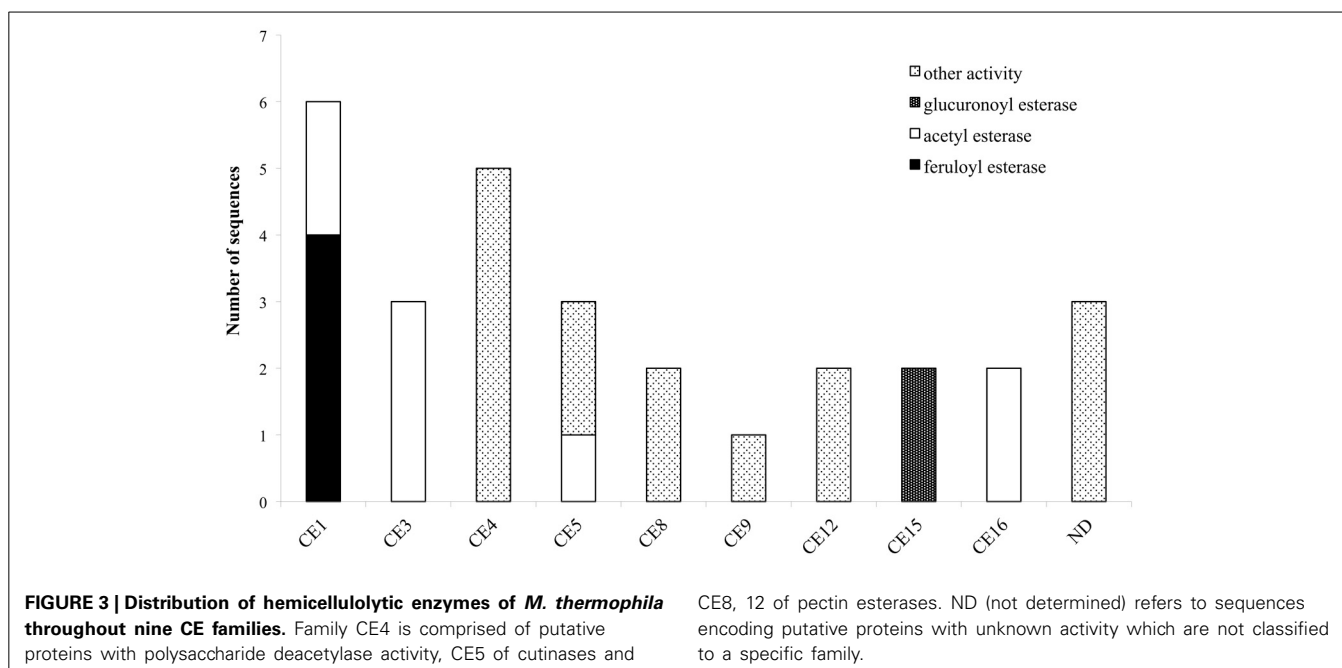
Recently, another type of specific activity was revealed. Xyloglucan specific exo- $\beta$ -1,4-glucanase (Xgl74A; EC 3.2.1.155) is classified to GH74 family and catalyzes the hydrolysis of (1-4)-D-glucosidic linkages in xyloglucans aiming in the successful removal of oligosaccharides from the chain end (Grishutin et al., 2004). Xyloglucan is a major structural polysaccharide found in the primary cell walls of higher plants that interact with cellulose microfibrils via hydrogen bonds to form a structural network that is assumed to play a key role in cell wall integrity. It consists of a cellulose-like backbone of  $\beta$ -1,4-linked D-glucopyranose (D-Glcp) residues, which most of them are substituted at C-6 with  $\alpha$ -D-Xylp-(1 $\rightarrow$ 6) residues, to which other saccharides may be attached (most frequently, d-Galp and

l-Fucp). *M. thermophila* was found to produce an exo- $\beta$ -1,4-glucanase (Xgl74A) (Grishutin et al., 2004) with high specific activity toward tamarind xyloglucan, and very low or absent activity against carboxymethylcellulose (CMC) and barley  $\beta$ -glucan. Due to its unique substrate specificity the enzyme was given a new number in the Enzyme Nomenclature (EC 3.2.1.155). Apart from Xgl74A, two out of the seven cellulases reported from *M. thermophila* (Cel12A and Cel45A) possess a notable activity against xyloglucan, together with their major activities toward CMC and barley  $\beta$ -glucan (Bukhtojarov et al., 2004).

## HEMICELLULOLYTIC SYSTEM

Hemicellulose polymers have a much more diverse structure than cellulose and consequently several enzymes are needed to completely degrade the polysaccharide into monosaccharides. Xylan that is the major component of hemicellulose in the plant cell wall, is consisted of a  $\beta$ -D-(1,4)-linked xylopyranosyl backbone, which, depending on the origin, can be substituted with arabinofuranosyl, 4-O-methylglucopyranosyl, feruloyl and acetyl groups (Shibuya and Iwasaki, 1985). Feruloyl groups can form strong networks through peroxidase-catalyzed oxidative coupling forming diferuloyl bridges (Topakas et al., 2007). The main enzymes needed for depolymerization are xylanases, assisted by accessory enzymes such as  $\beta$ -xylosidases and different arabinofuranosidases making the xylan backbone more accessible (Sørensen et al., 2007). Other accessible enzymes that enhance xylan degradation are acetyl-xylanesterases (Poutanen et al., 1990), ferulic acid esterases (Topakas et al., 2007), and  $\alpha$ -glucuronidases (De Vries et al., 1998). *M. thermophila*'s hemicellulase genes are organized in 10 GH families (3, 10, 11, 30, 43, 51, 62, and 67) (Figure 2) and nine carbohydrate esterase (CE) families (1, 3, 4, 5, 8, 9, 12, 15, and 16) (Figure 3). Many of the encoding proteins have been isolated from the WT culture supernatant or expressed in heterologous hosts and finally characterized in terms of specific





activity and physicochemical properties. The majority of them are predicted to follow the secretion pathway, while modified with *N*- and/or *O*- glucans, comprising a total amount of 66 enzymes that act synergistically for the degradation of hemicellulose.

#### XYLANASES/XYLOSIDASES

The degradation of xylan requires the concerted action of a number of powerful enzymes with varying specific activities, including xylanases and  $\beta$ -xylosidases. Xylanases (endo-1,4- $\beta$ -xylanases, EC 3.2.1.8) are enzymes hydrolyzing  $\beta$ -1,4-glycosidic linkages in the backbone of xylans, while most of them belong to GH family 10 or 11 based on amino acid similarities and structural features (Henrissat, 1991). GH10 xylanases exhibit less substrate specificity than GH11 enzymes and can hydrolyze different types of decorated xylans, while GH11 xylanases are highly specific and do not tolerate many decorations on the xylan backbone (Biely et al., 1997).  $\beta$ -Xylosidases (EC 3.2.1.37) hydrolyze the soluble xylo-oligosaccharides and xylobiose from the non-reducing end liberating xylose, produced by the activity of xylanases. These enzymes play an important role in xylan degradation by relieving the end product inhibition of endoxylanases (Knob et al., 2010). The genome of *M. thermophila* encodes totally 12 xylanases with endo- mode of activity, classified to GH 10 and 11 and four xylosidases, classified to GH3 and 43 families (Table 4). All aforementioned xylan-degrading translated sequences, apart from three, are predicted to exhibit a potential secretion signal. Xylanases possess 1-3 *N*-glycosylation and several *O*-glycosylation sites, whereas more *N*-sites are predicted for xylosidases, though not all of them are glycosylated during post-translational modifications. GH family 30 contains two genes encoding xylanolytic enzymes with endo-exo activity and one sequence for a characterized xylobiohydrolase, releasing xylobiose units from the substrate (Emalfarb et al., 2012).

Ten xylanases have been purified and characterized from multienzyme preparations of *M. thermophila* modified strains (Ustinov et al., 2008; van Gool et al., 2013). Four of them, belonging to GH10 family (Table 5), are the products of two genes, either with the presence of a family 1 CBM or displaying only the catalytic domain after partial proteolytic digestion (Ustinov et al., 2008). These enzymes, thought classified to the same family, can hydrolyze different types of decorated xylans. They differ in degradation of high and low substituted substrates and the substitution pattern seems to be an important factor influencing their efficiency (van Gool et al., 2012). Six xylanases, belonging to GH11 family, represent true xylanases, with high specific activities against glucuronoxylans and arabinoxylans. Four of these enzymes exhibit lower thermostability in comparison to GH10 xylanases, in which extended glycosylation has been noticed (Ustinov et al., 2008). One showed a substrate specificity pattern similar to GH10 enzymes and secreted in two forms, with or without CBM (van Gool et al., 2013).

#### ARABINOHYDROLASES

L-arabinose is widely present in various hemicellulosic biomass components, such as arabinoxylan, where the main  $\beta$ -D-(1,4)-linked xylopyranosyl backbone is substituted with arabinose residues.  $\alpha$ -L-arabinofuranosidases (AFase; EC 3.2.1.55) are enzymes that release arabinofuranose residues substituted at position *O*-2 or *O*-3 of mono or di-substituted xylose residues (Gruppen et al., 1993). Apart from that, AFases act in synergism with other arabinohydrolases, endo-(1,5)- $\alpha$ -L-arabinanases (ABNase; EC 3.2.1.99) for the decomposition of arabinan, a major pectin polysaccharide. Arabinan consists of a backbone of  $\alpha$ -(1,5)-linked L-arabinofuranosyl residues, some of which are substituted with  $\alpha$ -(1,2)- or  $\alpha$ -(1,3)-linked arabinofuranosides (Weinstein and Albersheim, 1979). Degradation of arabinan polymer to arabinose sugars is driven by the synergistic action of

Table 4 | Number of predicted sequences encoding enzymes with hemicellulolytic activity (β-1,4-xylanases and β-xylosidases).

Protein_ ID	InterPro_ ID	InterPro_ description	CAZy module	GenBank ID	Location	Secretion signal	Exons	pI/MW (kDa)	Length	N-gly positions	O-gly positions	Conserved domains	BLAST parameters
XYLs													
MYCTH_2139438	IPR013781	Glycoside hydrolase, catalytic domain	GH10	AE058598.1	4:852357–853828	19aa	3	6.42/39.65	356aa	1	–		84% identity with XynA from <i>Humicola insolens</i> (GenBank: AGG68962.1)
MYCTH_52904	IPR013781	Glycoside hydrolase, catalytic domain	GH10	AE059320.1	4:3932271–3933431	16aa	3	5.70/34.20	311aa	1	1		77% identity with endo-1,4-beta-xylanase protein from <i>Chaetomium thermophilum</i> [GenBank: EGS20178.1]
MYCTH_2125938 (fragment)	IPR013781	Glycoside hydrolase, catalytic domain	GH10	AE056947.1	2:4614086–4614433	ND	1	10.11/13.53	115aa	1	1		63% identity with XynA from <i>Humicola insolens</i> (GenBank: AGG68962.1)
MYCTH_112050	IPR013781	Glycoside hydrolase, catalytic domain	GH10	AE060457.1	5:4306110–4307709	17aa	2	5.52/42.86	396aa	–	22	CBM1	66% identity with endo-1,4-beta-xylanase A from <i>Gaeumannomyces graminis</i> [GenBank: EJT78185.1]
MYCTH_100068	IPR008985	Concanavalin A-like lectin/glucanases superfamily	GH11	AE057157.1	2:5285193–5266157	19aa	2	7.77/27.53	259aa	1	1	CBM1	74 % identity with endo-beta1,4-xylanase from <i>Chaetomium gracile</i> [GenBank: BAA08650.1]
MYCTH_89603	IPR008985	Concanavalin A-like lectin/glucanases superfamily	GH11	AE055365.1	1:9130476–9131260	20aa	2	6.71/23.12	208aa	1	1		87% identity with endo-beta-1,4-xylanase from <i>Chaetomium</i> sp. CQ31 [GenBank: ADW78258.1]
MYCTH_2121801	IPR008985	Concanavalin A-like lectin/glucanases superfamily	GH11	AE062054.1	7:3382576–3383357	18aa	2	7.81/22.59	200aa	1	2		86% identity with endo-1,4-beta-xylanase protein from <i>Chaetomium thermophilum</i> [GenBank: EGS23409.1]
MYCTH_99786	IPR008985	Concanavalin A-like lectin/glucanases superfamily	GH11	AE054512.1	1:5821400–5822224	21aa	2	5.50/22.17	205aa	1	2		56% identity from xylanase II from <i>Hypocrea orientalis</i> [GenBank: AFD50199.1]
MYCTH_2309574 (fragment)	IPR008985	Concanavalin A-like lectin/glucanases superfamily	GH11	AE060385.1	5:4039395–4039724	ND	1	9.19/7.10	66aa	1	–		63% identity with xyn11C from <i>Chaetomium thermophilum</i> [GenBank: CAD48751.1]

(Continued)



Table 4 | Continued

Protein_ ID	InterPro_ ID	InterPro_ description	CAZy module	GenBank ID	Location	Secretion signal	Exons	pI/MW (kDa)	Length	N-gly positions	O-gly positions	Conserved domains	BLAST parameters
MYCTH_ 2309575 (frag- ment)	IPR008985	Concanavalin A-like lectin/glucanases superfamily	GH11	AEO60386.1	5:4039732– 4040297	–	1	5.49/ 16.34	146aa	2	4		77% identity with endoxylanase 11C from <i>Chaetomium thermophilum</i> [GenBank: CAD48751.1]
MYCTH_ 56237	IPR008985	Concanavalin A-like lectin/glucanases superfamily	GH11	AEO59539.1	5:215065–215872	18aa	2	4.83/ 22.08	205aa	1	2		72% identity with xylanase from <i>Scytalidium thermophilum</i> [GenBank: BAD07040.1]
MYCTH_ 49824	IPR008985	Concanavalin A-like lectin/glucanases superfamily	GH11	AEO58284.1	3:4545212– 4546113	16aa	2	5.93/ 23.79	214aa	3	1		79% identity with endo-1,4-beta-xylanase from <i>Chaetomium cupreum</i> [GenBank: ABI48363.1]
<b>BXYLs</b>													
MYCTH_ 104628	IPR017853	Glycoside hydrolase, catalytic core	GH3	AEO60531.1	6:209038–211675	–	2	4.69/89.69	835aa	6	1	BglX	55% identity with 1,4-beta-xylosidase from <i>Fusarium fujikuroi</i> IMI 58289 [GenBank: CCT61496.1]
MYCTH_ 50705	IPR017853/ IPR026891	Glycoside hydrolase, catalytic core/fibronectin type III-like domain	GH3	AEO58346.1	3:4871063– 4873395	21aa	2	5.57/79.51	739aa	6	2		53% identity with 1,4-beta-xylosidase from <i>Neurospora crassa</i> [GenBank: CAB91343.2]
MYCTH_ 80104	IPR008985	Concanavalin A-like lectin/glucanases superfamily	GH43	AEO58399.1	3:5027172– 5028785	–	1	5.86/ 61.23	537aa	1	–	XynB	80% identity with beta-xylosidase from <i>Fusarium</i> <i>oxysporum</i> [GenBank: EMIT73385.1]
MYCTH_ 2072383	IPR008985	Concanavalin A-like lectin/glucanases superfamily	GH43	AEO61672.1	7:1863035– 1864780	19aa	2	4.77/53.52	494aa	4	3	XynB	45% identity with beta-xylosidase from <i>Micromonospora</i> sp. ATCC39149 [NCBI Ref: WP_ 007073468.1]

*β*-1,4-xylanases are classified to GH10 and GH11 families, while *β*-xylosidases to GH43 family. Most of them predicted to possess a secretion signal and several N- and O-glycosylation sites.

**Table 5 | Description of the characterized  $\beta$ -1,4-xylanases isolated from the culture broth of a *M. thermophila* C1 mutant strain.**

Enzyme	CAZy module	MW-monomer (kDa)	pH <sub>opt</sub>	T <sub>opt</sub> (°C)	pI	Gene	References
Xyn10A	GH10	42/31*	5.5–7.0	65–70	7.8/8.9	xy11	Ustinov et al., 2008; van Gool et al., 2012
Xyn10B	GH10	57/46*	5.5–7.0	80–85	4.4/4.3	xy13	Ustinov et al., 2008; van Gool et al., 2012
Xyn10C	GH11	40	5.0	80	4.8	xy14	Ustinov et al., 2008
Xyn11A	GH11	24	6.5	70	7.9	xy12	Ustinov et al., 2008
Xyn11B	GH11	23	6.0–6.5	65–70	8.4	xy16	Ustinov et al., 2008
Xyn11C	GH11	22	4.5	65	6.7	xy15	Ustinov et al., 2008
Xyl7	GH11	22/30*	5.5–6.5	50–60	7.3/7.6	xy17	van Gool et al., 2013
Xyl8	GH11	22	5.5–6.0	50–65	6.2	xy18	van Gool et al., 2013

They display optimal activity at temperatures between 50 and 70°C, increasing fungi's hydrolytic efficiency in various temperatures. Marked proteins (\*) were isolated in two different forms, with (high molecular weight enzyme) or without CBM (low molecular weight enzyme).

two major enzymes, AFases and ABNases (Kim, 2008). AFases specifically catalyze the hydrolysis of terminal non-reducing L-arabinofuranosyl residues from arabinan, while the resulting debranched backbone could be efficiently hydrolyzed by endo-acting ABNases, thus generating a variety of arabinooligosaccharides with an inverting mode of action (Beldman et al., 1997). Throughout CAZy families, arabinohydrolases belong to the GH family 43, 51, 54, 62, and 93 (Figure 3).

The genome of the *M. thermophila* encodes 14 enzymes that putatively release arabinose or arabinose oligomers from arabinan (Hinz et al., 2009). Eleven sequences contain a secretion signal peptide and produced as extracellular or cell-bounded proteins, while almost all of them exhibit isoelectric point around 4.6–5.6 (Table 6). Seven of them have been selectively overexpressed homologously in *M. thermophila* C1 host and found to release arabinose from wheat arabinoxylan polymers and oligomers (Hinz et al., 2009). *M. thermophila* arabinofuranosidases are selective in releasing arabinose from either single or double substituted xylose residues in arabinoxylans. Eight enzymes, belonging to GH families 43, 51, 62, and 93 with different type of arabinolytic activity have been purified and characterized (Hinz et al., 2009; Kühnel et al., 2011; Pouvreau et al., 2011b) (Table 7).

Abn7 and Abf3 are GH43 and GH51 arabinases respectively, which were selectively produced in C1 host. Abn7 was found to hydrolyze arabinofuranosyl residues at position O-3 of double substituted xylosyl residues in arabinoxylan-derived oligosaccharides, while Abf3 released arabinose from position O-2 or O-3 of single substituted xyloses. When these enzymes were incubated together, in combination with a GH10 endoxylanase for the hydrolysis of arabinoxylans, they resulted in a synergistic increase in arabinose release from the substrate (Pouvreau et al., 2011b). In addition,  $\alpha$ -L-arabinohydrolases Abn1, Abn2, and Abn4 were overexpressed in C1 and the produced culture supernatant has been shown to produce neutral branched arabinooligosaccharides from sugar beet arabinan by enzymatic degradation. As found by sugar analysis, neutral arabinooligosaccharides contained an  $\alpha$ -(1,5)-linked backbone of l-arabinosyl residues and carried single substituted  $\alpha$ -(1,3)-linked l-arabinosyl residues or consisted of a double substituted  $\alpha$ -(1,2,3,5)-linked arabinan structure within the molecule (Westphal et al., 2010). Enzyme Abn4 belongs to GH43 family and is more active toward branched polymeric arabinan substrate

that releases arabinose monomers from single substituted arabinose residues, while Abn1 and Abn2 are active toward linear arabinan (Kühnel et al., 2010). Abn2 is a member of GH93 family that consists of exoarabinases acting on linear arabinan, hydrolyzing the  $\alpha$ -1,5-linkages of arabinan polysaccharides presented as side chains of pectin. Their mode of action was studied with Abn2, which binds two arabinose units at the subsites –1 and –2 and releases arabinose. Three more arabinohydrolases were also overexpressed in C1 strain (Hinz et al., 2009). Abn5 was found to be specifically active toward arabinan, but not arabinoxylan. Arabinofuranosidases Abf1 and Abf2, members of GH62 family released O-2 or O-3 substituted arabinose or linked arabinofuranosyl from mono substituted xylose. GH family 62 arabinofuranosidases are reported to be predominantly active toward arabinoxylan and are, therefore, also called arabinoxylan arabinofuranohydrolases (Beldman et al., 1997). Several of these enzymes contain either a CBM1, like Abf1, or a CBM43 (xylan)-binding domain.

## ESTERASES

The role of esterases in the breakdown of lignocellulosic material is complex and includes the cleavage of bonds between the main hemicellulose part and many types of side chains. So, upon a closer examination of the genome sequences of *M. thermophila*, there is a wide distribution of enzymatic activities through CE families. These enzymes are classified into nine families and their main activities, among others, include the hydrolysis of feruloyl and acetyl ester bonds.

Feruloyl esterases (FAEs; EC 3.1.1.73) are enzymes responsible for cleaving the ester-link between the polysaccharide main chain of xylans and monomeric or dimeric ferulates. They act synergistically with xylanases to release ferulic acid from cell-wall material and can be divided into four groups, namely A–D. The main difference between groups A and D is their substrate specificity toward synthetic substrates and their capability of liberating diferuloyl bridges (Crepin et al., 2004). One of the first FAEs reported from thermophilic fungi, was produced from *M. thermophila* under solid-state fermentation (SSF) conditions. The esterase activity was isolated and partially characterized for its ability to release ferulic acid from complex substrate, destarched wheat bran (Topakas et al., 2003). Two other FAEs, StFaeB, a protein with molecular weight of 66 kDa

**Table 6 | Number of predicted sequences encoding enzymes with hemicellulolytic activity (endoarabinases and arabinofuranosidases).**

Protein_ID	InterPro_ID	InterPro_description	CAZy module	GenBank ID	Location	Secretion signal	Exons	Length	N-gly positions	O-gly positions	Conserved domains	BLAST parameters
MYCTH_39555	IPR006710	Glycoside hydrolase family 43	GH43	AE061077.1	6:3204367–3205350	–	1	4.87/37.12	327aa	–	1	82% identity with a-N-arabinofuranosidase/alpha-L-arabinofuranosidase from <i>Fusarium fujikuroi</i> [GenBank: CCT69715.1]
MYCTH_2303298	IPR006710	Glycoside hydrolase family 43	GH43	AE057303.1	3:514704–516772	23aa	1	4.91/59.16	535aa	1	1	55% identity with arabinase from <i>Auricularia delicata</i> [GenBank: EJD41599.1]
MYCTH_2305738	IPR006710	Glycoside hydrolase family 43	GH43	AE058423.1	4:27829–29545	19aa	4	4.90/33.18	301aa	1	2	52% identity with a-N-arabinofuranosidase 2 from <i>Fusarium oxysporum</i> [GenBank: EMT68952.1]
MYCTH_103032	IPR016840	Glycoside hydrolase family 43, endo-1,5-alpha-L-arabinosidase	GH43	AE058422.1	4:25889–27048	20aa	3	5.63/32.74	301aa	1	4	60% identity with arabinan endo-1,5-alpha-L-arabinosidase A from <i>Aspergillus kawachii</i> [GenBank: GAA90221.1]
MYCTH_2300677	IPR006710	Glycoside hydrolase family 43	GH43	AE056123.1	2:1449574–1451195	22aa	1	4.79/41.72	410aa	5	35	38% identity with endo-arabinase from <i>Colletotrichum gloeosporioides</i> [GenBank: ELA34066.1]
MYCTH_2064169	IPR006710	Glycoside hydrolase family 43	GH43	AE058916.1	4:2252824–2255361	16aa	11	6.42/64.17	573aa	7	5	65% identity with arabinofuranosidase from <i>Coprinopsis cinerea</i> [NCBI Ref: XP_001836293.2] and 52% with beta-xylosidase from <i>Aspergillus oryzae</i> [GenBank: EIT77121.1]
MYCTH_2301869	IPR006710	Glycoside hydrolase family 43	GH43	AE056692.1	2:3685597–3687704	18aa	1	4.89/65.71	613aa	3	1	55% identity with Xylosidase/arabinosidase (arabinofuranoside) from <i>Fusarium oxysporum</i> [GenBank: ENH67211.1]
MYCTH_2127683	IPR007934	a-L-arabinofuranosidase B	GH43	AE058631.1	4:1109767–1110916	–	3	4.91/36.35	339aa	1	3	65% with a-N-arabinofuranosidase from <i>Chaetomium</i> sp. [GenBank: AFU88757.1]

(Continued)

Table 6 | Continued

Protein_ID	InterPro_ID	InterPro_description	CAZy module	GenBank_ID	Location	Secretion signal	Exons	pI/MW (kDa)	Length	N-gly positions	O-gly positions	Conserved domains	BLAST parameters
MYCTH_230666	IPR023296	Glycosyl hydrolase, five-bladed-beta-propeller domain	GH43	AEO58919.1	4:2260967–2262690	25aa	4	5.26/35.76	329aa	1	1		72% identity with arabinosidase (arabinofuranosidase) from <i>Glomerella graminicola</i> [GenBank: EFQ32395.1]
MYCTH_42071	IPR010720	a-L-arabinofuranosidase, C-terminal	GH51	AEO53569.1	1:2109152–2111615	18aa	6	5.80/69.54	636aa	6	2		54% identity with a-L-arabinofuranosidase from <i>Penicillium chrysogenum</i> [GenBank: BAH70480.1]
MYCTH_83019	IPR010720	a-L-arabinofuranosidase, C-terminal	GH51	AEO58452.1	4:178906–181205	–	5	5.81/57.67	512aa	2	2		75% identity with a-L-arabinofuranosidase from <i>Glomerella graminicola</i> [GenBank: EFQ27215.1]
MYCTH_98003	IPR005193	Glycoside hydrolase, family 62, arabinosidase	GH62	AEO60934.1	6:2448019–2449251	22aa	2	5.52/37.84	353aa	–	3	CBM1	73% identity with a-L-arabinofuranosidase axhA-2 from <i>Colletotrichum higginsianum</i> [GenBank: CCF42219.1]
MYCTH_55982	IPR005193	Glycoside hydrolase, family 62, arabinosidase	GH62	AEO59813.1	5:1711636–1712601	19aa	1	4.90/33.1	302aa	1	4		77% identity with a-L-arabinofuranosidase from <i>Streptomyces ghanaensis</i> [NCBI Ref: WP_004979944.1]
MYCTH_104827	IPR011040	Sialidases	GH93	AEO55492.1	1:9532077–9534361	17aa	2	5.16/41.8	378aa	3			56% with exo-arabinanase from <i>Penicillium chrysogenum</i> [NCBI Ref: XP_002562032.1]

All of them are selectively produced in culture supernatant and contain a secretion signal peptide, being produced as extracellular or cell-bounded proteins.



**Table 7 | Description of the characterized endoarabinases and arabinofuranosidases isolated from the culture broth of a *M. thermophila* C1 mutant strain.**

Enzyme	CAZy module	Mode of action	MW-monomer (kDa)	pH <sub>opt</sub>	T <sub>opt</sub> (°C)	Uniprot/SwissProt ID	Gene	Source	References
Abf1	GH62	Arabinofuranosidase/ releases O-2 or O-3 arabinose from mono-substituted xylose					<i>abf1</i>	Selectively produced in C1-host	Hinz et al., 2009
Abf2	GH62	Arabinofuranosidase/ releases O-2 or O-3 linked arabinofuranosyl residues from mono-substituted xylose					<i>abf2</i>	Selectively produced in C1-host	Hinz et al., 2009
Abf3	GH51	Arabinofuranosidase/ releases arabinose from the non-reducing end of reduced arabinose oligomers	70	5	40	HQ324254	<i>abf3</i>	Selectively produced in C1-host	Pouvreau et al., 2011b
Abn7	GH43	Arabinofuranosidase/ releases O-3 linked arabinofuranosyl residues from di-substituted xylose	70	5	40	HQ324255	<i>abn7</i>	Selectively produced in C1-host	Pouvreau et al., 2011b
Abn1	GH43	Endoarabinase	36	5.5	60	HQ324251	<i>abn1</i>	Overexpressed in fermentation supernatant	Kühnel et al., 2011
Abn2	GH93	Exoarabinase/ arabinobiose from the non-reducing end of reduced arabinose oligomers	40	4	50		<i>abn2</i>	Overexpressed in fermentation supernatant	Kühnel et al., 2011
Abn4	GH43	Arabinofuranosidase/ releases arabinose from the non-reducing end of reduced arabinose oligomers	33	5.5	60	HQ324253	<i>abn4</i>	Overexpressed in fermentation supernatant	Kühnel et al., 2011
Abn5		Arabinofuranosidase					<i>abn5</i>	Selectively produced in C1-host	Hinz et al., 2009

(homodimers of 33 kDa) (Topakas et al., 2004) and StFaeC, 46 kDa (homodimers of 23 kDa) (Topakas et al., 2005), were purified to homogeneity from culture supernatants of *M. thermophila*. StFaeB hydrolyzed methyl *p*-coumarate, methyl caffeate and methyl ferulate and was active on substrates containing ferulic acid ester linked to the C-5 and C-2 linkages of arabinofuranose. StFaeC showed maximum catalytic efficiency on 4-hydroxy-3-methoxy cinnamate, a substrate with both hydroxyl

and methoxy substituents, indicating that it may be the most promising type of FAE as a biocatalyst for the enzymatic feruloylation of aliphatic alcohols, oligo- and polysaccharides. Properties of characterized FAEs are summarized in **Table 8**. Among the sequences registered to Genome Portal, there are four sequences encoding proteins with catalytic activity of FAE, all belonging to CE family 1. Two of them (MYCTH\_48379, MYCTH\_39279) seem to be identical with characterized FAEs secreted from *M.*

**Table 8 | Description of *M. thermophila* characterized esterases (FAEs, AcEs, and GEs).**

Enzyme	Type	MW-monomer (kDa)	pH <sub>opt</sub>	T <sub>opt</sub> (°C)	pI	Gene	Source	References
<b>FAEs</b>								
StFaeB	B	33*	6	55–60	3.5		Isolated from culture supernatant	Topakas et al., 2004
StFaeC	C	23*	6	55	<3.5		Isolated from culture supernatant	Topakas et al., 2005
Fae	ND	27	8	60	5		Isolated from culture supernatant	Topakas et al., 2003
FaeA1	A	29	6.5	45	≈5.5		Overexpressed in fermentation supernatant	Kühnel et al., 2012
FaeA2	A	36	7.5	40	≈5.2		Overexpressed in fermentation supernatant	Kühnel et al., 2012
FaeB2	B	33	7.5	45	≈6.0		Overexpressed in fermentation supernatant	Kühnel et al., 2012
MtFae1a	B	39	7	50	ND	<i>fae1a</i>	Expressed in <i>P. pastoris</i>	Topakas et al., 2012
<b>AcEs</b>								
MtAxe3	1	33.6	7	40	ND	<i>axe3</i>	Overexpressed in fermentation supernatant	Pouvreau et al., 2011a
MtAxe2	5	23.6	7	40	ND	<i>axe2</i>	Overexpressed in fermentation supernatant	Pouvreau et al., 2011a
<b>GEs</b>								
StGE2	15	43	7	55	ND	<i>ge2</i>	Expressed in <i>P. pastoris</i>	Topakas et al., 2010
StGE1	15	58	6	60	ND		Isolated from culture supernatant	Vafiadi et al., 2009

\*All proteins are monomeric, while in case of FaeB2, dimeric structures are detected, after comparing the results of SDS-PAGE and native electrophoresis.

*thermophila* C1 strain (Kühnel et al., 2012). One sequence (JDI ID: 96478) has been heterologously expressed in *P. pastoris* and encoded a 39 kDa protein (*fae1A*; MtFae1a), which showed high activity toward methyl caffeate and p-coumarate and a strong preference for the hydrolysis of n-butyl and iso-butyl ferulate (Topakas et al., 2012). In addition, MtFae1 esterase release ferulic acid from destarched wheat bran only by the synergistic action of an endo-xylanase (a maximum of 41% total ferulic acid released after 1 h incubation). MYCTH\_2302953 sequence has not yet been characterized, however it still shows 66% identity with a type B FAE from *Neurospora crassa* (CAC05587.1). All proteins encoded by the above sequences appear to be secreted and bring several N- and O-glycosylation sites, as shown in Table 9.

About 60–70% of the xylose residues in hardwood xylan are acetylated at the C2 and/or C3 positions (Lindberg et al., 1973). The complete degradation of acetylated xylans by microbes requires the action of acetyl esterases (AcEs; EC 3.1.1.72), which cleave acetyl side groups from the heteroxylan backbone, and act in synergy with other hemicellulases (Tenkanen et al., 1996). Eight sequences that encode proteins with AcE activity were detected in the genome of *M. thermophila* and showed identity with characterized enzymes. All of them are secreted, as predicted with SignalP and belong to CE families 1, 3, 5, 16 (Table 9). Two of them, Axe2 and Axe3, which bare members of CE5 and CE1 families, respectively, were isolated and characterized (Pouvreau et al., 2011a). Annotated genes, encoding the putative enzymes were cloned into the specially designed *M. thermophila* C1-expression host (Verdoes et al., 2010) and over-produced in the culture medium. Axe2 and Axe3 are able to hydrolyze acetyl groups when they are substituted to the O-2 and O-3 positions of acetylated xylo-oligosaccharides and complex insoluble polymeric substrates and had a preference for xylooligosaccharides (Pouvreau et al., 2011a).

Glucuronoyl esterases (GEs) are recently discovered enzymes that are suggested to play an important role in the dissociation of lignin from hemicellulose and cellulose by cleaving the ester bonds between the aromatic alcohols of lignin and the

carboxyl groups of 4-O-methyl-D-glucuronic acid residues in glucuronoxylan (Špáníková and Biely, 2006). Sequence alignment studies of these enzymes have revealed a novel conserved amino acid sequence G-C-S-R-X-G that features the characteristic serine residue involved in the mechanism of this esterase family. It has been shown that the mode of action probably involves a nucleophilic serine (Topakas et al., 2010). The genome of *M. thermophila* possesses two genes classified to family CE15 that encode proteins with activity of 4-O-methyl-glucuronoyl esterase. Both putative enzymes are secreted and have potential glycosylation sites. The first GE (StEG1), isolated from the culture filtrate of *M. thermophila*, was proved to be a thermophilic enzyme that presents a C-terminal CBM, which was active on substrates containing glucuronic acid methyl ester (Vafiadi et al., 2009). Another CE15 protein molecule, StGE2 was heterologously expressed in yeast *P. pastoris* and was used to prove that nucleophilic serine residue is responsible for catalytic action of GEs, through site-directed mutagenesis studies (Topakas et al., 2010) and crystal structure determination (Charavgi et al., 2013).

### MANNAN-DEGRADING ENZYMES

Mannan is a great component of hemicellulose, therefore, as expected, the lignocellulolytic toolbox of *M. thermophila* possesses a complete reservoir of genes encoding mannan degrading enzymes. Mannan polymer primarily consists of a backbone structure composed of β-1,4-bound mannose residues or combination of glucose and mannose residues and can be hydrolyzed to its monomers with the synergistic action of β-mannanases (EC 3.2.1.78), β-mannosidases (EC 3.2.1.25), α-galactosidases (EC 3.2.1.22), and acetylmannan esterases (E.C. 3.1.1.6) (McCleary, 1988). The genome of *M. thermophila* encodes three enzymes that putatively catalyze random cleavage of the mannan polysaccharide and belong to GH family 5 and 26. One of these enzymes has been isolated from culture supernatant, characterized and classified as GH5 endo-β-1,4-mannosidase (bMan2, Dotsenko et al., 2012). In addition, there are two genes encoding putative β-mannosidases belonging to GH2 family, while one of them has been characterized in terms of its specificity and physicochemical

Table 9 | Number of predicted sequences encoding enzymes with hemicellulolytic activity capable of hydrolyzing ester bonds (FAEs, AcEs and GEs).

Protein_ID	InterPro_ID	InterPro_description	CAZy module	GenBank ID	Location	Secretion signal	Exons	pI/MW (kDa)	Length	N-gly positions	O-gly positions
<b>FAEs</b>											
MYCTH_96478	IPR010126	Esterase, PHB depolymerase	CE1	AE062008.1	7:3150678–3151912	18aa	3	4.48/29.43	272aa	2	–
MYCTH_48379	SSF53474	Alpha/beta hydrolases	CE1	AE057132.1	2:5185595–5186503	26aa	1	5.05/29.11	276aa	1	8
MYCTH_39279	SSF53474	Alpha/beta hydrolases	CE1	AE061094.1	6:3261246–3262256	20aa	2	6.15/27.22	295aa	–	2
MYCTH_2302953	IPR010126	Esterase, PHB depolymerase	CE1	AE057203.1	2:5187587–5189039	19aa	3	5.82/29.4	276aa	1	–
<b>AcEs</b>											
MYCTH_80233	IPR010126	Esterase, PHB depolymerase	CE1	AE058247.1	3:4383828–4384913	21aa	2	5.93/31.8	294aa	2	–
MYCTH_2130973	SSF53474	Alpha/beta-Hydrolases	CE1	AE061799.1	7:2287063–2288403	23aa	5	5.03/24.7	233aa	3	–
MYCTH_49700	IPR000675	Cutinase	CE5	AE058391.1	3:5004208–5005095	17aa	4	5.01/23.4	226aa	1	–
MYCTH_43727	IPR001087	Lipase, GDSL	CE3	AE054377.1	1:5184409–5185412	18aa	3	4.72/23.2	214aa	–	1
MYCTH_53698	IPR001087	Lipase, GDSL	CE3	AE059370.1	4:4105354–4106226	18aa	3	5.09/23.7	217aa	1	1
MYCTH_105128	IPR001087	Lipase, GDSL	CE3	AE060969.1	6:2683598–2685715	21aa	2	9.42/29.6	279aa	–	2
MYCTH_40885	IPR001087/ IPR013830	Lipase, GDSL/Esterase, SGNH hydrolase-type	CE16	AE060664.1	6:1191196–1192286	15aa	2	6.40/38.1	343aa	3	–
MYCTH_84133	IPR001087/ IPR013830	Lipase, GDSL/Esterase, SGNH hydrolase-type	CE16	AE059602.1	5:700951–702131	21aa	2	4.41/32.5	302aa	4	–
<b>GEs</b>											
MYCTH_2308381	SSF53474	Alpha/beta-Hydrolases	CE15	AE060464.1	5:1737542–1739553	18aa	1	5.84/41.7	397aa	–	2
MYCTH_2119719	SSF53474	Alpha/beta-Hydrolases	CE15	AE060465.1	5:87107–88569	17aa	3	6.44/44.3	417aa	1	1

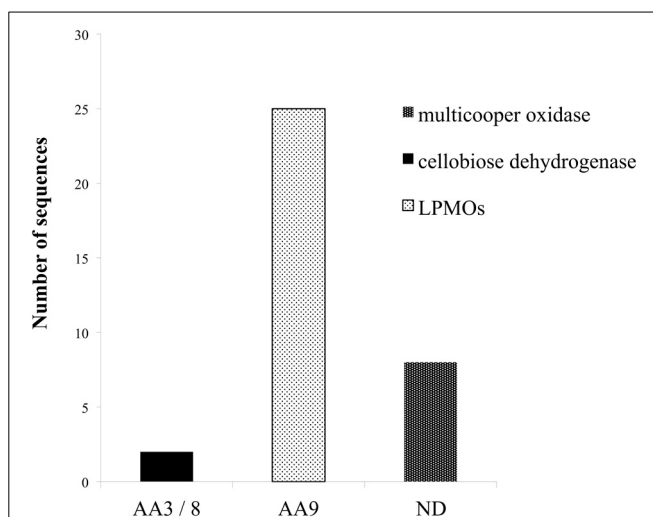
All of them contain a secretion signal peptide and have a theoretical molecular weight of  $28.51 \pm 4.1$  kDa, ranging between 23 and 39 kDa.

properties (bMann9, Dotsenko et al., 2012). Two GH27 and one GH26  $\alpha$ -galactosidases boost the efficiency of fungal culture supernatant against hydrolysis of mannan substrate (Emalfarb et al., 2012), while two CE12 family genes encoding proteins with high similarity to known acetyl-mannan esterases have been found.

## AUXILIARY ENZYMES

In spite of the cooperative activity exhibited by the cellulolytic and hemicellulolytic enzymes, the impressive hydrolytic ability of various microorganisms in nature cannot be attributed only to this endo–exo mechanism. Apart from the hydrolytic system responsible for carbohydrate degradation, it seems that an oxidative system catalyze lignin depolymerization and oxidation of plant cell wall components, yielding reactive molecules (e.g.,  $H_2O_2$ ). Recent evidence highlights the critical role of alternative enzymatic partners involved in the oxidation of cell wall components. Among these enzymes, outstanding role during hydrolysis exhibit the originally described as cellulases LPMO enzymes, CDH and multicopper enzymes such as laccases. The genome of *M. thermophila* possesses more than 30 genes that encode proteins with such auxiliary activities (Figure 4). Members of the LPMO family AA9, have been shown to be copper-dependent monooxygenases that enhance cellulose degradation in concert with classical cellulases, as aforementioned before and reviewed by Dimarogona et al. (2013). These enzymes catalyze the cleavage of cellulose by an oxidative mechanism provided that reduction equivalents are available. These equivalents either involve low molecular weight reducing agents (e.g., ascorbate) or are produced by CDH activity (Langston et al., 2011). CDHs are extracellular enzymes produced by various wood-degrading fungi that oxidize soluble celldextrins, mannodextrins and lactose efficiently to their corresponding lactones by a ping-pong mechanism using a wide spectrum of electron acceptors (Henriksson et al., 2000). Throughout the genome of *M. thermophila*, two genes encoding proteins classified to AA3 and 8 families have been identified (Figure 4). Both of them are predicted to be secreted in the culture supernatant and have potential glycosylation sites. The translated CDH MYCTH\_111388 exhibits a C-terminal CBM and a cDNA clone of this sequence has been isolated and biochemically characterized by screening an expression library of *M. thermophila* (Subramaniam et al., 1999). Canevascini et al. (1991) purified a monomeric (91 kDa) and a dimeric (192 kDa) form of CDH that differed not only in molecular weight, but amino acid composition and carbohydrate content. Both forms oxidized cellobiose in the presence of cytochrome c or dichlorophenol–indophenol.

Laccases (EC 1.10.3.1) are multicopper enzymes that catalyze the oxidation of a variety of phenolic compounds, with concomitant reduction of  $O_2$  to  $H_2O$ . These polyphenol oxidases are produced by most ligninolytic basidiomycetes (Baldrian, 2006) and can degrade lignin and other recalcitrant compounds in the presence of redox mediators (Ruiz-Dueñas and Martínez, 2009). The genome of the *M. thermophila* encodes eight putative enzymes with multicopper oxidase activity. Four of them have been annotated and one (MYCTH\_51627) matches the *lcc1* gene product encoding an extracellular laccase (Berka et al., 1997).



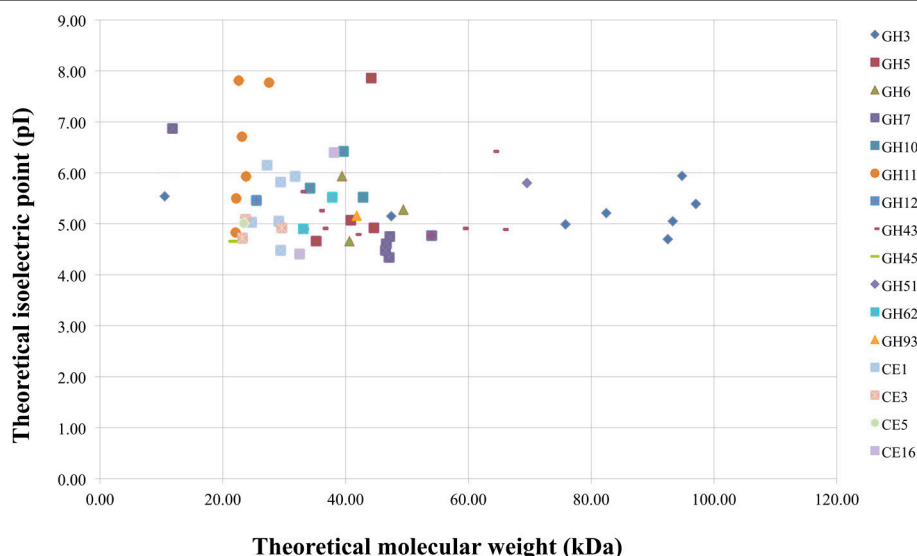
**FIGURE 4 | Distribution of enzymes of *M. thermophila* with auxiliary activities, classified to AA3/8, AA9 families and multicopper oxidases.** *M. thermophila* distinguishes itself from other cellulolytic fungi, exhibiting an impressive number of LPMOs accessory enzymes belonging to AA9 family (previously described as GH61).

Four sequences are predicted to possess a secretion signal, while one appears to remain membrane-bound. *Lcc1* gene has been isolated from fungi's genome, heterologously expressed in *A. oryzae* and the produced 85 kDa enzyme (MtL) was characterized as a thermostable low oxidation potential laccase with high reactivity in aqueous medium at room temperature and neutral pH. MtL was tested for its capacity to catalyze enzymatic oxidation of several phenolic and polyphenolic compounds (ferulic acid, gallic acid, caffeic acid, and catechin) (Mustafa et al., 2005). *M. thermophila* laccases have been reported to oxidize lignin surface, by increasing the amount of radicals during thermomechanical pulp fiber material bleaching (Grönqvist et al., 2003) and promote oxidative polymerization of Kraft lignin from black liquor, which is the main by-product of pulp and paper industry (Gouveia et al., 2013).

## LIGNOCELLULOSIC POTENTIAL—STATISTICS

*M. thermophila* is a powerful lignocellulolytic organism, which secretes a complex system of carbohydrate hydrolases for the breakdown of cellulose and hemicellulose, as well as oxidoreductases embedded in lignin degradation. Genome analysis in this review revealed 30 genes encoding cellulases classified to 10 GH families, 66 genes encoding hemicellulases classified to 10 GHs, 9 CEs and 35 genes encoding auxiliary enzymes. The latter include CDHs (AA3/AA8 family), LPMOs (AA9 family) and multicopper oxidases (laccases or laccase-like enzymes). Out of the total consortium of *M. thermophila* sequences encoding proteins with putative lignocellulosic activity, 80.2% are predicted to have a secretion signal peptide. Almost 76% of cellulases, hemicellulases and 88% of the accessory redox enzymes are targeted to secretion pathway, while only a very small amount remain inside the cell or represent membrane cell-bound macromolecules. Only 15.8% of the secreted enzymes in this





**FIGURE 5 | Theoretical molecular weight of secreted enzymes of *M. thermophila* classified in several GHs and CEs families, plotted against theoretical pI. The average molecular weight was calculated at  $51.05 \pm$**

**16.2 kDa** (range between 21 and 97 kDa) for cellulolytic enzymes,  $35.5 \pm 19.5$  kDa (range between 22 and 89 kDa) for hemicellulases (GHs/CEs), and  $28.51 \pm 4.1$  kDa (range between 23 and 39 kDa) for the fraction of esterases.

review are predicted to possess a CBM and the majority of them comprise of auxiliary enzyme activities. The theoretical average molecular weight of secreted enzymes is  $41.36 \pm 15.9$ , varying between 10 and 97 kDa. The majority of secreted enzymes have molecular weight varying between 20 and 50 kDa, whereas  $\beta$ -xylosidases and  $\beta$ -glucosidases (GH3 family), and arabinofuranosidases (GH43 and GH51) appear to be high molecular weight proteins (Figure 5). The theoretical average isoelectric point of secretory enzymes is calculated  $5.27 \pm 0.8$ , at a range 4.34–7.9. *In vivo* expression and study of these enzymes would give different results, as the proteins are glycosylated, so size and pI value tend to moderate.

### PROTEIN GLYCOSYLATION

A total proportion of 92.8% of secreted proteins have either *N*- or *O*- putative glycosylation sites. These proteins are often glycosylated due to the existence of many Asn-Xaa-Ser/Thr sequons, which are known to be a prerequisite for *N*-glycosylation post-translational modifications. The molecules of many GHs and accessory enzymes have a modular structure consisting of a catalytic module, flexible peptide linker, and CBM. Flexible linker peptides, which are rich in Ser and Thr residues, are typically *O*-glycosylated (Gilkes et al., 1991). The *N*-glycosylation seems to be restricted to the catalytic modules, and it is usually absent in other parts of enzyme molecules. Various *N*-linked glycan structures have been found in different enzymes from *M. thermophila*, belonging to different enzyme classes and protein families (Gusakov et al., 2008). It has been noticed that glycosylation follows a heterogeneity pattern, meaning that in some molecules, the same Asn residue was modified with oligosaccharides having different structure, while not all of the potential glycosylation sites were found to be occupied. The most frequently met *N*-linked glycan was (Man)<sub>3</sub>(GlcNAc)<sub>2</sub>, a pentasaccharide which

represents a well-known conserved core structure that forms mammalian-type high-mannose and hybrid/complex glycans in glycoproteins from different organisms (Dwek et al., 1993). Both types of glycosylation occur in 65% of secreted cellulases, 62.1% of secreted hemicellulases, while only *O*-glycosylation patterns appear in most of accessory enzymes. The presence of *N*-linked glycans is common for catalytic domain of the enzymes, while *O*-glycosylation usually occurs in linker region. Even though predicted to, non-secreted enzymes are not modified *in vivo* with glycans, since this procedure has been noticed as a post-translational modification in proteins targeted to the secretory pathway of the cell (Blom et al., 2004).

### CONCLUSIONS

Rapid depolymerization of lignocellulosic material is a distinguishing feature of thermophilic fungi, such as *M. thermophila*, which was isolated from soil and self-heating masses of composted vegetable matter (Domsch et al., 1993). However, the precise biochemical mechanisms and underlying genetics of this procedure are not completely understood. Systematic examination of the *M. thermophila* genome revealed a unique enzymatic system comprising of an unusual repertoire of auxiliary enzymes, especially those classified to AA9 family, and provided insights into its extraordinary capacity for protein secretion. The current review constitutes, to the best of our knowledge, the first genomic analysis of the lignocellulolytic system of *M. thermophila*. The genomic data, along with the observed enzymatic activity of several isolated and characterized enzymes suggest that this fungus possesses a complete set of enzymes, including 30 cellulases, 66 hemicellulases, and 35 proteins with auxiliary auxiliary enzymes, covering the most of the recognized CAZy families. From its cellulases to its oxido-reductases and multicopper enzymes, *M. thermophila* gene complement

represents several avenues for further research and its diverse array of enzymatic capabilities will contribute to the study of lignocellulose degradation and the subsequent ethanol biofuel production.

## ACKNOWLEDGMENTS

Anthi Karnaouri thanks the State Scholarships Foundation (Greece) for a Grant. Paul Christakopoulos thanks Bio4Energy, a strategic research environment appointed by the Swedish government, for supporting this work.

## REFERENCES

- Baldrian, P. (2006). Fungal laccases—occurrence and properties. *FEMS Microbiol. Rev.* 30, 215–242. doi: 10.1111/j.1574-4976.2005.00010.x
- Beldman, G., Schols, H. A., Pitson, S. M., Searle-van Leeuwen, M. F., and Voragen, A. G. J. (1997). “Arabinans and arabinan degrading enzymes,” in *Advances in Macromolecular Carbohydrate Research*, Vol. 1, ed R. Sturgeon (Stamford, CT: JAI Press Inc), 1–64.
- Berka, R. M., Grigoriev, I. V., Otilar, R., Salamov, A., Grimwood, J., Reid, I., et al. (2011). Comparative genomic analysis of the thermophilic biomass-degrading fungi *Myceliophthora thermophila* and *Thielavia terrestris*. *Nat. Biotechnol.* 29, 922–927. doi: 10.1038/nbt.1976
- Berka, R. M., Schneider, P., Golightly, E. J., Brown, S. H., Madden, M., Brown, K. M., et al. (1997). Characterization of the gene encoding an extracellular laccase of *Myceliophthora thermophila* and analysis of the recombinant enzyme expressed in *Aspergillus oryzae*. *Appl. Environ. Microbiol.* 63, 3151–3157.
- Bhat, K. M., and Maheshwari, R. (1987). *Sporotrichum thermophile* growth, cellulose degradation and cellulase activity. *Appl. Environ. Microbiol.* 53, 2175–2182.
- Biely, P., Vrsanská, M., Tenkanen, M., and Kluepfel, D. (1997). Endo-b-1,4-xylanase families: differences in catalytic properties. *J. Biotechnol.* 57, 151–166. doi: 10.1016/S0168-1656(97)00096-5
- Blom, N., Sicheritz-Ponten, T., Gupta, R., Gammeltoft, S., and Brunak, S. (2004). Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence. *Proteomics* 4, 1633–1649. doi: 10.1002/pmic.200300771
- Bukhtojarov, F. E., Ustinov, B. B., Salanovich, T. N., Antonov, A. I., Gusakov, A. V., Okunev, O. N., et al. (2004). Cellulase complex of the fungus *Chrysosporium lucknowense*: isolation and characterization of endoglucanases and cellobiohydrolases. *Biochemistry (Mosc.)* 69, 542–551. doi: 10.1023/B:BIRY.0000029853.34093.13
- Canevascini, G., Borer, P., and Dreyer, J. L. (1991). Cellobiose dehydrogenases of *Sporotrichum (Chrysosporium) thermophile*. *Eur. J. Biochem.* 198, 43–52. doi: 10.1111/j.1432-1033.1991.tb15984.x
- Charavgi, M. D., Dimarogona, M., Topakas, E., Christakopoulos, P., and Chrysina, E. D. (2013). The structure of a novel glucuronoyl esterase from *Myceliophthora thermophila* gives new insights into its role as a potential biocatalyst. *Acta Crystallogr. D Biol. Crystallogr.* 69, 63–73. doi: 10.1107/S0907444912042400
- Crepin, V. F., Faulds, C. B., and Connerton, I. F. (2004). Functional recognition of newclasses of feruloyl esterase. *Appl. Microbiol. Biotechnol.* 63, 647–652. doi: 10.1007/s00253-003-1476-3
- Davies, G., and Henrissat, B. (1995). Structures and mechanisms of glycosyl hydrolases. *Structure*. 3, 853–859. doi: 10.1016/S0969-2126(01)00220-9
- De Vries, R. P., Poulsen, C. H., Madrid, S., and Visser, J. (1998). *AguA*, the gene encoding an extracellular  $\alpha$ -glucuronidase from *Aspergillus tubingensis*, is specifically induced on xylose and not on glucuronic acid. *J. Bacteriol.* 180, 243–249.
- Dimarogona, M., Topakas, E., and Christakopoulos, P. (2013). Recalcitrant polysaccharide degradation by novel oxidative biocatalysts. *Appl. Microbiol. Biotechnol.* 97, 8455–8465. doi: 10.1007/s00253-013-5197-y
- Domsch, K. H., Gams, W., and Anderson, T. H. (1993). *Compendium of Soil Fungi*. (Reprint of 1980 Ed. Verlag) (New York, NY: Academic Press), 780–783.
- Dotsenko, G. S., Semenova, M. V., Sinitsyna, O. A., Hinz, S. W., Wery, J., Zorov, I. N., et al. (2012). Cloning, purification, and characterization of galactomannan-degrading enzymes from *Myceliophthora thermophila*. *Biochemistry (Mosc.)* 77, 1303–1311. doi: 10.1134/S0006297912110090
- Duff, S. J. B., and Murray, W. D. (1996). Bioconversion of forest products industry waste cellulose to fuel ethanol: a review. *Biores. Technol.* 55, 1–33. doi: 10.1016/0960-8524(95)00122-0
- Dwek, R. A., Edge, C. J., Harvey, D. J., Wormald, M. R., and Parekh, R. B. (1993). Analysis of glycoprotein-associated oligosaccharides. *Annu. Rev. Biochem.* 62, 65–100. doi: 10.1146/annurev.bi.62.070193.000433
- Emalfarb, M., Hinz, S., Joosten, V., Koetsier, M., Visser, J., Visser, J., et al. (2012). Novel fungal enzymes. US Patent No. US20120030838 A1.
- Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, R. C., and Warren, R. A. (1991). Domains in microbial beta 1,4-glycanases: sequence conservation, function and enzyme families. *J. Microbiol. Rev.* 55, 303–315.
- Gouveia, S., Fernández-Costas, C., Sanromán, M. A., and Moldes, D. (2013). Polymerisation of Kraft lignin from black liquors by laccase from *Myceliophthora thermophila*: effect of operational conditions and black liquor origin. *Bioresour. Technol.* 131, 288–294. doi: 10.1016/j.biortech.2012.12.155
- Grishutin, S. G., Gusakov, A. V., Markov, A. V., Ustinov, B. B., Semenova, M. V., and Sinitsyn, A. P. (2004). Specific xyloglucanases as a new class of polysaccharide-degrading enzymes. *Biochim. Biophys. Acta* 1674, 268–281. doi: 10.1016/j.bbagen.2004.07.001
- Grönqvist, S., Buchert, J., Rantanen, K., Viikari, L., and Suurnäkki, A. (2003). Activity of laccase on unbleached and bleached thermomechanical pulp. *Enzyme Microb. Technol.* 32, 439–445. doi: 10.1016/S0141-0229(02)00319-8
- Gruppen, H., Kormelink, F. J. M., and Voragen, A. G. J. (1993). Water-unextractable cell wall material from wheat flour. A structural model for arabinoxylans. *J. Cer. Sci.* 18, 111–128. doi: 10.1006/jcers.1993.1040
- Guillén, D., Sánchez, S., and Rodríguez-Sanoja, R. (2009). Carbohydrate-binding domains: multiplicity of biological roles. *Appl. Microbiol. Biotechnol.* 85, 1241–1249. doi: 10.1007/s00253-009-2331-y
- Gusakov, A. V., Antonov, A. I., and Ustinov, B. B. (2008). N-Glycosylation in *Chrysosporium lucknowense* enzymes. *Carbohydr. Res.* 343, 48–55. doi: 10.1016/j.carres.2007.10.014
- Gusakov, A. V., Punt, P. J., Verdoes, J. C., van der Meij, J., Sinitsyn, A. P., Vlasenko, E., et al. (2011). *Fungal Enzymes*. Jupiter, FL: Dyadic International (USA), Inc. US Patent No. 7,923,236.
- Gusakov, A. V., Salanovich, T. N., Antonov, A. I., Ustinov, B. B., Okunev, O. N., Burlingame, R., et al. (2007). Design of highly efficient cellulose mixtures for enzymatic hydrolysis of cellulose. *Biotechnol. Bioeng.* 97, 1028–1038. doi: 10.1002/bit.21329
- Gusakov, A. V., Sinitsyna, A. P., Salanovich, T. N., Bukhtojarova, F. E., Markova, A. V., Ustinova, B. B., et al. (2005). Purification, cloning and characterisation of two forms of thermostable and highly active cellobiohydrolase I (Cel7A) produced by the industrial strain of *Chrysosporium lucknowense*. *Enz. Microb. Technol.* 36, 57–69. doi: 10.1016/j.enzmictec.2004.03.025
- Harris, P. V., Welner, D., McFarland, K. C., Re, E., Poulsen, J. C. N., Brown, K., et al. (2010). Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: structure and function of a large, enigmatic family. *Biochemistry* 49, 3305–3316. doi: 10.1021/bi100009p
- Henriksson, G., Johansson, G., and Pettersson, G. (2000). A critical review of cellobiose dehydrogenases. *J. Biotechnol.* 78, 93–113. doi: 10.1016/S0168-1656(00)00206-6
- Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* 280, 309–316.
- Hinz, S. W. A., Pouvreau, L., Joosten, R., Bartels, J., Jonathan, M. C., Wery, J., et al. (2009). Hemicellulase production in *Chrysosporium lucknowense* C1. *J. Cer. Sci.* 50, 318–323. doi: 10.1016/j.jcs.2009.07.005
- Horn, S. J., Vaaje-Kolstad, G., Westereng, B., and Eijsink, V. G. (2012). Novel enzymes for the degradation of cellulose. *Biotechnol. Biofuels.* 5, 45. doi: 10.1186/1754-6834-5-45
- Jun, H., Kieselbach, T., and Jönsson, L. (2011). Enzyme production by filamentous fungi: analysis of the secretome of *Trichoderma reesei* grown on unconventional carbon source. *Microb. Cell Fact.* 10, 68. doi: 10.1186/1475-2859-10-68
- Karlsson, J., Saloheimo, M., Siika-aho, M., Tenkanen, M., Penttilä, M., and Tjerneld, F. (2001). Homologous expression and characterization of Cel61A (EG IV) of *Trichoderma reesei*. *Eur. J. Biochem.* 268, 6498–6507. doi: 10.1046/j.0014-2956.2001.02605.x
- Karnaouri, A., Topakas, E., Paschos, T., Taouki, I., and Christakopoulos, P. (2013). Cloning, expression and characterization of an ethanol tolerant GH3

- $\beta$ -glucosidase from *Myceliophthora thermophila*. *PeerJ* 1, e46. doi: 10.7717/peerj.46
- Karnaouri, A. C., Topakas, E., and Christakopoulos, P. (2014). Cloning, expression, and characterization of a thermostable GH7 endoglucanase from *Myceliophthora thermophila* capable of high-consistency enzymatic liquefaction. *Appl. Microbiol. Biotechnol.* 98, 231–242. doi: 10.1007/s00253-013-4895-9
- Kim, T. J. (2008). "Microbial exo- and endo-arabinosyl hydrolases: structure, function, and application in L-arabinose production," in *Carbohydrate-Active Enzymes*, ed K. H. Park (Cambridge: CRC Press; Woodhead Publishing Ltd.), 229–257.
- Knob, A., Terrasan, C. R. F., and Carmona, E. C. (2010).  $\beta$ -Xylosidases from filamentous fungi: an overview. *World J. Microbiol. Biotechnol.* 26, 389–407. doi: 10.1007/s11274-009-0190-4
- Kühnel, S., Hinz, S. W., Pouvreau, L., Wery, J., Schols, H. A., and Gruppen, H. (2010). *Chrysosporium lucknowense* arabinohydrolases effectively degrade sugar beet arabinan. *Bioresour. Technol.* 101, 8300–8307. doi: 10.1016/j.biortech.2010.05.070
- Kühnel, S., Pouvreau, L., Appeldoorn, M. M., Hinz, S. W., Schols, H. A., and Gruppen, H. (2012). The ferulic acid esterases of *Chrysosporium lucknowense* C1: purification, characterization and their potential application in biorefinery. *Enzyme Microb. Technol.* 50, 77–85. doi: 10.1016/j.enzmictec.2011.09.008
- Kühnel, S., Westphal, Y., Hinz, S. W., Schols, H. A., and Gruppen, H. (2011). Mode of action of *Chrysosporium lucknowense* C1  $\alpha$ -L-arabinohydrolases. *Bioresour. Technol.* 102, 1636–1643. doi: 10.1016/j.biortech.2010.09.029
- Langston, J. A., Shaghasi, T., Abbate, E., Xu, F., Vlasenko, E., and Sweeney, M. D. (2011). Oxidoreductive cellulose depolymerization by the enzymes cellobiose dehydrogenase and glycoside hydrolase 61. *Appl. Environ. Microbiol.* 77, 7007–7015. doi: 10.1128/AEM.05815-11
- Lindberg, B., Rossel, K. G., and Svensson, S. (1973). Position of the O-acetyl groups in birch xylan. *Sven. Papperstidn.* 76, 30–32.
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014). The Carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* 42, D490–D495. doi: 10.1093/nar/gkt1178
- Lynd, L. R., Weimer, P. J., van Zyl, W. H., and Pretorius, I. S. (2002). Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol. Mol. Biol. Rev.* 66, 506–577. doi: 10.1128/MMBR.66.3.506-577.2002
- Maheshwari, R., Bharadwaj, G., and Bhat, M. K. (2000). Thermophilic fungi: their physiology and enzymes. *Microbiol. Mol. Biol. Rev.* 64, 461–488. doi: 10.1128/MMBR.64.3.461-488.2000
- Margaritis, A., and Merchant, R. F. J. (1986). Thermostable cellulases from thermophilic microorganisms. *Crit. Rev. Biotechnol.* 4, 327–367. doi: 10.3109/07388558609150799
- McCleary, B. V. (1988).  $\beta$ -D-Mannanase. *Meth. Enzymol.* 160, 596–610. doi: 10.1016/0076-6879(88)60174-1
- Morgenstern, I., Powlowski, J., Ishmael, N., Darmond, C., Marquetteau, S., Moisan, M. C., et al. (2012). A molecular phylogeny of thermophilic fungi. *Fungal Biol.* 116, 489–502. doi: 10.1016/j.funbio.2012.01.010
- Moukoulis, M., Topakas, E., and Christakopoulos, P. (2010). "Optimized expression of a GH11 xylanase and a type C feruloyl esterase from *Fusarium oxysporum* in *Pichia pastoris* and study of their synergistic action," in *Final Workshop, COST Action 928, Control and exploitation of enzymes for added-value products, Poster Presentation* (Naples).
- Mustafa, R., Muniglia, L., Rovel, B., and Girardin, M. (2005). Phenolic colourants obtained by enzymatic synthesis using a fungal laccase in a hydro-organic biphasic system. *Food Res. Int.* 38, 995–1000. doi: 10.1016/j.foodres.2005.02.019
- Poutanen, K., Sundberg, M., Korte, H., and Puls, J. (1990). Deacetylation of xylans by acetyl esterases of *Trichoderma reesei*. *J. Appl. Microbiol. Biotechnol.* 33, 506–510. doi: 10.1007/BF00172542
- Pouvreau, L., Jonathan, M. C., Kabel, M. A., Hinz, S. W., Gruppen, H., and Schols, H. A. (2011a). Characterization and mode of action of two acetyl xylan esterases from *Chrysosporium lucknowense* C1 active towards acetylated xylans. *Enzyme Microb. Technol.* 49, 312–320. doi: 10.1016/j.enzmictec.2011.05.010
- Pouvreau, L., Joosten, R., Hinz, S. W., Gruppen, H., and Schols, H. A. (2011b). *Chrysosporium lucknowense* C1 arabinofuranosidases are selective in releasing arabinose from either single or double substituted xylose residues in arabinoxylans. *Enzyme Microb. Technol.* 48, 397–403. doi: 10.1016/j.enzmictec.2011.01.004
- Punta, M., Coghill, P. C., Eberhardt, R. Y., Mistry, J., Tate, J., Boursnell, C., et al. (2012). The Pfam protein families' database. *Nucleic Acids Res.* 40, D290–D301. doi: 10.1093/nar/gkr1065
- Ribeiro, D. A., Cota, J., Alvarez, T. M., Brühl, F., Bragato, J., Pereira, B. M., et al. (2012). The *Penicillium echinulatum* secretome on sugar cane bagasse. *PLoS ONE* 7:e50571. doi: 10.1371/journal.pone.0050571
- Ruiz-Dueñas, F. J., and Martínez, A. T. (2009). Microbial degradation of lignin: how a bulky recalcitrant polymer is efficiently recycled in nature and how we can take advantage of this. *Microb. Biotechnol.* 2, 164–177. doi: 10.1111/j.1751-7915.2008.00078.x
- Špániková, S., and Biely, P. (2006). Glucuronoyl esterase—novel carbohydrate esterase produced by *Schizophyllum commune*. *FEBS Lett.* 580, 4597–4601. doi: 10.1016/j.febslet.2006.07.033
- Shibuya, N., and Iwasaki, T. (1985). Structural features of rice bran hemicellulose. *Phytochemistry* 30, 488–494.
- Sørensen, H. R., Pedersen, S., Jørgensen, C. T., and Meyer, A. S. (2007). Enzymatic hydrolysis of wheat arabinoxylan by a recombinant "minimal" enzyme cocktail containing  $\alpha$ -xylosidase and novel endo-1,4- $\alpha$ -xylanase and  $\alpha$ -L-arabinofuranosidase activities. *Biotechnol. Prog.* 23, 100–107. doi: 10.1021/bp0601701
- Subramaniam, S. S., Nagalla, S. R., and Renganathan, V. (1999). Cloning and characterization of a thermostable cellobiose dehydrogenase from *Sporotrichum thermophile*. *Arch. Biochem. Biophys.* 365, 223–230. doi: 10.1006/abbi.1999.1152
- Sweeney, M. D., and Xu, F. (2012). Biomass converting enzymes as industrial biocatalysts for fuels and chemicals: recent developments. *Catalysts* 2, 244–263. doi: 10.3390/catal202044
- Szjártó, N., Horan, E., Zhang, J., Puranen, T., Siika-Aho, M., and Viikari, L. (2011). Thermostable endoglucanases in the liquefaction of hydrothermally pretreated wheat straw. *Biotechnol. Biofuels* 4, 2. doi: 10.1186/1754-6834-4-2
- Tambor, J. H., Ren, H., Ushinsky, S., Zheng, Y., Riemens, A., St-Francois, C., et al. (2012). Recombinant expression, activity screening and functional characterization identifies three novel endo-1,4- $\beta$ -glucanases that efficiently hydrolyse cellulosic substrates. *Appl. Microbiol. Biotechnol.* 93, 203–14. doi: 10.1007/s00253-011-3419-8
- Tenkanen, M., Siika-aho, M., Hausalo, T., Puls, J., and Viikari, L. (1996). "Synergism between xylanolytic enzymes of *Trichoderma reesei* in the degradation of acetyl-4-O-methylglucuronoxylan," in *Biotechnology in Pulp and Paper Industry Advances in Applied and Fundamental Research*, eds K. Messner and E. Srebotnik (Vienna: WUA Universitätsverlag), 503–508.
- Terri, T. T. (1997). Crystalline cellulose degradation: a new insight into the function of cellobiohydrolases. *Trends Biotechnol.* 15, 160–167. doi: 10.1016/S0167-7799(97)01032-9
- Tolan, J. S., and Foody, B. (1999). Cellulases from submerged fermentation. *Adv. Biochem. Eng. Biotechnol.* 65, 41–67. doi: 10.1007/3-540-49194-5\_3
- Topakas, E., Christakopoulos, P., and Faulds, C. B. (2005). Comparison of mesophilic and thermophilic feruloyl esterases: characterization of their substrate specificity for methyl phenylalkanoates. *J. Biotechnol.* 115, 355–366. doi: 10.1016/j.jbiotec.2004.10.001
- Topakas, E., Kalogeris, E., Kekos, D., Macris, B. J., and Christakopoulos, P. (2003). Production and partial characterisation of feruloyl esterase by *Sporotrichum thermophile* in solid-state fermentation. *Proc. Biochem.* 38, 1539–1543. doi: 10.1016/S0032-9592(03)00044-X
- Topakas, E., Moukoulis, M., Dimarogona, M., and Christakopoulos, P. (2012). Expression, characterization and structural modelling of a feruloyl esterase from the thermophilic fungus *Myceliophthora thermophila*. *Appl. Microbiol. Biotechnol.* 94, 399–411. doi: 10.1007/s00253-011-3612-9
- Topakas, E., Moukoulis, M., Dimarogona, M., Vafiadi, C., and Christakopoulos, P. (2010). Functional expression of a thermophilic glucuronoyl esterase from *Sporotrichum thermophile*: identification of the nucleophilic serine. *Appl. Microbiol. Biotechnol.* 87, 1765–1772. doi: 10.1007/s00253-010-2655-7
- Topakas, E., Stamatis, H., Biely, P., and Christakopoulos, P. (2004). Purification and characterization of a type B feruloyl esterase (StFAE-A) from the thermophilic fungus *Sporotrichum thermophile*. *Appl. Microbiol. Biotechnol.* 63, 686–690. doi: 10.1007/s00253-003-1481-6
- Topakas, E., Vafiadi, C., and Christakopoulos, P. (2007). Microbial production, characterization and applications of feruloyl esterases. *Proc. Biochem.* 42, 497–509. doi: 10.1016/j.procbio.2007.01.007

- Ustinov, B. B., Gusakov, A. V., Antonov, A. I., and Sinitsyn, A. P. (2008). Comparison of properties and mode of action of six secreted xylanases from *Chrysosporium lucknowense*. *Enzyme Microb. Technol.* 43, 56–65. doi: 10.1016/j.enzmictec.2008.01.017
- Vafiadi, C., Topakas, E., Biely, P., and Christakopoulos, P. (2009). Purification, characterization and mass spectrometric sequencing of a thermophilic glucuronoyl esterase from *Sporotrichum thermophile*. *FEMS Microbiol. Lett.* 296, 178–184. doi: 10.1111/j.1574-6968.2009.01631.x
- van Gool, M. P., van Muiswinkel, G. C., Hinz, S. W., Schols, H. A., Sinitsyn, A. P., and Gruppen, H. (2012). Two GH10 endo-xylanases from *Myceliophthora thermophila* C1 with and without cellulose binding module act differently towards soluble and insoluble xylans. *Bioresour. Technol.* 119, 123–132. doi: 10.1016/j.biortech.2012.05.117
- van Gool, M. P., van Muiswinkel, G. C., Hinz, S. W., Schols, H. A., Sinitsyn, A. P., and Gruppen, H. (2013). Two novel GH11 endo-xylanases from *Myceliophthora thermophila* C1 act differently toward soluble and insoluble xylans. *Enzyme Microb. Technol.* 53, 25–32. doi: 10.1016/j.enzmictec.2013.03.019
- Verdoes, J. C., Punt, P. J., Burlingame, R. P., Pynnonen, C. M., Olson, P. T., and Wery, J. (2010). New fungal production system. Patent No. WO2010107303 A2.
- Visser, H., Joosten, V., Punt, P. J., Gusakov, A. V., Olson, P. T., Joosten, R., et al. (2011). Development of a mature fungal technology and production platform for industrial enzymes based on a *Myceliophthora thermophila* isolate, previously known as *Chrysosporium lucknowense* C1. *Ind. Biotechnol.* 7, 214–223. doi: 10.1089/ind.2011.7.214
- Weinstein, L., and Albersheim, P. (1979). Structure of plant cell walls: IX. Purification and partial characterization of a wall-degrading endo-arabanase and an arabinosidase from *Bacillus subtilis*. *Plant Physiol.* 63, 425–432. doi: 10.1104/pp.63.3.425
- Westphal, Y., Kühnel, S., de Waard, P., Hinz, S. W., Schols, H. A., Voragen, A. G., et al. (2010). Branched arabino-oligosaccharides isolated from sugar beet arabinan. *Carbohydr. Res.* 345, 1180–1189. doi: 10.1016/j.carres.2010.03.042
- Zhang, J., Pakarinen, A., and Viikari, L. (2013). Synergy between cellulases and pectinases in the hydrolysis of hemp. *Bioresour. Technol.* 29, 302–307. doi: 10.1016/j.biortech.2012.11.064
- Zhang, Y. H. P., Himmel, M. E., and Mielenz, J. R. (2006). Outlook for cellulose improvement: screening and selection strategies. *Biotechnol. Adv.* 24, 452–481. doi: 10.1016/j.biotechadv.2006.03.003

**Conflict of Interest Statement:** The Review Editor Ulrika Rova declares that, despite being affiliated to the same institution as authors Anthi Karnaouri, Io Antonopoulou, and Paul Christakopoulos, the review process was handled objectively and no conflict of interest exists. The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 22 April 2014; accepted: 22 May 2014; published online: 18 June 2014.

Citation: Karnaouri A, Topakas E, Antonopoulou I and Christakopoulos P (2014) Genomic insights into the fungal lignocellulolytic system of *Myceliophthora thermophila*. *Front. Microbiol.* 5:281. doi: 10.3389/fmicb.2014.00281

This article was submitted to Microbial Physiology and Metabolism, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Karnaouri, Topakas, Antonopoulou and Christakopoulos. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Comparative genomics and evolution of regulons of the LacI-family transcription factors

Dmitry A. Ravcheev<sup>1†</sup>, Matvei S. Khoroshkin<sup>1†</sup>, Olga N. Laikova<sup>1</sup>, Olga V. Tsoy<sup>1,2</sup>, Natalia V. Sernova<sup>1</sup>, Svetlana A. Petrova<sup>1,2</sup>, Aleksandra B. Rakhmaninova<sup>2</sup>, Pavel S. Novichkov<sup>3</sup>, Mikhail S. Gelfand<sup>1\*</sup> and Dmitry A. Rodionov<sup>1,4\*</sup>

<sup>1</sup> Research Scientific Center for Bioinformatics, A.A. Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia

<sup>2</sup> Faculty of Bioengineering and Bioinformatics, Moscow State University, Moscow, Russia

<sup>3</sup> Lawrence Berkeley National Laboratory, Genomics Division, Berkeley, CA, USA

<sup>4</sup> Department of Bioinformatics, Sanford-Burnham Medical Research Institute, La Jolla, CA, USA

## Edited by:

Katherine M. Pappas, University of Athens, Greece

## Reviewed by:

Paul Alan Hoskisson, University of Strathclyde, UK

John Alan Gerlt, University of Illinois, USA

## \*Correspondence:

Mikhail S. Gelfand, Research Scientific Center for Bioinformatics, A.A. Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, 19-1, Bolshoi Karetny pereulok, Moscow 127994, Russia  
e-mail: gelfand@iitp.ru;  
Dmitry A. Rodionov, Sanford-Burnham Medical Research Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA  
e-mail: rodionov@sanfordburnham.org

† These authors have contributed equally to this work.

DNA-binding transcription factors (TFs) are essential components of transcriptional regulatory networks in bacteria. LacI-family TFs (LacI-TFs) are broadly distributed among certain lineages of bacteria. The majority of characterized LacI-TFs sense sugar effectors and regulate carbohydrate utilization genes. The comparative genomics approaches enable *in silico* identification of TF-binding sites and regulon reconstruction. To study the function and evolution of LacI-TFs, we performed genomics-based reconstruction and comparative analysis of their regulons. For over 1300 LacI-TFs from over 270 bacterial genomes, we predicted their cognate DNA-binding motifs and identified target genes. Using the genome context and metabolic subsystem analyses of reconstructed regulons, we tentatively assigned functional roles and predicted candidate effectors for 78 and 67% of the analyzed LacI-TFs, respectively. Nearly 90% of the studied LacI-TFs are local regulators of sugar utilization pathways, whereas the remaining 125 global regulators control large and diverse sets of metabolic genes. The global LacI-TFs include the previously known regulators CcpA in Firmicutes, FruR in Enterobacteria, and PurR in Gammaproteobacteria, as well as the three novel regulators—GluR, GapR, and PckR—that are predicted to control the central carbohydrate metabolism in three lineages of Alphaproteobacteria. Phylogenetic analysis of regulators combined with the reconstructed regulons provides a model of evolutionary diversification of the LacI protein family. The obtained genomic collection of *in silico* reconstructed LacI-TF regulons in bacteria is available in the RegPrecise database (<http://regprecise.lbl.gov>). It provides a framework for future structural and functional classification of the LacI protein family and identification of molecular determinants of the DNA and ligand specificity. The inferred regulons can be also used for functional gene annotation and reconstruction of sugar catabolic networks in diverse bacterial lineages.

**Keywords:** bacteria, transcription factors, regulons, sugar metabolism, comparative genomics

## INTRODUCTION

Evolution of regulatory interactions in bacteria can be approached from three directions. The first approach is the comparative analysis of regulation of a functional system, e.g., a metabolic pathway, in a variety of species. Such analysis demonstrates high flexibility of regulatory interactions even in closely related species, with expansion, contraction, and merging of regulons or a complete change of regulators (Manson McGuire and Church, 2000; McCue et al., 2001; Tan et al., 2001; Gelfand, 2006; Rodionov et al., 2006; Ravcheev et al., 2007; Kazakov et al., 2009; Suvorova et al., 2011, 2012). The second approach is to consider a taxon of a relatively low level (genus or family) and to use comparative genomics to predict as many regulatory interactions as possible. This has been done for  $\gamma$ -Proteobacteria from the *Shewanella* genus (Rodionov et al.,

2011); Firmicutes closely related to *Bacillus subtilis* (Leyn et al., 2013) and *Staphylococcus aureus* (Ravcheev et al., 2011); two families of lactic acid bacteria from the Lactobacillales order (Ravcheev et al., 2013b); hyperthermophilic bacteria related to *Thermotoga maritima* (Rodionov et al., 2013); human gut habitant *Bacteroides thetaiotaomicron*; and related organisms (Ravcheev et al., 2013a). An important side product of such studies is functional annotation of hypothetical proteins by assigning them, via co-regulation, to known metabolic pathways and other functional subsystems (Rodionov, 2007; Gelfand and Rodionov, 2008).

The third approach, implemented here, is to consider a family of transcription factors (TFs) and then identify binding motifs for as many TFs as possible. This is mainly motivated by the desire to analyze the structure of protein–DNA interactions and



the co-evolution of TFs and the motifs they recognize (Desai et al., 2009; Huang et al., 2009; Camas et al., 2010; Leyn et al., 2011; Ravcheev et al., 2012). An important issue in such studies is to connect TFs to the cognate TF binding sites (TFBSs) identified by phylogenetic footprinting and other computational techniques (Conlan et al., 2005; Wels et al., 2006; Liu et al., 2008). This problem is either solved experimentally or addressed computationally, for instance for regulons controlled by local TF from specific protein families (Rigali et al., 2004; Francke et al., 2008; Sahota and Stormo, 2010; Ahn et al., 2012; Kazakov et al., 2013). Phylogenetic profiling of TF genes and motifs upstream of candidate regulon members is an alternative bioinformatics approach for assigning TFs to putative regulons (Rodionov and Gelfand, 2005). The comparative analysis of ligand-binding domains in TFs also helps identify ligand specificity determinants and propose models of functional diversification within large and functionally heterogeneous families of TFs (Kazanov et al., 2013).

Here we study the LacI family of bacterial transcription factors (LacI-TFs). The namesake of the family, the lactose repressor LacI of *E. coli*, has been the model object for the analysis of bacterial transcriptional regulation since the classical papers of Jacob and Monod (1959, 1961). The family was established by the analysis of similarity of protein sequences, and simultaneously the similarity of DNA motifs recognized by the family members was noted (Weickert and Adhya, 1992). At the same time, it was observed that its DNA-binding domains of LacI-family regulators are similar to the helix-turn-helix domains of other TFs (Nguyen and Saier, 1995), whereas the ligand-binding domain of LacI-TFs is homologous to the periplasmic proteins of ABC-transporters (Mauzy and Hermodson, 1992; Fukami-Kobayashi et al., 2003). Interestingly, this domain was also seen in combination with another DNA-binding domain, winged helix-turn-helix of the GntR family (Franco et al., 2006). While early observations on a limited number of sequences suggested that the history of the family involved a series of duplications at the early stage and a low level of duplications later on (Nguyen and Saier, 1995), more data, derived from bacterial genomes becoming available, demonstrated that the duplications were occurring throughout the history of the family (Fukami-Kobayashi et al., 2003). Given its large size and high level of structural similarity, yielding reliable multiple alignments, this family was widely used as a model for algorithms for identification of functionally important residues both uniformly conserved and specificity-determining (Mirny and Gelfand, 2002; Fukami-Kobayashi et al., 2003; Kalinina et al., 2004; Pei et al., 2006; Tungtur et al., 2007; Parente and Swint-Kruse, 2013). Some of these predictions were further tested in experiment (Meinhardt and Swint-Kruse, 2008; Camas et al., 2010; Tungtur et al., 2010, 2011).

The structural similarity of the DNA motifs recognized by the LacI-TFs was used to identify regulons in *Lactobacillus plantarum* (Francke et al., 2008) and *Dickeya dadantii* [*Erwinia chrysanthemi*] (Van Gijsegem et al., 2008). The structural aspects of interactions of the LacI-TFs with DNA and ligands and between subunits in dimers have been summarized in a recent review (Swint-Kruse and Matthews, 2009).

Here we report results of a large-scale, manual comparative genomics analysis of LacI-TFs aimed at identification of their

binding sites, motifs, and regulons in bacterial genomes. By analyzing the genomic and metabolic context of the reconstructed regulons and by combining this analysis with information gathered from literature, we inferred the biological roles and molecular effectors for a large number of the studied LacI-TFs. As result, we made a number of observations on the distribution of orthologous LacI-TFs in genomes, the statistics of the binding sites' arrangement in regulatory regions, and the number and functional characteristics of the regulated genes. By combining the functional annotations with phylogenetic analysis, we proposed evolutionary models of functional diversification for a number of LacI-TF groups. The obtained reference dataset of 1281 regulons in 272 genomes was deposited in the RegPrecise database (Novichkov et al., 2013).

## MATERIALS AND METHODS

The genomes were downloaded from the MicrobesOnline database (Dehal et al., 2010). TFs from the LacI family were identified by similarity searches and domain predictions in the Pfam database (Finn et al., 2014). LacI-TFs consist of two characteristic domains, an N-terminal HTH DNA-binding domain (PF00356) and a C-terminal effector-binding domain, which is homologous to periplasmic binding proteins of sugar ABC transporters (PF00532, PF13377, or PF13407). Gene orthology was defined by the bidirectional best-hit criterion implemented in the GenomeExplorer software (Mironov et al., 2000) and validated by phylogenetic trees from the MicrobesOnline database (Dehal et al., 2010). Genes were considered as orthologs if they (i) formed a mono- or paraphyletic branch of the phylogenetic tree and (ii) demonstrated conserved chromosomal gene context. For genes in the reconstructed regulons, correspondence to both of these criteria was sufficient for these genes to be considered as orthologs. For the studied LacI-family regulators, an additional criterion of orthology was used. Thus, orthologous groups of LacI-TFs should (i) form a mono- or paraphyletic group on the phylogenetic tree; (ii) have a conserved gene context; (iii) have highly similar TFBS motifs; and (iv) have the same effector specificity (known or predicted based on the regulon content).

For the regulon reconstruction, we used a previously established comparative genomics approach (Rodionov, 2007) implemented in the RegPredict Web server (Novichkov et al., 2010). This approach includes prediction of putatively regulated genes, inference of TFBSs, construction of positional weight matrices (PWMs) for TFBS motifs, and a further search for additional regulon members on the basis of predicted TFBSs in gene promoter regions. Overall, three main strategies were used for the reconstruction of regulons: (1) construction of PWMs on the basis of known TFBSs, for regulons being previously analyzed in model organisms; (2) prediction of novel TFBS motifs in promoter regions of regulated genes, for regulons with only regulated genes but not TFBSs known; and (3) prediction of putatively co-regulated genes followed by the inference of putative TFBS motifs in their promoter regions and attribution of a candidate TF to a putative regulon. Presumably, regulated genes were predicted by the analysis of conserved gene neighborhoods around an analyzed LacI-TF gene. Data about known regulated genes and LacI-TFBS motifs were extracted from the literature and from the

RegTransBase (Cipriano et al., 2013), RegulonDB (Salgado et al., 2013), DBTBS (Sierro et al., 2008), and CoryneRegNet (Pauling et al., 2012) databases.

Candidate motifs in upstream regions of regulated operons were identified by the Discover Profile tool in RegPredict (Novichkov et al., 2010). A search for palindromic DNA motifs of 14- to 24-bp length was carried out within putative promoter regions from −400 to +100 bp relative to the translational gene start. Motifs were further manually validated by phylogenetic footprinting, that is, analysis of conserved islands in multiple alignments of DNA fragments (Shelton et al., 1997). The constructed PWMs were further used to search for additional regulon members using the Run Profile tool in RegPredict. The lowest score observed in the training set of known and/or predicted TFBSs was used as the threshold for a site search in genomes. To eliminate false-positive TFBS predictions, the consistency check approach (Ravcheev et al., 2007; Rodionov, 2007) and/or functional relatedness of candidate target operons were used. In this approach, an operon can be considered as regulated when its upstream region contains putative TFBSs with a score higher than the threshold, and such sites can be found in a number of related genomes. Operons were defined as groups of genes satisfying the following criteria: same direction of transcription, intergenic distance up to 100 bp, absence of internal TF-binding sites, and conservation of the locus structure in a number of related genomes. All predicted TFs, motifs, and sites are available in the RegPrecise database (<http://regprecise.lbl.gov/>) (Novichkov et al., 2013), where they are publicly available within the TF family collections of regulons.

Functional gene annotations were extracted from the literature and uploaded from the SEED (Disz et al., 2010), UniProt (Magrane and Consortium, 2011), MicrobesOnline (Dehal et al., 2010), and KEGG (Kanehisa et al., 2012) databases. Known functional annotations for a particular gene were expanded to all orthologous genes. For prediction of gene functions, both the comparative genomics and context-based methods were used (reviewed in Osterman and Overbeek, 2003; Overbeek et al., 2005; Rodionov, 2007). Multiple alignments of protein and DNA sequences were built by MUSCLE (Edgar, 2004). Phylogenetic trees were constructed using a maximum-likelihood algorithm implemented in PhyML 3.0 (Guindon et al., 2010) and visualized via Dendroscope (Huson et al., 2007) and iTOL (Letunic and Bork, 2011). Sequence logos for DNA motifs were drawn with WebLogo (Crooks et al., 2004).

## RESULTS

The comparative genomics workflow for regulon reconstruction implemented in the RegPredict Web server (Novichkov et al., 2010) and the RegPrecise database (Novichkov et al., 2013) includes three steps: (i) selection of a taxonomic group of related bacteria; (ii) selection of a subset of diverse genomes that represent a given group; and (iii) reconstruction of regulons in the selected genomes. For the analysis of LacI-TF regulons, we selected a set of 344 representative genomes from 39 taxonomic groups from 7 bacterial phyla (Table S1 in Supplementary Material). Among the analyzed lineages, there are 19 taxonomic groups of Proteobacteria (183

genomes), 9 groups of Firmicutes (72 genomes), and 7 groups of Actinobacteria (57 genomes). The Bacteroides, Chloroflexi, Deinococcus-Thermus, and Thermotogae phyla are each represented by a single taxonomic group and have 32 genomes in total.

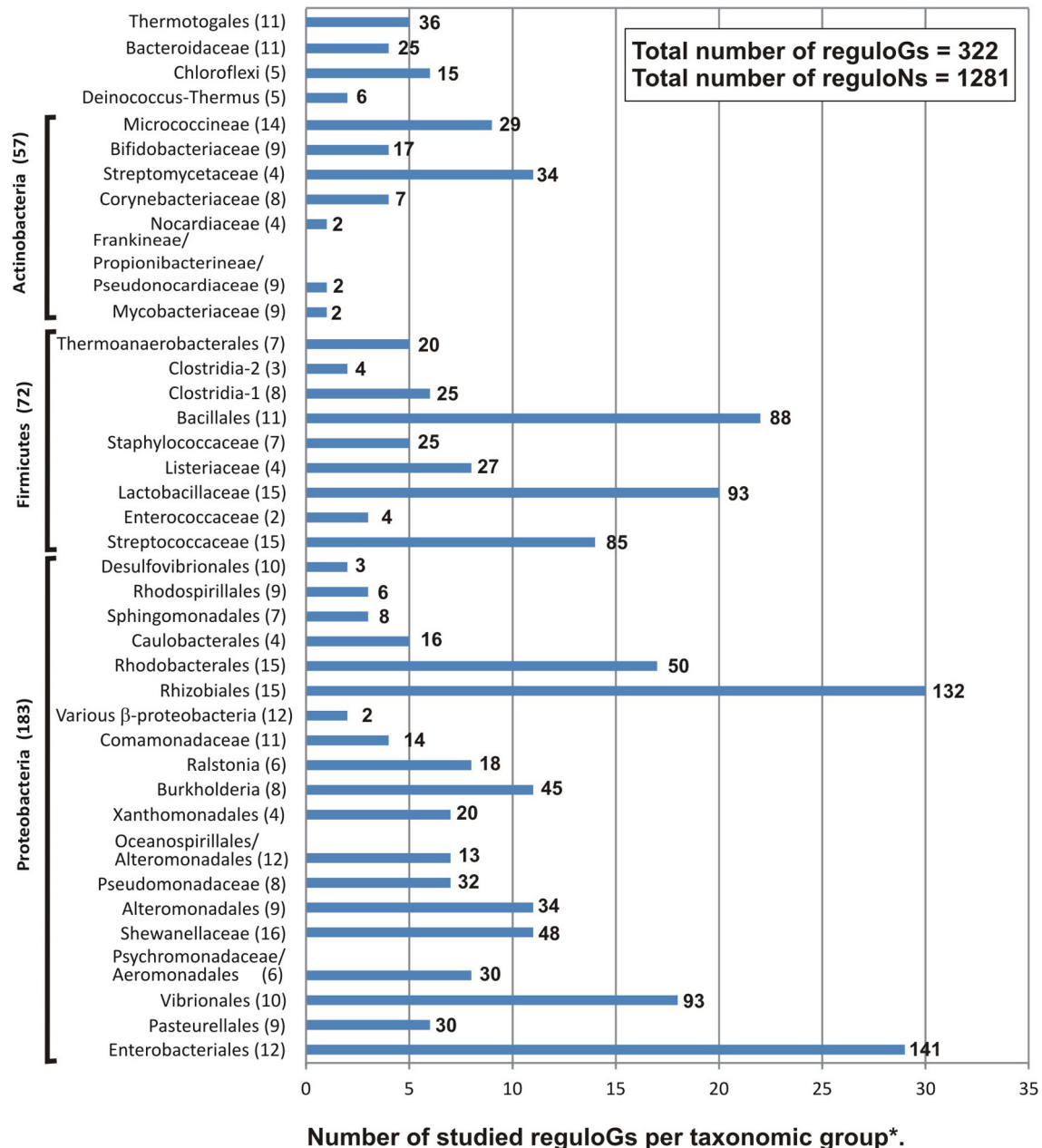
## REPERTOIRE OF LacI-TF GENES IN BACTERIAL GENOMES

To estimate the abundance of LacI-family TFs (LacI-TFs) in the studied genomes, we collected primary LacI-TF sets using a similarity search and the existing prokaryotic TF compilations. In total, 2572 proteins were found unevenly distributed in most (309/344; 90%) of the studied genomes, whereas 10% of the genomes do not encode putative LacI-TFs (Table S1 in Supplementary Material). The largest average numbers of LacI-TFs per genome were found in several lineages of the Actinobacteria phylum including Streptomycetaceae and Bifidobacteriaceae (from 32 and to 17 regulators), in two lineages of Proteobacteria–Rhizobiales and Enterobacteriales (15 regulators in each group), and in two lineages of Firmicutes–Bacillales and Enterococcaceae (12 regulators in each group). The remaining taxonomic groups possess less than 10 LacI-TFs per genome on average. Noteworthy, the Methylophilales, Neisseriales, Nitrosomonadales, Oceanospirillales, Magnetospirillum/Rhodospirillum, and Desulfotribionales groups completely lack LacI-TFs in their genomes. The absence of LacI-TFs in these taxonomic groups of Proteobacteria can be related to (i) relatively small proportion of sugar catabolic genes in their genomes (as LacI-TFs mostly control sugar catabolism, see below); (ii) increasing usage of TFs from other families to compensate the contraction of the LacI-TF pool.

## STATISTICS OF RECONSTRUCTED REGULONS AND REGULOGS

The entire set of identified LacI-TFs was broken into taxonomic group-specific orthologous groups that were subjected to further comparative genomics analysis using the RegPredict Web server (Novichkov et al., 2010). Normally, an orthologous group contained no more than one TF per genome. However, in some cases TFs formed by recent, mainly genome-specific, duplication were assigned to the same orthologous group (Table S1 in Supplementary Material). By analyzing orthologous groups of regulators in each taxonomic group, candidate motifs and binding sites were predicted for 1303 LacI-TFs (50% of all putative LacI-TFs) in 272 bacterial genomes (80% of studied genomes). The main outcome of this analysis is an annotated regulog, which is defined as a set of genome-specific regulons controlled by orthologous TFs. Overall, we inferred 1281 LacI-TF regulons that constitute 322 populated regulogs unevenly distributed across 39 studied taxonomic groups of genomes (Tables S1, S2 in Supplementary Material). The reconstructed regulons included 7465 candidate sites, 6076 operons, and 13,558 genes.

The taxonomical distribution of the reconstructed LacI-TF regulons is highly uneven but generally follows the distribution of all LacI-family TFs (**Figure 1**): 57% of regulons were from Proteobacteria, about 30% from Firmicutes, 7% from Actinobacteria, about 1–2% from each of Thermotogales, Bacteroides, Chloroflexi, and Deinococcus/Thermus. Yet, compared to the genomic distribution of all putative LacI-TFs (Table



\*Total number of reconstructed regulons is shown by a number next to bar.

**FIGURE 1 | Distribution of LacI-TF regulons and regulos in the analyzed taxonomic groups of bacteria.**

S1 in Supplementary Material), the Actinobacteria phylum is underrepresented. Based on the phylogenetic analysis of LacI-TF proteins, regulators from the reconstructed regulos were merged into larger orthologous groups that were consistent with the taxonomy, had regulated orthologous genes, and had similar binding motifs. TFs were assigned to an orthologous group if they formed a mono- or paraphyletic branch (see below) in the phylogenetic tree (Figure S1 in Supplementary Material). As in most gene families containing multiple paralogs resulting from

frequent duplications, losses, and horizontal transfers, the resolution of orthology in some cases was difficult and required arbitrary decisions that are supported by the genomic context and/or functional attributes of the reconstructed regulons.

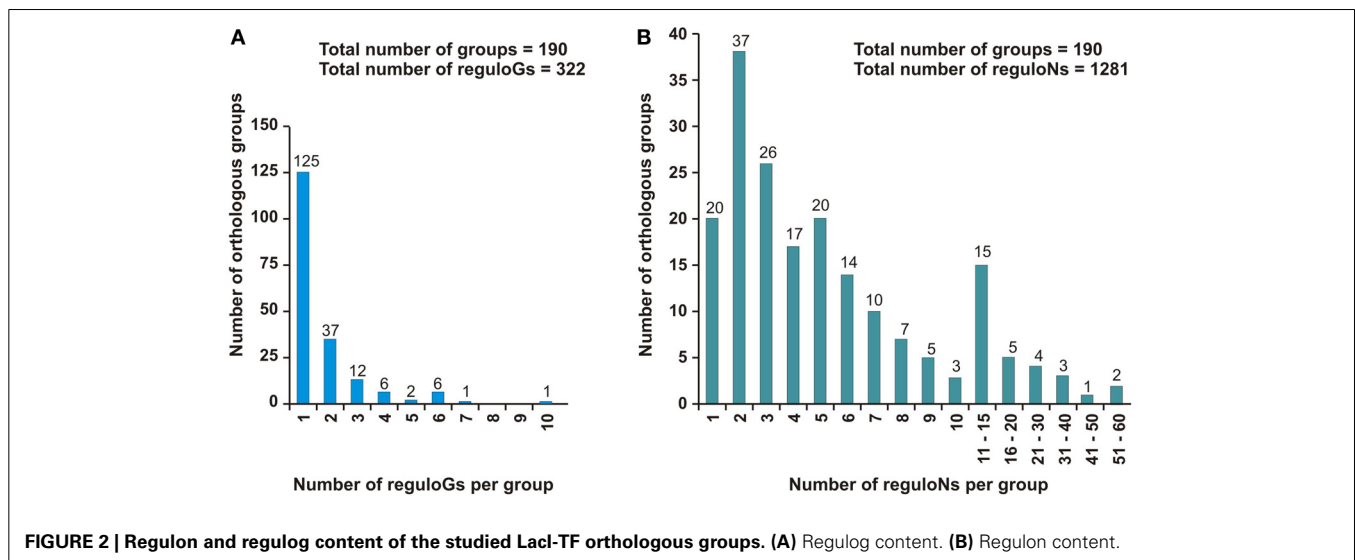
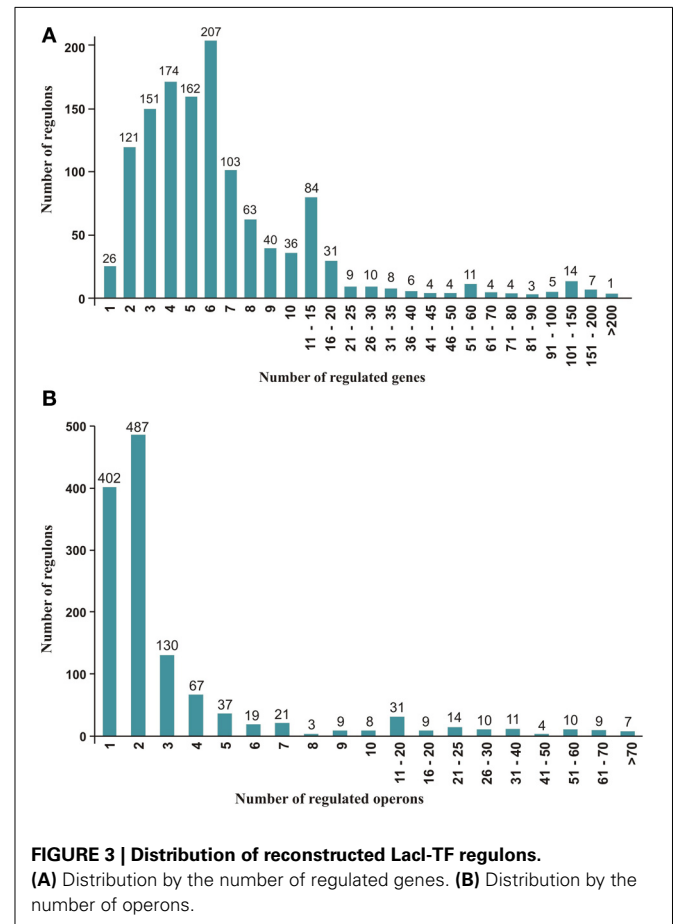
As a result, the studied LacI-TFs were classified into 190 orthologous groups characterized by conserved DNA motifs and regulated pathways (Table S2 in Supplementary Material). Two-thirds of the obtained orthologous groups (125/190) contain TFs from a single regulos, which is defined as a set of orthologous

regulons in a group of closely related genomes. Thirty-seven orthologous groups include two regulogs, whereas the remaining 28 groups were assigned to three or more regulogs (**Figure 2A**). The total number of regulons (and corresponding TFs) per orthologous group of LacI-TFs varies between 1 and 59, with the average being 6.7 (**Figure 2B**). The maximal number of groups was observed for the groups including two TFs (37 groups). Orthologous groups containing up to 5 and between 6 and 10 regulons constitute 60 and 20% of all groups, respectively. The most populated groups of LacI-TF orthologs were found for the global catabolite control regulator CcpA in Firmicutes (59 regulons, 6 regulogs), the ribose repressor RbsR in Proteobacteria (52 regulons, 10 regulogs), the maltose repressor MalR in Firmicutes (42 regulons, 6 regulogs), as well as sugar catabolism regulators FruR (40 regulons, 6 regulogs), GntR (37 regulons, 7 regulogs), and GalR (27 regulons, 5 regulogs) in  $\gamma$ -Proteobacteria.

### GLOBAL AND LOCAL REGULONS

The reconstructed LacI-TF regulons demonstrate drastic differences in the numbers of predicted target genes and operons. The majority of regulons (1198/1288, 93%) include 20 or fewer genes (**Figure 3A**), and further, three-fourths of these regulons contain between 2 and 7 genes, whereas 26 regulons have only 1 target gene. With respect to the number of regulated operons (**Figure 3B**), the largest portion of the studied LacI-TFs regulates one (31%) or two (38%) operons. We divided all reconstructed LacI-TF regulons into two main categories depending on their size (number of regulated genes and operons) and functional diversity (number of regulated pathways). A total of 125 regulons (12 regulogs) were classified as global, since each of them (i) contained more than 15 target genes that were arranged in at least 7 operons and (ii) controlled multiple metabolic pathways. The remaining LacI-TF regulons (1163/1288, 90%) were classified as local, each having a smaller number of targets and controlling a single metabolic pathway, usually a particular carbohydrate catabolic pathway.

Almost one-half of the identified global regulons (59 regulons, 6 regulogs) are operated by orthologs of the *B. subtilis* catabolite control regulator CcpA in all analyzed taxonomic groups of Firmicutes. CcpA regulons were previously described in detail for bacteria from the Bacillaceae (Sonenshein, 2007;





Fujita, 2009; Leyn et al., 2013), Staphylococcaceae (Seidl et al., 2009; Ravcheev et al., 2011), Lactobacillales (Mahr et al., 2000; Zheng et al., 2011, 2012; Zotta et al., 2012; Ravcheev et al., 2013b), and Clostridiaceae (Antunes et al., 2012) lineages. Another large group of global regulons (31 regulons, 3 regulogs) are operated by orthologs of the *E. coli* purine repressor PurR in three related taxonomic groups of  $\gamma$ -Proteobacteria—Enterobacteriales, Pasteurellales, and Vibrionales. PurR is a global transcriptional regulator of *E. coli*, controlling biosynthesis of purines, some steps of biosynthesis of pyrimidines, polyamine metabolism, and nitrogen assimilation (Ravcheev et al., 2002; Cho et al., 2011). The remaining three global regulogs—FruR in Enterobacteriales, PckR in Rhizobiales, and GapR in Rhodobacteriales (totaling 35 regulons)—that control the central and periphery carbohydrate catabolic pathways in diverse groups of Proteobacteria are described in more detail below.

Autoregulation was observed for 72% (943 TFs) of the studied LacI-TFs. The autoregulation of TFs was more typical for local regulons, as expected. Thus, less than one-half of the studied global regulators (55 TFs) demonstrated autoregulation, whereas more than three-fourths of the analyzed local regulators (888 TFs) were autoregulated.

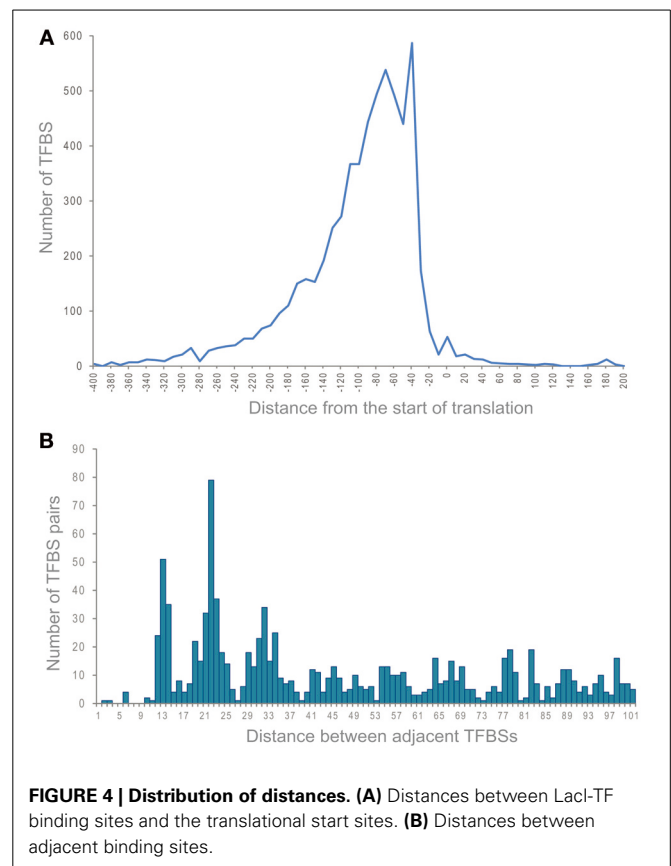
### BINDING SITES AND MOTIFS

Binding motifs of the considered LacI-TFs are palindromes formed by highly conserved inverted repeats, which is consistent with previous studies (Francke et al., 2008; Camas et al., 2010; Milk et al., 2010). The distance between the repeats is usually constant for a given orthologous group of TFs, although in rare cases there is some flexibility. The overwhelming majority of LacI-TF binding motifs (1251/1303 TFs) are even palindromes (16, 18, 20, or 22 nt long), whereas non-canonical palindromes (17, 19, or 21 nt long) were found for only 4% of regulators (e.g., LacR, GalR, and EbGR). The characteristic feature of even palindromes is the presence of a consensus CG pair in the center of the palindrome (1173 TFs). Nonetheless, in some even palindromes the central pair can be either different (49 TFs) or degenerate (29 TFs).

More than 75% of identified LacI-TFBSs are located within the area between 140 and 30 bp upstream of the start codon (Figure 4A). Less than 1% of sites are localized within coding regions, including experimentally demonstrated *E. coli* sites of LacI deep within the *lacZ* gene (Lewis, 2005) and PurR in the *purB* gene (He et al., 1992). Approximately 7% of sites are localized far upstream (>200 bp), and while some of them might in fact regulate divergently transcribed genes, experimental examples of such localization are known, e.g., the PurR site upstream of *prsA*, again in *E. coli* (He et al., 1993).

About 20% of regulated operons are preceded by more than one binding site. Thus, tandem sites were observed upstream of 1118 operons. The multiplicity of sites within regulatory regions of divergently transcribed genes (i.e., divergons) increases with the divergon length. Groups of three and four adjacent sites were found for 116 and 5 regulated operons, respectively.

The histogram of intersite distances for pairs of sites localized in the same intergenic region has pronounced peaks at 13, 22, and 32 bp (Figure 4B). The latter two lengths are multiples of the DNA helix step and hence are clearly indicative of cooperative



binding of two TF dimers. Several LacI-TF regulogs, e.g., RafR from Enterobacteriales and ScrR from Burkholderiales, have only double sites at the distance of 21–22 bp, suggesting that the cooperative binding in this case is obligatory. Overlapping sites, situated at a distance of about 13 bp, were observed upstream of 110 operons. Such an arrangement may be functional, as has been demonstrated for GntR-binding sites upstream of *gntKU* of *E. coli* (Tsunedomi et al., 2003a) and it is conserved for more than one-half of the operons regulated by GntR and its orthologs in  $\gamma$ -Proteobacteria.

A trivial explanation—that these observations are an artifact—does not seem plausible for two reasons. Firstly, such multiple sites are conserved and even preferred in some orthologous groups. Secondly, the artifact hypothesis does not explain the preferred distance of 13 bp. In the available tertiary structures of LacI-TFs complexed with DNA (Schumacher et al., 1994; Barbier et al., 1997; Kalodimos et al., 2002), seven base-pairs nearest to the site center form contacts between bases and side residues. Hence, the site-overlap region strongly overlaps with the zone of specific TF-DNA contacts. While tetramer binding has been suggested, it is difficult to reconcile this with the structural data, as dimers bound at these distances cannot interact. It is possible that there exists a specific binding mode involving partial unwinding of the DNA strands.

### REGULATED METABOLIC PATHWAYS AND EFFECTORS

By assessing the functional content of the reconstructed regulons, we tentatively predicted possible biological functions and



effectors for 190 orthologous groups of LacI-TFs. As result, metabolic pathways were predicted for 182 groups of LacI-TFs. These include 54 groups that were only assigned to the general category of sugar metabolism, and for them, a specific sugar catabolic pathway remained unknown (Table S2 in Supplementary Material). We compared the predicted regulon functions with previous results of experimental studies available for 24 selected LacI-TFs. The previously established functions of these regulators are in good agreement with the target-regulated pathways that were predicted in this work. Based on the metabolic pathway reconstruction and the knowledge of pathway metabolites, a range of possible molecular effectors was suggested for 108 groups of LacI-TFs (Table S2 in Supplementary Material). Of these, effectors were previously known for regulators from 21 LacI-TF groups.

As expected, the overwhelming majority of the studied LacI-TF orthologous groups control carbohydrate metabolism (176/182; 96% of groups with assigned pathways). At that, most of the orthologous groups containing local LacI-TFs are assigned to specific carbohydrate utilization pathways. In contrast, five orthologous groups of local regulators including AdeR, HpxR, and UriR control the nucleoside utilization pathways, whereas a local regulator NtdR controls the neotrehalosadiamine biosynthesis. Most global regulators from the LacI family also are involved in the control of carbohydrate metabolism (FruR, CcpA, PckR, GapR), whereas PurR regulates several key metabolic pathways including purine and pyrimidine biosynthesis.

In agreement with the observed tendency of LacI-TF to control sugar metabolism, we report that carbohydrates constitute the largest class of effectors for these regulators (103/107; 96% of orthologous groups with assigned effectors). The majority of carbohydrate effectors assigned to the LacI-TF groups are monosaccharides and their derivatives (26/45), including hexoses (e.g., glucose, galactose, mannose), pentoses (e.g., ribose, xylose), sugar phosphates (e.g., fructose-1-phosphate, allose-6-phosphate), sugar acids (e.g., gluconate, galacturonate), sugar alcohols (e.g., ribitol), and amino sugars (e.g., N-acetylglucosamine). The second-largest category of carbohydrate effectors contains various oligosaccharides (14/26), including common disaccharides cellobiose, maltose, sucrose, and trehalose or their phosphorylated derivatives (e.g., cellobiose-6-phosphate, sucrose-6-phosphate). Finally, non-carbohydrate effectors of LacI-TFs include nucleobases (e.g., guanine and hypoxanthine are co-repressors of PurR in *E. coli* Schumacher et al., 1994), nucleosides (e.g., cytidine and adenosine are inducers of CytR in *E. coli* Barbier et al., 1997), and proteins (e.g., phosphoprotein HPr-Ser46-P is a co-repressor of CcpA in *B. subtilis* Schumacher et al., 2011).

Functional analysis of reconstructed LacI-TF regulons revealed that many sugar utilization pathways are regulated by two or more non-orthologous regulators. These include catabolic pathways for at least 10 distinct types of carbohydrates that are controlled by more than 90 orthologous groups of LacI-TFs (Table 1; Table S1 and Figure S1 in Supplementary Material). The observed large numbers of non-orthologous regulators for the glucoside and galactoside catabolic pathways correlate with structural diversity of glucose- and galactose-containing oligosaccharides that can be

**Table 1 | Sugar utilization pathways controlled by non-orthologous LacI-TFs.**

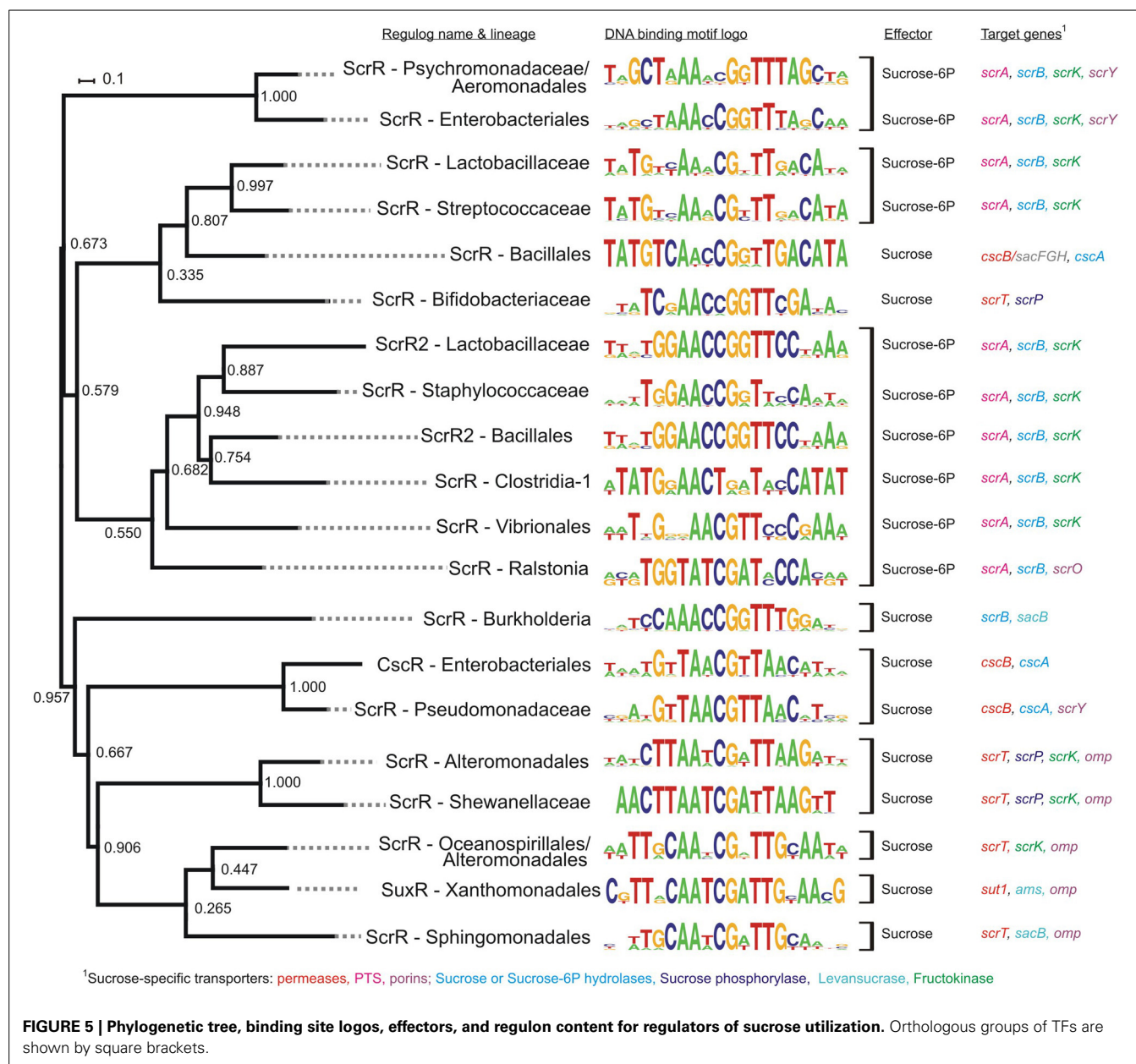
Sugar utilization pathway	LacI-TF orthologous groups	
	Number of groups	Regulator examples
Glucose and glucosides	20	BglR, BglZ, CelR, AscG, KojR
Sucrose	11	CscR, ScrR, SuxR
Galactose and galactosides	10	BgaR, GalR, GanR, MsmR, EbgR, LacI, LacR
Ribose	9	RbsR
Maltose and maltodextrins	8	MalR, MdxR, Mall
Inositol	7	IolR
Gluconate and idonate	7	GntR, IdnR
Gluconate and galacturonate	6	ExuR, KdgR, UxaR, UxuR
Mannose and mannoses	5	ManR
Fructose and fructooligosaccharides	5	FruR, BfrR
Trehalose	4	TreR, ThuR

utilized by bacteria in diverse natural habitats. On the other hand, the diversity of regulators for several other sugars including ribose and sucrose suggest a high frequency of convergent evolutionary events when the same ligand specificity has evolved independently in different branches of the LacI family.

For example, the sucrose utilization pathway is regulated by LacI-TFs from at least 11 orthologous groups from the phyla of Proteobacteria (11 lineages) and Firmicutes (5 lineages), as well as a single lineage of Actinobacteria (Figure 5). Analysis of the respective sucrose regulons revealed multiple distinct combinations of sucrose uptake transporters including permeases (*scrT*, *cscB*, *sut1*), phosphotransferase systems (PTSs) (*scrA*) and porins (*scrY*, *scrO*, *omp*), and sucrose catabolic enzymes including phosphorylases (*scrP*), hydrolases (*scrB*), and fructokinases (*scrK*). The effectors sucrose and sucrose-6-phosphate were assigned to the respective groups of ScrR regulators based on the type of regulated sucrose-specific transporters, i.e., a permease or a PTS, respectively. Interestingly, some families such as Bacillales and Enterobacteriales contain non-orthologous ScrR regulators with different effectors. As expected, DNA motifs of TFs controlling the sucrose catabolic pathway are well conserved within orthologous groups but are clearly different between non-orthologous regulators (Figure 5).

#### GLOBAL REGULONS FOR CENTRAL SUGAR METABOLISM

The LacI family contains a number of global regulators for central carbohydrate metabolic pathways, in particular the previously known regulators CcpA in Firmicutes and FruR in Enterobacteria. Here, we report a comparative genomics reconstruction of orthologous FruR regulons in  $\gamma$ -Proteobacteria, while the reconstructions of CcpA regulons in different lineages of Firmicutes has been reported previously (Ravcheev et al., 2011, 2013b; Antunes et al., 2012; Leyn et al., 2013) and is available in



the RegPrecise database. Further, we report identification of three novel non-orthologous regulons (named PckR, GapR, and GluR) that control central carbohydrate metabolism in three lineages of  $\alpha$ -Proteobacteria, namely, Rhizobiales, Rhodobacterales, and Caulobacteriales. Below we provide functional analysis of regulons for each of these four LacI-family regulators in Proteobacteria.

The fructose repressor FruR (also known as the catabolism repressor and activator Cra) is a global regulator of central metabolism in *E. coli* (Saier and Ramseier, 1996). FruR/Cra coordinates the carbon flow by repressing glycolytic genes involved in the Embden–Meyerhof, Entner–Doudoroff, and pentose-phosphate pathways and by activating gluconeogenesis genes. The comparative genomics reconstruction of orthologous FruR regulons in  $\gamma$ -Proteobacteria revealed that the regulon size

correlates with the taxonomy of studied groups and with the phylogeny of the FruR proteins (Figure 6). In Vibrionales and Pseudomonadales, the fructose utilization operon *fruBKA* and the *fruR* gene are the only members of the reconstructed FruR regulons; therefore, FruR operates as a local regulator in these lineages. In the Enterobacteriales, the FruR regulon is expanded to cover genes of the central glycolytic pathways, a part of the TCA cycle, and several fermentation and respiration pathways. Further regulon expansion to the genes of glyoxylate bypass is observed in the *Escherichia*, *Salmonella*, *Citrobacter*, *Enterobacter*, and *Klebsiella* species. Finally, in the closely related *Escherichia* and *Salmonella* species, the FruR/Cra regulon is expanded to include the Entner–Doudoroff pathway genes. Similar trends in the evolution of global regulons in closely-related bacterial species





were previously demonstrated for the PhoP regulon in enterobacteria (Perez and Groisman, 2009). On the other hand, in Pasteurellales, the regulon is degrading: FruR is absent in the *Haemophilus* spp. and *Actinobacillus pleuropneumoniae*, whereas in *Pasteurella multocida*, *Mannheimia succiniciproducens*, *A. succinogenes*, and *A. aphrophilus* the *fruR* gene is seemingly intact, but no candidate FruR-binding sites could be detected.

The hypothetical TF PckR (SMc02975 in *Sinorhizobium meliloti*) was previously annotated as a putative regulator of the phosphoenolpyruvate carboxykinase *pckA* (EMBL accession number AF004316.1); however, it had not yet been studied experimentally. Orthologs of PckR were found in 10 out of 15 analyzed genomes from the Rhizobiales order including species from the Rhizobiaceae, Brucellaceae, Phyllobacteriaceae, and Xanthobacteraceae families. By using the comparative genomics approach, we identified the putative PckR binding motif and reconstructed the PckR regulons in each of these genomes (Figure 7). Furthermore, by analyzing the binding-site position within promoter regions, we predicted a negative or positive mode of PckR regulation. As result, PckR was predicted to function as a dual transcriptional regulator that represses glycolytic genes from the Embden–Meyerhof and Entner–Doudoroff pathways (*glk*, *fba*, *pykA*, *zwf-pgl-edd*, *eda*) and activates genes from the gluconeogenesis and TCA cycle (*pckA*, *mdh-sucCDAB*, *sdhABCD*).

In the Xanthobacteraceae family, the reconstructed PckR regulon includes a minimal number of genes (*pckA*, *pckR*, *edd*, and *hpr*), whereas it is significantly expanded in the other three families of Rhizobiales and contains from 14 to 28 genes per genome. The most conserved members of the reconstructed PckR regulons include the *pckA* and *edd* genes (in 10 and 9 genomes, respectively), the *zwf-pgl* and *mdh-sucCDAB* operons (in 8 genomes), and the *fba*, *glk*, and *eda* genes (in 7 genomes). In most of the analyzed genomes, the *pckR* genes are not clustered with their target genes, which is a common feature of many global regulators in bacteria. However, in the genome of *Xanthobacter autotrophicus*, *pckR* is divergently transcribed with the 6-phosphogluconate dehydrogenase gene *gnd*, which is preceded by a putative PckR-binding site.

Orthologs of PckR and their cognate DNA motifs were found in Rhizobiales but not in other lineages, suggesting that the PckR regulon was introduced relatively recently in the evolution of  $\alpha$ -Proteobacteria. PckR from Rhizobiales can be considered as a partial functional replacement of the Enterobacterial Cra/FruR (see above) and the *Shewanella* HexR regulators (Leyn et al., 2011) that both play a pleiotropic role modulating the direction of carbon flow through different carbohydrate metabolic pathways. The molecular effector of PckR is unknown. A plausible hypothesis is that PckR dissociates from its DNA sites in response to an intermediate of the central glycolytic pathway in rhizobia.

A novel global transcriptional regulon for carbohydrate metabolism genes named GapR (RSP\_1663 in *Rhodobacter capsulatus*) was identified in all 13 studied genomes from the Rhodobacteraceae family (Figure 7). GapR was predicted to recognize a 20-bp palindromic DNA consensus, which is distinct from the PckR-binding consensus. The reconstructed GapR regulons contain from 7 to 18 genes per genome organized in

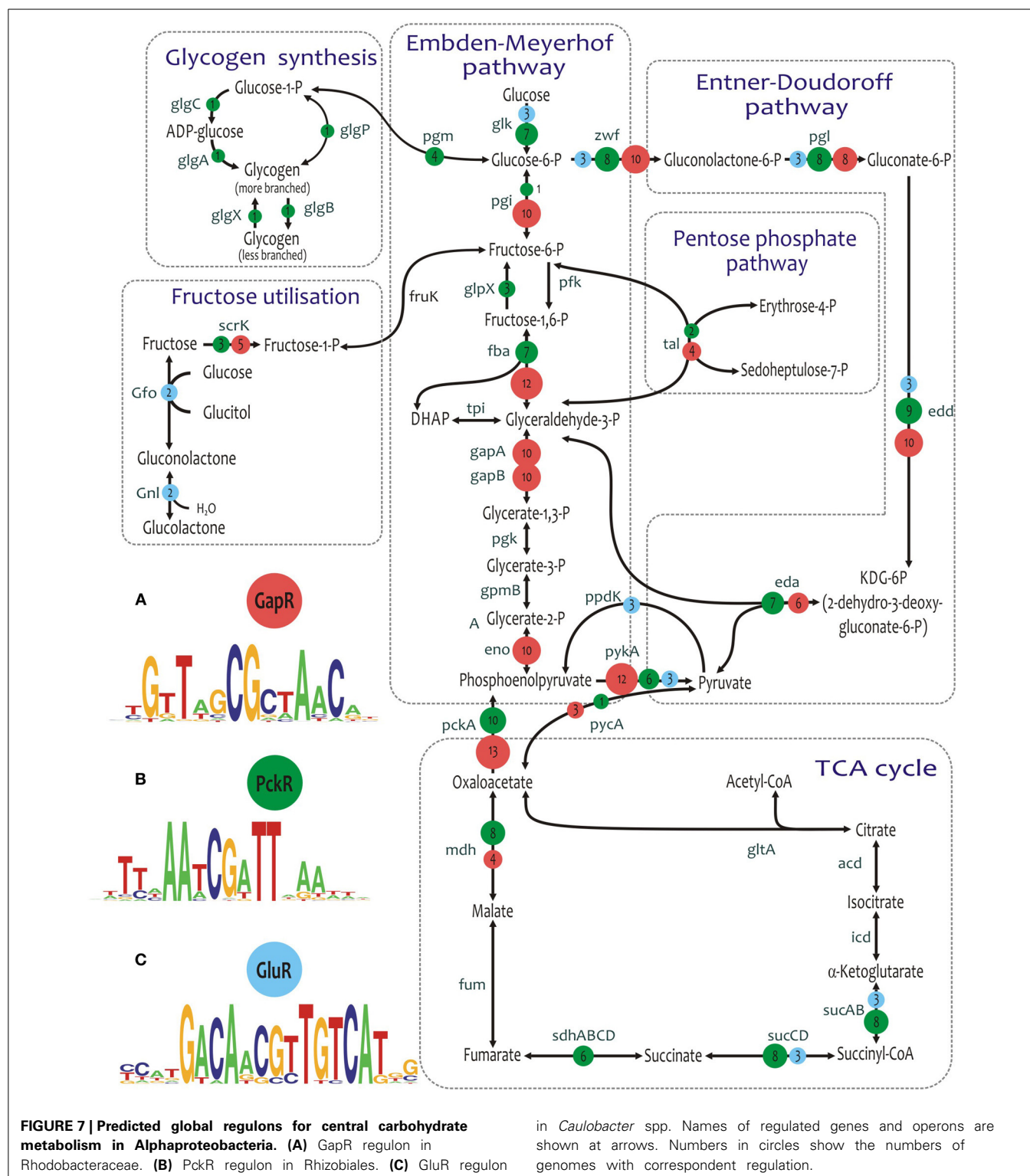
5–13 operons. GapR regulates glycolytic genes involved in the Embden–Meyerhof and Entner–Doudoroff pathways (*zwf-pgl-pgi*, *edd-eda*, *fba*, *gapA*, *gapB*, *eno*, *pykA*), gluconeogenesis (*pckA*, *pycA*), fructose utilization (*scrK*), pentose phosphate pathway (*tal*), and the TCA cycle (*mdh*). A similar DNA motif was identified upstream of the *gapR* genes in five genomes, suggesting their autoregulation. Similarly to PckR in Rhizobiales, the *gapR* genes do not cluster on the chromosome with their target genes in most genomes. However, *Roseobacter* possess two copies of *gapR*. One of these paralogs is divergently transcribed with the *gapB* gene, which is preceded by a putative GapR-binding site. The molecular mechanism and effector for GapR regulators remain to be elucidated.

Although PckR and GapR regulate overlapping sets of genes from the central carbohydrate metabolism in two distinct lineages of  $\alpha$ -Proteobacteria, these regulators are not orthologous to each other (Figure S1 in Supplementary Material) and recognize different DNA motifs (Figure 7). Another non-orthologous regulator from the LacI family named GluR (CC2053 in *Caulobacter crescentus*) was predicted to control the central carbohydrate metabolism in the Caulobacteraceae family (Figure 7). GluR recognizes a conserved 20-bp palindromic consensus, different from the PckR- and GapR-binding motifs above. The reconstructed GluR regulon in the *Caulobacter* spp. is composed of the glycolytic genes *zwf-pgl-edd-glK*, *pykA*, and *gnl-gfo*, as well as the gluconeogenic gene *ppdK* encoding pyruvate-phosphate dikinase and the *sucABCD* operon involved in the TCA cycle. The predicted regulatory gene *gluR* is located immediately downstream of the target operon *zwf-pgl-edd-glK* and is preceded by a conserved GluR-binding site in all three analyzed *Caulobacter* genomes.

Orthologs of GluR were not found in other  $\alpha$ -Proteobacteria, although bacteria from the Caulobacterales lineage have three paralogs (BglR1-3) that are predicted to control the cognate operons involved in the  $\beta$ -glucoside utilization (Figure S1 in Supplementary Material). The molecular effector of GluR is not known. Based on its close similarity to the predicted BglR repressors that possibly respond to  $\beta$ -glucosides and/or glucose, we propose that GluR dissociates from its DNA sites in response to glucose. In confirmation of our hypothesis, it has been demonstrated that glucose induces expression of the Entner–Doudoroff pathway genes and that the *edd* and *glk* genes are essential for the glucose utilization in *C. crescentus* (Hottes et al., 2004).

## DISCUSSION

We used the comparative genomics reconstruction of regulons for analysis of the LacI family of bacterial TFs. This choice was based on the following features. First, the LacI family is large, varied, and broadly distributed in bacteria. Second, proteins from this family have a rigid domain structure and a highly conserved structure of TFBS motifs. This study resulted in a detailed reconstruction of 1281 LacI-TF regulons in 272 bacterial genomes. Most (~90%) of the analyzed TF-LacI regulons are local, i.e., they control a small number of genes and operons that are involved in only one metabolic pathway. However, some LacI regulons are global, controlling tens to hundreds of genes involved in multiple metabolic pathways. In addition to the reconstruction of previously known global regulons, such as FruR/Cra, PurR,



and CcpA, we identified three novel regulators for the central carbohydrate metabolism in  $\alpha$ -Proteobacteria, PckR, GluR, and GapR, and reconstructed their corresponding regulons. For two of these global regulons, FruR and GluR, we reconstructed their possible evolutionary histories. Both these TFs likely originated

from local regulators during a process of gradual regulon expansion.

A large-scale phylogenetic analysis of LacI-TFs reveals numerous examples of various evolutionary processes for regulators and their regulons including divergent evolution (diversification of

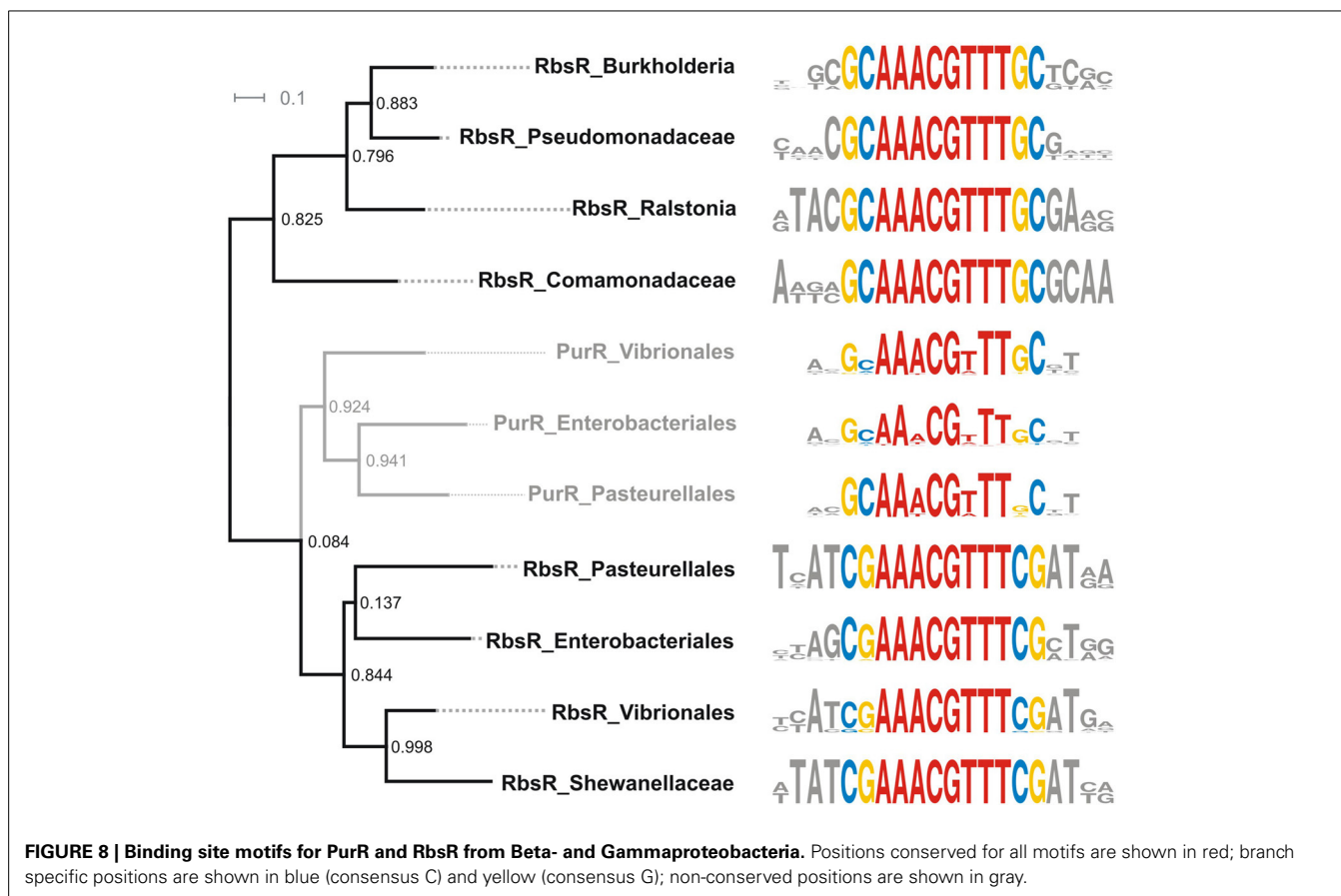


TF functions and binding specificities after duplication), convergent evolution (appearance of the same function in distantly related branches of a phylogenetic tree), and formation of paraphyletic groups (origin of novel functions and specificities, non-characteristic for a given branch of TFs). Below we discuss these evolutionary processes in more detail and provide examples of functional diversification within the LacI-TF family.

The LacI-TF phylogenetic tree demonstrates that some orthologous groups of TFs form branches consistent with the taxonomy, with TFs regulating orthologous genes, and with recognizing similar motifs, but those branches included an internal clade with demonstrated differences in the motif structure, effector specificity, or regulon content, i.e., these TFs formed so-called paraphyletic groups. The most interesting example of paraphyly is the branch of ribose repressors RbsR in  $\beta$ - and  $\gamma$ -Proteobacteria that have the purine repressor PurR as an excluded clade. Phylogenetic analysis revealed the presence of PurR orthologs in only three bacterial orders, Enterobacteriales, Pasteurellales, and Vibrionales. The closest PurR paralogs in these groups are ribose repressors (RbsRs) (Figure S1 in Supplementary Material). Most probably, PurR was originated by duplication of RbsR in the common ancestor of Enterobacteriales, Pasteurellales, and Vibrionales. The Pseudomonadales and  $\beta$ -Proteobacteria with a single RbsR repressor seemingly feature the ancestral state. The RbsR orthologs from the above three orders retain the ligand and specificity but have a slightly modified DNA binding motif,

compared to RbsR of Pseudomonadales and  $\beta$ -Proteobacteria (Figure 8). On the contrary, PurR retained the motif but changed the ligand and the default state, as it binds DNA in the presence of its ligand, whereas RbsR, like the majority of the LacI-TFs, binds DNA in the absence of the ligand.

This example illustrates a possible way of formation of paraphyletic branches by duplication of an ancestral gene with subsequent conversion of one copy of the copies, resulting in the origin of a novel function. In the case of RbsR and PurR, duplication and conversion are rather ancient, and we observe only the result of such evolution but cannot observe the intermediate states of the evolutionary process. On the contrary, for the  $\alpha$ -glucoside utilization regulon AglR and the trehalose regulon ThuR, such an intermediate state is clearly observable. AglR regulators from Rhizobiales and Rhodobacteriales form a paraphyletic branch with ThuR regulators from Rhizobiales as an excluded clade (Figure S1 in Supplementary Material). Both AglR and ThuR are local regulators, each controlling expression of the regulator genes and one other operon, divergently transcribed with the latter. The regulated operons contain homologous genes for kinases and ABC transporters, but non-homologous genes for hydrolases. The TFBS motif for ThuR (natcnAAAnCGnTTTngatt) is different from the one for AglR (nnntcAAAGCGCTTTgannn). Thus, during diversification, ThuR changed both the ligand and motif specificity. For the paraphyletic group CelR in  $\gamma$ -Proteobacteria, the AscG regulator in Enterobacteriales is an excluded clade. In



the case of the CelR and AscG regulators, their TFBS motifs and sets of regulated genes are drastically changed after duplication, but the effector specificity (cellobiose-6P) and the target metabolic pathway (cellobiose utilization) are retained.

The phylogenetic tree for the analyzed LacI-TFs (Figure S1 in Supplementary Material) is patchy, complicating the reconstruction of the evolutionary history. However, in some cases we can observe two or more adjacent branches, each corresponding to one orthologous group. A natural explanation is that these groups appeared as a result of initial duplication of a TF gene followed by diversification of copies. The LacI-TF tree contains multiple examples of monophyletic taxon-specific branches that consist of proteins from a single bacterial lineage. One such branch includes the *Bacteroides* UxaR, UxuR, and KdgR regulators that control the catabolic pathways for galacturonate, glucuronate, and 2-keto-3-deoxygluconate, respectively. Two other monophyletic branches include (i) kojibiose regulator KojR and the unknown sugar regulator Caur\_3448 from *Chloroflexus* and (ii) ribose regulator RbsR and uridine regulator UriR from *Corynebacteria* spp. Previously, a similar situation was observed for the ROK family of sugar-specific regulators in the deeply branched Thermotogales lineage (Kazanov et al., 2013).

Other examples of divergent evolution of regulator specificity are often demonstrated by adjacent branches in the phylogenetic tree. Thus, idonate repressor IdnR in Enterobacteriales and glucuronate repressor GntR in multiple lineages of  $\gamma$ -Proteobacteria are the closest paralogs (Figure S1 in Supplementary Material). In the IdnR regulons, the *idnK* and *idnT* genes are the closest homologs of the GntR-regulated genes *gntK* and *gntU*, respectively. Thus, duplication affected not only a TF gene, but also some of the regulated genes. The ability of GntR to recognize IdnR-binding sites upstream of the *idnK* and *idnDOTR* operons in *E. coli* (Tsunedomi et al., 2003b) also confirms a recent duplication of these regulators. Structural similarity of sugar effectors for GntR and IdnR also points to a recent duplication and further specialization of IdnR in Enterobacteriales.

Another scenario of TF diversification is duplication of an ancestral TF gene followed by acquisition of novel regulated genes and, accordingly, new effector specificity. An example is provided by a branch containing fructose (FruR) and sucrose (ScrR) orthologous groups in  $\gamma$ -Proteobacteria. Sucrose is a fructose-containing disaccharide. Because the FruR- and ScrR-regulated genes are functionally and structurally different, the regulator was most probably duplicated alone, and then one copy acquired new regulated genes. The acquisition of a novel regulatory function was coupled with the changes in the cognate TFBS motif. Here, divergent TFs have structurally similar effectors, fructose-1,6-biphosphate for FruR and sucrose-6-phosphate for ScrR.

Based on the analysis of paraphyletic branches and adjacent monophyletic branches, three main types of the origin of TF with novel functions can be described. The first type is duplication of both the TF gene and a regulated gene or operon followed by diversification, as in the case of ThuR in Rhizobiales. During diversification, the TF and regulated genes change their specificities, some regulated genes may be lost, and novel genes may be included in the regulon. The second type is duplication of a regulator followed by acquisition of novel regulated genes, as for PurR

in  $\gamma$ -Proteobacteria or for AscG in Enterobacteriales. The third type is rare, with only one example—the SCO5692 regulon in Streptomycetaceae. In this case, novel specificities for TFs originated without duplication, probably resulting from the loss of regulated genes.

In the process of acquisition of a new function, three characteristics of a TF regulon can be changed: (i) a set of regulated genes, (ii) effector specificity, and (iii) a TFBS motif structure. In most cases of TFs with a novel function, we observed the change of at least two of these characteristics. Change of all three characteristics is rarer and is usually observed in TFs from deeply branched lineages of bacteria such as Thermotogales. Most probably, in these cases change of all three characteristics is a result of a long evolutionary history. Change of only one of these characteristics is observed for only recently duplicated TFs, for which the diversification process is not yet complete.

In summary, the obtained extensive dataset for the LacI-TF family provides numerous examples of various evolutionary processes for regulators and their regulons. These data are publicly available in the RegPrecise database within the LacI family collection, which will enable further detailed analysis of signature residues in both DNA- and ligand-binding domains of regulators and establishment of the correlations between these residues and specificities toward the DNA motifs and molecular effectors they recognize.

## AUTHOR CONTRIBUTIONS

Dmitry A. Rodionov, Pavel S. Novichkov, and Mikhail S. Gelfand conceived and designed the research project. Dmitry A. Ravcheev, Dmitry A. Rodionov, and Mikhail S. Gelfand wrote the manuscript. Dmitry A. Ravcheev, Dmitry A. Rodionov, Matvei S. Khoroshkin, Olga N. Laikova, Olga V. Tsoy, Natalia V. Sernova, and Svetlana A. Petrova performed comparative genomics analysis to reconstruct TF regulons. Dmitry A. Rodionov provided the quality control of annotated regulons in the RegPrecise database. Matvei S. Khoroshkin analyzed statistical properties of TF regulons. All authors read and approved the final manuscript.

## ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (14-04-00870, 14-04-91154) and by the Russian Academy of Sciences via the programs “Molecular and Cellular Biology” and “Living Nature.” This research was also partially supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), and U.S. Department of Energy (DOE) and is a contribution of the Pacific Northwest National Laboratory (PNNL) Foundational Scientific Focus Area. Preliminary reconstruction of regulons was done by students of the Spring 2013 Bioinformatics course at the Faculty of Bioengineering and Bioinformatics in the Lomonosov Moscow State University as a part of regular coursework.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fmicb.2014.00294/abstract>

## REFERENCES

- Ahn, S. K., Cuthbertson, L., and Nodwell, J. R. (2012). Genome context as a predictive tool for identifying regulatory targets of the TetR family transcriptional regulators. *PLoS ONE* 7:e50562. doi: 10.1371/journal.pone.0050562
- Antunes, A., Camiade, E., Monot, M., Courtois, E., Barbut, F., Sernova, N. V., et al. (2012). Global transcriptional control by glucose and carbon regulator CcpA in *Clostridium difficile*. *Nucleic Acids Res.* 40, 10701–10718. doi: 10.1093/nar/gks864
- Barbier, C. S., Short, S. A., and Seneor, D. F. (1997). Allosteric mechanism of induction of CytR-regulated gene expression. CytR repressor-cytidine interaction. *J. Biol. Chem.* 272, 16962–16971. doi: 10.1074/jbc.272.27.16962
- Camas, F. M., Alm, E. J., and Poyatos, J. F. (2010). Local gene regulation details a recognition code within the LacI transcriptional factor family. *PLoS Comput. Biol.* 6:e1000989. doi: 10.1371/journal.pcbi.1000989
- Cho, B. K., Federowicz, S. A., Embree, M., Park, Y. S., Kim, D., and Palsson, B. O. (2011). The PurR regulon in *Escherichia coli* K-12 MG1655. *Nucleic Acids Res.* 39, 6456–6464. doi: 10.1093/nar/gkr307
- Cipriano, M. J., Novichkov, P. N., Kazakov, A. E., Rodionov, D. A., Arkin, A. P., Gelfand, M. S., et al. (2013). RegTransBase—a database of regulatory sequences and interactions based on literature: a resource for investigating transcriptional regulation in prokaryotes. *BMC Genomics* 14:213. doi: 10.1186/1471-2164-14-213
- Conlan, S., Lawrence, C., and McCue, L. A. (2005). *Rhodopseudomonas palustris* regulons detected by cross-species analysis of alphaproteobacterial genomes. *Appl. Environ. Microbiol.* 71, 7442–7452. doi: 10.1128/AEM.71.11.7442-7452.2005
- Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. doi: 10.1101/gr.849004
- Dehal, P. S., Joachimiak, M. P., Price, M. N., Bates, J. T., Baumohl, J. K., Chivian, D., et al. (2010). MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Res.* 38, D396–D400. doi: 10.1093/nar/gkp919
- Desai, T. A., Rodionov, D. A., Gelfand, M. S., Alm, E. J., and Rao, C. V. (2009). Engineering transcription factors with novel DNA-binding specificity using comparative genomics. *Nucleic Acids Res.* 37, 2493–2503. doi: 10.1093/nar/gkp079
- Disz, T., Akhter, S., Cuevas, D., Olson, R., Overbeek, R., Vonstein, V., et al. (2010). Accessing the SEED genome databases via Web services API: tools for programmers. *BMC Bioinformatics* 11:319. doi: 10.1186/1471-2105-11-319
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Finn, R. D., Bateman, A., Clements, J., Coghill, P., Eberhardt, R. Y., Eddy, S. R., et al. (2014). Pfam: the protein families database. *Nucleic Acids Res.* 42, D222–D230. doi: 10.1093/nar/gkt1223
- Francke, C., Kerkhoven, R., Wels, M., and Siezen, R. J. (2008). A generic approach to identify Transcription Factor-specific operator motifs; Inferences for LacI-family mediated regulation in *Lactobacillus plantarum* WCF51. *BMC Genomics* 9:145. doi: 10.1186/1471-2164-9-145
- Franco, I. S., Mota, L. J., Soares, C. M., and De Sa-Nogueira, I. (2006). Functional domains of the *Bacillus subtilis* transcription factor AraR and identification of amino acids important for nucleoprotein complex assembly and effector binding. *J. Bacteriol.* 188, 3024–3036. doi: 10.1128/JB.188.8.3024-3036.2006
- Fujita, Y. (2009). Carbon catabolite control of the metabolic network in *Bacillus subtilis*. *Biosci. Biotechnol. Biochem.* 73, 245–259. doi: 10.1271/bbb.80479
- Fukami-Kobayashi, K., Tateno, Y., and Nishikawa, K. (2003). Parallel evolution of ligand specificity between LacI/GalR family repressors and periplasmic sugar-binding proteins. *Mol. Biol. Evol.* 20, 267–277. doi: 10.1093/molbev/msg038
- Gelfand, M. S. (2006). Evolution of transcriptional regulatory networks in microbial genomes. *Curr. Opin. Struct. Biol.* 16, 420–429. doi: 10.1016/j.sbi.2006.04.001
- Gelfand, M. S., and Rodionov, D. A. (2008). Comparative genomics and functional annotation of bacterial transporters. *Phys. Life Rev.* 5, 22–49. doi: 10.1016/j.plrev.2007.10.003
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010
- He, B., Choi, K. Y., and Zalkin, H. (1993). Regulation of *Escherichia coli* *glnB*, *prsA*, and *speA* by the purine repressor. *J. Bacteriol.* 175, 3598–3606.
- He, B., Smith, J. M., and Zalkin, H. (1992). *Escherichia coli* *purB* gene: cloning, nucleotide sequence, and regulation by *purR*. *J. Bacteriol.* 174, 130–136.
- Hottes, A. K., Meewan, M., Yang, D., Arana, N., Romero, P., McAdams, H. H., et al. (2004). Transcriptional profiling of *Caulobacter crescentus* during growth on complex and minimal media. *J. Bacteriol.* 186, 1448–1461. doi: 10.1128/JB.186.5.1448-1461.2004
- Huang, N., De Ingeniis, J., Galeazzi, L., Mancini, C., Korostelev, Y. D., Rakhmaninova, A. B., et al. (2009). Structure and function of an ADP-ribose-dependent transcriptional regulator of NAD metabolism. *Structure* 17, 939–951. doi: 10.1016/j.str.2009.05.012
- Huson, D. H., Richter, D. C., Rausch, C., Dezulian, T., Franz, M., and Rupp, R. (2007). Dendroscope: an interactive viewer for large phylogenetic trees. *BMC Bioinformatics* 8:460. doi: 10.1186/1471-2105-8-460
- Jacob, F., and Monod, J. (1959). Genes of structure and genes of regulation in the biosynthesis of proteins. *C. R. Hebd. Seances Acad. Sci.* 249, 1282–1284.
- Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* 3, 318–356. doi: 10.1016/S0022-2836(61)80072-7
- Kalinina, O. V., Mironov, A. A., Gelfand, M. S., and Rakhmaninova, A. B. (2004). Automated selection of positions determining functional specificity of proteins by comparative analysis of orthologous groups in protein families. *Protein Sci.* 13, 443–456. doi: 10.1110/ps.03191704
- Kalodimos, C. G., Bonvin, A. M., Salinas, R. K., Wechselberger, R., Boelens, R., and Kaptein, R. (2002). Plasticity in protein-DNA recognition: lac repressor interacts with its natural operator O1 through alternative conformations of its DNA-binding domain. *EMBO J.* 21, 2866–2876. doi: 10.1093/emboj/cdf318
- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 40, D109–D114. doi: 10.1093/nar/gkr988
- Kazakov, A. E., Rodionov, D. A., Alm, E., Arkin, A. P., Dubchak, I., and Gelfand, M. S. (2009). Comparative genomics of regulation of fatty acid and branched-chain amino acid utilization in proteobacteria. *J. Bacteriol.* 191, 52–64. doi: 10.1128/JB.01175-08
- Kazakov, A. E., Rodionov, D. A., Price, M. N., Arkin, A. P., Dubchak, I., and Novichkov, P. S. (2013). Transcription factor family-based reconstruction of singleton regulons and study of the Crp/Fnr, ArsR, and GntR families in *Desulfovibrionales* genomes. *J. Bacteriol.* 195, 29–38. doi: 10.1128/JB.01977-12
- Kazanov, M. D., Li, X., Gelfand, M. S., Osterman, A. L., and Rodionov, D. A. (2013). Functional diversification of ROK-family transcriptional regulators of sugar catabolism in the Thermotogae phylum. *Nucleic Acids Res.* 41, 790–803. doi: 10.1093/nar/gks1184
- Letunic, I., and Bork, P. (2011). Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy. *Nucleic Acids Res.* 39, W475–W478. doi: 10.1093/nar/gkr201
- Lewis, M. (2005). The *lac* repressor. *C. R. Biol.* 328, 521–548. doi: 10.1016/j.crv.2005.04.004
- Leyn, S. A., Kazanov, M. D., Sernova, N. V., Ermakova, E. O., Novichkov, P. S., and Rodionov, D. A. (2013). Genomic reconstruction of the transcriptional regulatory network in *Bacillus subtilis*. *J. Bacteriol.* 195, 2463–2473. doi: 10.1128/JB.00140-13
- Leyn, S. A., Li, X., Zheng, Q., Novichkov, P. S., Reed, S., Romine, M. F., et al. (2011). Control of Proteobacterial central carbon metabolism by the HexR transcriptional regulator. A case study in *Shewanella oneidensis*. *J. Biol. Chem.* 286, 35782–35794. doi: 10.1074/jbc.M111.267963
- Liu, J., Xu, X., and Stormo, G. D. (2008). The cis-regulatory map of *Shewanella* genomes. *Nucleic Acids Res.* 36, 5376–5390. doi: 10.1093/nar/gkn515
- Magrane, M., and Consortium, U. (2011). UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)* 2011:bar009. doi: 10.1093/database/bar009
- Mahr, K., Hillen, W., and Titgemeyer, F. (2000). Carbon catabolite repression in *Lactobacillus pentosus*: analysis of the *ccpA* region. *Appl. Environ. Microbiol.* 66, 277–283. doi: 10.1128/AEM.66.1.277-283.2000
- Manson McGuire, A., and Church, G. M. (2000). Predicting regulons and their cis-regulatory motifs by comparative genomics. *Nucleic Acids Res.* 28, 4523–4530. doi: 10.1093/nar/28.22.4523
- Mauzy, C. A., and Hermodson, M. A. (1992). Structural homology between *rbS* repressor and ribose binding protein implies functional similarity. *Protein Sci.* 1, 843–849. doi: 10.1002/pro.5560010702
- McCue, L., Thompson, W., Carmack, C., Ryan, M. P., Liu, J. S., Derbyshire, V., et al. (2001). Phylogenetic footprinting of transcription factor binding sites

- in proteobacterial genomes. *Nucleic Acids Res.* 29, 774–782. doi: 10.1093/nar/29.3.774
- Meinhardt, S., and Swint-Kruse, L. (2008). Experimental identification of specificity determinants in the domain linker of a LacI/GalR protein: bioinformatics-based predictions generate true positives and false negatives. *Proteins* 73, 941–957. doi: 10.1002/prot.22121
- Milk, L., Daber, R., and Lewis, M. (2010). Functional rules for lac repressor-operator associations and implications for protein-DNA interactions. *Protein Sci.* 19, 1162–1172. doi: 10.1002/pro.389
- Mirny, L. A., and Gelfand, M. S. (2002). Using orthologous and paralogous proteins to identify specificity determining residues. *Genome Biol.* 3:PREPRINT0002. doi: 10.1186/gb-2002-3-3-preprint0002
- Mironov, A. A., Vinokurova, N. P., and Gelfand, M. S. (2000). Software for analyzing bacterial genomes. *Mol. Biol.* 34, 253–262.
- Nguyen, C. C., and Saier, M. H. Jr. (1995). Phylogenetic, structural and functional analyses of the LacI-GalR family of bacterial transcription factors. *FEBS Lett.* 377, 98–102. doi: 10.1016/0014-5793(95)01344-X
- Novichkov, P. S., Kazakov, A. E., Ravcheev, D. A., Leyn, S. A., Kovaleva, G. Y., Sutormin, R. A., et al. (2013). RegPrecise 3.0 – A resource for genome-scale exploration of transcriptional regulation in bacteria. *BMC Genomics* 14:745. doi: 10.1186/1471-2164-14-745
- Novichkov, P. S., Rodionov, D. A., Stavrovskaya, E. D., Novichkova, E. S., Kazakov, A. E., Gelfand, M. S., et al. (2010). RegPredict: an integrated system for regulon inference in prokaryotes by comparative genomics approach. *Nucleic Acids Res.* 38, W299–W307. doi: 10.1093/nar/gkq531
- Osterman, A., and Overbeek, R. (2003). Missing genes in metabolic pathways: a comparative genomics approach. *Curr. Opin. Chem. Biol.* 7, 238–251. doi: 10.1016/S1367-5931(03)00027-9
- Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H. Y., Cohoon, M., et al. (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* 33, 5691–5702. doi: 10.1093/nar/gki866
- Parente, D. J., and Swint-Kruse, L. (2013). Multiple co-evolutionary networks are supported by the common tertiary scaffold of the LacI/GalR proteins. *PLoS ONE* 8:e84398. doi: 10.1371/journal.pone.0084398
- Pauling, J., Rottger, R., Tauch, A., Azevedo, V., and Baumbach, J. (2012). CoryneRegNet 6.0—Updated database content, new analysis methods and novel features focusing on community demands. *Nucleic Acids Res.* 40, D610–D614. doi: 10.1093/nar/gkr883
- Pei, J., Cai, W., Kinch, L. N., and Grishin, N. V. (2006). Prediction of functional specificity determinants from protein sequences using log-likelihood ratios. *Bioinformatics* 22, 164–171. doi: 10.1093/bioinformatics/bti766
- Perez, J. C., and Groisman, E. A. (2009). Transcription factor function and promoter architecture govern the evolution of bacterial regulons. *Proc. Natl. Acad. Sci. U.S.A.* 106, 4319–4324. doi: 10.1073/pnas.0810343106
- Ravcheev, D. A., Best, A. A., Sernova, N. V., Kazanov, M. D., Novichkov, P. S., and Rodionov, D. A. (2013b). Genomic reconstruction of transcriptional regulatory networks in lactic acid bacteria. *BMC Genomics* 14:94. doi: 10.1186/1471-2164-14-94
- Ravcheev, D., Godzik, A., Osterman, A., and Rodionov, D. (2013a). Polysaccharides utilization in human gut bacterium *Bacteroides thetaiotaomicron*: comparative genomics reconstruction of metabolic and regulatory networks. *BMC Genomics* 14:873. doi: 10.1186/1471-2164-14-873
- Ravcheev, D. A., Best, A. A., Tintle, N., Dejongh, M., Osterman, A. L., Novichkov, P. S., et al. (2011). Inference of the transcriptional regulatory network in *Staphylococcus aureus* by integration of experimental and genomics-based evidence. *J. Bacteriol.* 193, 3228–3240. doi: 10.1128/JB.00350-11
- Ravcheev, D. A., Gelfand, M. S., Mironov, A. A., and Rakhmaninova, A. B. (2002). Purine regulon of gamma-proteobacteria: a detailed description. *Genetika* 38, 1203–1214. doi: 10.1023/A:1020231513079
- Ravcheev, D. A., Gerasimova, A. V., Mironov, A. A., and Gelfand, M. S. (2007). Comparative genomic analysis of regulation of anaerobic respiration in ten genomes from three families of gamma-proteobacteria (Enterobacteriaceae, Pasteurellaceae, Vibrionaceae). *BMC Genomics* 8:54. doi: 10.1186/1471-2164-8-54
- Ravcheev, D. A., Li, X., Latif, H., Zengler, K., Leyn, S. A., Korostelev, Y. D., et al. (2012). Transcriptional regulation of central carbon and energy metabolism in bacteria by redox responsive repressor Rex. *J. Bacteriol.* 194, 1145–1157. doi: 10.1128/JB.06412-11
- Rigali, S., Schlicht, M., Hoskisson, P., Nothaft, H., Merzbacher, M., Joris, B., et al. (2004). Extending the classification of bacterial transcription factors beyond the helix-turn-helix motif as an alternative approach to discover new cis/trans relationships. *Nucleic Acids Res.* 32, 3418–3426. doi: 10.1093/nar/gkh673
- Rodionov, D. A. (2007). Comparative genomic reconstruction of transcriptional regulatory networks in bacteria. *Chem. Rev.* 107, 3467–3497. doi: 10.1021/cr068309+
- Rodionov, D. A., and Gelfand, M. S. (2005). Identification of a bacterial regulatory system for ribonucleotide reductases by phylogenetic profiling. *Trends Genet.* 21, 385–389. doi: 10.1016/j.tig.2005.05.011
- Rodionov, D. A., Gelfand, M. S., Todd, J. D., Curson, A. R., and Johnston, A. W. (2006). Computational reconstruction of iron- and manganese-responsive transcriptional networks in alpha-proteobacteria. *PLoS Comput. Biol.* 2:e163. doi: 10.1371/journal.pcbi.0020163
- Rodionov, D. A., Novichkov, P. S., Stavrovskaya, E. D., Rodionova, I. A., Li, X., Kazanov, M. D., et al. (2011). Comparative genomic reconstruction of transcriptional networks controlling central metabolism in the *Shewanella* genus. *BMC Genomics* 12(Suppl. 1):S3. doi: 10.1186/1471-2164-12-S1-S3
- Rodionov, D. A., Rodionova, I. A., Li, X., Ravcheev, D. A., Tarasova, Y., Portnoy, V. A., et al. (2013). Transcriptional regulation of the carbohydrate utilization network in *Thermotoga maritima*. *Front. Microbiol.* 4:244. doi: 10.3389/fmicb.2013.00244
- Sahota, G., and Stormo, G. D. (2010). Novel sequence-based method for identifying transcription factor binding sites in prokaryotic genomes. *Bioinformatics* 26, 2672–2677. doi: 10.1093/bioinformatics/btq501
- Saier, M. H., and Ramseier, T. M. (1996). The catabolite repressor/activator (Cra) protein of enteric bacteria. *J. Bacteriol.* 178, 3411–3417.
- Salgado, H., Peralta-Gil, M., Gama-Castro, S., Santos-Zavaleta, A., Muniz-Rascado, L., Garcia-Sotelo, J. S., et al. (2013). RegulonDB v8.0: omics data sets, evolutionary conservation, regulatory phrases, cross-validated gold standards and more. *Nucleic Acids Res.* 41, D203–D213. doi: 10.1093/nar/gks1201
- Schumacher, M. A., Choi, K. Y., Zalkin, H., and Brennan, R. G. (1994). Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices. *Science* 266, 763–770. doi: 10.1126/science.7973627
- Schumacher, M. A., Sprehe, M., Bartholomae, M., Hillen, W., and Brennan, R. G. (2011). Structures of carbon catabolite protein A (HPr-Ser46-P) bound to diverse catabolite response element sites reveal the basis for high-affinity binding to degenerate DNA operators. *Nucleic Acids Res.* 39, 2931–2942. doi: 10.1093/nar/gkq1177
- Seidl, K., Muller, S., Francois, P., Kriebitzsch, C., Schrenzel, J., Engelmann, S., et al. (2009). Effect of a glucose impulse on the CcpA regulon in *Staphylococcus aureus*. *BMC Microbiol.* 9:95. doi: 10.1186/1471-2180-9-95
- Shelton, D. A., Stegman, L., Hardison, R., Miller, W., Bock, J. H., Slightom, J. L., et al. (1997). Phylogenetic footprinting of hypersensitive site 3 of the beta-globin locus control region. *Blood* 89, 3457–3469.
- Sierro, N., Makita, Y., De Hoon, M., and Nakai, K. (2008). DBTBS: a database of transcriptional regulation in *Bacillus subtilis* containing upstream intergenic conservation information. *Nucleic Acids Res.* 36, D93–D96. doi: 10.1093/nar/gkm910
- Sonenshein, A. L. (2007). Control of key metabolic intersections in *Bacillus subtilis*. *Nat. Rev. Microbiol.* 5, 917–927. doi: 10.1038/nrmicro1772
- Suvorova, I. A., Ravcheev, D. A., and Gelfand, M. S. (2012). Regulation and evolution of the malonate and propionate catabolism in the proteobacteria. *J. Bacteriol.* 194, 3234–3240. doi: 10.1128/JB.00163-12
- Suvorova, I. A., Tutukina, M. N., Ravcheev, D. A., Rodionov, D. A., Ozoline, O. N., and Gelfand, M. S. (2011). Comparative genomic analysis of the hexuronate metabolism genes and their regulation in gamma-proteobacteria. *J. Bacteriol.* 193, 3228–3240. doi: 10.1128/JB.00277-11
- Swint-Kruse, L., and Matthews, K. S. (2009). Allostery in the LacI/GalR family: variations on a theme. *Curr. Opin. Microbiol.* 12, 129–137. doi: 10.1016/j.mib.2009.01.009
- Tan, K., Moreno-Hagelsieb, G., Collado-Vides, J., and Stormo, G. D. (2001). A comparative genomics approach to prediction of new members of regulons. *Genome Res.* 11, 566–584. doi: 10.1101/gr.149301
- Tsunedomi, R., Izu, H., Kawai, T., Matsushita, K., Ferenci, T., and Yamada, M. (2003a). The activator of GntII genes for gluconate metabolism, GntH, exerts negative control of GntR-regulated GntI genes in *Escherichia coli*. *J. Bacteriol.* 185, 1783–1795. doi: 10.1128/JB.185.6.1783-1795.2003

- Tsunedomi, R., Izu, H., Kawai, T., and Yamada, M. (2003b). Dual control by regulators, GntH and GntR, of the GntII genes for gluconate metabolism in *Escherichia coli*. *J. Mol. Microbiol. Biotechnol.* 6, 41–56. doi: 10.1159/000073407
- Tungtur, S., Egan, S. M., and Swint-Kruse, L. (2007). Functional consequences of exchanging domains between LacI and PurR are mediated by the intervening linker sequence. *Proteins* 68, 375–388. doi: 10.1002/prot.21412
- Tungtur, S., Meinhardt, S., and Swint-Kruse, L. (2010). Comparing the functional roles of nonconserved sequence positions in homologous transcription repressors: implications for sequence/function analyses. *J. Mol. Biol.* 395, 785–802. doi: 10.1016/j.jmb.2009.10.001
- Tungtur, S., Parente, D. J., and Swint-Kruse, L. (2011). Functionally important positions can comprise the majority of a protein's architecture. *Proteins* 79, 1589–1608. doi: 10.1002/prot.22985
- Van Gijsegem, F., Wlodarczyk, A., Cornu, A., Reverchon, S., and Hugouvieux-Cotte-Pattat, N. (2008). Analysis of the LacI family regulators of *Erwinia chrysanthemi* 3937, involvement in the bacterial phytopathogenicity. *Mol. Plant Microbe Interact.* 21, 1471–1481. doi: 10.1094/MPMI-21-11-1471
- Weickert, M. J., and Adhya, S. (1992). A family of bacterial regulators homologous to Gal and Lac repressors. *J. Biol. Chem.* 267, 15869–15874.
- Wels, M., Francke, C., Kerkhoven, R., Kleerebezem, M., and Siezen, R. J. (2006). Predicting cis-acting elements of *Lactobacillus plantarum* by comparative genomics with different taxonomic subgroups. *Nucleic Acids Res.* 34, 1947–1958. doi: 10.1093/nar/gkl138
- Zheng, L., Chen, Z., Itzek, A., Ashby, M., and Kreth, J. (2011). Catabolite control protein A controls hydrogen peroxide production and cell death in *Streptococcus sanguinis*. *J. Bacteriol.* 193, 516–526. doi: 10.1128/JB.01131-10
- Zheng, L., Chen, Z., Itzek, A., Herzberg, M. C., and Kreth, J. (2012). CcpA regulates biofilm formation and competence in *Streptococcus gordonii*. *Mol. Oral Microbiol.* 27, 83–94. doi: 10.1111/j.2041-1014.2011.00633.x
- Zotta, T., Ricciardi, A., Guidone, A., Sacco, M., Muscariello, L., Mazzeo, M. F., et al. (2012). Inactivation of *ccpA* and aeration affect growth, metabolite production and stress tolerance in *Lactobacillus plantarum* WCFS1. *Int. J. Food Microbiol.* 155, 51–59. doi: 10.1016/j.ijfoodmicro.2012.01.017

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 06 April 2014; paper pending published: 12 May 2014; accepted: 28 May 2014; published online: 11 June 2014.

Citation: Ravcheev DA, Khoroshkin MS, Laikova ON, Tsoy OV, Sernova NV, Petrova SA, Rakhmaninova AB, Novichkov PS, Gelfand MS and Rodionov DA (2014) Comparative genomics and evolution of regulons of the LacI-family transcription factors. *Front. Microbiol.* 5:294. doi: 10.3389/fmicb.2014.00294

This article was submitted to *Microbial Physiology and Metabolism*, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Ravcheev, Khoroshkin, Laikova, Tsoy, Sernova, Petrova, Rakhmaninova, Novichkov, Gelfand and Rodionov. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Connecting lignin-degradation pathway with pre-treatment inhibitor sensitivity of *Cupriavidus necator*

Wei Wang<sup>1\*†</sup>, Shihui Yang<sup>2\*†</sup>, Glendon B. Hunsinger<sup>2</sup>, Philip T. Pienkos<sup>2</sup> and David K. Johnson<sup>1</sup>

<sup>1</sup> National Renewable Energy Laboratory, Biosciences Center, Golden, CO, USA

<sup>2</sup> National Renewable Energy Laboratory, National Bioenergy Center, Golden, CO, USA

## Edited by:

Katherine M. Pappas, University of Athens, Greece

## Reviewed by:

Harold J. Schreier, University of Maryland Baltimore County, USA  
Claudio Avignone-Rossa, University of Surrey, UK

## \*Correspondence:

Wei Wang and Shihui Yang,  
15013 Denver West Parkway,  
Golden, 80401 CO, USA  
e-mail: wei.wang@nrel.gov;  
shihui.yang@nrel.gov

<sup>†</sup> These authors have contributed equally to this work.

To produce lignocellulosic biofuels economically, the complete release of monomers from the plant cell wall components, cellulose, hemicellulose, and lignin, through pre-treatment and hydrolysis (both enzymatic and chemical), and the efficient utilization of these monomers as carbon sources, is crucial. In addition, the identification and development of robust microbial biofuel production strains that can tolerate the toxic compounds generated during pre-treatment and hydrolysis is also essential. In this work, *Cupriavidus necator* was selected due to its capabilities for utilizing lignin monomers and producing polyhydroxybutyrate (PHB), a bioplastic as well as an advanced biofuel intermediate. We characterized the growth kinetics of *C. necator* in pre-treated corn stover slurry as well as individually in the pre-sence of 11 potentially toxic compounds in the saccharified slurry. We found that *C. necator* was sensitive to the saccharified slurry produced from dilute acid pre-treated corn stover. Five out of 11 compounds within the slurry were characterized as toxic to *C. necator*, namely ammonium acetate, furfural, hydroxymethylfurfural (HMF), benzoic acid, and p-coumaric acid. Aldehydes (e.g., furfural and HMF) were more toxic than the acetate and the lignin degradation products benzoic acid and p-coumaric acid; furfural was identified as the most toxic compound. Although toxic to *C. necator* at high concentration, ammonium acetate, benzoic acid, and p-coumaric acid could be utilized by *C. necator* with a stimulating effect on *C. necator* growth. Consequently, the lignin degradation pathway of *C. necator* was reconstructed based on genomic information and literature. The efficient conversion of intermediate catechol to downstream products of cis,cis-muconate or 2-hydroxymuconate-6-semialdehyde may help improve the robustness of *C. necator* to benzoic acid and p-coumaric acid as well as improve PHB productivity.

**Keywords:** *Cupriavidus necator*, pre-treatment inhibitor, saccharified slurry, deacetylation, lignin degradation, biofuel, polyhydroxybutyrate (PHB), genomics

## INTRODUCTION

Lignocellulosic biomass is considered a renewable and sustainable source for energy production. The great environmental, energy security, and economic benefits of biofuels have driven research on biomass throughout the world, resulting in several pilot and demonstration scale projects, based primarily on cellulosic ethanol. One of the leading routes for cellulosic ethanol production is based on the deconstruction of biomass to a monomeric sugar solution produced by chemical pre-treatment and enzymatic saccharification termed a saccharified slurry (hereafter referred to as slurry) followed by fermentation of the sugars to ethanol. However, the physico-chemical properties of ethanol limit its further penetration into the current petroleum-based transportation fuel infrastructure (Serrano-Ruiz and Dumesic, 2011). The lower energy density of ethanol due to its higher oxygen content compared to conventional petroleum hydrocarbons, is a significant disadvantage to using bioethanol as a fuel alternative.

Current transportation fuel infrastructure has been established for hydrocarbon-based petroleum. Many advanced biofuels under development exhibit advantages such as higher energy

densities, longer carbon chains and lower oxygen numbers. For example, there are some fuels derived from fatty acids with carbon numbers in the range from C<sub>12</sub> to C<sub>22</sub> that can be upgraded to high-energy-density hydrocarbons via hydrotreating. Therefore, development of hydrocarbon fuels, which are compatible with the current fuel infrastructure as replacements or blendstocks for gasoline, jet, and diesel, has recently received great attention.

The primary routes leading to hydrocarbon synthesis from biomass are thermochemical, biochemical, and hybrid approaches which employ elements of both. Bacteria, yeasts, or fungi can naturally synthesize fatty acids, isoprenoids, and polyalkanoates for energy storage. Although these compounds are mostly exploited in pharmaceutical, nutritional, and packaging sectors, they also have great potential for production of hydrocarbon fuels (Zhang et al., 2011). However, several obstacles need to be solved for future hydrocarbon production at the commercial scale. Firstly, although compounds, such as acetate, furfural, and HMF that are produced during chemical pre-treatments (e.g., dilute acid pre-treatment) and have been investigated extensively due to their toxicity to ethanologens (Olsson and Hahn-Hagerdal, 1996; Liu et al., 2004, 2005, 2008; Gorsich et al., 2006; Endo et al.,

2008; Franden et al., 2009, 2013; Allen et al., 2010; Bowman et al., 2010; Yang et al., 2010a,b, 2012b; He et al., 2012; Ask et al., 2013; Bajwa et al., 2013; Wilson et al., 2013), there are few reports on the effect of these compounds on hydrocarbon producers (Huang et al., 2012). Various mechanisms and approaches have been proposed and applied for decreasing inhibition in cellulosic hydrolysates (Olsson and Hahn-Hagerdal, 1996; Larsson et al., 1999; Zaldivar and Ingram, 1999; Zaldivar et al., 1999; Petersson et al., 2006; Endo et al., 2008; Franden et al., 2009, 2013; Mills et al., 2009; Yang et al., 2010a,b, 2012a,b; Ask et al., 2013; Bajwa et al., 2013; Iwaki et al., 2013), but strategies for strain improvement have yet to be applied (Dunlop, 2011; Dunlop et al., 2011; Kang and Chang, 2012).

Moreover, the efficient utilization of lignin, a major cell component, would greatly improve the economics of advanced biofuel production. Due to its highly recalcitrant aromatic structure, lignin is seen largely as an impediment to biochemical processes due to its reinforcement of the cellulose/hemicellulose matrix complicating the saccharification process (Zeng et al., 2014). Owing to the high energy state of its aromatic ring structures, lignin could have great value as a substrate for energy production, but so far very limited opportunities for microbial transformation of lignin to high value-added fuels or fuel intermediates have been identified. Some bacteria are known to degrade lignin monomers via the  $\beta$ -ketoadipate pathway (Harwood and Parales, 1996). Currently lignin is almost untouched in lignocellulosic ethanol fermentations and remains behind in the fermentation residue after distillation. These residues are mostly burned to provide power in current cellulosic ethanol production processes, which places a low value on this material. To reach the goal of utilizing lignin as a carbon source for advanced biofuel production, the composition of lignin degradation compounds as well as the toxic effects of these compounds on hydrocarbon-producing microorganisms should be investigated.

In this study we are seeking a biological understanding of the impact of model inhibitors, which include lignin degradation products as well as furans and acetate, on hydrocarbon-producing microorganisms. *Cupriavidus necator* (Syn. *Alcaligenes eutrophus*, *Ralstonia eutropha*) has been extensively studied for production of polyhydroxybutyrate (PHB), which consists of a C<sub>4</sub> repeating unit that can be thermally depolymerized and then decarboxylated to propene (Fischer et al., 2011; Pilath et al., 2013), an intermediate, which can be upgraded to hydrocarbon fuels via commercial oligomerization technologies. *C. necator* has been reported to be able to utilize lignin monomers as a carbon source (Pérez-Pantoja et al., 2008).

The genome sequence of *C. necator* H16 has been published with *in silico* genome modeling and a developed genetics system (Pohlmann et al., 2006; Park et al., 2011; Brigham et al., 2012). In addition, several transcriptomic studies have recently been reported (Peplinski et al., 2010; Brigham et al., 2012), and the genome sequences for a number of other *Cupriavidus* spp. are also now available (Amadou et al., 2008; Pérez-Pantoja et al., 2008; Janssen et al., 2010; Lykidis et al., 2010; Poehlein et al., 2011; Cserhati et al., 2012; Hong et al., 2012; Van Houdt et al., 2012; Li et al., 2013). This information will facilitate future comparative genomics and systems biology studies to develop *C. necator* H16

as a robust and metabolically diverse hydrocarbon-intermediate production strain. Genomics is applied in this study to explore the metabolic pathways related to lignin utilization and response to toxic compounds in slurries, which will provide perspectives for strain metabolic engineering toward future economic hydrocarbon production using lignin.

## MATERIALS AND METHODS

### STRAINS AND MEDIA

The strain used in this study is a glucose-utilizing mutant of *C. necator* H16 (wild-type H16 is not able to metabolize glucose) (Orita et al., 2012), *C. necator* 11599, which was purchased from NCIMB culture collection. It is routinely cultured in LB at 37°C. A minimal medium recipe was selected for this study (Cavalheiro et al., 2009). Specifically, the defined minimal medium for *C. necator* (per liter, pH 6.8) was: 10 g glucose, 1.0 g (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 1.5 g KH<sub>2</sub>PO<sub>4</sub>, 9 g Na<sub>2</sub>HPO<sub>4</sub> 12H<sub>2</sub>O, 0.2 g MgSO<sub>4</sub> 7H<sub>2</sub>O, 1.0 mL trace element solution. The Trace Element Solution (per liter): 10 g FeSO<sub>4</sub> 7H<sub>2</sub>O, 2.25 g ZnSO<sub>4</sub> 7H<sub>2</sub>O, 0.5 g MnSO<sub>4</sub> 5H<sub>2</sub>O, 2 g CaCl<sub>2</sub> 2H<sub>2</sub>O, and 1 g CuSO<sub>4</sub> 5H<sub>2</sub>O, 0.23 g Na<sub>2</sub>B<sub>4</sub>O<sub>7</sub> 10H<sub>2</sub>O, 0.1 g (NH<sub>4</sub>)<sub>6</sub>Mo<sub>7</sub>O<sub>24</sub>, 10 mL 35% HCl.

### PRODUCTION OF SACCHARIFIED SLURRY AND MOCK MEDIA

A deacetylated saccharified slurry, which was produced from the modified sulfuric acid pre-treatment and enzymatic hydrolysis of corn stover including an added deacetylation step before pre-treatment, was used in this study (Chen et al., 2012). The composition of the mock sugar media simulating the saccharified slurry is summarized in **Table 1**. The composition is based on the composition of the saccharified slurry in fermentation media at the level of 20% total solids.

### GROWTH OF *C. NECATOR* ON SACCHARIFIED SLURRY

*Cupriavidus necator* was first grown in 5 mL of LB in 125 mL baffled flasks, cultured at 200 rpm, and 37°C. After 1 day, a 10% inoculum was added to 50 mL of fermentation media in a 250 mL flask and incubated in a shaker at 37°C and 180 rpm for 4 days. The fermentation media contained either mock sugar slurry as shown in **Table 1** or saccharified slurry supplemented with tryptone (10 g/L) and yeast extract (5 g/L) as nutrients. Mock slurry was added at a level to achieve the same sugar concentrations (e.g., the glucose concentration in the 2X-diluted mock medium was 50 g/L). All experiments were run in duplicate.

### PHB ANALYSIS

The PHB content of the bacterial cells was determined by a quantitative method that used HPLC analysis to measure the crotonic acid formed by acid-catalyzed depolymerization of PHB (Karr et al., 1983). Cell mass samples were freeze-dried before analysis. PHB-containing dried bacterial cells (15–50 mg) were then digested in 96% H<sub>2</sub>SO<sub>4</sub> (1 mL) at 90°C for 1 h. The reaction

**Table 1 | Composition of mock saccharified hydrolysate slurry.**

	Glucose (g/L)	Xylose (g/L)	Acetate (mM)
Mock slurry (1X)	100.4	51.4	36.6

vials were then cooled on ice, after which, ice-cold 0.01N H<sub>2</sub>SO<sub>4</sub> (4 mL) was added followed by rapid mixing. The samples were further diluted 20- to 150-fold with 0.01N H<sub>2</sub>SO<sub>4</sub> before analysis by HPLC.

The concentration of crotonic acid was measured at 210 nm using an HPLC equipped with a photodiode array detector (Agilent 1100, Agilent Technologies, Palo Alto, CA). A Rezex RFQ Fast Acids column (100 × 7.8 mm, 8 μm particle size, Phenomenex, Torrance, CA) and Cation H+ guard column (BioRad Laboratories, CA) operated at 85°C were used to separate the crotonic acid present in the reaction solutions. The eluent was 0.01N H<sub>2</sub>SO<sub>4</sub> at a flow rate of 1.0 mL min<sup>-1</sup>. Samples and crotonic acid standards were filtered through 0.45 μm pore size nylon membrane syringe filters (Pall Corp., NY) prior to injection onto the column. The HPLC was controlled and data were analyzed using Agilent ChemStation software (Rev.B.03.02).

### CHARACTERIZATION OF POTENTIAL INHIBITORS

The selection of chemicals for sensitivity assays was based on analyses of saccharified slurry by GC-MS and LC-MS analyses performed at NREL as well as ICP-MS analyses carried out by Huffman Labs (Golden, CO). The compounds identified with high concentrations in the slurry and their potential derivatives were selected for further investigation. The compounds included ammonium (added to neutralize pre-treated corn stover) with two common anions of acetate (released by hydrolysis of hemicelluloses) and sulfate (from the sulfuric acid pre-treatment); sugar degradation products furfural and HMF; lignin monomers vanillin, coumaric acid, ferulic acid, and 4-hydroxybenzaldehyde as well as benzoic acid, a common intermediate from lignin monomer (coumarate and cinnamate) degradation (Figure 5). We also included products from the oxidation of lignin monomers, vanillic acid and 4-hydroxybenzoic acid, in a similar concentration range to their aldehyde forms (Table 2). The 1X concentrations for some chemicals (e.g., benzoic acid) in the slurry used for testing are based on concentrations used in a previous study.

### BIOSCREEN C HIGH THROUGHPUT TOXICITY ASSAY

The high-throughput Bioscreen C assay was carried out as reported previously (Franden et al., 2009, 2013). Briefly, *C. necator* cells were revived from overnight LB culture with OD<sub>600nm</sub> adjusted to 3.0 using minimal medium. This cell suspension was used as seed culture to inoculate Bioscreen C plates containing 290 μL minimal medium per well at an initial OD<sub>600nm</sub> of 0.1. Growth was then monitored using the Bioscreen C instrument (GrowthCurves USA, NJ) with three technical replicates. The experiments were repeated at least two times.

*Cupriavidus necator* grown in the absence of potential inhibitory compounds was used as the control, and inhibition studies utilized cultures of *C. necator* challenged with different concentrations of each compound ranging from 0.1- to 10-fold (0.1X to 10X, Table 2). Stock solutions of compounds at 10X concentrations were prepared by dissolving the compounds to be tested in the minimal medium. These stock solutions were then diluted in minimal medium for testing at lower concentrations. For certain compounds with low aqueous solubility, incubation

at 55–60°C in a hot dry bath for several hours was needed for complete dissolution. The pH of the stock solutions was adjusted to the desired point of 6.8 using ammonium hydroxide (NH<sub>4</sub>OH) or sulfuric acid (H<sub>2</sub>SO<sub>4</sub>) and then filter sterilized before using.

### GENOMIC INVESTIGATION FOR LIGNIN DEGRADATION PATHWAY RECONSTRUCTION

To identify the enzymes related to lignin degradation, the protein sequences of *C. necator* H16 were extracted and reannotated functionally. Briefly, 6626 protein sequences were downloaded from NCBI (Genbank#: AM260479) and imported into CLC Genomics Workbench (V5.5) as the reference protein sequences for Blast search. In addition, the protein sequences were also imported into Blast2GO for the functional annotation (Gotz et al., 2008). The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were then extracted as well as the information of euKaryotic Orthologous Groups (KOG), enzyme code, and the reaction substrate(s) and product(s). The potential homologous gene(s) in *C. necator* H16 were identified by re-iterated BlastP searches. The information on protein product and conserved domains were examined and the pathway was reconstructed with the enzyme and pathway information from a literature search.

### Gas chromatographic analysis

Analysis of samples was performed on an Agilent 7890 GC equipped with a 5975 MS (Agilent Technologies, Palo Alto, CA). Sample compounds were separated using a 30 m × 0.25 mm × 0.25 mm DB-FFAP column (Agilent). HP MSD Chemstation software (Agilent) equipped with NIST database Rev. D.03.00 was used to determine the identity of the unknown compounds found within the samples.

Each sample was placed on an auto-sampler (Agilent) and injected at a volume of 1 μL into the GC-MS (Agilent). The GC-MS method consisted of a front inlet temperature of 250°C, MS transfer line temperature of 280°C, and a scan range from 35 to 550 m/z. A starting temperature of 50°C was held for 1.5 min and then ramped at 11°C/min to a temperature of 165°C with no hold time, then continued at a ramped rate of 35°C/min to 250°C and held for 9.617 min. The method resulted in a run time of 28 min for each sample.

## RESULTS AND DISCUSSIONS

### GROWTH KINETICS OF *C. NECATOR* IN THE SACCHARIFIED SLURRY

*Cupriavidus necator* did not grow in the original deacetylated saccharified slurry (Figure 1A) even in the 2-fold diluted saccharified slurry (Figure 1B) though it was able to grow in media containing the mock slurry. The sugars provided by the mock slurry allowed *C. necator* for superior growth over the LB alone in both 1-fold concentration and 2-fold dilution conditions with fast growth and high final OD<sub>600nm</sub> value. This complete inhibition of growth in real slurry suggests that *C. necator* is not tolerant to the combination of all toxic compounds present in the saccharified slurry.

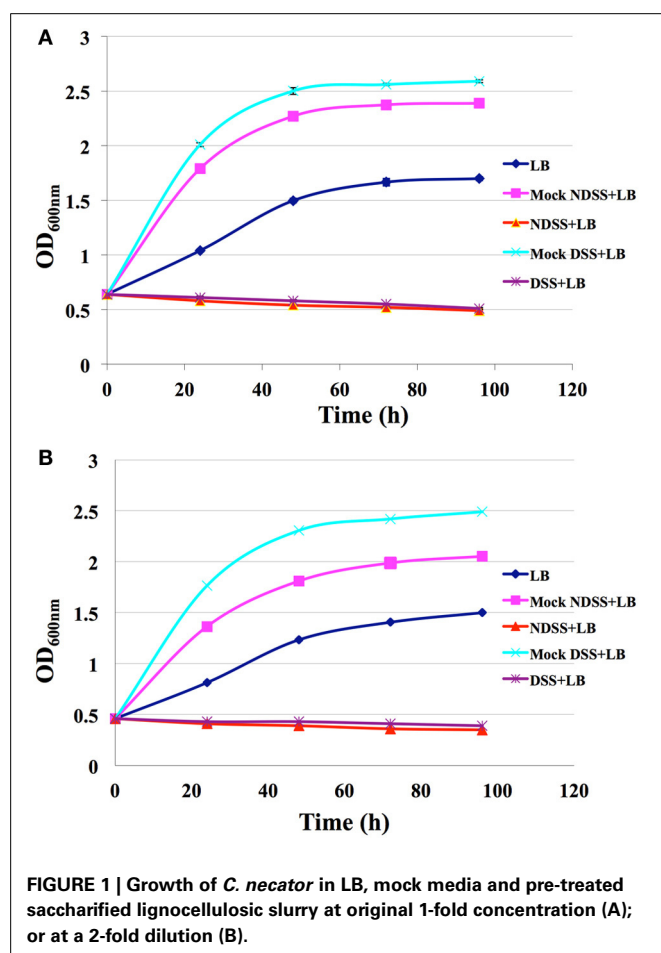
### KINETICS OF *C. NECATOR* IN THE SACCHARIFIED SLURRY AFTER ACTIVATED CARBON (AC) TREATMENT

It is well-known that acid pre-treatment of corn stover can generate various kinds of potential toxic inhibitors including organic

**Table 2 | The chemicals and their concentration ranges (mM) selected for toxicity testing.**

Concentration Range tested (Fold)	0.1×	0.25×	0.5×	1×	2.5×	5×	10×
<b>AMMONIUM SALTS</b>							
Ammonium Acetate (AA, 36.6 mM)	4.64	11.6	23.2	46.4	116.1	232.2	464.4
Ammonium Sulfate (AS, 21 mM)	4.0	10.0	20.0	40.0	100.0	200.0	400.0
<b>FURANS</b>							
Furfural (F, 12.3 mM)	1.26	3.15	6.30	12.59	31.48	62.97	125.94
HMF (1.3 mM)	0.14	0.36	0.71	1.43	3.57	7.14	14.27
<b>LIGNIN MONOMERS</b>							
4-Hydroxybenzaldehyde (HBA, 0.08 mM)	0.01	0.02	0.04	0.08	0.20	0.41	0.82
Vanillin (V, 0.13 mM)	0.01	0.03	0.07	0.13	0.33	0.66	1.31
Benzoic Acid (B, 0.037 mM)	0.08	0.20	0.41	0.82	2.05	4.09	8.19
p-Coumaric Acid (CA, 0.73 mM)	0.07	0.18	0.37	0.73	1.83	3.65	7.31
Ferulic Acid (FA, 0.46 mM)	0.05	0.12	0.23	0.46	1.16	2.32	4.63
4-Hydroxybenzoic Acid (HB, 0.20 mM)	0.01	0.04	0.07	0.14	0.36	0.72	1.45
Vanillic Acid (VA, 0.054 mM)	0.01	0.03	0.06	0.12	0.30	0.59	1.19

The concentration (mM) following the chemical abbreviation is the concentration in the slurry calculated from analytical results of GC-MS, LC-MS, or ICP-MS.

**Table 3 | The compositions of deacetylated saccharified slurry before and after activated carbon (AC) treatment.**

Composition	Untreated slurry	AC-treated slurry
Cellobiose (g/L)	17.4	9.3
Glucose (g/L)	100.4	95.1
Xylose (g/L)	51.4	48.3
Arabinose (g/L)	10.1	10.4
Glycerol (g/L)	0.8	0.7
Acetic acid (mM)	36.6	26.6
HMF (mM)	ND	ND
Furfural (mM)	12.5	ND

ND, Not detected; the detection limitation for furfural is 2.6 mM, HMF is 2.8 mM.

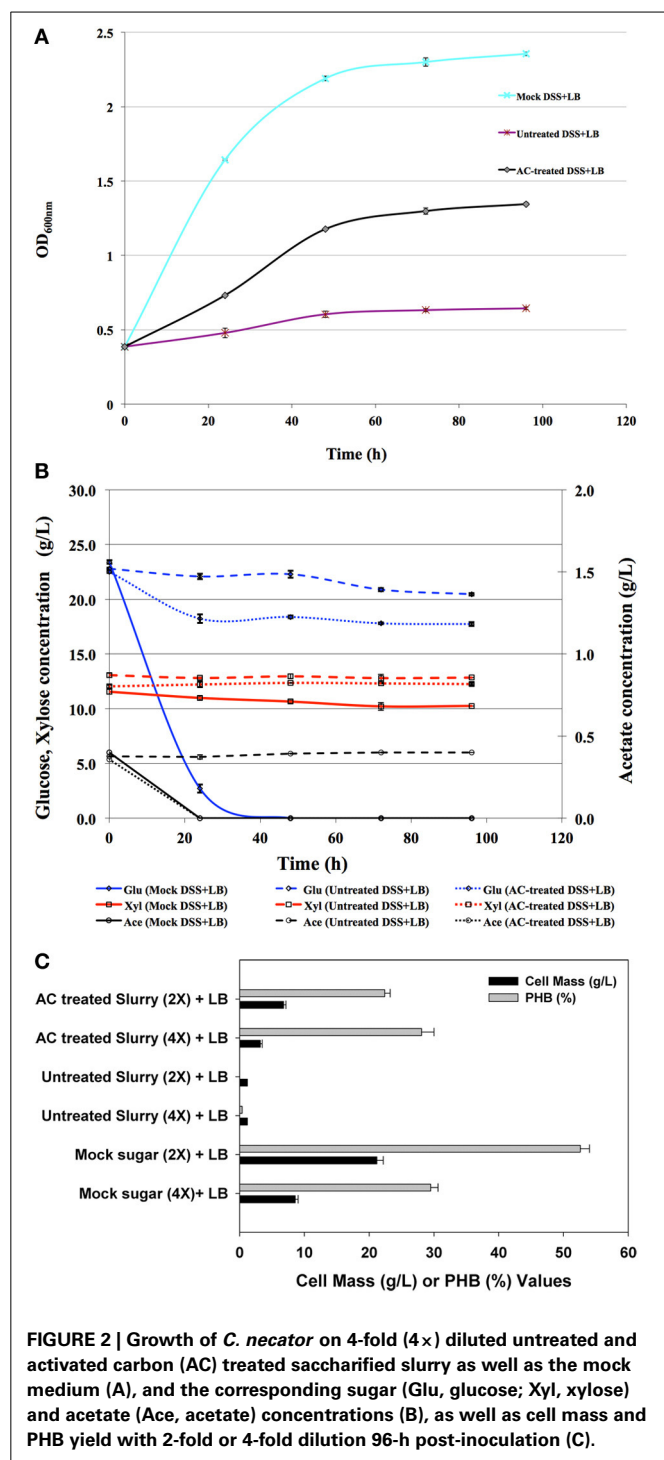
and furfural). Furfural is generally considered the most potent inhibitor to various microbial catalysts such as *E. coli*, *Z. mobilis*, and yeast (Zaldivar et al., 1999; Liu et al., 2004, 2005, 2008; Gorsich et al., 2006; Allen et al., 2010; Bowman et al., 2010; He et al., 2012; Huang et al., 2012; Franden et al., 2013). The maximum amount of furfural was 2 g/L in the slurries in this study. Shake flask growth experiments showed that furfural greatly inhibited cell viability when the concentration was above 2 g/L (data not shown).

A chemical treatment process using activated carbon (AC) was applied in this study in order to improve growth by removing potential inhibitors such as furfural from the slurry. AC was added to the slurry at 0.1 g/mL loading and incubated for 2 h at 130 rpm, 24°C. The compositions of saccharified slurry before and after AC-treatment indicated that all the furfural and about 30% of the acetate was removed by the AC treatment (Table 3).

The impact of AC treatment on the toxicity of diluted saccharified slurry (4-fold dilution), which provided about 25 g/L of glucose to the growth media, was tested. As shown in Figure 2A, untreated saccharified slurry was very toxic to *C. necator*, while AC-treated saccharified slurry was much less inhibitory to growth, although it was still inhibitory compared to the

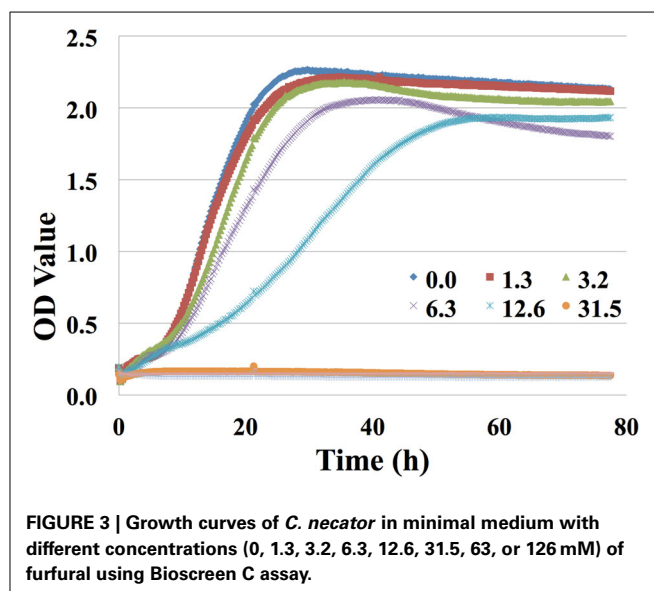
acids and aldehydes, and can also add to the toxicity through the build up of inorganic salts from the sulfuric acid used for pre-treatment and ammonium hydroxide added for neutralization (e.g., organic and inorganic acids of levulinic acid, vanillic acid, hydroxybenzoic acid, sulfuric acid, and aldehydes of HMF





**FIGURE 2 |** Growth of *C. necator* on 4-fold (4x) diluted untreated and activated carbon (AC) treated saccharified slurry as well as the mock medium (A), and the corresponding sugar (Glu, glucose; Xyl, xylose) and acetate (Ace, acetate) concentrations (B), as well as cell mass and PHB yield with 2-fold or 4-fold dilution 96-h post-inoculation (C).

mock slurry control. It is noteworthy that the consumption of glucose from AC-treated slurry was far slower than that from mock slurry and at 24 h, only about 25% of the total glucose was removed from the culture medium compared to 85% in the mock slurry (Figure 2B). This indicates that inhibitors remained in AC-treated slurry although furfural was completely removed. The remaining inhibitors are either unidentified residual components in the slurry (other than furfural) or



**FIGURE 3 |** Growth curves of *C. necator* in minimal medium with different concentrations (0, 1.3, 3.2, 6.3, 12.6, 31.5, 63, or 126 mM) of furfural using Bioscreen C assay.

components released from pre-treated corn stover during saccharification. It is noteworthy that besides the removal of all furfural by AC-treatment, our preliminary data indicated that a significant fraction of the other potentially toxic compounds such as acetic acid (Table 3) and lignin degradation products (e.g., vanillin, 4-hydroxybenzaldehyde, and p-coumaric acid) were also removed by AC-treatment. The impact of these individual toxic compounds in the saccharified slurry will be evaluated in the following section.

The cell yield from the AC-treated slurry (4-fold diluted, 4X) was about one third that of the culture grown on the mock slurry (4X), with similar amounts of PHB accumulated in the cells from both growth conditions 96 h post-inoculation (Figure 2C). However, when 2-fold diluted slurry was used, increased inhibition led to a lower cell mass and accordingly a lower PHB yield even with the carbon treated slurry (Figure 2C). All these results indicate that the toxicity of the biomass-derived slurry had a major impact on PHB production. Moreover, the low cell viability on 2x diluted slurry indicates that the inhibitory effect on cells is likely due to the combined contribution of several inhibitors besides furfural. Although the efficacy of the AC-treatment is obvious in term of cell growth, it only partially mitigated the inhibition. Identification of the other inhibitors in the AC-treated slurry is needed to help eliminate or mitigate the observed inhibition.

#### INHIBITOR SENSITIVITY INVESTIGATION USING THE HIGH-THROUGHPUT BIOSCREEN C ASSAY

To systematically explore the impact of inhibitors in the saccharified slurry on *C. necator*, a high-throughput Bioscreen C growth assay was performed using growth medium augmented with the potentially toxic compounds that had been identified at concentration ranges discussed above (Table 2). Briefly, growth curves were generated by subtracting OD readings from the test wells in the Bioscreen C from the background wells containing blank medium only. Typical growth curves with *C.*



**Table 4 | The growth rates and responses of *C. necator* to various furfural concentrations (mM) supplemented in the minimal medium.**

Conc. (mM)	0	1.3	3.2	6.3	12.6	31.5	63	126
Growth rate ( $\mu$ )	0.19	0.18	0.13	0.12	0.06	0	0	0
Response (%)	100.0	97.5	66.6	61.1	32.1	0	0	0

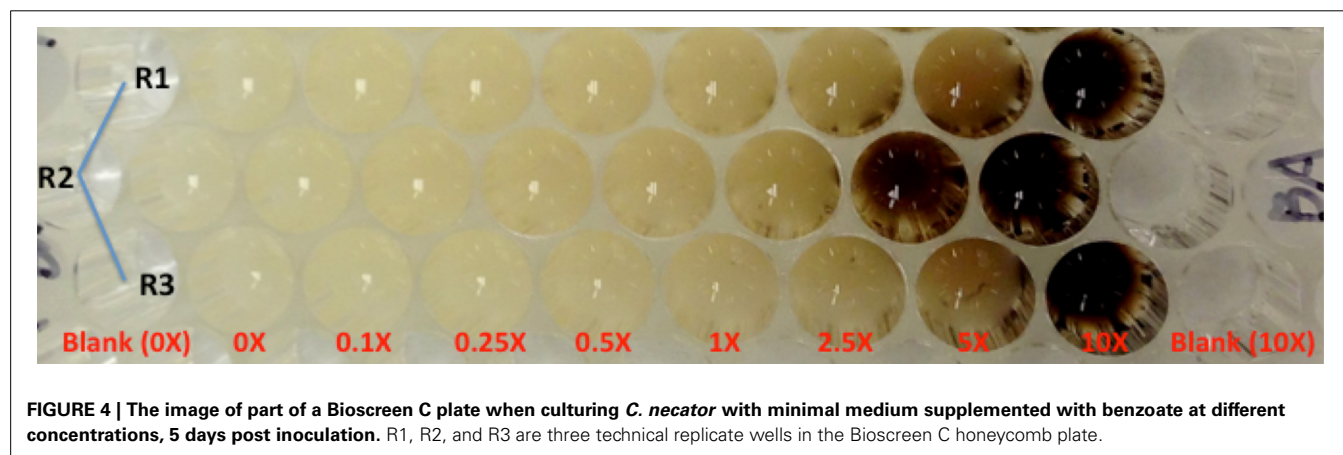
**Table 5 | The IC50 values and the ratio of the IC50 value to Conc. value (the concentration of the toxic compound identified in the saccharified slurry, mM) for *C. necator*.**

	Ammonium sulfate	Ammonium acetate	Furfural	HMF	Benzoic acid	p-Coumaric acid
Conc. (mM)	21	36.6	12.6	1.43	0.037	0.73
IC50 (mM)	388	210	9	2.9	0.44	3.2
Ratio	18.5	5.7	0.7	2.0	11.9	4.4

**Table 6 | Response values of *C. necator* to 11 potentially toxic compounds identified in the saccharified slurry at different concentrations (mM) supplemented in the minimal medium.**

Conc. Range (Folds)	0x	0.1x	0.25x	0.5x	1x	2.5x	5x	10x
Ammonium acetate	100	<b>112</b>	<b>154</b>	<b>145</b>	<b>115</b>	72	48	38
Ammonium sulfate	100	<b>122</b>	<b>122</b>	<b>114</b>	<b>107</b>	104	84	34
Furfural	100	98	67	61	32	0	0	0
4-Hydroxybenzaldehyde	100	100	100	77	93	105	91	108
HMF	100	87	76	73	59	48	43	31
Vanillin	100	97	98	101	98	92	68	58
Benzoic acid	100	<b>115</b>	<b>114</b>	<b>107</b>	96	60	38	29
Coumaric acid	100	101	<b>112</b>	<b>119</b>	<b>107</b>	79	46	26
Ferulic acid	100	108	99	99	101	103	101	89
4-Hydroxybenzoic acid	100	100	94	96	102	98	88	83
Vanillic acid	100	77	73	96	95	87	68	61

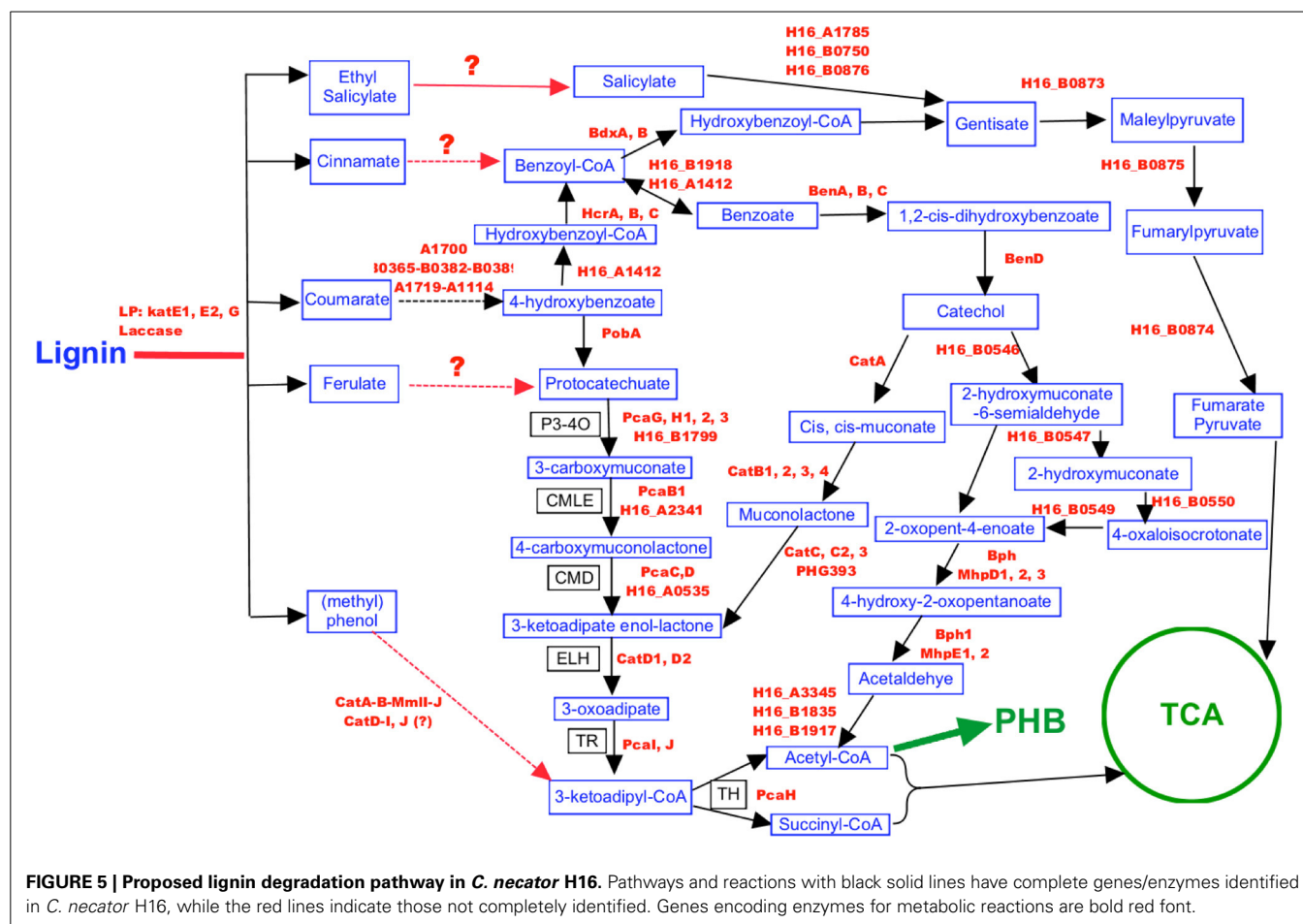
The values in bold font indicate stimulation of growth by the compound at the corresponding concentration.

**FIGURE 4 | The image of part of a Bioscreen C plate when culturing *C. necator* with minimal medium supplemented with benzoate at different concentrations, 5 days post inoculation. R1, R2, and R3 are three technical replicate wells in the Bioscreen C honeycomb plate.**

*necator* grown with varied concentrations of furfural are shown in **Figure 3**. The high-throughput nature of this assay allowed for utilization of both biological replicates on multiple plates and technical replicates on the same plate. Generally speaking, correlation between growth curves in both technical and biological replicates was quite high (data not shown), emphasizing the power of this method. The results were consistent with previous shake flask experiments showing that furfural greatly

inhibited cell growth when the concentration was above 2 g/L (ca. 20 mM).

Growth rates ( $\mu$  in terms of  $h^{-1}$ ) for each growth curve were then calculated as described previously (Franden et al., 2009, 2013). The response values, given as the percentage of the growth rate compared with the control in the absence of supplemented inhibitor, were then calculated for each concentration (**Table 4**). The response values were then used to determine the

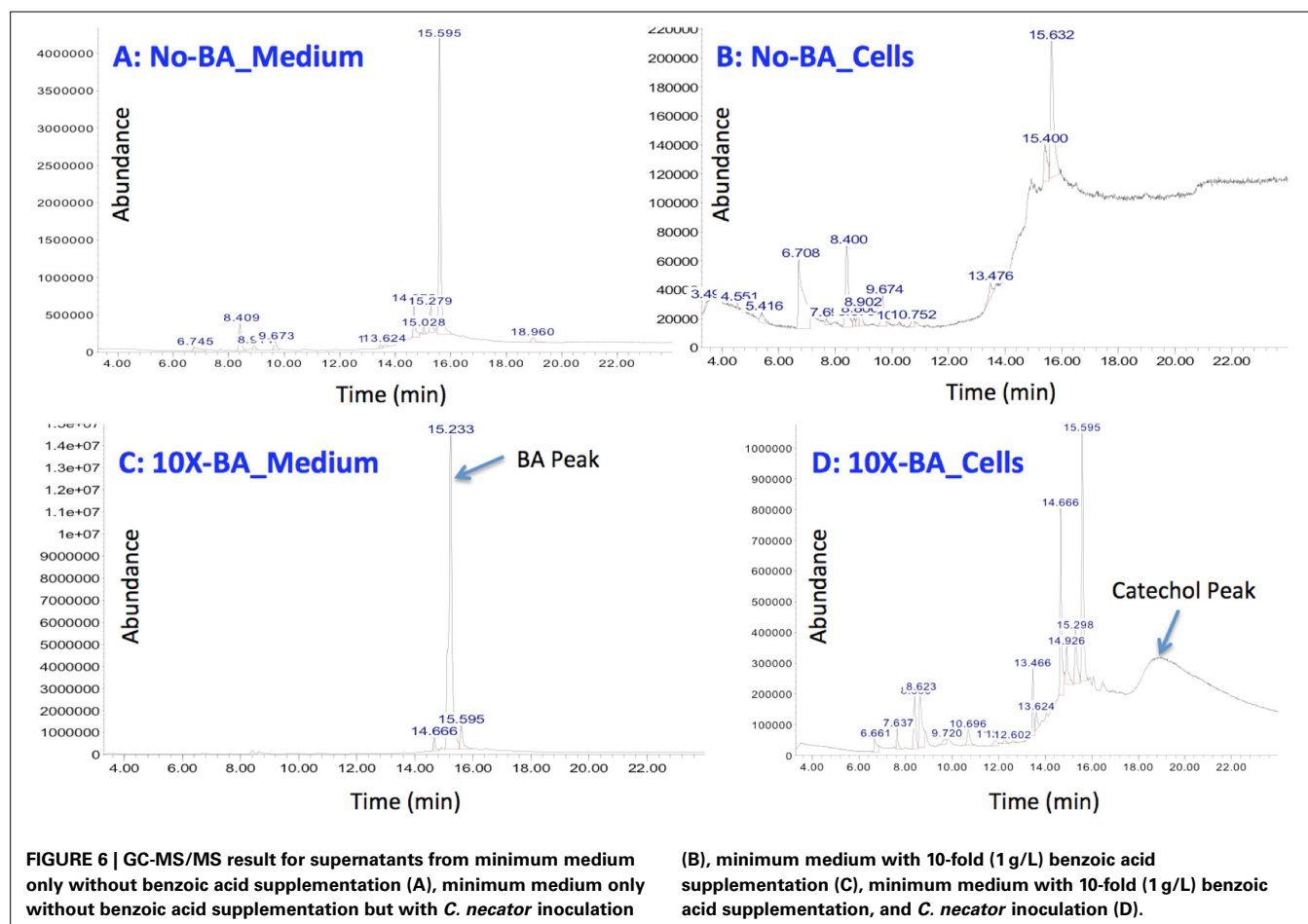


concentration of inhibitor (IC<sub>50</sub>) that resulted in 50% growth compared to the control. The application of IC<sub>50</sub> in this study was used to determine the top inhibitors in the slurries by comparing the IC<sub>50</sub> of different compounds. This approach has frequently been used as a general toxicity indicator for potential inhibitors (Franden et al., 2013). IC<sub>50</sub> values for 4-hydroxybenzaldehyde, vanillin, ferulic acid, 4-hydroxybenzoic acid, and vanillic acid were above the highest values tested, and the IC<sub>50</sub> values for the remaining compounds were listed in **Table 5**. The growth of *C. necator* was inhibited by 5 out of the 11 compounds tested. The aldehydes, furfural and HMF, were the most toxic compounds in the slurry. The IC<sub>50</sub> value of 9 mM for furfural was lower than the concentration detected in the saccharified slurry (12.6 mM or 1.2 g/L). Although the IC<sub>50</sub> for benzoic acid had the lowest value, its concentration in the slurry was only 0.037 mM and less than 10% of the IC<sub>50</sub> value making it the least toxic component in the slurry. The lignin degradation products were minimally toxic in the relevant concentration range except for p-coumaric acid with an IC<sub>50</sub> value about 4 times higher than the concentration in the saccharified slurry. Ammonium acetate was found to be more toxic than ammonium sulfate (**Table 5**).

A few key points, however, must be made that aren't captured in **Table 4**. One major observation is that some of the compounds found in saccharified slurry stimulate growth by as

much as 50% (**Table 6**). Growth stimulation with some of these compounds (e.g., ammonium sulfate, ammonium acetate, benzoic acid, and coumaric acid) was seen at low concentrations, but these compounds became inhibitory at higher concentrations. This indicated that *C. necator* might utilize these compounds as supplemental carbon or nitrogen sources at low levels. The utilization of acetate by *C. necator* in shake flask experiments (data not shown) supports this hypothesis that the higher response values could be an indicator of growth stimulation by utilizing the supplemented substrates (**Figure 2B**), which is consistent with previous report that addition of acetate as a supplementary substrate improved the cell growth and PHB production in *C. necator* DSMZ 545 (Sharifzadeh Baei et al., 2009). The combination of growth stimulation by some compounds and inhibition by others complicated analysis of the impact of saccharified slurry on growth and productivities and should be further explored in the future.

An additional point concerning the bacterial toxicity profiles is that they can provide additional information that could be used to understand microbial physiology and to propose genetic targets for metabolic engineering. For example, as the concentration of benzoic acid increased and the response changed from growth stimulation to inhibition, the color of the culture medium also became darker (**Figure 4**). This color change was not seen



in un-inoculated control wells. The availability of the *C. necator* genome sequence will facilitate our understanding of this phenomenon using a genomics approach that will be explained in the following section.

### C. NECATOR LIGNIN-DEGRADATION PATHWAY RECONSTRUCTION

*Cupriavidus necator* is capable of utilizing lignin monomers and recently a *Cupriavidus* sp. (*C. basiliensis* B-8) has been reported to be able to utilize kraft lignin (Shi et al., 2013). With that information in hand, we then set out to reconstruct the lignin degradation pathway of *C. necator* based on genome information and literature reports to help understand the bottleneck for further development of *C. necator* as a lignocellulosic biofuel-production strain (Figure 5).

Although the enzymatic reaction from lignin to lignin monomers and most reactions from lignin monomers to the metabolic intermediates are not yet fully revealed, homologous genes encoding the enzyme BenA, B, C, D were identified in *C. necator* that would convert benzoate to catechol. When the concentration of the supplemented benzoate is lower than the concentration characterized in the slurry, catechol will be further converted into cis,cis-muconate or 2-hydroxymuconate-6-semialdehyde through two different metabolic pathways leading to the accumulation of the key intermediate of acetyl-CoA for PHB production (Figure 5). Previous study already indicated that

the accumulation of catechol in *Pseudomonas mendocina* can lead to its conversion of catechol into 1, 2-benzoquinone (Parulekar and Mavinkurve, 2006), which caused the culture medium to turn dark and at the same time inhibited the cellular growth since benzoquinone is a toxic agent against microorganisms such as *E. coli*, *Pseudomonas fluorescens*, and *Erwinia amylovora* (Beckman and Siedow, 1985). Our GC/MS study also indicated the correlation between the catechol appearance and benzoic acid disappearance when 1 g/L benzoic acid was supplemented into the minimum medium (Figure 6). Although further experimental data are needed to quantify the disappearance of benzoic acid and the appearance of catechol and 1,2-benzoquinone, our current study suggest that at high benzoate concentrations, *C. necator* can not convert catechol completely to cis,cis-muconate or 2-hydroxymuconate-6-semialdehyde and catechol began to accumulate which potentially caused the formation of 1, 2-benzoquinone and then turned medium into dark brown color and inhibited cell growth (Figures 4–6).

In addition, homologous genes were also identified in *C. necator* to carry out the reaction from p-coumarate to 4-hydrobenzoate and then feed into TCA cycle (Figure 5), which indicates that *C. necator* can utilize lignin monomer coumarate as carbon source for cell growth and may explain the stimulus effect of coumarate on *C. necator* growth in the lower concentration range. When high concentration of p-coumarate was

supplemented into the medium, high concentration of catechol may be produced and the bottleneck reaction from catechol to cis,cis-muconate or 2-hydroxymuconate-6-semialdehyde (**Figure 5**) caused the inhibitory effect as discussed above. However, it is possible that other metabolic pathways that we have not been covered will also involved in and even play a key role on lignin monomer utilization and pre-treatment inhibitor sensitivity.

This study is an attempt and shows the feasibility to connect genomic information to explain microbial physiological phenomena. Although it could provide genetic targets for metabolic engineering to improve strain robustness (e.g., overexpression catechol dioxygenase to drive catechol into TCA instead of catechol accumulation), other experimental approaches such as systems biology study are needed to completely understand the mechanism of pre-treatment inhibitor sensitivity and genetic studies are especially required to confirm the hypothesis generated by bioinformatics study.

## AUTHOR CONTRIBUTIONS

Wei Wang, Shihui Yang, Philip T. Pienkos, and David K. Johnson designed the experiments. Wei Wang carried out the flask assay and AC treatment. Shihui Yang carried out the Bioscreen C assay and genomic study. Glendon B. Hunsinger performed the GC/MS. Shihui Yang and Wei Wang did the spectrum scanning. Wei Wang, Shihui Yang, Philip T. Pienkos, and David K. Johnson analyzed the data, and wrote the manuscript.

## ACKNOWLEDGMENTS

This work was supported by the US Department of Energy, Bioenergy Technology Office (BETO) under contract number DE-AC36-08-GO28308 to NREL. NREL is a national laboratory of the US Department of Energy, Office of Energy Efficiency and Renewable Energy, operated by the Alliance for Sustainable Energy, LLC. We thank Michael Guarnieri and Thieny Trinh for preliminary toxicity experiments with *C. necator*, Ashutosh Mittal and William E. Michener for PHB and GC/MS analyses.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fmicb.2014.00247/abstract>

## REFERENCES

- Allen, S. A., Clark, W., McCaffery, J. M., Cai, Z., Lancot, A., Slininger, P. J., et al. (2010). Furfural induces reactive oxygen species accumulation and cellular damage in *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* 3:2. doi: 10.1186/1754-6834-3-2
- Amadou, C., Pascal, G., Mangelot, S., Glew, M., Bontemps, C., Capela, D., et al. (2008). Genome sequence of the beta-rhizobium *Cupriavidus taiwanensis* and comparative genomics of rhizobia. *Genome Res.* 18, 1472–1483. doi: 10.1101/gr.076448.108
- Ask, M., Bettiga, M., Mapelli, V., and Olsson, L. (2013). The influence of HMF and furfural on redox-balance and energy-state of xylose-utilizing *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* 6:22. doi: 10.1186/1754-6834-6-22
- Bajwa, P. K., Ho, C. Y., Chan, C. K., Martin, V. J., Trevors, J. T., and Lee, H. (2013). Transcriptional profiling of *Saccharomyces cerevisiae* T2 cells upon exposure to hardwood spent sulphite liquor: comparison to acetic acid, furfural and hydroxymethylfurfural. *Antonie Van Leeuwenhoek* 103, 1281–1295. doi: 10.1007/s10482-013-9909-1
- Beckman, J. S., and Siedow, J. N. (1985). Bactericidal agents generated by the peroxidase-catalyzed oxidation of para-hydroquinones. *J. Biol. Chem.* 260, 14604–14609.
- Bowman, M. J., Jordan, D. B., Vermillion, K. E., Braker, J. D., Moon, J., and Liu, Z. L. (2010). Stereochemistry of furfural reduction by a *Saccharomyces cerevisiae* aldehyde reductase that contributes to in situ furfural detoxification. *Appl. Environ. Microbiol.* 76, 4926–4932. doi: 10.1128/AEM.00542-10
- Brigham, C. J., Speth, D. R., Rha, C., and Sinskey, A. J. (2012). Whole-genome microarray and gene deletion studies reveal regulation of the polyhydroxyalkanoate production cycle by the stringent response in *Ralstonia eutropha* H16. *Appl. Environ. Microbiol.* 78, 8033–8044. doi: 10.1128/AEM.01693-12
- Cavalheiro, J. M. B. T., de Almeida, M. C. M. D., Grandfils, C., and da Fonseca, M. M. R. (2009). Poly(3-hydroxybutyrate) production by *Cupriavidus necator* using waste glycerol. *Process Biochem.* 44, 509–515. doi: 10.1016/j.procbio.2009.01.008
- Chen, X., Shekiri, J., Franden, M. A., Wang, W., Zhang, M., Kuhn, E., et al. (2012). The impacts of deacetylation prior to dilute acid pretreatment on the bioethanol process. *Biotechnol. Biofuels* 5:8. doi: 10.1186/1754-6834-5-8
- Cserhati, M., Kriszt, B., Szoboszlai, S., Toth, A., Szabo, I., Tancsics, A., et al. (2012). *De novo* genome project of *Cupriavidus basilensis* OR16. *J. Bacteriol.* 194, 2109–2110. doi: 10.1128/JB.06752-11
- Dunlop, M. J. (2011). Engineering microbes for tolerance to next-generation biofuels. *Biotechnol. Biofuels* 4:32. doi: 10.1186/1754-6834-4-32
- Dunlop, M. J., Dossani, Z. Y., Szmidt, H. L., Chu, H. C., Lee, T. S., Keasling, J. D., et al. (2011). Engineering microbial biofuel tolerance and export using efflux pumps. *Mol. Syst. Biol.* 7:487. doi: 10.1038/msb.2011.21
- Endo, A., Nakamura, T., Ando, A., Tokuyasu, K., and Shima, J. (2008). Genome-wide screening of the genes required for tolerance to vanillin, which is a potential inhibitor of bioethanol fermentation, in *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* 1:3. doi: 10.1186/1754-6834-1-3
- Fischer, C. R., Peterson, A. A., and Tester, J. W. (2011). Production of C(3) hydrocarbons from biomass via hydrothermal carboxylate reforming. *Ind. Eng. Chem. Res.* 50, 4420–4424. doi: 10.1021/ie1023386
- Franden, M. A., Pienkos, P. T., and Zhang, M. (2009). Development of a high-throughput method to evaluate the impact of inhibitory compounds from lignocellulosic hydrolysates on the growth of *Zymomonas mobilis*. *J. Biotechnol.* 144, 259–267. doi: 10.1016/j.jbiotec.2009.08.006
- Franden, M. A., Pilath, H. M., Mohagheghi, A., Pienkos, P. T., and Zhang, M. (2013). Inhibition of growth of *Zymomonas mobilis* by model compounds found in lignocellulosic hydrolysates. *Biotechnol. Biofuels* 6:99. doi: 10.1186/1754-6834-6-99
- Gorsich, S. W., Dien, B. S., Nichols, N. N., Slininger, P. J., Liu, Z. L., and Scory, C. D. (2006). Tolerance to furfural-induced stress is associated with pentose phosphate pathway genes ZWF1, GND1, RPE1, and TKL1 in *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* 71, 339–349. doi: 10.1007/s00253-005-0142-3
- Gotz, S., Garcia-Gomez, J. M., Terol, J., Williams, T. D., Nagaraj, S. H., Nueda, M. J., et al. (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* 36, 3420–3435. doi: 10.1093/nar/gkn176
- Harwood, C. S., and Parales, R. E. (1996). The  $\beta$ -ketoadipate pathway and the biology of self-identity. *Annu. Rev. Microbiol.* 50, 553–590. doi: 10.1146/annurev.micro.50.1.553
- He, M. X., Wu, B., Shui, Z. X., Hu, Q. C., Wang, W. G., Tan, F. R., et al. (2012). Transcriptome profiling of *Zymomonas mobilis* under furfural stress. *Appl. Microbiol. Biotechnol.* 95, 189–199. doi: 10.1007/s00253-012-4155-4
- Hong, K. W., Thinagaran, D., Gan, H. M., Yin, W. F., and Chan, K. G. (2012). Whole-genome sequence of *Cupriavidus* sp. strain BIS7, a heavy-metal-resistant bacterium. *J. Bacteriol.* 194:6324. doi: 10.1128/JB.01608-12
- Huang, C., Wu, H., Smith, T. J., Liu, Z. J., Lou, W. Y., and Zong, M. H. (2012). *In vivo* detoxification of furfural during lipid production by the oleaginous yeast *Trichosporon fermentans*. *Biotechnol. Lett.* 34, 1637–1642. doi: 10.1007/s10529-012-0948-x
- Iwaki, A., Kawai, T., Yamamoto, Y., and Izawa, S. (2013). Biomass conversion inhibitors furfural and 5-hydroxymethylfurfural induce formation of messenger RNP granules and attenuate translation activity in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 79, 1661–1667. doi: 10.1128/AEM.02797-12
- Janssen, P. J., Van Houdt, R., Moors, H., Monsieus, P., Morin, N., Michaux, A., et al. (2010). The complete genome sequence of *Cupriavidus metallidurans* strain CH34, a master survivalist in harsh and anthropogenic environments. *PLoS ONE* 5:e10433. doi: 10.1371/journal.pone.0010433



- Kang, A., and Chang, M. W. (2012). Identification and reconstitution of genetic regulatory networks for improved microbial tolerance to isooctane. *Mol. Biosyst.* 8, 1350–1358. doi: 10.1039/c2mb05441h
- Karr, D. B., Waters, J. K., and Emerich, D. W. (1983). Analysis of poly- $\beta$ -Hydroxybutyrate in *Rhizobium japonicum* bacteroids by ion-exclusion high-pressure liquid chromatography and UV detection. *Appl. Environ. Microbiol.* 46, 1339–1344.
- Larsson, S., Reimann, A., Nilvebrant, N. O., and Jonsson, L. J. (1999). Comparison of different methods for the detoxification of lignocellulose hydrolyzates of spruce. *Appl. Biochem. Biotechnol.* 77, 91–103. doi: 10.1385/ABAB:77:1-3:91
- Li, L. G., Cai, L., and Zhang, T. (2013). Genome of *Cupriavidus* sp. HMR-1, a heavy metal-resistant bacterium. *Genome Announc.* 1:e00202–e00212. doi: 10.1128/genomeA.00202-12
- Liu, Z. L., Moon, J., Andersh, B. J., Slininger, P. J., and Weber, S. (2008). Multiple gene-mediated NAD(P)H-dependent aldehyde reduction is a mechanism of *in situ* detoxification of furfural and 5-hydroxymethylfurfural by *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* 81, 743–753. doi: 10.1007/s00253-008-1702-0
- Liu, Z. L., Slininger, P. J., Dien, B. S., Berhow, M. A., Kurtzman, C. P., and Gorsich, S. W. (2004). Adaptive response of yeasts to furfural and 5-hydroxymethylfurfural and new chemical evidence for HMF conversion to 2,5-bis-hydroxymethylfuran. *J. Ind. Microbiol. Biotechnol.* 31, 345–352. doi: 10.1007/s10295-004-0148-3
- Liu, Z. L., Slininger, P. J., and Gorsich, S. W. (2005). Enhanced biotransformation of furfural and hydroxymethylfurfural by newly developed ethanologenic yeast strains. *Appl. Biochem. Biotechnol.* 121–124, 451–460. doi: 10.1385/ABAB:121:1-3:0451
- Lykidis, A., Perez-Pantoja, D., Ledger, T., Mavromatis, K., Anderson, I. J., Ivanova, N. N., et al. (2010). The complete multipartite genome sequence of *Cupriavidus necator* JMP134, a versatile pollutant degrader. *PLoS ONE* 5:e9729. doi: 10.1371/journal.pone.0009729
- Mills, T. Y., Sandoval, N. R., and Gill, R. T. (2009). Cellulosic hydrolysate toxicity and tolerance mechanisms in *Escherichia coli*. *Biotechnol. Biofuels* 2:26. doi: 10.1186/1754-6834-2-26
- Olsson, L., and Hahn-Hagerdal, B. (1996). Fermentation of lignocellulosic hydrolysates for ethanol production. *Enzyme Microb. Technol.* 18, 312–331. doi: 10.1016/0141-0229(95)00157-3
- Orita, I., Iwazawa, R., Nakamura, S., and Fukui, T. (2012). Identification of mutation points in *Cupriavidus necator* NCIMB 11599 and genetic reconstitution of glucose-utilization ability in wild strain H16 for polyhydroxyalkanoate production. *J. Biosci. Bioeng.* 113, 63–69. doi: 10.1016/j.jbiosc.2011.09.014
- Park, J. M., Kim, T. Y., and Lee, S. Y. (2011). Genome-scale reconstruction and *in silico* analysis of the *Ralstonia eutropha* H16 for polyhydroxyalkanoate synthesis, lithoautotrophic growth, and 2-methyl citric acid production. *BMC Syst. Biol.* 5:101. doi: 10.1186/1752-0509-5-101
- Parulekar, C., and Mavinkurve, S. (2006). Formation of ortho-benzoquinone from sodium benzoate by *Pseudomonas mendocina* P2d. *Indian J. Exp. Biol.* 44, 157–162.
- Peplinski, K., Ehrenreich, A., Doring, C., Bomeke, M., Reinecke, F., Huttmacher, C., et al. (2010). Genome-wide transcriptome analyses of the “Knallgas” bacterium *Ralstonia eutropha* H16 with regard to polyhydroxyalkanoate metabolism. *Microbiology* 156, 2136–2152. doi: 10.1099/mic.0.038380-0
- Pérez-Pantoja, D., De la Iglesia, R., Pieper, D. H., and Gonzalez, B. (2008). Metabolic reconstruction of aromatic compounds degradation from the genome of the amazing pollutant-degrading bacterium *Cupriavidus necator* JMP134. *FEMS Microbiol. Rev.* 32, 736–794. doi: 10.1111/j.1574-6976.2008.00122.x
- Petersson, A., Almeida, J. R., Modig, T., Karhumaa, K., Hahn-Hagerdal, B., Gorwa-Grauslund, M. F., et al. (2006). A 5-hydroxymethyl furfural reducing enzyme encoded by the *Saccharomyces cerevisiae* ADH6 gene conveys HMF tolerance. *Yeast* 23, 455–464. doi: 10.1002/yea.1370
- Pilath, H., Mittal, A., Moens, L., Vinzant, T., Wang, W., and Johnson, D. (2013). “A route from biomass to hydrocarbons via depolymerization and decarboxylation of microbially produced polyhydroxybutyrate,” in *Direct Microbial Conversion of Biomass to Advanced Biofuels*, ed H. E. Himmel (Waltham, MA: Elsevier-Science and Technology Books).
- Poehlein, A., Kusian, B., Friedrich, B., Daniel, R., and Bowien, B. (2011). Complete genome sequence of the type strain *Cupriavidus necator* N-1. *J. Bacteriol.* 193:5017. doi: 10.1128/JB.05660-11
- Pohlmann, A., Fricke, W. F., Reinecke, F., Kusian, B., Liesegang, H., Cramm, R., et al. (2006). Genome sequence of the bioplastic-producing “Knallgas” bacterium *Ralstonia eutropha* H16. *Nat. Biotechnol.* 24, 1257–1262. doi: 10.1038/nbt1244
- Serrano-Ruiz, J. C., and Dumesic, J. A. (2011). Catalytic routes for the conversion of biomass into liquid hydrocarbon transportation fuels. *Energy Environ. Sci.* 4, 83–99. doi: 10.1039/c0ee00436g
- Sharifzadeh Baei, M., Najafpour, G. D., Younesi, H., Tabandeh, F., and Eisazadeh, H. (2009). Poly(3-hydroxybutyrate) synthesis by *Cupriavidus necator* DSMZ 545 utilizing various carbon sources. *World Appl. Sci. J.* 7, 157–161.
- Shi, Y., Chai, L., Tang, C., Yang, Z., Zhang, H., Chen, R. H., et al. (2013). Characterization and genomic analysis of kraft lignin biodegradation by the beta-proteobacterium *Cupriavidus basilensis* B-8. *Biotechnol. Biofuels* 6:1. doi: 10.1186/1754-6834-6-1
- Van Houdt, R., Monsieurs, P., Mijndonckx, K., Provoost, A., Janssen, A., Megeay, M., et al. (2012). Variation in genomic islands contribute to genome plasticity in *Cupriavidus metallidurans*. *BMC Genomics* 13:111. doi: 10.1186/1471-2164-13-111
- Wilson, C. M., Yang, S., Rodriguez, M. Jr., Ma, Q., Johnson, C. M., Dice, L., et al. (2013). *Clostridium thermocellum* transcriptomic profiles after exposure to furfural or heat stress. *Biotechnol. Biofuels* 6:131. doi: 10.1186/1754-6834-6-131
- Yang, J., Ding, M. Z., Li B. Z., Liu, Z. L., Wang, X., and Yuan, Y. J. (2012a). Integrated phospholipidomics and transcriptomics analysis of *Saccharomyces cerevisiae* with enhanced tolerance to a mixture of acetic acid, furfural, and phenol. *OMICS* 16, 374–386. doi: 10.1089/omi.2011.0127
- Yang, S., Keller, M., and Brown, S. (2012b). “Genomics on pretreatment inhibitor tolerance of *Zymomonas mobilis*,” in *Microbial Stress Tolerance for Biofuels*, ed Z. L. Liu (Berlin Heidelberg: Springer), 161–175. doi: 10.1007/978-3-642-21467-7\_7
- Yang, S., Land, M. L., Klingeman, D. M., Pelletier, D. A., Lu, T. Y., Martin, S. L., et al. (2010a). Paradigm for industrial strain improvement identifies sodium acetate tolerance loci in *Zymomonas mobilis* and *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.* 107, 10395–10400. doi: 10.1073/pnas.0914506107
- Yang, S., Pelletier, D. A., Lu, T. Y., and Brown, S. D. (2010b). The *Zymomonas mobilis* regulator *hfq* contributes to tolerance against multiple lignocellulosic pretreatment inhibitors. *BMC Microbiol.* 10:135. doi: 10.1186/1471-2180-10-135
- Zaldivar, J., and Ingram, L. O. (1999). Effect of organic acids on the growth and fermentation of ethanologenic *Escherichia coli* LY01. *Biotechnol. Bioeng.* 66, 203–210. doi: 10.1002/(SICI)1097-0290(1999)66:4<203::AID-BIT1>3.0.CO;2-#
- Zaldivar, J., Martinez, A., and Ingram, L. O. (1999). Effect of selected aldehydes on the growth and fermentation of ethanologenic *Escherichia coli*. *Biotechnol. Bioeng.* 65, 24–33. doi: 10.1002/(SICI)1097-0290(19991005)65:1<24::AID-BIT4>3.0.CO;2-2
- Zeng, Y., Zhao, S., Yang, S., and Ding, S.-Y. (2014). Lignin plays a negative role in the biochemical process for producing lignocellulosic biofuels. *Curr. Opin. Biotechnol.* 27:8. doi: 10.1016/j.copbio.2013.09.008
- Zhang, F., Rodriguez, S., and Keasling, J. D. (2011). Metabolic engineering of microbial pathways for advanced biofuels production. *Curr. Opin. Biotechnol.* 22, 775–783. doi: 10.1016/j.copbio.2011.04.024

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 09 January 2014; accepted: 06 May 2014; published online: 27 May 2014.  
Citation: Wang W, Yang S, Hunsinger GB, Pienkos PT and Johnson DK (2014) Connecting lignin-degradation pathway with pre-treatment inhibitor sensitivity of *Cupriavidus necator*. *Front. Microbiol.* 5:247. doi: 10.3389/fmicb.2014.00247  
This article was submitted to *Microbial Physiology and Metabolism*, a section of the journal *Frontiers in Microbiology*.  
Copyright © 2014 Wang, Yang, Hunsinger, Pienkos and Johnson. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Elucidation of *Zymomonas mobilis* physiology and stress responses by quantitative proteomics and transcriptomics

Shihui Yang<sup>1,2,3\*</sup>, Chongle Pan<sup>4,5</sup>, Gregory B. Hurst<sup>5</sup>, Lezlee Dice<sup>1,2</sup>, Brian H. Davison<sup>1,2</sup> and Steven D. Brown<sup>1,2\*</sup>

<sup>1</sup> Biosciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA

<sup>2</sup> BioEnergy Science Center, Oak Ridge National Laboratory, Oak Ridge, TN, USA

<sup>3</sup> National Bioenergy Center, National Renewable Energy Laboratory, Golden, CO, USA

<sup>4</sup> Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA

<sup>5</sup> Chemical Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA

## Edited by:

Katherine M. Pappas, University of Athens, Greece

## Reviewed by:

Harold J. Schreier, University of

Maryland Baltimore County, USA

Claudio Avignone-Rossa, University of Surrey, UK

## \*Correspondence:

Shihui Yang, National Bioenergy Center, National Renewable Energy Laboratory, 15013 Denver West Parkway, Golden, CO 80401, USA  
e-mail: shihui.yang@nrel.gov;

Steven D. Brown, BioEnergy Science Center, Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, TN 37831, USA  
e-mail: brownstd@ornl.gov

*Zymomonas mobilis* is an excellent ethanologenic bacterium. Biomass pretreatment and saccharification provides access to simple sugars, but also produces inhibitors such as acetate and furfural. Our previous work has identified and confirmed the genetic change of a 1.5-kb deletion in the sodium acetate tolerant *Z. mobilis* mutant (AcR) leading to constitutively elevated expression of a sodium proton antiporter encoding gene *nhaA*, which contributes to the sodium acetate tolerance of AcR mutant. In this study, we further investigated the responses of AcR and wild-type ZM4 to sodium acetate stress in minimum media using both transcriptomics and a metabolic labeling approach for quantitative proteomics the first time. Proteomic measurements at two time points identified about eight hundreds proteins, or about half of the predicted proteome. Extracellular metabolite analysis indicated AcR overcame the acetate stress quicker than ZM4 with a concomitant earlier ethanol production in AcR mutant, although the final ethanol yields and cell densities were similar between two strains. Transcriptomic samples were analyzed for four time points and revealed that the response of *Z. mobilis* to sodium acetate stress is dynamic, complex, and involved about one-fifth of the total predicted genes from all different functional categories. The modest correlations between proteomic and transcriptomic data may suggest the involvement of posttranscriptional control. In addition, the transcriptomic data of forty-four microarrays from four experiments for ZM4 and AcR under different conditions were combined to identify strain-specific, media-responsive, growth phase-dependent, and treatment-responsive gene expression profiles. Together this study indicates that minimal medium has the most dramatic effect on gene expression compared to rich medium followed by growth phase, inhibitor, and strain background. Genes involved in protein biosynthesis, glycolysis and fermentation as well as ATP synthesis and stress response play key roles in *Z. mobilis* metabolism with consistently strong expression levels under different conditions.

**Keywords:** *Zymomonas mobilis*, microarray, proteomics and metabolomics, acetate, pretreatment inhibitor, stress responses, quantitative proteomics, systems biology

## BACKGROUND

Yeast strains are among the leading current generation industrial biocatalyst microorganisms for fuel production (Hahn-Hagerdal et al., 2006). However, engineered bacteria such as *Zymomonas mobilis*, *E. coli*, *Bacillus subtilis* are also being developed and deployed to address commercial biofuel catalyst requirements (Dien et al., 2003; Inui et al., 2004; Romero et al., 2007; Alper and Stephanopoulos, 2009). *Z. mobilis* is an ethanologenic bacterium with many desirable industrial characteristics such as high-specific productivity and yield, high ethanol tolerance, and wide pH range (Gunasekaran and Raj, 1999; Panesar et al., 2006; Rogers et al., 2007). Recently, transformation efficiency has been improved by modifying the DNA restriction-modification systems (Kerr et al., 2010), and the inhibitor tolerance genes have been identified to improve the pretreatment inhibitor

tolerance using genes from *Z. mobilis* (Yang et al., 2010a,b) or from *Deinococcus radiodurans* (Zhang et al., 2010). The genome sequences for strains ZM4, NCIMB 11163, 10988, 29291 and 29292 have been determined (Seo et al., 2005; Kouvelis et al., 2009, 2011; Pappas et al., 2011; Desiniotis et al., 2012), and the ZM4 genome annotation was improved recently (Yang et al., 2009a). Genome-scale *in silico* metabolic modeling analysis have been reported (Lee et al., 2010; Widiastuti et al., 2011; Rutkis et al., 2013) and recombinant strains have been engineered to express and secrete cellulase (Linger et al., 2010) or ferment hexoses and pentose sugars such as xylose and arabinose (Zhang et al., 1995; Deanda et al., 1996).

A core challenge in next-generation biomass-based cellulosic biofuel is the recalcitrance of biomass to breakdown into sugars (Himmel et al., 2007; Alper and Stephanopoulos, 2009).

Biomass pretreatment regimes are required to release the sugars, which can create inhibitors such as furfural, acetate, and vanillin (Almeida et al., 2007; Pienkos, 2009). The existence of pretreatment inhibitors increases production costs due to lower production rates and decreased yields. The development and deployment of robust inhibitor-tolerant biocatalysts for efficient fermentation of biomass to biofuel will be a critical component for successful production of biofuels at industrial-scale quantities to meet sustainability and energy security challenges associated with fossil fuels (Almeida et al., 2007). Several recent transcriptomics studies have used microarray approach to characterize mutant strain or *Z. mobilis* stress responses (Yang et al., 2009b, 2013; Hayashi et al., 2011; He et al., 2012a,b; Jeon et al., 2012; Skerker et al., 2013).

Acetic acid is an important inhibitor produced by the de-acetylation of hemicelluloses during biomass pretreatment. Unlike another major inhibitor furfural which is volatile and can be converted into a less toxic product of furfural alcohol (Liu et al., 2005; Heer and Sauer, 2008; Franden et al., 2009; Agrawal and Chen, 2011), acetate is stable during fermentation and poses a constitutive stress on the growth and ethanol production of *Z. mobilis* (Yang et al., 2010a,b). A *Z. mobilis* mutant strain, designated AcR, generated by chemical mutagenesis and selection, is able to produce ethanol efficiently in the presence of 20 g/L sodium acetate (NaAc), while the parent ZM4 is inhibited significantly above 12 g/L (Joachimstahl et al., 1998). Through comparative genome sequencing and next-generation sequencing (NGS)-based genome resequencing, we characterized the AcR mutant and identified a 1.5-kb deletion in strain AcR, which likely truncated the promoter of the *nhaA* gene encoding a sodium proton antiporter. We have carried out genetics study to confirm the association of 1.5-kb deletion in AcR mutant with its sodium acetate tolerance phenotype, we further performed microarray study to identify the differentially expressed genes between wild-type ZM4 and AcR mutant, and identified that *nhaA* gene is consistently upregulated in AcR mutant background compared to wild-type ZM4 under different conditions of NaCl and NaAc stress (Yang et al., 2010b).

Although we confirmed the 1.5-kb deletion in AcR mutant background leading to *nhaA* gene overexpression for enhanced sodium acetate tolerance phenotype (Yang et al., 2010b), we haven't systematically explored the global transcriptional profile difference between AcR mutant and wild-type ZM4 especially in the condition of minimal medium (MM), which is potentially more relevant to industrial fermentation conditions and more stressful to *Z. mobilis* than rich media (RM) we used for previous study (Yang et al., 2010b). In addition, there is no systems biology for this important industrial strain in minimal medium yet. To further explore the acetate-tolerance differences between ZM4 and AcR in MM condition, comprehensive microarray-based transcriptomic profiles and  $^{14}/^{15}\text{N}$ -labelled quantitative proteomic data were generated for ZM4 and AcR in MM with a large number of *Z. mobilis* proteins detected and quantified, which will be useful for future biocatalyst development. In addition, data collected in this and previous studies were used to hypothesize condition-responsive (strain, media, growth phase, or treatment) genes.

## RESULTS

### PHYSIOLOGICAL RESPONSE OF *Z. MOBILIS* TO SODIUM ACETATE IN MINIMAL MEDIUM (MM)

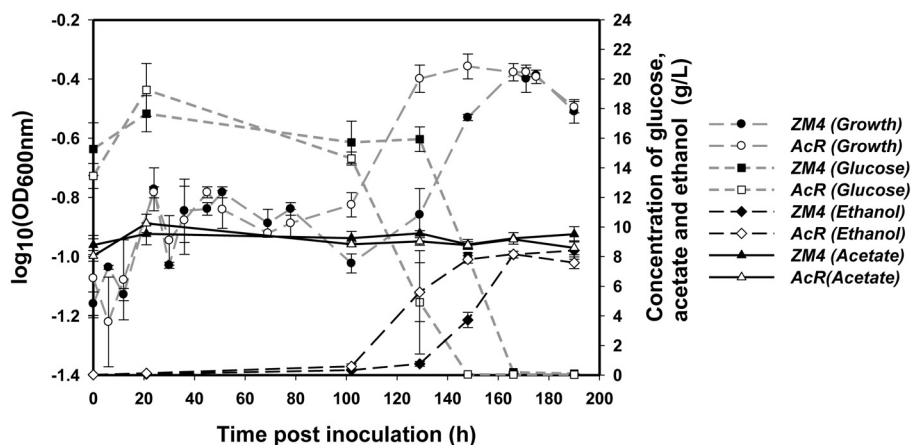
The growth of *Z. mobilis* wild-type ZM4 and AcR mutant in MM supplemented with 0, 12, or 16 g/L NaAc as well as 8.65 or 11.4 g/L NaCl with same corresponding  $\text{Na}^+$  molar concentrations as NaAc was assessed using a Bioscreen C instrument (Growth Curves USA, NJ) under anaerobic conditions to determine the effect of NaCl and NaAc on *Z. mobilis* growth and to decide on an appropriate NaAc concentration for subsequent systems biology studies. *Z. mobilis* grew more slowly in MM and attained lower cell densities (Additional File 1) compared to that in RM conditions (Yang et al., 2010b). Consistent with earlier RM results, wild-type ZM4 growth was arrested when NaAc was added to minimum medium at 16 g/L, and differences were observed between ZM4 and AcR with NaAc supplemented at concentrations of 12 g/L (Additional File 1). The concentration of 10 g/L was chosen for NaAc treatment in subsequent systems biology studies.

Similar to previous reports in RM (Joachimstahl et al., 1998; Yang et al., 2010b), mutant strain AcR also outperformed wild-type strain ZM4 in MM supplemented with NaAc (Figure 1, Additional File 1). Strain AcR overcame the acetate stress and reached stationary phase after about 130 h post-inoculation while ZM4 reached stationary phase after approximately 166 h. Both strains achieved similar final cell densities based upon  $\text{OD}_{600\text{nm}}$  readings. Strain ZM4 had a longer lag phase and began to consume glucose much later than AcR, with most of the glucose in AcR cultures consumed at 148 h post-inoculation while cultures of ZM4 took 166 h to use the majority of the glucose in MM (Figure 1). As glucose was consumed earlier by AcR cultures, there was also a concomitant earlier onset for ethanol production compared to ZM4. Both strains produced similar yield of ethanol by the end of the experiment with acetate concentration kept steady during the whole experiment (Figure 1).

### *Z. MOBILIS* WILD-TYPE ZM4 AND AcR MUTANT PROTEOMIC PROFILING DIFFERENCES IN MM

Quantitative proteomics was used to compare the proteomic differences between *Z. mobilis* wild-type ZM4 and AcR mutant cells at 148 and 190 h post-inoculation. We identified 705 chromosomal encoded proteins and 6 plasmid derived proteins for the 148 h comparison, and 728 chromosomal proteins and 7 plasmid proteins for the 190 h comparison. We thus identified 638 chromosomal proteins and 4 plasmid proteins in common for the two time points. Altogether, this study identified 795 unique chromosomal proteins and 9 unique plasmid proteins, which is about 46% of the predicted *Z. mobilis* ZM4 proteome based upon a reannotation of the genome (Yang et al., 2009a) (Additional Files 2, 3A). The PI and MW distributions for the identified proteins were similar to that of the predicted theoretical PI and MW distributions for all proteins (Additional File 4). This study represents, to our knowledge, the largest quantitative measurement of the *Z. mobilis* proteome published to date.

The 148 h time point comparison between ZM4 and AcR strains revealed 120 proteins that had at least 1.5-fold differences in abundance and were significantly different based upon



**FIGURE 1 |** The growth, glucose consumption, ethanol production and acetate amount of *Z. mobilis* wild-type ZM4 and acetate-tolerant mutant AcR in the presence of ~10 g/L NaAc.

the ProRata likelihood algorithm for quantitative shotgun proteomics and a confidence interval of 90% as described previously (Pan et al., 2006). This included 55 upregulated proteins and 65 downregulated proteins in AcR compared to ZM4, of which 11 were upregulated and 25 were downregulated more than 2-fold (Additional File 2). There were 107 proteins with a significant difference and at least a 1.5-fold change in levels for stationary phase comparison (190 h) using the same criteria. In strain AcR, 47 proteins were upregulated and 60 downregulated compared to ZM4, of which 10 were upregulated and 26 were downregulated and had changes of at least 2-fold (Additional File 2). Approximately half of the proteins that were identified as being different for one time point were also statistically different in the other time point (Additional Files 2, 3B–C).

The interactions among the down-regulated or up-regulated proteins with at least a 1.5-fold change were also analyzed for their previously documented interactions using STRINGS database (Jensen et al., 2009). The 14 proteins consistently upregulated in AcR for both time points compared to ZM4 had fewer interactions (Additional Files 2, 3D). Among the 39 proteins downregulated in AcR at both the 148 and 190 h time points post-inoculation, stress-responsive proteins such as catalase (ZMO0918), glutathione synthetase (ZMO1913), glutaredoxin 2 (ZMO0070), and a glutaredoxin-related protein (ZMO1873) were downregulated and most of them have been shown to interact with one another (Additional Files 2, 3E). Other proteins downregulated and connected included preprotein translocase (ZMO1896-8), cofactor synthesis of biotin synthase (ZMO0094), and amino acid biosynthesis such as shikimate 5-dehydrogenase (AroB, ZMO0041), threonine synthase (ThrC, ZMO1891), and 3-isopropylmalate dehydratase large subunit (LeuC, ZMO0105) (Additional Files 2, 3E).

Although our intention was to identify differentially expressed proteins more confidently by combining datasets from both time points than merely the second time point of stationary at 190 h to exclude the growth effect, the first comparison however will be potentially problematical due to the growth difference between

AcR mutant and wild-type ZM4 at 148 h time point (Figure 1). Therefore, although comparison for second time point of stationary phase at 190 h is reasonably accurate, caution should be taken to interpret proteomic comparison result of second time point due to potential growth effect impact. To overcome this problem, samples from same growth phase between AcR mutant and ZM4 wild-type were used for our transcriptomic comparisons as discussed below.

#### **Z. MOBILIS WILD-TYPE ZM4 AND AcR MUTANT TRANSCRIPTOMIC PROFILING DIFFERENCES IN MM**

Transcriptomic profiles for ZM4 and AcR in the presence of NaAc in MM were examined using NimbleGen high density expression arrays, essentially as reported previously (Yang et al., 2010b). The expression profiles generated in this study have been deposited in the GEO database (accession number GSE25443). About seventeen hundred genes were identified to be significantly differentially expressed using ANOVA modeling with strain (ZM4 and AcR) and time points post-inoculation as variables in MM (Additional File 5), which covered nearly all of the reannotated *Z. mobilis* ZM4 genes (Yang et al., 2009a). Hence dynamic gene expression changes were observed and for NaAc responses, 474 genes were significantly differentially expressed with at least a 2-fold change in expression values (Additional File 5).

Nine differentially expressed genes from different functional categories with a broad range of expression ratios were chosen for real-time quantitative PCR (RT-qPCR) validation (Additional File 6). RT-qPCR results indicated a high degree of concordance between microarray and RT-qPCR data with R-squared correlation coefficient values of 0.88, 0.87, 0.71, or 0.78 for the time points of 130, 148, 166, or 190 h respectively (Additional File 7).

Four genes were upregulated and 30 genes downregulated significantly with at least 2-fold changes in AcR when comparing AcR expression profiles to ZM4 profiles for all the time points (Additional File 5). Similar to the transcriptomics study in RM as reported previously (Yang et al., 2010b), the sodium/proton antiporter gene (ZMO0119) was also upregulated and ZMO0117

was downregulated in AcR in MM conditions. The other three genes that were consistently upregulated in AcR were L-asparaginase (ZMO1683), beta-fructofuranosidase (ZMO0375), and aldose 1-epimerase (ZMO0889). The 29 remaining genes that were consistently downregulated encoded mostly ribosomal proteins as well as proteins related to chemotaxis and flagellar biosynthesis, electron transport, fatty acid biosynthesis and protein export (Additional File 5).

#### CORRELATIONS BETWEEN PROTEOMIC AND TRANSCRIPTOMIC DATA OF ZM4 AND AcR IN MM

We also examined the correlations between the gene or protein expression ratios at two time points of 148 and 190 h post-inoculation. There were 632 common proteins identified in both the 148 and 190 h time points. When the relative  $\log_2$  ratios (AcR/ZM4) were compared, a moderate correlation ( $R^2 = 0.47$ ) was observed between the different sampling times (Figure 2A). Similarly, the correlation of transcriptomic profiles (AcR/ZM4) between 148 and 190 h was moderately high ( $R^2 = 0.59$ ) for the 530 common genes that showed statistically significant differential expression, but had not undergone any fold-change filtering (Figure 2B). When transcriptomic and proteomic data from the same time point (either 148 or 190 h) were compared, there are 460 gene-protein pairs for the 148 h time point that correlated poorly ( $R^2 = 0.09$ ) (Figure 2C), and the correlation between the 302 gene-protein pairs at 190 h post-inoculation is also relatively low ( $R^2 = 0.17$ ) (Figure 2D).

We have reported that the correlation between transcriptomic and proteomic data increases when only the significant genes or proteins were included for comparison, and the correlation even increases more when the ratios of transcriptomics or proteomics comparison increases from our recent *Z. mobilis* ethanol stress experiment of high-density microarray and shot-gun proteomics (Yang et al., 2013). Similarly, in this study, when only the genes or proteins with a significant and a  $\geq 1.5$ -fold change were used for comparison purposes, the correlation between transcriptomic and proteomic data increased to  $R^2 = 0.30$  ( $N = 35$ ) and  $R^2 = 0.25$  ( $N = 32$ ) for 148 and 190 h comparisons, respectively (Figure 2E). Eighteen common gene-protein pairs were identified between proteomics and transcriptomics data in both 148 and 190 h post-inoculation with at least a 1.5-fold significant change, and all were downregulated in AcR (Table 1). The correlation for the proteomic study between 148 and 190 h went up to  $R^2 = 0.79$  ( $N = 18$ ) when significance and fold change were considered (Table 1, Figure 2F). However, the correlation for the transcriptomic study between 148 and 190 h were decreased to 0.01 ( $N = 18$ ) when significance and fold change criteria were applied (Table 1, Figure 2F). This result may suggest that change at transcriptional level is rapid which is consistent with our transcriptomic study in *Clostridium thermocellum* that genes were differentially expression even after several minutes post ethanol or furfural shock (Yang et al., 2012; Wilson et al., 2013). Moreover, this study also suggested that the change at protein level is relatively steady with potential post-transcriptional modification existing.

#### METABOLIC PATHWAY ANALYSES OF PROTEOMIC AND TRANSCRIPTOMIC DATA OF ZM4 AND AcR IN MM

The PathwayTools Omics Viewer (Karp et al., 2002, 2011) was used to further examine transcriptomics and proteomics data and their relationships. Genes in the fatty acid and hopanoid biosynthetic pathways are upregulated in ZM4 compared to AcR (Additional File 8A). Hopanoids have been found in a variety of bacteria including *Z. mobilis* and are reported to protect against the toxic effects of ethanol (Flesch and Rohmer, 1989; Hermans et al., 1991; Horbach et al., 1991; Shigeri et al., 1991; Welander et al., 2009). For example, they have also been reported to play a role in *Rhodospseudomonas palustris* membrane integrity and pH homeostasis (Welander et al., 2009). However, there is debate of their role in *Z. mobilis* (Moreau et al., 1997) and we didn't identify obvious differential gene expression when *Z. mobilis* encountered ethanol stress (Yang et al., 2013). Further study is required to elucidate their role in *Z. mobilis*, and their relative contribution, if any, to sodium acetate sensitivity or tolerance to the strains used in this experiment.

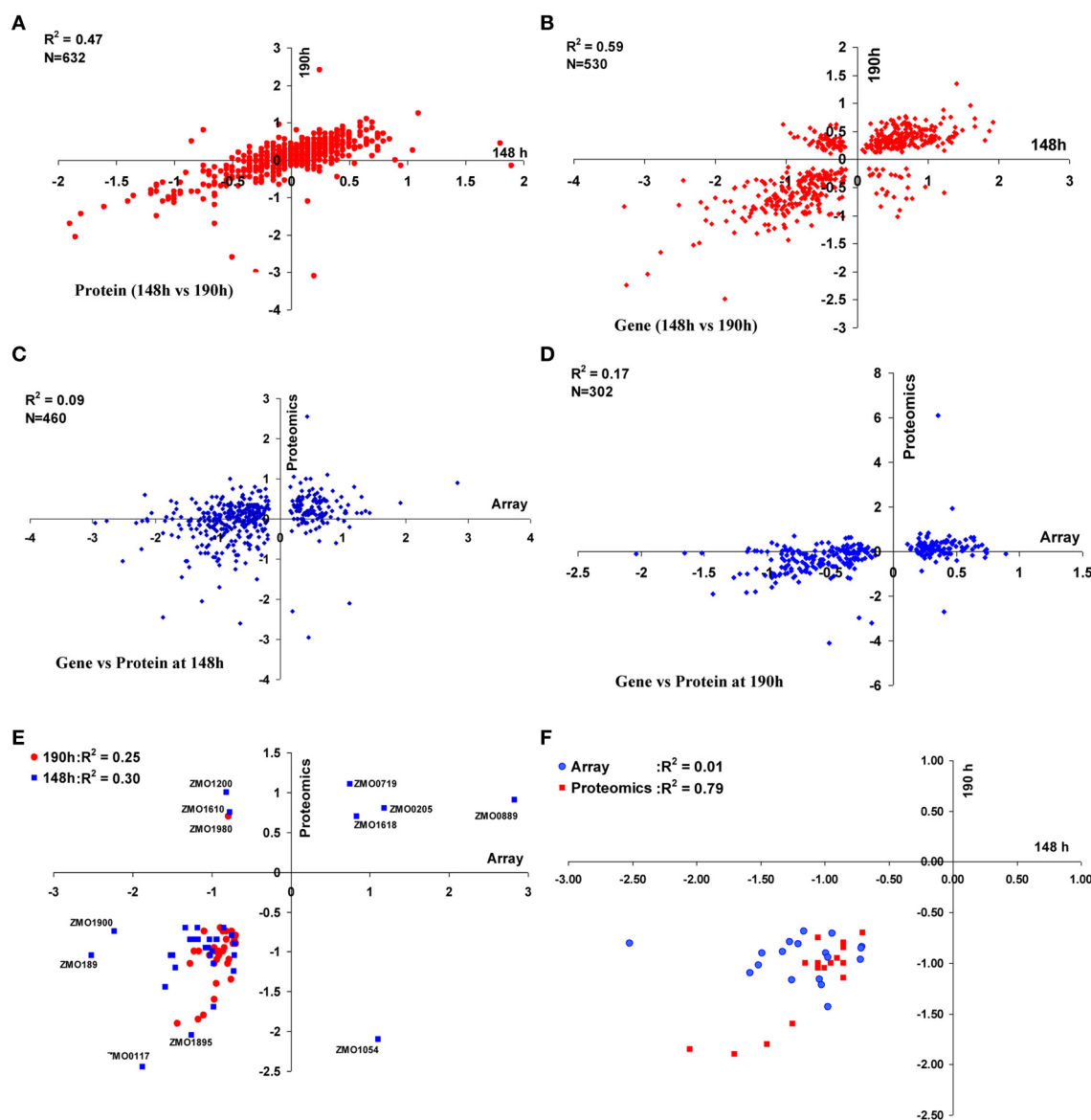
In other pathway comparisons of AcR to ZM4 at 148 and 190 h, the strains had similar patterns of protein detection, with 120 or 107 differentially expressed proteins with  $\geq 1.5$ -fold change at each time point respectively were used for analyses (Additional Files 8B,C). We identified 509 chromosomal gene/protein pairs that have no significant changes between AcR and ZM4 in MM with NaAc in both 148 and 190 h post-inoculation (Additional File 2), the majority of which are involved in the ED and mixed-acid fermentation pathways, and tRNA charging (Additional File 8A–C). When the  $\log_2$  transformed peptide hits of all the proteins identified in 148 and 190 h post-inoculation were used for pathways analyses, proteins involved in ED pathway and ethanol production are among the most abundant proteins (Additional Files 2, 8D). The abundance of ED pathway and ethanol production enzymes and the relatively stability of protein levels between ZM4 and AcR indicates that this pathway for carbon and electron flow are core elements in the physiology of both strains.

Recently, a number of studies have added important details to our understanding of *Z. mobilis* physiology (Kouvelis et al., 2009; Yang et al., 2009a, 2010a,b; Kerr et al., 2010; Linger et al., 2010; Zhang et al., 2010; Widiastuti et al., 2011). However, many genes encode proteins for which we now have proteomics evidence of their expression, their functions remain unknown. The largest quantitative measurement of the *Z. mobilis* proteome obtained in this study may help elucidate *Z. mobilis* genes, proteins, their functions and regulation.

#### ZM4 AND AcR TRANSCRIPTOMIC PROFILES IN DIFFERENT CONDITIONS

In this study, we also combined *Z. mobilis* microarray data collected from previous studies with expression data collected in this experiment to provide greater insights in *Z. mobilis* physiology and gene regulation. Forty-four microarrays were used in the analysis (Additional File 9) to compare the gene expression differences between the variables of strain (AcR and ZM4), growth phase (exponential and stationary), media (RM and MM), and treatment (NaCl, NaAc, and control of RM only). Considering the growth differences between ZM4 and AcR in this





**FIGURE 2 | Correlations between proteomic and transcriptomic data of ZM4 and AcR in MM.** Correlation within proteomics or transcriptomics shared at time point of 148 and 190 h between 632 proteins (A) or 530 genes (B) differentially expressed between AcR and ZM4 (AcR/ZM4) in MM; and the correlation between proteomics or transcriptomics at different time points of 148 h (C) or 190 h (D)

post-inoculation; as well as correlations between significant gene-protein pairs with at least 1.5-fold changes between transcriptomics and proteomics (E); and the correlation between common genes or proteins with at least 1.5-fold significant changes shared between the time point of 148 h and 190 h post-inoculation within transcriptomics or proteomics (F). Numbers of X-axis and Y-axis are  $\log_2$ -based values.

experiment (Figure 1) and the fact that *Z. mobilis* has dramatic transcriptional differences between exponential and stationary phase (Additional Files 5, 10) (Yang et al., 2009b), the relative large number of genes significantly differentially expressed at time point 148 h post-inoculation between AcR and ZM4 as well as the low transcriptomic correlation between 148 and 190 h post-inoculation is likely associated with the growth phase differences. To ensure compatible comparisons were being made among different time points in RM or MM under different conditions, the correlations among all the forty-four microarrays were investigated using the JMP Genomics 4.0 (SAS, NC) and

based upon a hierarchical clustering and growth curve analyses (Figure 3), time points were assigned to either exponential or stationary phase for further statistical analyses (Figure 3, Additional File 9). Differences between the variables of strain, growth phase, media, and treatment were then analyzed by ANOVA, as described previously (Yang et al., 2009b, 2010b). The ANOVA identified that nearly every gene in the *Z. mobilis* genome was differentially expressed under one or more of the many conditions tested in this global analysis (Additional File 11), leading to an opportunity to identify the strain, media, growth phase or treatment responsive genes, as discussed below.



**Table 1 | Gene-protein pairs identified in both transcriptomic and proteomic studies at both 148 and 190 h post-inoculation, and those only identified in either 148 or 190 h post-inoculation.**

Gene	Product	A148	P148	A190	P190
18 TRANSCRIPTOMICS-PROTEOMICS PAIRS IDENTIFIED IN BOTH 148 AND 190 h					
ZMO0009	Sulfite reductase [NADPH] flavoprotein, alpha chain	−1.20	−0.85	−0.81	−0.85
ZMO0020	Protein of unknown function DUF162	−1.58	−1.45	−1.10	−1.80
ZMO0021	Protein of unknown function DUF162	−0.97	−1.70	−1.43	−1.90
ZMO0027	IMP cyclohydrolase	−1.52	−1.05	−1.02	−1.05
ZMO0035	Ankyrin	−0.72	−1.25	−0.96	−1.60
ZMO0041	Shikimate 5−dehydrogenase	−0.71	−0.90	−0.84	−0.95
ZMO0079	Response regulator receiver protein	−0.71	−1.05	−0.85	−1.00
ZMO0090	Short−chain dehydrogenase/reductase SDR	−0.99	−1.00	−0.90	−1.05
ZMO1372	Short−chain dehydrogenase/reductase SDR	−1.27	−0.85	−0.79	−1.15
ZMO1891	Threonine synthase	−1.04	−0.95	−1.16	−1.00
ZMO1895	PfkB domain protein	−1.26	−2.05	−1.17	−1.85
ZMO1896	Protein−export membrane protein SecF	−1.33	−0.70	−0.89	−0.70
ZMO1897	Protein−export membrane protein SecD	−2.52	−1.05	−0.81	−0.75
ZMO1912	GTP−binding protein YchF	−0.97	−1.15	−0.95	−1.00
ZMO1913	Glutathione synthetase	−1.16	−0.85	−0.69	−0.80
ZMO1927	Signal peptide peptidase SppA, 67K type	−1.49	−1.05	−0.90	−1.00
ZMO1955	Malate dehydrogenase	−0.94	−0.85	−0.71	−0.85
ZMO1964	Glutamyl−tRNA synthetase	−1.02	−0.85	−1.22	−1.00
Gene	Product	A148	P148		
17 UNIQUE TRANSCRIPTOMICS-PROTEOMICS PAIRS IDENTIFIED ONLY IN 148 h					
ZMO1905	Glycerol−3−phosphate dehydrogenase	−0.93	−0.85		
ZMO1900	Fatty acid/phospholipid synthesis protein PlsX	−2.22	−0.75		
ZMO1899	3−oxoacyl−(acyl−carrier−protein) synthase III	−1.46	−1.20		
ZMO1898	Preprotein translocase, YajC subunit	−1.07	−0.95		
ZMO1875	Protein of unknown function DUF1476	−0.84	−0.70		
ZMO1618	Carbamoyl−phosphate synthase, small subunit	0.84	0.70		
ZMO1610	Hypothetical protein	−0.77	0.75		
ZMO1200	Sugar−phosphate isomerase, RpiB/LacA/LacB family	−0.81	1.00		
ZMO1054	DNA topoisomerase IV, A subunit	1.10	−2.10		
ZMO0889	Aldose 1−epimerase	2.83	0.90		
ZMO0719	Lytic transglycosylase catalytic	0.75	1.10		
ZMO0205	NAD−dependent epimerase/dehydratase	1.18	0.80		
ZMO0117	hybrid cluster protein	−1.87	−2.45		
ZMO0094	Biotin synthase	−1.18	−0.70		
ZMO0080	CheD, stimulates methylation of MCP proteins	−0.70	−0.90		
ZMO0022	Protein of unknown function	−1.01	−1.05		
ZMO0013	Purine NTP pyrophosphatase, rdgB/HAM1 family	−0.74	−0.80		
Gene	Product	A190	P190		
14 UNIQUE TRANSCRIPTOMICS−PROTEOMICS PAIRS IDENTIFIED ONLY IN 190 h					
ZMO0017	Fmu	−0.73	−0.90		
ZMO0056	Glucosamine/fructose−6−phosphate aminotransferase	−0.78	−1.10		
ZMO0062	Aldo/keto reductase	−0.97	−0.95		
ZMO0070	Glutaredoxin, GrxB family	−0.75	−1.35		
ZMO0106	3−isopropylmalate dehydratase, small subunit	−1.09	−0.75		
ZMO1885	NADH:flavin oxidoreductase/NADH oxidase	−0.74	−0.75		
ZMO1956	DNA repair protein RecN	−0.81	−0.75		
ZMO1963	Citrate synthase I	−0.86	−0.75		
ZMO1970	3−methyl−2−oxobutanoate hydroxymethyltransferase	−0.69	−0.90		

(Continued)

Table 1 | Continued

Gene	Product	A148	P148	A190	P190
ZMO1980	Methyltransferase GidB	−0.78	0.70		
ZMO1984	Aldo/keto reductase	−0.96	−1.15		
ZMO1989	Methylated–DNA/protein–cysteine methyltransferase	−0.93	−1.10		
ZMO1992	Carboxymethylenebutenolidase	−0.94	−1.40		
ZMO1993	Alcohol dehydrogenase GroES domain protein	−1.27	−1.15		

A148 and A190: array result comparing AcR to ZM4 at time point 148 and 190 h post-inoculation respectively; P148 and P190: proteomics result comparing AcR to ZM4 at time point 148 and 190 h post-inoculation respectively. All the gene-protein pairs are statistically significant with at least 1.5-fold difference between AcR and ZM4 in both transcriptomic and proteomic studies. Negative number indicates gene/protein was downregulated comparing AcR to ZM4 and positive number indicates upregulation.

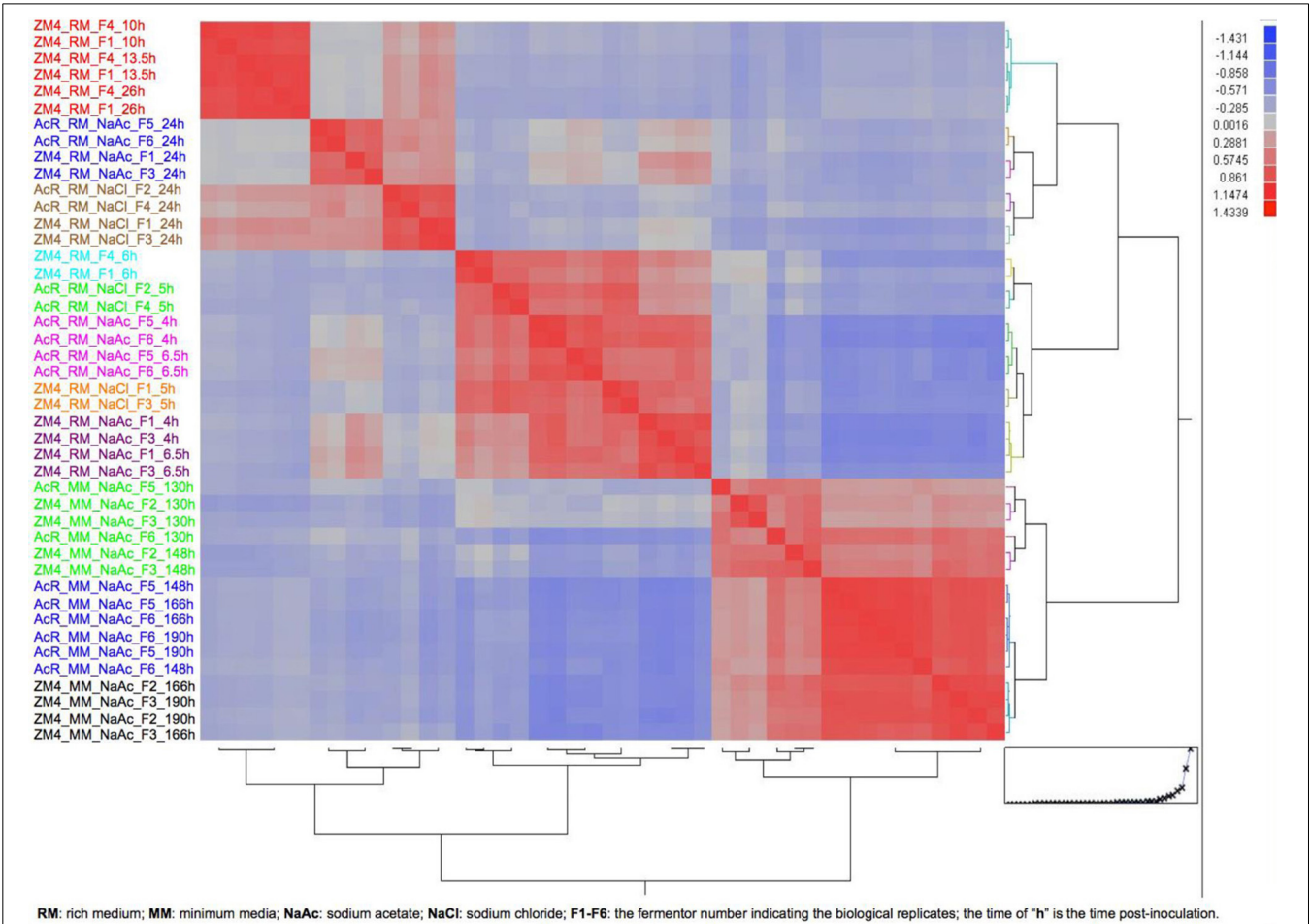


FIGURE 3 | The heat map and dendrogram of hierarchical clustering analysis on correlation among 44 arrays for growth phase determination. The information about each array is listed on figure left and the color from top to bottom indicates different growth media and phase.

Strain-specific genes

Gene expression profiles for ZM4 and AcR were compared and when all the conditions (Additional File 9) were taken into account, only five genes were significantly differentially expressed with at a least 2-fold change between strains. The significant differentially expressed strain specific genes included two genes with increased expression and three genes down-regulated in strain AcR (Additional Files 11, 12). We have

previously reported that *nhaA* (ZMO0119) is upregulated and ZMO0117 is downregulated in AcR compared to ZM4 in exponential and stationary growth conditions for RM (Yang et al., 2010b). In this study, which considered more conditions, we identified that that another gene (hypothetical protein ZMO1787, 82aa) was consistently upregulated in AcR under the conditions tested and two AcR genes encoding a predicted permease (ZMO0055, 263 aa) and a conserved hypothetical protein

(ZMO0025, 234 aa) were consistently downregulated (Additional File 11).

### Media-responsive genes

Gene expression differences between MM and RM in the presence of NaAc were compared (Additional File 9). When the strain responses of both AcR and ZM4 in both exponential and stationary phases were combined to examine media differences, 232 genes were identified as being upregulated and 247 downregulated in MM compared to RM (Additional File 13). Almost half of upregulated genes encode hypothetical proteins with unknown functions, and many of the remaining upregulated genes encode proteins involved in stress sensing and responses. These included two-component signal transduction genes (7), transcriptional regulators (11), sigma factors (*rpoD*, *rpoH*), flagellar and chemotaxis genes, and other stress responsive genes such as an putative operon containing heat-inducible transcription repressor *hrcA* (ZMO0015) and chaperone gene *grpE* (ZMO0016). Other genes included cold shock gene (ZMO0925), DnaJ-class chaperone (ZMO1069) and chaperone *hspD* (ZMO0989), three glutaredoxin-related genes, three thioredoxin-like genes, and genes associated with DNA repair such as UvrABC system gene *uvrB* (ZMO0362), *mutM* (ZMO1187), *mutS* (ZMO1907) and ZMO1426 (Additional File 13).

There are more previously documented interactions among the downregulated genes in MM relative to RM compared to the upregulated genes (Additional File 14). These MM-downregulated genes are mostly related to the central carbon metabolism, which included genes such as 6-phosphogluconolactonase gene *pgl* (ZMO1478), phosphoglycerate mutase gene *pgmA* (ZMO1240), gluconolactonase gene *gnl* (ZMO1649), pyruvate decarboxylase gene *pdC* (ZMO1360), and pyruvate dehydrogenase gene *pdhA* and *pdhB* (ZMO1606, ZMO1605) (Additional Files 13, 14B). Genes involved in amino acid biosynthesis, encoding ribosomal proteins and genes for nucleotide biosynthesis, as well as genes related to energy metabolism such as electron transport system and ATP generation were also mainly down-regulated in MM (Additional Files 13, 14B).

### Treatment-responsive genes

The following conditions were used to examine treatment-specific responses: ZM4 grown in RM; both ZM4 and AcR growth in RM with NaCl, RM with NaAc, or MM with NaAc (Additional File 9). These conditions were used to make the following comparisons: treatment of NaCl or NaAc in RM vs. RM for strain ZM4; and as the treatment of NaAc vs. NaCl in RM for both ZM4 and AcR strains (Additional File 13).

The NaCl-responsive genes in the ZM4 wild-type background included 88 downregulated genes and 47 upregulated genes (Additional Files 13, 15A,B). More genes were responsive in a ZM4 background grown in RM supplemented with NaAc, including 159 downregulated and 103 upregulated genes. This was approximately twice the number as responsive to the less severe NaCl treatment (Additional Files 13, 15C,D). Differentially expressed genes shared between NaCl and NaAc in RM compared to RM included 27 upregulated and 66 downregulated genes

(Additional Files 13, 15E,F). Genes upregulated in both NaCl and NaAc treatment related to cysteine synthesis, with 6 genes clustered together (*cysD*, *I*, *J*, *K*, *N*, ZMO0006 and ZMO0055) and tryptophan synthesis (*trpB*, *F*), and most of the downregulated genes encoded hypothetical proteins with unknown function with the exception of several related to flagellar biosynthesis and sucrose metabolism (Additional Files 13, 15).

When the responses of *Z. mobilis* in RM with NaAc were compared to that of NaCl, there were 37 upregulated and 41 downregulated genes (Additional Files 13, 15I,J). Genes related to ribosomal proteins and amino acid biosynthesis were downregulated and upregulated genes related to the Entner–Doudoroff (ED) pathway (ZMO1518 and *pgi*), energy metabolism [e.g., electron transport (ZMO0021, ZMO1851, ZMO1885) and ATP synthesis gene *atpC* (ZMO0242)], cell wall formation (ZMO1724), regulator gene *zrp* (ZMO0372), and several transporter related genes such as signal peptidase I gene *lepB* (ZMO1710), Sec-independent protein translocase *tatA/E* homolog gene (ZMO1220), TolQ biopolymer transport gene (ZMO0161) and biopolymer transport gene *exbD* (ZMO1715), Fe<sup>2+</sup> transport system gene *feoB* (ZMO1541), and potassium transport system gene *kup* (ZMO1209).

### Growth phase-dependent genes

When strain profiles in the presence of NaAc were considered together and a  $\geq 2$ -fold change was applied, 331 genes were significantly differentially expressed in a phase-dependent manner for RM (Additional Files 11, 12). When the RM NaCl profiles were analyzed, 661 genes were identified (Additional Files 11, 12). Similar patterns were identified for the comparisons when variable of strain was taken into consideration separately (Additional Files 11, 12). In the presence of NaAc, the stationary phase upregulated genes were less correlated than the downregulated ones (Additional Files 16). The genes downregulated in stationary phase are related to ribosomal proteins, chemotaxis and flagellar systems, amino acid and nucleotide biosynthesis, and electron transport (Additional Files 11, 16A). Except for those encoding hypothetical proteins, genes upregulated in stationary phase are related to stress responses such as the catalase gene (ZMO0918), glutaredoxin gene (ZMO0070), thioredoxin gene (ZMO1705), ferredoxin gene *fdxN* (ZMO1818), organic hydroperoxide resistance gene (ZMO0693), ATP-dependent Clp protease gene *clpB* (ZMO1424), RNA polymerase sigma-32 factor gene *rpoH* (ZMO0749), and integration host factor gene *ihfB* (ZMO1801) (Additional Files 11, 16B).

## DISCUSSION

### GENES CONTRIBUTING TO SODIUM ACETATE TOLERANCE PHENOTYPE OF AcR

AcR mutant had advantages under NaAc stress in both RM (Yang et al., 2010b) and MM conditions, with shorter lag phase leading to earlier glucose consumption and ethanol production than ZM4 cells (Figure 1). We have reported previously that the truncation within gene ZMO0117 in AcR causes the consistent upregulation of the sodium proton antiporter gene *nhaA* (ZMO0119), which is the determining factor of AcR mutant for sodium acetate tolerance in RM (Yang et al., 2010b). In this study, the *nhaA* gene also

had higher expression levels in strain AcR compared to ZM4 in MM (Additional File 11). ZM4 *nhaA* expression in unamended RM was higher (~2-fold) in stationary phase compared to exponential phase (Additional File 11) and differential expression was also measured for other conditions and strain comparisons (Additional File 13). Acetate accumulates in stationary phase *Z. mobilis* fermentations (Yang et al., 2009b) and greater expression of *nhaA* gene may ameliorate cellular function and activity under such stressful conditions.

In addition, compared to ZM4, the ZMO0117 gene in AcR showed a similar profile of down-regulation in MM (Additional File 11) to that of in RM (Yang et al., 2010b). Proteomic data also showed that the peptide hits for ZMO0117 were less abundant in AcR compared to ZM4 at 148 h post-inoculation (Additional File 2). However, the NhaA protein was not detected in any sample in this experiment (Additional File 2) and yet differential expression was observed for this gene, which demonstrates the value of conducting integrated experiments and may point to a technical challenge in detecting hydrophobic proteins like NhaA that have many transmembrane spanning domains (Poetsch and Wolters, 2008; Gilmore and Washburn, 2010).

As discussed above, this study also identified other AcR specific genes. For example, ZMO0055 was significantly downregulated in AcR compared to ZM4 in all conditions with NaAc and responded to other conditions too (Additional File 11). ZMO0055 is predicted to belong to the sulfite exporter TauE/SafE family protein (pfam01925), a family of integral membrane proteins involved in transporting anions across the cytoplasmic membrane during taurine metabolism as an exporter of sulfoacetate (Weinitschke et al., 2007). As discussed above, ZMO0055 is closely related to CysC, D, I, N (ZMO0003, 4, 5, 8) based on protein analysis using String pre-computed database (Additional File 15), which were upregulated in both NaCl and NaAc treatment. In addition, recent work by Skerker et al. (2013) also indicated that cysteine synthases, which are required for L-cysteine biosynthesis from sulfate, are related to hydrolysate tolerance. They proposed that the increased demand for sulfite and cysteine is due to the increased needs of glutathione, which is formed from glutamate and cysteine, during *Z. mobilis* growth in stressful condition (Skerker et al., 2013). Therefore, the downregulation of sulfite exporter potentially could help *Z. mobilis* retain sulfite for enhanced sodium acetate tolerance phenotype. However, further detailed study is needed to understand the role and contribution of these AcR specific genes, which will help reveal the complete difference between AcR and ZM4 and at the same time help assign the function for these AcR-specific genes with hypothetical function (ZMO1787 and ZMO0025).

#### RELATIONSHIP AMONG MEDIA, INHIBITOR, GROWTH PHASE, AND GENE EXPRESSION DYNAMICS

The gene expression intensity mean value of the significantly expressed genes (Additional File 17) were further used for hierarchical clustering to compare each gene at 14 different conditions, the result indicated that minimum medium had the most dramatic effect on gene expression compared to those in RM, followed by growth phase, inhibitors (NaAc>NaCl), and strain (Additional File 18A). Gene expression profiles in MM were

clustered together and separated from the rest in RM (Additional File 18A). *Z. mobilis* growth in MM has a longer lag phase and generates a lower final cell density compared to that in RM (Additional File 1) (Yang et al., 2010b), which indicates that one or more factors are limiting *Z. mobilis* growth in MM, and MM is possibly a more stressful environment for *Z. mobilis* cells which is further supported by the upregulation of stress-responsive genes in MM compared to that in RM (Additional Files 13, 14).

The second level of clusters was between exponential and stationary growth phases, consistent with an earlier global profiling report for strain ZM4 that growth phase tends to have more dramatic effect on gene expression than a single stressor (Yang et al., 2009b). The greatest number of differentially expressed genes between exponential and stationary phases is observed for strain ZM in RM, followed by strain ZM in RM with NaCl, RM with NaAc and then ZM4 in MM (Additional File 12). In RM, *Z. mobilis* growth is inhibited by NaCl but NaAc is more inhibitory (Yang et al., 2010b), and a similar trend was observed in MM (Additional File 1). In this experiment, the inhibitor (NaAc) was added into the medium before the fermentation allowing the strains to respond and adapt to the conditions for several generations before samples were taken for systems biology studies. Hence, when transcriptomic and proteomic profiles between strains are compared under the same growth condition, the differences likely represent differences in homeostasis that allow AcR to function better than ZM4. The two strains were in different growth phases for the first proteomics time point, which confounds strain comparisons for tolerance but does still allow gene-protein relationships to be examined. In general, the large datasets reported in this study will allow others to investigate aspects such as codon bias and RNA secondary structures.

In addition, the microarray results for each gene at different conditions can help us understand the microbial physiology and identify consistently strong or inducible promoters for metabolic engineering application. For example, in this study, 52 genes and 6 other genetic features were among the top 2.5% of all genetic features with consistently strong expression intensity. Except for a dozen of genes encoding hypothetical proteins, they involve in protein biosynthesis, glycolysis and fermentation as well as ATP synthesis and stress response indicating their key role on *Z. mobilis* metabolism (Additional File 17). Several papers have been published recently to study the stress responses of *Z. mobilis* to different inhibitors (e.g., ethanol, furfural) using systems biology approaches (Yang et al., 2009b, 2010b, 2013; He et al., 2012a,b; Jeon et al., 2012; Skerker et al., 2013), we are currently working on several transcriptomics studies using microarray and next-generation sequencing based strand-specific RNA-Seq technique. Combining all these systems biology datasets, we will revisit this topic to investigate the impact of different variables of media, carbon source (glucose or xylose), different inhibitors, and growth phase on *Z. mobilis*. The transcriptional profiles will be further compared to help identify the condition specific promoters more confidently, and the discrepancy between transcriptomics and proteomics at different conditions will help understand the post-transcriptional regulation mechanism.



## CONCLUSIONS

Our study has provided the global profiling of the model ethanogenic bacterium *Z. mobilis* at both transcriptomic and proteomic levels in minimal medium (MM) for the first time. The results indicated that AcR had similar advantage in MM as in RM. AcR mutant can overcome acetate stress earlier with shorter lag phase, earlier glucose utilization and ethanol production than wild-type ZM4, and sodium proton antiporter gene (*nhaA*) also plays an important role in sodium acetate tolerance in MM as that of in RM we reported previously (Yang et al., 2010b). In addition, this is also the first attempt that we are aware of to combine massive transcriptomic data for condition-specific gene identification. The proteomic and transcriptomic data generated in this study and the one we have reported will provide massive datasets for future metabolic modeling and strain improvement.

## MATERIALS AND METHODS

### STRAINS AND GROWTH CONDITIONS

Wild-type *Z. mobilis* ZM4 was obtained from the American Type Culture Collection (ATCC 31821). *Z. mobilis* acetate tolerant strain AcR has been described previously (Joachimstahl et al., 1998). ZM4 and AcR were cultured in RM (Glucose, 20.0 g/L; Yeast Extract, 10.0 g/L;  $\text{KH}_2\text{PO}_4$ , 2.0 g/L, pH5.0) (Yang et al., 2009b) for routine strain maintenance. MM was similar to the one reported to isolate auxotrophic *Z. mobilis* mutants (Goodman et al., 1982): 20 g Glucose, 1 g  $\text{KH}_2\text{PO}_4$ , 1 g  $\text{K}_2\text{HPO}_4$ , 0.5 g NaCl, 1 g  $(\text{NH}_4)_2\text{SO}_4$  was dissolved in 986.5 mL distilled  $\text{H}_2\text{O}$  with pH adjusted to pH 5.0 before autoclave. The sterilized MM broth was then cooled to below 55°C before 500 mg  $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$  (2.5 mL 200 g/L stock), 25 mg  $\text{Na}_2\text{MoO}_4 \cdot 2\text{H}_2\text{O}$  (1 mL 25 g/L stock), and 10 mL Vitamin (ATCC MD-VS) were added to make a 1-L MM.

Growth assays were used to identify the effects of sodium acetate (NaAc) and sodium chloride (NaCl) on the growth of *Z. mobilis* in MM for subsequent fermentation experiments using a Bioscreen C instrument (GrowthCurves USA, NJ). Fermentations were conducted in 7.5-L BioFlo110 bioreactors (New Brunswick Scientific, NJ) fitted with agitation, pH, and temperature probes and controls, and bacterial growth was monitored turbidometrically by measuring optical density at 600<sub>nm</sub> with a model 8453 spectrophotometer (Hewlett-Packard, CA.) as described previously (Yang et al., 2009b), except that the fermentation volume was 4 L. Samples were harvested during fermentation at different time points (Figure 1) as described previously (Yang et al., 2009b).

### HPLC

HPLC analysis was used for the measurements of the concentration of glucose, acetate, and ethanol in 0.2  $\mu\text{m}$ -filtered samples taken at different time points during fermentation (Figure 1) and analyzed as described previously (Yang et al., 2009b).

### PROTEOME SAMPLE PREPARATION

Duplicate mixtures of microbial cells that were metabolically labeled with either  $^{15}\text{N}$  ammonium sulfate for the wild-type *Z. mobilis* ZM4 culture or  $^{14}\text{N}$  ammonium sulfate for the AcR mutant culture were prepared by mixing equal weights of cell pellets from duplicate cultures for each strain. Cell mixtures were lysed by sonication in ice-cold 50 mM Tris-HCl (pH 7.5) buffer,

and unbroken cells were removed by centrifugation at  $5000 \times g$  for 10 min. Protein concentration for each sample was determined with the RC DC™ protein assay (Bio-Rad Lab, CA). The two fractions from each cell mixture were digested using the following protocol. The proteins were denatured and reduced with 6 M guanidine and 10 mM dithiothreitol (DTT) (Sigma-Aldrich, MO) at 60°C for 1 h. The denatured proteome fractions were diluted 6-fold with 50 mM Tris/10 mM  $\text{CaCl}_2$  (pH 7.6), and sequencing grade trypsin was added at the ratio of 1:100 (wt:wt). The first digestion was run overnight at 37°C and, after adding additional trypsin, the second digestion was run for 5 h at 37°C. The samples were then reduced with 20 mM DTT for 1 h at 60°C and were desalted using Sep-Pak Plus C-18 solid-phase extraction (Waters Co, MA).

### QUANTITATIVE PROTEOMICS MEASUREMENT

The protein digests were examined with LC-MS/MS using twelve-step split-phase MudPIT (MacCoss et al., 2002; McDonald et al., 2002) in duplicate. The samples were loaded via a pressure cell (New Objective, MA) onto a 250- $\mu\text{m}$ -I.D. fused silica front column fritted into an M-520 filter union (Upchurch Scientific, WA). The column packing consisted of 2 cm strong cation exchange resin Luna® and 2 cm C18 reverse-phase resin Aqua (Phenomenex, CA). A 100- $\mu\text{m}$ -I.D. PicoFrit column (New Objective, MA) was packed with 15 cm C18 reverse-phase resin. The front column was connected with the PicoFrit column and then placed in-line with a Dionex Ultimate quaternary HPLC. Two-dimensional LC separation was performed with twelve salt pulses, each of which was followed by a 2-h reverse-phase gradient. MS/MS analysis was performed on an LTQ linear ion trap instrument (ThermoFinnigan, CA) with dynamic exclusion enabled. Each full scan (400–1700  $m/z$ ) was followed by three data-dependent MS/MS scans at 35% normalized collision energy. The full scans were averaged from five microscans and the MS/MS scans were averaged from two microscans.

### QUANTITATIVE PROTEOMICS DATA ANALYSIS

All MS/MS scans were searched in two iterations against the FASTA database containing all annotated *Z. mobilis* proteins using the SEQUEST program (Eng et al., 1994). In the first iteration, the molecular masses of amino acids containing  $^{14}\text{N}$  were used, and, in the second iteration, the masses of amino acids containing  $^{15}\text{N}$  were used. The peptide identifications from the two iterations were merged. The DTASelect program (Tabb et al., 2002) was used to filter the peptide identifications and to assemble the peptides into proteins using the following parameters: retaining the duplicate MS/MS spectra for each peptide sequence (DTASelect option: -t 0), fully tryptic peptides only, with a delCN of at least 0.08 and cross-correlation scores (Xcorr) of at least 1.8 (for parent ion charge state,  $z = +1$ ), 2.5 ( $z = +2$ ), or 3.5 ( $z = +3$ ). Selected ion chromatogram extraction, peptide abundance ratio estimation and protein abundance ratio estimation were completed with the ProRata program as described previously (Pan et al., 2006, 2008).

### MICROARRAY ANALYSIS AND qRT-PCR VALIDATION

Microarray analysis was conducted essentially as described previously (Yang et al., 2010b). Briefly, total cellular RNA was



extracted using the TRIzol reagent (Invitrogen, CA) followed by RNase-free DNase I (Ambion, TX) digestion. RNA quality and quantity were tested with a NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, DE) and Agilent Bioanalyzer (Agilent, CA) before ds-cDNA synthesis using Invitrogen ds-cDNA synthesis kit (Invitrogen, CA). The ds-cDNA was sent to NimbleGen for labeling, hybridization, and scanning following company's protocols. Quality assessments, normalization, and statistical analyses were conducted using JMP Genomics 4.0 software (SAS Institute, Cary, NC) as described earlier (Yang et al., 2010b). An analysis of variance (ANOVA) determined differential expression levels between strains and time points using the FDR testing method ( $p < 0.05$ ). The interaction among differentially-regulated genes was investigated using the String 8.2 database (Jensen et al., 2009), available at <http://string.embl.de/>. The transcriptomic and proteomic data are also mapped to predicted metabolic pathway using PathwayTools Omics Viewer (Karp et al., 2002, 2011) at <http://biocyc.org/expression.html>.

Microarray data were validated using real-time qPCR as described previously (Yang et al., 2009b, 2010b), except that the Bio-Rad MyiQ2 Two-Color Real-Time PCR Detection System (Bio-Rad Lab, CA) and Roche FastStart SYBR Green Master (Roche Applied Science, IN) were used for this study. Nine genes representing different functional categories and a range of gene expression values based on microarray hybridizations were analyzed using qPCR from cDNA derived from different time point samples. Primer pairs were designed as described previously (Yang et al., 2009b), and the oligonucleotide sequences of the nine genes selected for qPCR analysis are listed in Additional File 6. The qRT-PCR ratios were plotted against the microarray ratios and a regression analysis was conducted to generate R-squared correlation coefficient value.

## AUTHOR CONTRIBUTIONS

Steven D. Brown, Shihui Yang, Gregory B. Hurst, and Chongle Pan designed the experiment. Shihui Yang carried out the fermentation, RNA extraction, and sample preparation for HPLC, microarray, and proteomics. Lezlee Dice performed the qRT-PCR. Chongle Pan performed the proteomic runs and generated proteomic raw data. Shihui Yang, Steven D. Brown, Chongle Pan, and Gregory B. Hurst analyzed the data. Shihui Yang and Steven D. Brown wrote the manuscript, and Brian H. Davison, Gregory B. Hurst, and Chongle Pan provided inputs on manuscript revision.

## ACKNOWLEDGMENTS

We thank Miguel Rodriguez Jr. for assistance with HPLC analyses. This work was sponsored by the Laboratory Directed Research and Development Program of Oak Ridge National Laboratory (ORNL) and concluded under the BioEnergy Science Center which is a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science. This manuscript has been authored by UT-Battelle, LLC, under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United

States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fmicb.2014.00246/abstract>

## REFERENCES

- Agrawal, M., and Chen, R. R. (2011). Discovery and characterization of a xylose reductase from *Zymomonas mobilis* ZM4. *Biotechnol. Lett.* 33, 2127–2133. doi: 10.1007/s10529-011-0677-6
- Almeida, J. R. M., Modig, T., Petersson, A., Hahn-Hagerdal, B., Liden, G., and Gorwa-Grauslund, M. F. (2007). Increased tolerance and conversion of inhibitors in lignocellulosic hydrolysates by *Saccharomyces cerevisiae*. *J. Chem. Tech. Biotech.* 82, 340–349. doi: 10.1002/jctb.1676
- Alper, H., and Stephanopoulos, G. (2009). Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential? *Nat. Rev. Microbiol.* 7, 715–723. doi: 10.1038/nrmicro2186
- Deanda, K., Zhang, M., Eddy, C., and Picataggio, S. (1996). Development of an arabinose-fermenting *Zymomonas mobilis* strain by metabolic pathway engineering. *Appl. Environ. Microbiol.* 62, 4465–4470.
- Desinotiis, A., Kouvelis, V. N., Davenport, K., Bruce, D., Detter, C., Tapia, R., et al. (2012). Complete genome sequence of the ethanol-producing *Zymomonas mobilis* subsp. *mobilis* Centrotype ATCC 29191. *J. Bacteriol.* 194, 5966–5967. doi: 10.1128/JB.01398-12
- Dien, B. S., Cotta, M. A., and Jeffries, T. W. (2003). Bacteria engineered for fuel ethanol production: current status. *Appl. Microbiol. Biotechnol.* 63, 258–266. doi: 10.1007/s00253-003-1444-y
- Eng, J. K., McCormack, A. L., and Yates, J. R. (1994). An approach to correlate tandem mass-spectral data of peptides with amino-acid-sequences in a protein database. *J. Am. Soc. Mass Spectr.* 5, 976–989. doi: 10.1016/1044-0305(94)80016-2
- Flesch, G., and Rohmer, M. (1989). Prokaryotic triterpenoids. A novel hopanoid from the ethanol-producing bacterium *Zymomonas mobilis*. *Biochem. J.* 262, 673–675.
- Franden, M. A., Pienkos, P. T., and Zhang, M. (2009). Development of a high-throughput method to evaluate the impact of inhibitory compounds from lignocellulosic hydrolysates on the growth of *Zymomonas mobilis*. *J. Biotechnol.* 144, 259–267. doi: 10.1016/j.jbiotec.2009.08.006
- Gilmore, J. M., and Washburn, M. P. (2010). Advances in shotgun proteomics and the analysis of membrane proteomes. *J. Proteomics* 73, 2078–2091. doi: 10.1016/j.jprot.2010.08.005
- Goodman, A. E., Rogers, P. L., and Skotnicki, M. L. (1982). Minimal medium for isolation of auxotrophic *Zymomonas* mutants. *Appl. Environ. Microbiol.* 44, 496–498.
- Gunasekaran, P., and Raj, K. C. (1999). Ethanol fermentation technology: *Zymomonas mobilis*. *Curr. Sci.* 7, 56–68.
- Hahn-Hagerdal, B., Galbe, M., Gorwa-Grauslund, M. F., Liden, G., and Zacchi, G. (2006). Bio-ethanol—the fuel of tomorrow from the residues of today. *Trends Biotechnol.* 24, 549–556. doi: 10.1016/j.tibtech.2006.10.004
- Hayashi, T., Furuta, Y., and Furukawa, K. (2011). Respiration-deficient mutants of *Zymomonas mobilis* show improved growth and ethanol fermentation under aerobic and high temperature conditions. *J. Biosci. Bioeng.* 111, 414–419. doi: 10.1016/j.jbiosc.2010.12.009
- He, M. X. P., Wu, B. D., Shui, Z. X. D., Hu, Q. C. P., Wang, W. G. D., Tan, F. R. D., et al. (2012b). Transcriptome profiling of *Zymomonas mobilis* under ethanol stress. *Biotechnol. Biofuels* 5, 75. doi: 10.1186/1754-6834-5-75
- He, M. X., Wu, B., Shui, Z. X., Hu, Q. C., Wang, W. G., Tan, F. R., et al. (2012a). Transcriptome profiling of *Zymomonas mobilis* under furfural stress. *Appl. Microbiol. Biotechnol.* 95, 189–199. doi: 10.1007/s00253-012-4155-4
- Heer, D., and Sauer, U. (2008). Identification of furfural as a key toxin in lignocellulosic hydrolysates and evolution of a tolerant yeast strain. *Microb. Biotechnol.* 1, 497–506. doi: 10.1111/j.1751-7915.2008.00050.x

- Hermans, M. A., Neuss, B., and Sahm, H. (1991). Content and composition of hopanoids in *Zymomonas mobilis* under various growth conditions. *J. Bacteriol.* 173, 5592–5595.
- Himmel, M. E., Ding, S. Y., Johnson, D. K., Adney, W. S., Nimlos, M. R., Brady, J. W., et al. (2007). Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* 315, 804–807. doi: 10.1126/science.1137016
- Horbach, S., Neuss, B., and Sahm, H. (1991). Effect of azasqualene on hopanoid biosynthesis and ethanol tolerance of *Zymomonas mobilis*. *FEMS Microbiol. Lett.* 79, 347–350. doi: 10.1111/j.1574-6968.1991.tb04553.x
- Inui, M., Kawaguchi, H., Murakami, S., Vertes, A. A., and Yukawa, H. (2004). Metabolic engineering of *Corynebacterium glutamicum* for fuel ethanol production under oxygen-deprivation conditions. *J. Mol. Microbiol. Biotechnol.* 8, 243–254. doi: 10.1159/000086705
- Jensen, L. J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., et al. (2009). STRING 8—a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Res.* 37, D412–D416. doi: 10.1093/nar/gkn760
- Jeon, Y. J., Xun, Z., Su, P., and Rogers, P. L. (2012). Genome-wide transcriptomic analysis of a flocculent strain of *Zymomonas mobilis*. *Appl. Microbiol. Biotechnol.* 93, 2513–2518. doi: 10.1007/s00253-012-3948-9
- Joachimstahl, E., Haggett, K. D., Jang, J. H., and Rogers, P. L. (1998). A mutant of *Zymomonas mobilis* ZM4 capable of ethanol production from glucose in the presence of high acetate concentrations. *Biotechnol. Lett.* 20, 137–142. doi: 10.1023/A:1005320306410
- Karp, P. D., Paley, S. M., Krummenacker, M., Latendresse, M., Dale, J. M., Lee, T. J., et al. (2011). Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology. *Brief. Bioinform.* 11, 40–79. doi: 10.1093/bib/bbp043
- Karp, P. D., Paley, S., and Romero, P. (2002). The Pathway Tools software. *Bioinformatics (Oxford, England)* 18(Suppl. 1), S225–S232. doi: 10.1093/bioinformatics/18.suppl\_1.S225
- Kerr, A. L., Jeon, Y. J., Svenson, C. J., Rogers, P. L., and Neilan, B. A. (2010). DNA restriction-modification systems in the ethanologen, *Zymomonas mobilis* ZM4. *Appl. Microbiol. Biotechnol.* 89, 761–769. doi: 10.1007/s00253-010-2936-1
- Kouvelis, V. N., Davenport, K. W., Brettin, T. S., Bruce, D., Detter, C., Han, C. S., et al. (2011). Genome sequence of the ethanol-producing *Zymomonas mobilis* subsp. pomaceae lectotype strain ATCC 29192. *J. Bacteriol.* 193, 5049–5050. doi: 10.1128/JB.05273-11
- Kouvelis, V. N., Saunders, E., Brettin, T. S., Bruce, D., Detter, C., Han, C., et al. (2009). Complete genome sequence of the ethanol producer *Zymomonas mobilis* NCIMB 11163. *J. Bacteriol.* 191, 7140–7141. doi: 10.1128/JB.01084-09
- Lee, K. Y., Park, J. M., Kim, T. Y., Yun, H., and Lee, S. Y. (2010). The genome-scale metabolic network analysis of *Zymomonas mobilis* ZM4 explains physiological features and suggests ethanol and succinic acid production strategies. *Microb. Cell Fact.* 9, 94. doi: 10.1186/1475-2859-9-94
- Linger, J. G., Adney, W. S., and Darzins, A. (2010). Heterologous expression and extracellular secretion of cellulolytic enzymes by *Zymomonas mobilis*. *Appl. Environ. Microbiol.* 76, 6360–6369. doi: 10.1128/AEM.00230-10
- Liu, Z. L., Slininger, P. J., and Gorsich, S. W. (2005). Enhanced biotransformation of furfural and hydroxymethylfurfural by newly developed ethanologenic yeast strains. *Appl. Biochem. Biotechnol.* 121–124, 451–460. doi: 10.1385/ABAB:121:1-3:0451
- MacCoss, M. J., McDonald, W. H., Saraf, A., Sadygov, R., Clark, J. M., Tasto, J. J., et al. (2002). Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc. Natl. Acad. Sci. U.S.A.* 99, 7900–7905. doi: 10.1073/pnas.122231399
- McDonald, W. H., Ohi, R., Miyamoto, D. T., Mitchison, T. J., and Yates, J. R. (2002). Comparison of three directly coupled HPLC MS/MS strategies for identification of proteins from complex mixtures: single-dimension LC-MS/MS, 2-phase MudPIT, and 3-phase MudPIT. *Int. J. Mass Spectrom.* 219, 245–251. doi: 10.1016/S1387-3806(02)00563-8
- Moreau, R. A., Powell, M. J., Fett, W. F., and Whitaker, B. D. (1997). The effect of ethanol and oxygen on the growth of *Zymomonas mobilis* and the levels of hopanoids and other membrane lipids. *Curr. Microbiol.* 35, 124–128. doi: 10.1007/s002849900224
- Pan, C., Kora, G., McDonald, W. H., Tabb, D. L., VerBerkmoes, N. C., Hurst, G. B., et al. (2006). ProRata: a quantitative proteomics program for accurate protein abundance ratio estimation with confidence interval evaluation. *Anal. Chem.* 78, 7121–7131. doi: 10.1021/ac060654b
- Pan, C., Oda, Y., Lankford, P. K., Zhang, B., Samatova, N. F., Pelletier, D. A., et al. (2008). Characterization of anaerobic catabolism of p-coumarate in *Rhodospseudomonas palustris* by integrating transcriptomics and quantitative proteomics. *Mol. Cell. Proteomics* 7, 938–948. doi: 10.1074/mcp.M700147-MCP200
- Panesar, P. S., Marwaha, S. S., and Kennedy, J. F. (2006). *Zymomonas mobilis*: an alternative ethanol producer. *J. Chem. Technol. Biotechnol.* 81, 623–635. doi: 10.1002/jctb.1448
- Pappas, K. M., Kouvelis, V. N., Saunders, E., Brettin, T. S., Bruce, D., Detter, C., et al. (2011). Genome sequence of the ethanol-producing *Zymomonas mobilis* subsp. mobilis lectotype strain ATCC 10988. *J. Bacteriol.* 193, 5051–5052. doi: 10.1128/JB.05395-11
- Pienkos, P. Z. (2009). Min: role of pretreatment and conditioning processes on toxicity of lignocellulosic biomass hydrolysates. *Cellulose* 16, 743–762. doi: 10.1007/s10570-009-9309-x
- Poetsch, A., and Wolters, D. (2008). Bacterial membrane proteomics. *Proteomics* 8, 4100–4122. doi: 10.1002/pmic.200800273
- Rogers, P. L., Jeon, Y. J., Lee, K. J., and Lawford, H. G. (2007). *Zymomonas mobilis* for fuel ethanol and higher value products. *Biofuels* 108, 263–288. doi: 10.1007/10\_2007\_060
- Romero, S., Merino, E., Bolivar, E., Gosset, G., and Martinez, A. (2007). Engineering of *Bacillus subtilis* for ethanol production: lactate dehydrogenase plays a key role in fermentative metabolism. *Appl. Environ. Microbiol.* 73, 5190–5198. doi: 10.1128/AEM.00625-07
- Rutkis, R., Kalnenieks, U., Stalidzans, E., and Fell, D. A. (2013). Kinetic modelling of the *Zymomonas mobilis* Entner-Doudoroff pathway: insights into control and functionality. *Microbiology* 159(Pt 12), 2674–2689. doi: 10.1099/mic.0.071340-0
- Seo, J. S., Chong, H. Y., Park, H. S., Yoon, K. O., Jung, C., Kim, J. J., et al. (2005). The genome sequence of the ethanologenic bacterium *Zymomonas mobilis* ZM4. *Nat. Biotechnol.* 23, 63–68. doi: 10.1038/nbt1045
- Shigeri, Y., Nishino, T., Yumoto, N., and Tokushige, M. (1991). Hopanoid biosynthesis of *Zymomonas mobilis*. *Agric. Biol. Chem.* 55, 589–591. doi: 10.1271/bbb1961.55.589
- Skerker, J. M., Leon, D., Price, M. N., Mar, J. S., Tarjan, D. R., Wetmore, K. M., et al. (2013). Dissecting a complex chemical stress: chemogenomic profiling of plant hydrolysates. *Mol. Syst. Biol.* 9, 674. doi: 10.1038/msb.2013.30
- Tabb, D. L., McDonald, W. H., and Yates, J. R. (2002). DTASelect and contrast: tools for assembling and comparing protein identifications from shotgun proteomics. *J. Proteome Res.* 1, 21–26. doi: 10.1021/pr015504q
- Weinitschke, S., Denger, K., Cook, A. M., and Smits, T. H. (2007). The DUF81 protein TauE in *Cupriavidus necator* H16, a sulfite exporter in the metabolism of C2 sulfonates. *Microbiology* 153(Pt 9), 3055–3060. doi: 10.1099/mic.0.2007/009845-0
- Welander, P. V., Hunter, R. C., Zhang, L. C., Sessions, A. L., Summons, R. E., and Newman, D. K. (2009). Hopanoids play a role in membrane integrity and pH homeostasis in *Rhodospseudomonas palustris* TIE-1. *J. Bacteriol.* 191, 6145–6156. doi: 10.1128/JB.00460-09
- Widiastuti, H., Kim, J. Y., Selvarasu, S., Karimi, I. A., Kim, H., Seo, J. S., et al. (2011). Genome-scale modeling and *in silico* analysis of ethanologenic bacterium *Zymomonas mobilis*. *Biotechnol. Bioeng.* 108, 655–665. doi: 10.1002/bit.22965
- Wilson, C. M., Yang, S., Rodriguez, M. Jr., Ma, Q., Johnson, C. M., Dice, L., et al. (2013). *Clostridium thermocellum* transcriptomic profiles after exposure to furfural or heat stress. *Biotechnol. Biofuels* 6, 131. doi: 10.1186/1754-6834-6-131
- Yang, S., Giannone, R. J., Dice, L., Yang, Z. K., Engle, N. L., Tschaplinski, T. J., et al. (2012). *Clostridium thermocellum* ATCC27405 transcriptomic, metabolomic and proteomic profiles after ethanol stress. *BMC Genomics* 13:336. doi: 10.1186/1471-2164-13-336
- Yang, S., Land, M. L., Klingeman, D. M., Pelletier, D. A., Lu, T.-Y. S., Martin, S. L., et al. (2010b). Paradigm for industrial strain improvement identifies sodium acetate tolerance loci in *Zymomonas mobilis* and *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.* 107, 10395–10400. doi: 10.1073/pnas.0914506107
- Yang, S., Pan, C., Tschaplinski, T. J., Hurst, G. B., Engle, N. L., Zhou, W., et al. (2013). Systems biology analysis of *Zymomonas mobilis* ZM4 ethanol stress responses. *PLoS ONE* 8:e68886. doi: 10.1371/journal.pone.0068886

- Yang, S., Pappas, K. M., Hauser, L. J., Land, M. L., Chen, G.-L., Hurst, G. B., et al. (2009a). Improved genome annotation for *Zymomonas mobilis*. *Nat. Biotechnol.* 27, 893–894. doi: 10.1038/nbt1009-893
- Yang, S., Pelletier, D. A., Lu, T. Y., and Brown, S. D. (2010a). The *Zymomonas mobilis* regulator hfq contributes to tolerance against multiple lignocellulosic pretreatment inhibitors. *BMC Microbiol.* 10:135. doi: 10.1186/1471-2180-10-135
- Yang, S., Tschaplinski, T. J., Engle, N. L., Carroll, S. L., Martin S. L., Davison, B. H. Jr., et al. (2009b). Transcriptomic and metabolomic profiling of *Zymomonas mobilis* during aerobic and anaerobic fermentations. *BMC Genomics* 10:34. doi: 10.1186/1471-2164-10-34
- Zhang, M., Eddy, C., Deanda, K., Finkestein, M., and Picataggio, S. (1995). Metabolic engineering of a pentose metabolism pathway in ethanologenic *Zymomonas mobilis*. *Science* 267, 240–243. doi: 10.1126/science.267.5195.240
- Zhang, Y., Ma, R., Zhao, Z., Zhou, Z., Lu, W., Zhang, W., et al. (2010). irrE, an exogenous gene from *Deinococcus radiodurans*, improves the growth of and ethanol production by a *Zymomonas mobilis* strain under ethanol and acid stress. *J. Microbiol. Biotechnol.* 20, 1156–1162. doi: 10.4014/jmb.0912.12036

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 28 February 2014; accepted: 06 May 2014; published online: 22 May 2014.

Citation: Yang S, Pan C, Hurst GB, Dice L, Davison BH and Brown SD (2014) Elucidation of *Zymomonas mobilis* physiology and stress responses by quantitative proteomics and transcriptomics. *Front. Microbiol.* 5:246. doi: 10.3389/fmicb.2014.00246

This article was submitted to Microbial Physiology and Metabolism, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Yang, Pan, Hurst, Dice, Davison and Brown. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Metabolic engineering of yeasts by heterologous enzyme production for degradation of cellulose and hemicellulose from biomass: a perspective

William Kricka, James Fitzpatrick and Ursula Bond\*

School of Genetics and Microbiology, Department of Microbiology, Trinity College Dublin, Dublin, Ireland

## Edited by:

Katherine M. Pappas, University of Athens, Greece

## Reviewed by:

Biswarup Sen, Feng Chia University, Taiwan

Ed Louis, University of Leicester, UK

## \*Correspondence:

Ursula Bond, School of Genetics and Microbiology, Department of Microbiology, Trinity College Dublin, College Green, Dublin 2, Ireland  
e-mail: ubond@tcd.ie

This review focuses on current approaches to metabolic engineering of ethanologenic yeast species for the production of bioethanol from complex lignocellulose biomass sources. The experimental strategies for the degradation of the cellulose and xylose-components of lignocellulose are reviewed. Limitations to the current approaches are discussed and novel solutions proposed.

**Keywords: recombinant yeasts, cellulases, xylose-utilizing enzymes**

## INTRODUCTION

With dwindling fossil fuel resources and the necessity to combat climate change by reducing greenhouse gas emissions, there is a growing need to identify alternative environmentally sustainable energy sources. One potential green energy source is that derived from biomass, which can be converted to useable energy such as biofuels. The two most common examples of biofuels are bio-ethanol and bio-diesel. Biofuels are considered the cleanest liquid fuel alternative to fossil fuel and it is estimated that replacing fossil fuels with biofuels could decrease CO<sub>2</sub> emissions by 60–90% (Hasunuma and Kondo, 2012). Currently, over 100 billion liters of biofuels are produced annually, yet this accounts for only a fraction (2.7%) of total energy used in transportation. Bioethanol production reached 86 billion liters in 2010, with the United States and Brazil as the world's top producers, accounting together for 90% of global production (Buijs et al., 2013).

## GENERATIONS OF BIOETHANOL

Bioethanol production has been classified into different generations based on biotechnology developments and also in terms of the feedstock used. First generation bioethanol is produced from either corn or sugarcane. The released sugars, glucose and sucrose, respectively, are readily fermentable into ethanol by microorganisms such as the yeast *Saccharomyces cerevisiae* (Buijs et al., 2013). A disadvantage to using these substrates as energy sources is the fact that they are food crops and hence controversy regarding the ethics of exploiting “food for fuel” surrounds the established bioethanol industry. Currently over 90% of the world's bioethanol is produced from food crops, however governmental directives are setting exacting targets on the limitation of generating renewable energy from food-based crops. The US Renewable Fuel Standards mandate (RFS; US Energy Policy Act (EPA) 2005) requires that 44% (16 billion gallons) of renewable-fuel to be blended into gasoline by 2022 be derived from non-food cellulosic biomass. Current negotiations at the European Council aim to limit

the amount of biofuels from food-based crops that can be counted toward the 10% target for renewable energy in the transport sector by 2020 (EU Renewable Energy Directive (RED; 2009/28/EC).

Second generation biofuels are derived from more complex non-food based biomass, which can be grouped into roughly four categories, namely, wood residues, municipal solid waste, agricultural waste, and dedicated energy crops. The most abundant renewable form of biomass is lignocellulose comprising between 50 and 90% of all plant matter. The global production of plant biomass amounts to approximately  $2 \times 10^{11}$  Mt per annum of which  $8\text{--}20 \times 10^9$  Mt is potentially accessible for processing. Lignocellulose is composed of three major components, cellulose, hemicellulose, and lignin. The relative amount of each component varies in different plant types, with an average composition of cellulose 30–50%, hemicellulose 20–30%, and lignin 15–25%.

Cellulose is the most abundant polysaccharide on earth (Chandel and Singh, 2011). In its simplest chemical form, cellulose is a  $\beta$ -glucan linear polymer of D-glucose linked by  $\beta$ -1,4-glycosidic bonds. The basic repeating subunit is cellobiose, consisting of two glucose molecules. At the macroscopic level, cellulose exists as two distinct forms, tightly packed crystalline and non-organized amorphous regions. The amorphous regions of cellulose are believed to result from surface shaving caused by natural erosion. Both the amorphous and crystalline forms are made up of cellulosic fibers comprising microfibrils, which are composed of approximately 30  $\beta$ -glucan chains. The highly accessible amorphous regions account for approximately 1% of the structure of cellulose (Ruel et al., 2012).

Hemicellulose is the second most abundant polysaccharide within lignocellulose (Peng et al., 2012). It is a highly branched heteropolymer composed of pentoses and hexoses such as xylose, arabinose, mannose, glucose, and galactose as well as sugar acids. The composition of hemicellulose is variable in nature and depends upon the plant source, however the most abundant component of hemicellulose is generally xylose.

The third component of lignocellulose is lignin, a polymer of three aromatic alcohols, coniferyl, p-coumaryl, and sinapyl. Lignin links both hemicelluloses and cellulose together forming a physical barrier in the plant cell wall. Lignin is recalcitrant to degradation and is resistant to most microbial attacks and oxidative stress (Dashtban et al., 2009). Unlike cellulose and hemicellulose, hydrolysis of lignin does not generate fermentable sugars. Furthermore, phenolic compounds produced during lignin hydrolysis actively inhibit fermentation.

Biomass is not readily fermentable and expensive pre-treatments, both physical (milling and steam explosion) and chemical (acid and alkaline hydrolysis), are required to increase access to the sugars within the biomass. The sugars released from pre-treatment must be further hydrolyzed by enzymatic actions to yield fermentable glucose.

Several different approaches have been developed to generate bioethanol from biomass. One such process, called separate hydrolysis and fermentation (SHF), as its name suggests, requires a two-step process in which biomass hydrolysis and fermentation of released sugars are performed in separate reaction vessels. The main advantage of this method is that the two processes can be performed under their own individual optimal conditions. Hydrolysis of cellulose by enzymes, referred to as cellulases, is most efficient at temperatures between 50 and 60°C, whereas fermentation reactions are generally performed at 25–35°C, the optimum temperature range for yeast growth and metabolism. Cellulase enzymes are not found naturally in fermenting microorganisms and must be supplied *ex vivo*. A disadvantage of SHF is that glucose and cellobiose released by the action of cellulases inhibit the subsequent activity of enzymes.

To streamline production, an alternative process of simultaneous saccharification and fermentation (SSF) was devised. In this process the saccharification and fermentation are performed together in a single vessel and glucose released by the action of cellulases is immediately metabolized, thereby reducing enzyme inhibition. However, a disadvantage with SSF is the need to use a compromise temperature that is neither optimal for hydrolysis or fermentation.

The requirement for maximum efficiency to ensure economic viability of bioethanol production led to the development of consolidated bioprocessing (CBP) or third generation biofuels. In CBP, all steps are performed in the same reaction vessel by a single organism capable of both producing biomass hydrolyzing enzymes and fermenting the resultant sugars to ethanol. Thus, CBP avoids the need to supply exogenously produced cellulase enzymes.

## THE CELLULASES

Cellulases belong to the O-glycoside hydrolases group of enzymes. There are three major classes of cellulases, endoglucanases (EG), cellobiohydrolases (CBH) or exoglucanases, and  $\beta$ -glucosidases (BGL). The concerted action of all three cellulases is required to efficiently convert cellulose into glucose. The generally accepted view is that cellulases act sequentially and synergistically. Endoglucanases randomly cleave the cellulose backbone at amorphous sites along the cellulose fiber. This leads to a rapid decrease in the degree of polymerization of the cellulose fiber and exposes

new chain ends. Cellobiohydrolases act processively on reducing and non-reducing chain ends to release mainly cellobiose.  $\beta$ -Glucosidases hydrolyze the  $\beta$ -1,4 glycosidic bond of cellobiose and oligosaccharides to release glucose units.

Several novel classes of enzymes such as the copper-requiring polysaccharide monoxygenases, for example GH61, contribute to cellulose degradation by acting in synergy with the exo- and endoglucanases (Leggio et al., 2012; Žifčáková and Baldrian, 2012). Elastin-like proteins such as swollenin and other cellulase-enhancing proteins contribute to the hydrolysis of cellulose by increasing access of the cellulase enzymes to the cellulose chains ends (Kubicek, 2013; Nakatani et al., 2013). Cellulase enzymes are naturally produced by a variety of filamentous fungi of different genera such as *Trichoderma*, *Aspergillus*, *Talaromyces*, and several anaerobic bacteria. Two different cellulase systems, referred to as complexed and non-complexed have been described. Complexed cellulase systems are multi-enzyme complexes, referred to as the cellulosome, that remain tethered to the cell wall of cellulolytic bacteria (Lynd et al., 2002; Fontes and Gilbert, 2010). These cell wall tethered complexes are primarily encountered in anaerobic bacteria such as species of the *Clostridium* and *Ruminococcus* genera. Filamentous fungi and certain actinomycete bacteria such as *Cellulomonas* species use non-complexed cellulase systems. Non-complexed cellulases are secreted into the extracellular environment and are not attached to the cell surface. The most extensively studied cellulolytic organism is the filamentous fungus *Trichoderma reesei*. *T. reesei* synthesizes an array of cellulases, including at least five EGs, two CBHs, and two BGLs (Foreman et al., 2003; Martinez et al., 2008; Kubicek, 2013).

## STRATEGIES FOR HYDROLYSIS OF CELLULOSE AND FERMENTATION OF RELEASED SUGARS TO ETHANOL

While clearly capable of degrading cellulose, *T. reesei* is not an efficient fermenter of sugars to alcohol, a prerequisite for a CBP microorganism. *T. reesei* can generate up to 4.8 g/L ethanol from growth in glucose while *Aspergillus oryzae* can produce 24.4 g/L ethanol (Skory et al., 1997). Natural ethanologens such as *Zymomonas mobilis* or *Saccharomyces cerevisiae* can produce 130–200 g/L under the right environmental conditions. One major disadvantage with using mycelial fungi for ethanol production is the slow bioconversion rate compared to that observed in yeasts. Furthermore, filamentous fungi show intolerance to high concentrations of ethanol. An analysis of conversion of glucose to ethanol by 19 strains of *Aspergillus* revealed efficiencies of 21–98% after 6 days compared to 100% conversion by *S. cerevisiae* within 48 h (Skory et al., 1997).

To date, no naturally occurring microorganism capable of CBP at the desired efficiency for industrial bioethanol production has been identified. Therefore, researchers have pursued two strategies (native and recombinant) to generate the ideal microorganism for CBP. The native strategy focuses on modifying natural cellulolytic organisms to improve ethanol yields. Several approaches have been pursued, including directed evolution using error-prone Polymerase Chain Reaction-based mutagenesis of cellulase genes, adaptive evolution using natural selection to specific environmental conditions or rational protein design to improve the enzymatic activity of cellulases or to expand the



physiological conditions at which the enzymes are active (Elkins et al., 2010; Voutilainen et al., 2010; Liang et al., 2011; Anbar et al., 2012; Gefen et al., 2012; Wang et al., 2012). Challenges still remain for these types of approaches in order to upscale to industrial fermentation conditions.

The recombinant strategy involves genetic engineering of native cellulase-producing species to improve ethanol yields or heterologous expression of cellulase genes in natural ethanologens. Ethanol yields in *A. niger* can be increased by expression of a pyruvate decarboxylase gene from *Z. mobilis* (Skory et al., 1997). Likewise, heterologous gene expression of pyruvate decarboxylase and alcohol dehydrogenase genes from *Z. mobilis* in the cellulolytic bacterium *C. cellulolyticum* was found to increase ethanol production by 53% (Guedon et al., 2002).

The vast majority of research into producing a CBP candidate has however followed the recombinant strategy of heterologous expression of cellulase genes in natural ethanologens. In order to achieve complete hydrolysis of cellulose, at least one copy of each of the three classes of cellulase genes must be expressed in the host cell. The most commonly used host for heterologous expression of cellulase genes for CBP is the baker's yeast, *S. cerevisiae* (Fujita et al., 2004; Den Haan et al., 2007a; Tsai et al., 2010; Wen et al., 2010; Yamada et al., 2011; Fan et al., 2012; Nakatani et al., 2013), although the three classes of cellulase genes have also been expressed in other *Saccharomyces* species such as *S. pastorianus* (Fitzpatrick et al., 2014) as well as in bacterial species such as *Escherichia coli* (Ryu and Karim, 2011) (Table 1). Since the native promoters of cellulase genes are repressed by glucose, strategies of using inducible or constitutive promoters of the host have been pursued. Inducible promoters such as *S. cerevisiae* GAL1 or CUP1 promoters are extremely efficient but require the addition of an inducer, galactose or copper, respectively. The requirement for such inducers can be expensive and often incompatible with fermentation conditions for ethanol production. Moreover, the GAL promoters are repressed in the presence of glucose and therefore not suited for industrial CBP. Partow et al. (2010) tested the performance of several constitutive and inducible promoters for heterologous gene expression in *S. cerevisiae*. Their findings indicated that the constitutive promoters TEF1 and PGK1 produced the most constant expression profiles. These promoters have been used for heterologous cellulase gene expression in *S. cerevisiae* (Den Haan et al., 2007b; Yamada et al., 2011), however no more than a 2-fold difference in expression was observed in genes driven by these two promoters, with TEF1 generating the highest levels (Fitzpatrick et al., 2014).

Another strategy for increasing cellulase activity is to increase the gene copy number. Episomal plasmids have been extensively used to express cellulase genes (Fujita et al., 2004; Den Haan et al., 2007b; Tsai et al., 2010; Wen et al., 2010; Fan et al., 2012), however there is an issue with their genetic stability. Under non-selection, plasmids are lost in several generations and constant selection such as culturing in auxotrophic or antibiotic media must be applied. This set up is not suited to industrial production due to the cost of such selection reagents. The expression of *T. reesei* endoglucanase gene EGI on a high copy number 2  $\mu$  episomal vector (pRSH-series) was 50-fold greater than when expressed from an ARS/CEN vector (pGREG-series) (Fitzpatrick

et al., 2014). A preferred solution is the integration of cellulase gene cassettes directly into the chromosome of the host microbe. Classic integration methods ensure stability of the genes, however, enzyme production is limited by gene copy number (Du Plessis et al., 2010; Yamada et al., 2013). Multi-copy integration of cellulase genes offers a solution. Yamada and co-workers constructed *S. cerevisiae* strains containing multiple-copies of cellulase genes integrated into the delta ( $\delta$ ) repeat sites of transposable elements (Tn) in the host chromosome, leading to increased ethanol yields (Yamada et al., 2010).

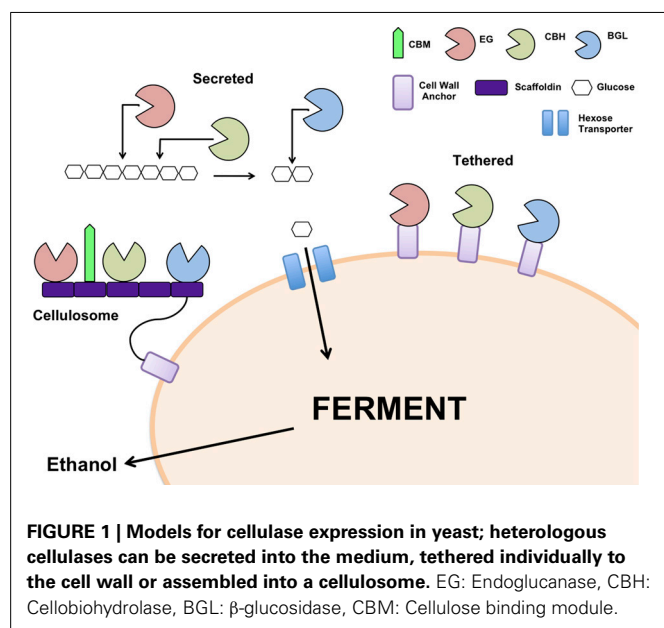
Defining the optimum ratio of the different classes of cellulases is also important to achieve maximum cellulose hydrolysis. The true optimum ratio of cellulases used by natural cellulolytic microorganisms is not fully known (Yamada et al., 2013) but it has been estimated that the total secreted protein of *T. reesei* under inducing conditions is 60% CBHI, 20% CBHII, 10% EG, and 1% BGL (Takashima et al., 1998). Hence, a 1:1:1 ratio of the three classes of enzymes may not yield an optimum hydrolysis synergy. Yamada and colleagues performed repeated rounds of integration at delta sites to generate an *S. cerevisiae* strain with 24 cellulase genes integrated into the genome in a ratio of 16:6:2 for *egl2*, *cbh2*, and *bgl1*, respectively (Tien-Yang et al., 2012). Cellulose degradation activity of this strain was lower than in a strain which had the cellulase genes *egl2*, *cbh2*, and *bgl1* in a ratio of 13:6:1, respectively. Interestingly, a strain with *egl2*, *cbh2*, and *bgl1* in a ratio of 5:9:6, respectively produced 1.3-fold less cellulose hydrolysis activity than a strain containing the genes in a 13:6:1 ratio, suggesting that the ratio of the cellulase enzymes as well as the copy number is crucial to ensure efficient cellulose hydrolysis.

The choice of cellulase system for integration into the recipient ethanologenic host is important to consider. The cellulosome from *C. thermocellum* has been reconstituted in *S. cerevisiae* (Wen et al., 2010) as have the secretory cellulases from several filamentous fungi including *A. aculeatus*, *A. oryzae*, *Saccharomycopsis fibuligera*, *Thermoascus aurantiacus*, and *T. reesei* (Table 1). Genes encoding the different classes of cellulases from a single species or from different fungal origins have been co-expressed in *S. cerevisiae* (Table 1). Secretion of the expressed cellulases has been achieved using the native secretory signals or host signals, for example, the  $\alpha$ -mating type secretory signal. In some cases, the recombinant cellulases have been tethered to the cell surface (Table 1, Figure 1) using anchor proteins such as *S. cerevisiae*  $\alpha$ -agglutinin (Yanase et al., 2010) or cell wall protein (Van Rooyen et al., 2005). Cell wall tethering leads to an effective increase in concentration of the cellulases. On the other hand, expressing cellulases so that the proteins are untethered facilitates the binding of the enzymes at multiple sites along the length of the cellulose chain. Interesting, very little difference in ethanol yields was observed in yeast strains in which the cellulases were secreted or cell wall tethered, indicating that one approach is not superior to the other (Yanase et al., 2010).

To date, haploid *S. cerevisiae* has been the host of choice for heterologous expression of cellulase genes, although a few studies have examined the expression of cellulase genes in other hosts. Yamada and colleagues constructed a diploid *S. cerevisiae* yeast strain with cellulase genes integrated at  $\delta$ -integration sites. The diploid strain displayed 6-fold higher phosphoric acid swollen

**Table 1 | Ethanol production from PASC by recombinant yeast strains expressing cellulases.**

Host strain	Cellulase enzyme	Cellulase system	PASC concentration (g/L)	Ethanol concentration (g/L)	Fermentation time (h)	References
<i>S. cerevisiae</i>	<i>T. reesei</i> EGII, CBHII, <i>A. aculeatus</i> BGLI	Tethered	10	2.9	40	Fujita et al., 2004
<i>S. cerevisiae</i>	<i>T. reesei</i> EGI, <i>S. fibuligera</i> bgl	Secretion	10	1	192	Den Haan et al., 2007b
<i>S. cerevisiae</i>	<i>T. reesei</i> EGII, CBHII, <i>A. aculeatus</i> BGLI	Secretion	10	1.6	60	Yanase et al., 2010
<i>S. cerevisiae</i>	<i>T. reesei</i> EGII, CBHII, <i>A. aculeatus</i> BGLI	Tethered	10	2.1	60	Yanase et al., 2010
<i>S. cerevisiae</i>	<i>C. thermocellum</i> endoglucanase, <i>C. cellulolyticum</i> exoglucanase, <i>T. reesei</i> CBHII, <i>T. aurantiacus</i> BGLI	Minicellulosome	10	1.9	48	Tsai et al., 2010
<i>S. cerevisiae</i>	<i>T. reesei</i> EGII, CBHII, <i>A. aculeatus</i> BGLI, <i>C. thermocellum</i> miniscaffoldin	Minicellulosome	10	1.8	72	Wen et al., 2010
<i>S. cerevisiae</i>	<i>T. reesei</i> EGII, CBHII, <i>A. aculeatus</i> BGLI	Tethered	20	7.6	72	Yamada et al., 2011
<i>E. coli</i>	<i>C. cellulolyticum</i> Cel5A, Cel9E, BGL	Tethered	10	3.6	60	Ryu and Karim, 2011
<i>S. cerevisiae</i>	<i>C. cellulolyticum</i> celCCA, celCCE, Ccel_2454	Minicellulosome	10	1.1	96	Fan et al., 2012
<i>S. cerevisiae</i>	<i>T. reesei</i> EGII, CBHII, <i>A. aculeatus</i> BGLI, <i>A. oryzae</i> AoelpI	Tethered	20	3.4	96	Nakatani et al., 2013
<i>S. pastorianus</i>	<i>T. reesei</i> EGI, CBHII, BGLI	Secretion	25	16.5	240	Fitzpatrick et al., 2014



cellulose (PASC) degradation than the parental haploid (Yamada et al., 2011). Cellulases have also been expressed in the polyploid lager yeast *S. pastorianus* (Fitzpatrick et al., 2014). The latter study compared the expression of the three classes of cellulases from *T. reesei* in strains of *S. pastorianus* and *S. cerevisiae*. Enzymatic activity for all three classes of enzymes was up to 10-fold higher in *S. pastorianus* strains compared to *S. cerevisiae* strains. Thus for CBP, it will be important to explore the use of yeast species other than *S. cerevisiae* for cellulase production. This point is particularly important in light of the temperature difference between optimal growth conditions for *Saccharomyces* species (25–35°C)

and for cellulases activity (50–60°C). Cellulase activity at 30°C is 3-fold less than at 50°C and therefore cellulose hydrolysis will always be compromised when *Saccharomyces* species are used as hosts. An organism that could bridge the gap between saccharification and fermenting temperatures is *Kluyveromyces* sp., which are more thermotolerant than *Saccharomyces* sp. (Fonseca et al., 2008). Expression of all three classes of cellulases in *K. marxianus* produced up to 43 g/L ethanol using the simple di-saccharide cellobiose as a sole carbohydrate source (Hong et al., 2007). The hydrolysis of more complex cellulose substrates such as PASC has yet to be tested with this recombinant host. One disadvantage with *K. marxianus* is that it is less ethanol tolerant than *S. cerevisiae*.

### THE CHICKEN AND EGG PROBLEM

Ultimately, a key requirement of a recombinant host species for CBP is the ability to utilize complex cellulose substrates such as lignocellulose biomass as the sole carbohydrate source. It remains unclear how the hydrolysis of cellulose by recombinant microbes can be initiated. The host must produce enzymes to degrade cellulose into glucose to ensure cell growth while at the same time, the cell must be growing in order to produce and secrete enzymes, thus the classic chicken and egg conundrum. Cellulose cannot be transported directly into the cell and must be hydrolyzed extracellularly. In order to produce the cellulases, the cells must be actively dividing to ensure transcription from inducible or constitutive promoters. It may be therefore necessary to supply a residual amount of fermentable sugars, such as glucose or sucrose to kick-start the process. We have previously shown that yeast cell growth can be maintained with as little as 0.5 g/L glucose, although the addition of sugars to CBP will ultimately add costs to the process. The approach taken by several groups is to use very high cell numbers to carry out fermentations of cellulose

substrates. The idea here is that residual cellulases synthesized in the pre-fermentation starter cultures will be released from cells upon incubation with fresh medium containing the cellulose substrate. A disadvantage to this approach is that the use of high cell densities limits the continued growth of the culture, which inevitably will quickly enter stationary phase. Tethering the cellulases to the cell surface may solve this dilemma, as cellulases produced in pre-fermentation starter cultures may remain active once the cells are switched to the biomass fermentation process.

Taking the known problems associated with CBP into account, we favor an SHF *in situ* approach in which a pre-hydrolysis of the cellulose substrate is performed prior to fermentation. Pre-hydrolysis can be carried out with recombinant cellulases secreted into the spent medium in yeast starter cultures. This spent medium is then used in a pre-fermentation step to begin the process of hydrolysis of the cellulose substrate at the optimal temperature of 50–60°C to release glucose. The medium can then be cooled, replenished with essential nutrients and cellulase-expressing yeasts, and fermentations carried out at 25–30°C (Figure 3). Interestingly, the majority of industrial fermentation facilities used for the production of potable alcohols (beer, ales, lagers, and spirits) incorporate a pre-fermentation hydrolysis step at high temperatures (50–63°C) in order to release fermentable sugars (maltose, sucrose etc.) from complex carbohydrate substrates (starches) such as wheat and barley. Therefore, a SHF *in situ* approach could easily be incorporated into current fermentation processes. SHF *in situ* differs from SHF in eliminating the requirement for the addition of costly commercially produced cellulases.

The ultimate test of any CBP host candidate, whether generated by native or recombinant strategies, is the amount of ethanol the microorganism can produce from a cellulosic substrate. The most commonly used cellulosic substrate tested by research groups is phosphoric acid swollen cellulose (PASC). Table 1 summarizes the ethanol yields from fermentation of hydrolyzed PASC reported to date. Despite the myriad of approaches undertaken, ethanol levels are in the range of 7–16.5 g/L, far below what is achievable with more conventional complex carbohydrate substrates such as grains where yields of up to 200 g/L are commonplace. For industrial bioethanol production from cellulosic biomass, far greater yields are required. The low yields most likely reflect the slow bioconversion rate of cellulose-based substrates such as PASC. The results summarized in Table 1 highlight the need for further improvements in this process.

It should be noted that it is difficult to compare ethanol yields from the various studies due to variations in experimental parameters such as starting substrate concentrations, incubation times, and the density of cells used for fermentations. Furthermore, the unit of activity for cellulase hydrolysis differs widely. For comparative purposes, going forward, it would be useful for research groups to standardize the experimental conditions used for PASC hydrolysis and to adopt a standard definition of a unit of cellulase activity, for example, grams of ethanol produced per gram of theoretical glucose present in the cellulose substrate per hour ( $\text{g ethanol g glucose}^{-1} \text{ h}^{-1}$ ). Furthermore, while current studies almost exclusively use PASC as an experimental cellulose substrate, it will be essential to expand this analysis to the use of more

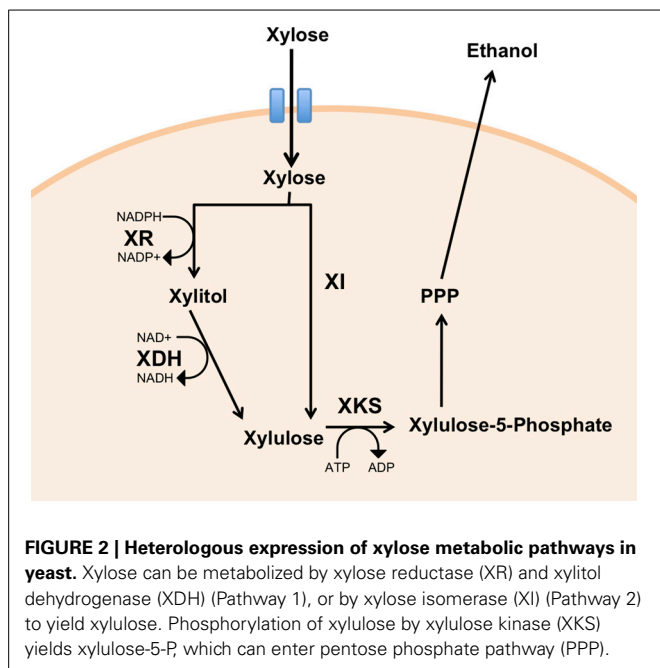
complex lignocellulose substrates such as straw, spent grains, and grasses.

## METABOLISM OF XYOSE FROM HEMICELLULOSE FOR BIOETHANOL PRODUCTION

The pentose sugar xylose is a major component of hemicellulose. Its abundance and relative ease of extraction from biomass by pre-treatments described above make it an attractive source of fermentable sugar for the production of bioethanol. In nature, xylose can be utilized by various yeast species and bacteria that are found associated directly or indirectly with lignocellulose. The mining and isolation of yeast from environments where xylose is likely to be an abundant natural sugar has developed greatly over recent years. Several xylose-utilizing yeast species such as *Spathaspora passalidarum* are indirectly associated with lignocellulose, via their symbiotic relationship with wood boring beetles such as *Odontotaenius disjunctus* (Hou, 2012) or wood roaches such as *Cryptocercus* sp. (Urbina et al., 2013). Additionally xylose-utilizing species including *Candida* sp., *Geotrichum* sp., *Sporopachydermis* sp., *Trichosporon* sp., *Pichia* sp., and *Sugiyamaella* sp. have been isolated from buffalo feces (Wanlapa et al., 2013) or from soil (Zhang et al., 2014). Although some success has been achieved using natural xylose fermenting species, their industrial relevance has yet to be demonstrated.

Two major xylose utilizing pathways have been identified. In xylose-fermenting fungi and yeasts, xylose utilization involves the action of two oxidoreductases, xylose reductase (XR) and xylitol dehydrogenase (XDH), each requiring the co-factors NADPH and  $\text{NAD}^+$ , respectively in the forward reactions (Figure 2, pathway 1). Most bacterial species use an alternative pathway requiring just a single enzyme, xylose isomerase (Figure 2, pathway 2). The product of both pathways, xylulose, is phosphorylated by xylulose kinase, which can enter the Pentose Phosphate Pathway (PPP), thus generating intermediates of the glycolytic pathway.

While a few xylose utilizing *Saccharomyces* sp. have been identified (Wenger et al., 2010; Schwartz et al., 2012), in general, yeasts belonging to the *Saccharomyces stricto sensu* group cannot metabolize xylose despite harboring genes encoding putative xylose-utilizing enzymes. To facilitate the incorporation of xylose utilization into CBP, efforts have focussed on introducing xylose metabolic pathways from other species into natural ethanologenic *Saccharomyces* sp. (Karhumaa et al., 2007; Matsushika et al., 2009a,b; Fernandes and Murray, 2010; Bera et al., 2011; Hasunuma et al., 2011; Hector et al., 2011; Usher et al., 2011; Xiong et al., 2011; Cai et al., 2012; Fujitomi et al., 2012; Kim et al., 2012; Tien-Yang et al., 2012; De Figueiredo Vilela et al., 2013; Demeke et al., 2013; Hector et al., 2013; Ismail et al., 2013; Kato et al., 2013; Kim et al., 2013). This topic has been reviewed recently (Matsushika et al., 2009a; Fernandes and Murray, 2010; Cai et al., 2012) and so will not be extensively covered here. Instead, we summarize the various experimental approaches used, the heterologous genes that have been expressed and the ethanol yields generated (Table 2). As with the data on ethanol yields from cellulose (Table 1), it is difficult to compare results on ethanol yields from xylose, with different studies representing ethanol yields as ethanol ( $\text{g/L}^{-1}$ ), grams ethanol/ $\text{g}^{-1}$  xylose or % yield (Table 2). Despite the variety of approaches, the levels of



ethanol produced are remarkably consistent ranging from 4 to 25 g/L, far below that which is achievable by native yeasts using carbohydrate sources such as glucose, maltose, or sucrose. These low yields reflect the poor uptake of xylose by *S. cerevisiae*, the species of choice for most of these studies and the slow metabolism of the sugar due to redox imbalances in the cell. Below we summarize some of the approaches that have been taken to improve these processes.

Pentose sugars can naturally be transported into *S. cerevisiae* by hijacking various hexose sugar transporters (Hxt7/Hxt5/Hxt4/Hxt2 and Gal2). The major problem encountered here is the greater affinity of these transporters for their natural substrate glucose. Therefore, in media containing both hexose and pentose sugars, such as that generated by pre-treatment of lignocellulose biomass, glucose is preferentially transported while xylose uptake is inhibited. Two distinct strategies, the overexpression of genes encoding native transporters from *S. cerevisiae* and the heterologous expression of genes encoding transporters from xylose utilizing species have been pursued. The genes encoding hexose transporters Hxt1, Hxt7, Hxt13, and Gal2 have been individually overexpressed in *S. cerevisiae*, with growth rates and xylose consumption varying between individual studies. Xylose transport was increased by the over expression of HXT1 and GAL2 but not by overexpression of HXT7 (Tanino et al., 2012). These studies revealed a trade off between transporter specificity and transporter efficiency, with transporters displaying high specificity (low  $K_m$ ) being less efficient at xylose transport (Young et al., 2012).

Based on these findings, the use of specific xylose transporters from xylose-utilizing species would seem like a more logical approach. Jaewoong et al. (2013) expressed genes encoding 6 xylose transporters from *Scheffersomyces stipitis* in *S. cerevisiae*, all of which demonstrated improved growth rates and ethanol production when grown solely on xylose, with the genes XUT7,

RGT2, and SUT4 showing the most benefit. A directed mutagenesis of xylose transporters GX1 and XUT3 from *Candida intermedia* and *S. stipitis*, respectively led to improved growth rate on xylose (Young et al., 2012). As might be expected, the expression of heterologous xylose specific transporter genes shows improved xylose uptake in mixed hexose and pentose medium (Runquist et al., 2010), however, in some cases, strains where xylose consumption was improved, overall sugar consumption (glucose and xylose) was lower than that achieved by overexpression of native hexose transporters such as HXT7 and GAL2 (Young et al., 2011).

Identifying the best xylose transporters for heterologous expression also poses problems. A comparative study on the heterologous expression of genes encoding different sugar transporters in *S. cerevisiae* demonstrated that of an initial 23 transporters tested, only five conferred the ability to grow on xylose (Young et al., 2011). Likewise, an analysis of 18 genes encoding putative xylose transporters led to the cloning of just two xylose-specific transporters, XYP29 and AN25 from *Pichia stipitis* and *Neurospora crassa*, respectively (Du et al., 2010).

In addition to problems associated with xylose transport, xylose utilization in *S. cerevisiae* is hampered by bottlenecks in the metabolism of xylose through the PPP (Fiaux et al., 2003). Several studies have been conducted to improve the metabolic fluxes through the PPP, including the overexpression of genes encoding key enzymes in the pathway such as transaldolase (TAL1), transketolase (TKL1), ribose-5-phosphate isomerase (RK11), and ribulose 5-phosphate epimerase (RPE1) (Karhumaa et al., 2005). While over expression of these genes individually produced mixed outcomes, the over expression of all 4 PPP enzymes produced a 30-fold increase in growth rate on xylose, however ethanol yields only increased fractionally (Bera et al., 2011). Although over expression of TAL1 did not greatly improve ethanol production from synthetic xylose, improved ethanol yields were achieved when detoxified hemicellulosic hydrosylate was used as a sole carbohydrate source (Hasunuma et al., 2014).

Genome-wide expression analysis identified additional host genes required for optimum xylose utilization including the PPP-associated SOL3 and GND1 as well as non PPP-associated genes such GAL1, 7, and 10 (Wahlbom et al., 2003; Bengtsson et al., 2008). Deletion of YLR042C, MNI1, and RPA49, resulted in an improvement in growth rates on xylose (Bengtsson et al., 2008), while deletion of PHO13, ALP1, ISC1, RPL20B, BUD21, NQM1, TKL2 led to increased ethanol yields from xylose (Van Vleet et al., 2008; Usher et al., 2011).

It is clear that PPP involves complex regulation involving many gene products. Thus, major alterations in the PPP may be required to optimize xylose utilization by *S. cerevisiae*. While mutational and over expression approaches have helped to identify key genes regulating the PPP, the industrial relevance of any such modifications must be considered. Given the complexity of the system, an adaptive evolutionary approach involving continued rounds of cell conditioning on xylose might be a more beneficial approach to increasing xylose utilization in *S. cerevisiae*.

## CONCLUSIONS AND PERSPECTIVES

The ultimate goal of research into third generation biofuel is to create an organism capable of CBP. For the efficient exploitation



**Table 2 | Ethanol production from xylose by recombinant yeast strains expressing xylose utilizing enzymes.**

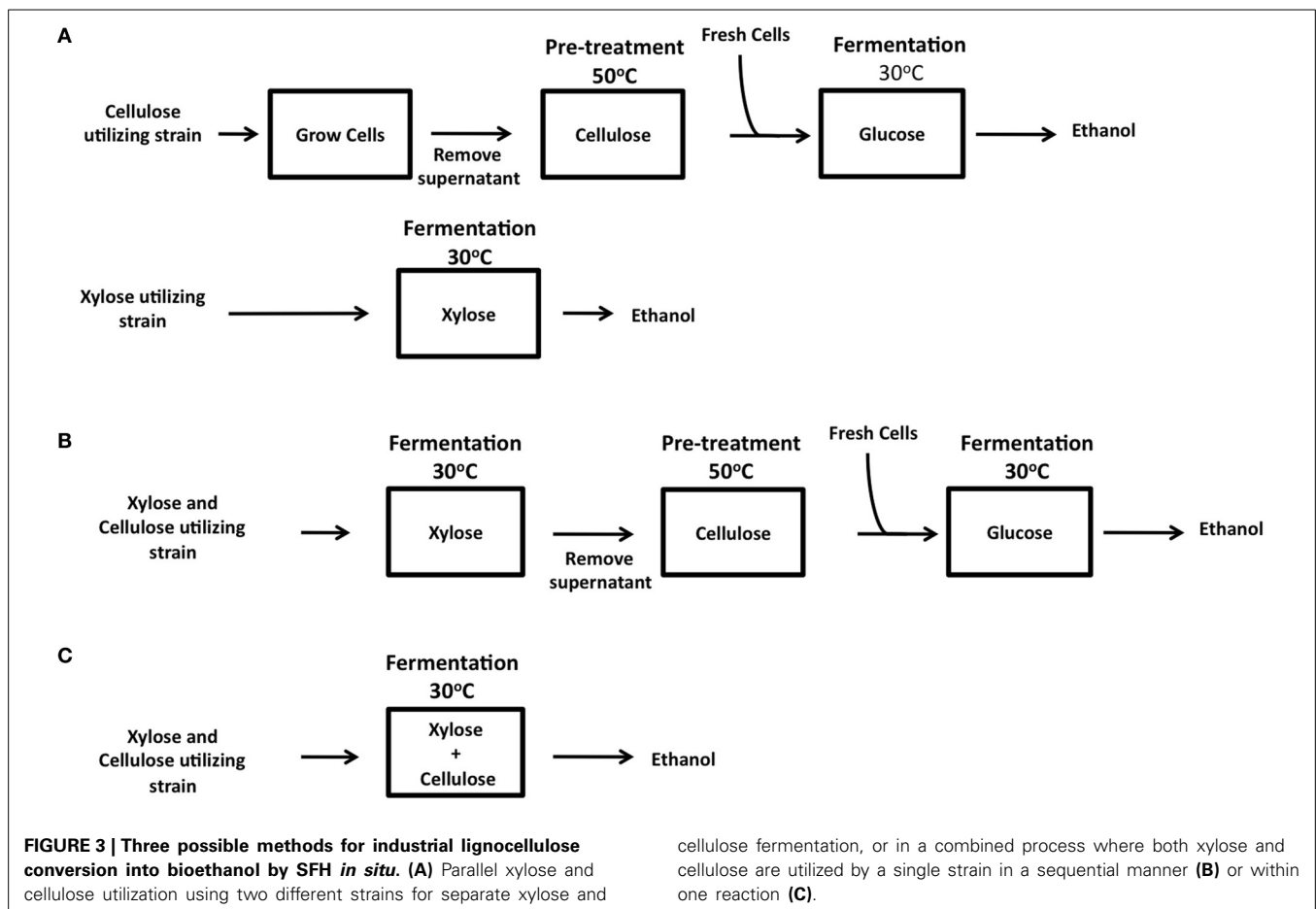
Host strain	XR	XDH	XI	XK	Ethanol (g/L)	Ethanol yield (g/g Xylose)	Ethanol yield (%)	References
<i>S. cerevisiae</i> (4124)	<i>Pichia stipitis</i>	<i>Pichia stipitis</i>		<i>S. cerevisiae</i>			76.8	Bera et al., 2011
<i>S. cerevisiae</i> (4124)	<i>Neurospora crassa</i>	<i>Pichia stipitis</i>		<i>S. cerevisiae</i>			78.6	
<i>S. cerevisiae</i> (4124A)	<i>Candida parapsilosis</i>	<i>Pichia stipitis</i>		<i>S. cerevisiae</i>			34.2	
<i>S. cerevisiae</i> (YRH400)	<i>Pichia stipitis</i>	<i>Pichia stipitis</i>		<i>S. cerevisiae</i>	6.9		48.3	Hector et al., 2013
<i>S. cerevisiae</i> (YRH631)			<i>Prevotella ruminicola</i>	<i>Prevotella ruminicola</i>	4.1		68.6	
<i>S. cerevisiae</i> (YRH1114)			<i>Prevotella ruminicola</i>	<i>Prevotella ruminicola</i>	13.6		82.9	
<i>S. cerevisiae</i> (YY5A)	<i>Candida guilliermondii</i>	<i>Pichia stipitis</i>		<i>Pichia stipitis</i>	6.1	0.33		Tien-Yang et al., 2012
<i>S. cerevisiae</i> (MA-D4)	<i>Pichia stipitis</i>	<i>Pichia stipitis</i>		<i>S. cerevisiae</i>	14.1	0.35		Matsushika et al., 2009b
<i>S. cerevisiae</i> (MA-T4)	<i>Pichia stipitis</i>	<i>Pichia stipitis</i>		<i>S. cerevisiae</i>	15.7	0.35		
<i>S. cerevisiae</i> (Y7092)			<i>Piromyces</i> sp. E2	<i>S. cerevisiae</i>	5.3	0.35		Usher et al., 2011
<i>S. cerevisiae</i> (YPH499XU)	<i>Scheffersomyces stipitis</i>	<i>Scheffersomyces stipitis</i>		<i>S. cerevisiae</i>		0.24		Kato et al., 2013
<i>S. cerevisiae</i> (8X-70)	<i>Scheffersomyces stipitis</i>	<i>Scheffersomyces stipitis</i>		<i>S. cerevisiae</i>		0.36		
<i>S. cerevisiae</i> (8X-27)	<i>Scheffersomyces stipitis</i>	<i>Scheffersomyces stipitis</i>		<i>S. cerevisiae</i>		0.35		
<i>S. cerevisiae</i> (TMB 3057)	<i>Pichia stipitis</i>	<i>Pichia stipitis</i>		<i>S. cerevisiae</i>	13.3	0.33		Karhumaa et al., 2007
<i>S. cerevisiae</i> (TMB 3066)			<i>Piromyces</i> sp. E2	<i>S. cerevisiae</i>	7.3	0.43		
<i>S. cerevisiae</i> (TMB 3400)	<i>Pichia stipitis</i>	<i>Pichia stipitis</i>		<i>S. cerevisiae</i>	12.1	0.34		
<i>S. cerevisiae</i> (BY4741X)	<i>Scheffersomyces stipitis</i>	<i>Scheffersomyces stipitis</i>		<i>S. cerevisiae</i>	25.4	0.28		Fujitomi et al., 2012
<i>S. cerevisiae</i> (HDY.GUF5)			<i>Clostridium phytofermentans</i>	<i>S. cerevisiae</i>		0.23		Demeke et al., 2013

of biomass as a source of bioenergy, it is obvious that both the cellulosic and hemicellulosic fractions of biomass must be used. The degradation of cellulose by cellulase expressing ethanologenic yeast strains is now well established, except for the caveat that to date only synthetic forms of cellulose have been tested (Table 1). The main stumbling block here is the poor enzyme activity of recombinant cellulases at fermentation conditions (30°C). Until activity can be increased at this low temperature, it is likely that there will be a need for an enzymatic pre-hydrolysis of biomass prior to fermentation, i.e., SHF *in situ* (Figure 3). To move the

field forward, it is essential that the degradation of more complex biomass by engineered strains be tested, be it straw, grasses, or waste products such as spent grains. Progress on xylose fermentation has also been made, however both fields have been developing in parallel rather than merging into one. To date only cellobiose and xylose co-utilizing strains having been generated (Katahira et al., 2006; Ha et al., 2011; Aeling et al., 2012).

To achieve CBP, many criteria have to be satisfied. The ideal host should be capable of such metabolizing both cellulose and xylose, maintaining maximum heterologous expression of





enzymes at the optimum ratio, be resistant to high ethanol concentrations and ideally be thermotolerant. Ideally, a single organism should possess all of these attributes, however the heterologous production of so many enzymes can reduce the fitness of the host cell (Tsai et al., 2010). Co-culturing yeast strains expressing individual cellulase or xylanases offers a easy way to vary the ratio of cellulase and xylanase enzymes by simply altering the number of cells expressing each enzyme in the fermentation. With the proposed SHF *in situ* approach, various scenarios can be envisioned. Cellulose and xylose can be fermented separately from pre-treated biomass (Figure 3A). Alternatively, yeast strains co-expressing xylose and cellulose utilizing genes together or co-cultures of strains expressing individual genes can be cultured on hydrolysates derived from pre-treatment of biomass. The available xylose and glucose in the hydrolysate can be used as energy sources to allow the growth of yeast and thus the production and secretion of cellulases into the medium (Figure 3B). The spent medium (supernatant) can then be used for enzymatic pre-hydrolysis of cellulose followed by a fermentation of released glucose. This step-wise process is compatible with current industrial fermentation processes. Ultimately, the most efficient CBP would require a single microorganism capable of fermentation of both xylose and cellulose in a single step at 30°C (Figure 3C). We eagerly await the development of this super strain!

In conclusion, every component involved in lignocellulosic bioethanol generation such as the host organism, hydrolyzing enzymes and even biomass substrate are now the focus of bio-engineering research in the pursuit for greater efficiency to reduce production costs. Thus, the future of biofuel utilization is dependent on the economics of its production, which is itself reliant on the science used to generate it.

## REFERENCES

- Aeling, K. A., Salmon, K. A., Laplaza, J. M., Li, L., Headman, J. R., Hutagalung, A. H., et al. (2012). Co-fermentation of xylose and cellobiose by an engineered *Saccharomyces cerevisiae*. *J. Ind. Microbiol. Biotechnol.* 39, 1597–1604. doi: 10.1007/s10295-012-1169-y
- Anbar, M., Gul, O., Lamed, R., Sezerman, U. O., and Bayer, E. A. (2012). Improved thermostability of *Clostridium thermocellum* endoglucanase Cel8A by using consensus-guided mutagenesis. *Appl. Environ. Microbiol.* 78, 3458–3464. doi: 10.1128/AEM.07985-11
- Bengtsson, O., Jeppsson, M., Sonderegger, M., Parachin, N., Sauer, U., Hahn-Hägerdal, B., et al. (2008). Identification of common traits in improved xylose-growing *Saccharomyces cerevisiae* for inverse metabolic engineering. *Yeast* 25, 835–847. doi: 10.1002/yea.1638
- Bera, A., Ho, N., Khan, A., and Sedlak, M. (2011). A genetic overhaul of *Saccharomyces cerevisiae* 424A(LNH-ST) to improve xylose fermentation. *J. Ind. Microbiol. Biotechnol.* 38, 617–626. doi: 10.1007/s10295-010-0806-6
- Buijs, N. A., Siewers, V., and Nielsen, J. (2013). Advanced biofuel production by the yeast *Saccharomyces cerevisiae*. *Curr. Opin. Chem. Biol.* 17, 480–488. doi: 10.1016/j.cbpa.2013.03.036

- Cai, Z., Zhang, B., and Li, Y. (2012). Engineering *Saccharomyces cerevisiae* for efficient anaerobic xylose fermentation: reflections and perspectives. *Biotechnol. J.* 7, 34–46. doi: 10.1002/biot.201100053
- Chandel, A. K., and Singh, O. V. (2011). Weedy lignocellulosic feedstock and microbial metabolic engineering: advancing the generation of 'Biofuel'. *Appl. Microbiol. Biotechnol.* 89, 1289–1303. doi: 10.1007/s00253-010-3057-6
- Dashtban, M., Schraft, H., and Qin, W. (2009). Fungal bioconversion of lignocellulosic residues; opportunities & perspectives. *Int. J. Biol. Sci.* 5, 578–595. doi: 10.7150/ijbs.5.578
- De Figueiredo Vilela, L., De Mello, V., Reis, V., Bon, E., Gonçalves Torres, F., Neves, B., et al. (2013). Functional expression of *Burkholderia cenocepacia* xylose isomerase in yeast increases ethanol production from a glucose-xylose blend. *Bioresour. Technol.* 128, 792–796. doi: 10.1016/j.biortech.2012.10.014
- Demeke, M., Dietz, H., Li, Y., Foulquié-Moreno, M., Mutturi, S., Deprez, S., et al. (2013). Development of a D-xylose fermenting and inhibitor tolerant industrial *Saccharomyces cerevisiae* strain with high performance in lignocellulose hydrolysates using metabolic and evolutionary engineering. *Biotechnol. Biofuels* 6:89. doi: 10.1186/1754-6834-6-89
- Den Haan, R., McBride, J. E., La Grange, D. C., Lynd, L. R., and Van Zyl, W. H. (2007a). Functional expression of cellobiohydrolases in *Saccharomyces cerevisiae* towards one-step conversion of cellulose to ethanol. *Enzyme Microb. Technol.* 40, 1291–1299. doi: 10.1016/j.enzmictec.2006.09.022
- Den Haan, R., Rose, S. H., Lynd, L. R., and Van Zyl, W. H. (2007b). Hydrolysis and fermentation of amorphous cellulose by recombinant *Saccharomyces cerevisiae*. *Metab. Eng.* 9, 87–94. doi: 10.1016/j.ymben.2006.08.005
- Du, J., Li, S., and Zhao, H. (2010). Discovery and characterization of novel d-xylose-specific transporters from *Neurospora crassa* and *Pichia stipitis*. *Mol. Biosyst.* 6, 2150–2156. doi: 10.1039/c0mb00007h
- Du Plessis, L., Rose, S. H., and Van Zyl, W. H. (2010). Exploring improved endoglucanase expression in *Saccharomyces cerevisiae* strains. *Appl. Microbiol. Biotechnol.* 86, 1503–1511. doi: 10.1007/s00253-009-2403-z
- Elkins, J. G., Raman, B., and Keller, M. (2010). Engineered microbial systems for enhanced conversion of lignocellulosic biomass. *Curr. Opin. Biotechnol.* 21, 657–662. doi: 10.1016/j.copbio.2010.05.008
- Fan, L.-H., Zhang, Z.-J., Yu, X.-Y., Xue, Y.-X., and Tan, T.-W. (2012). Self-surface assembly of cellulosomes with two miniscaffolds on *Saccharomyces cerevisiae* for cellulosic ethanol production. *Proc. Natl. Acad. Sci. U.S.A.* 109, 13260–13265. doi: 10.1073/pnas.1209856109
- Fernandes, S., and Murray, P. (2010). Metabolic engineering for improved microbial pentose fermentation. *Bioeng. Bugs* 1, 424–428. doi: 10.4161/bbug.1.6.12724
- Fiaux, J., Cakar, Z. P., Sonderegger, M., Wuthrich, K., Szyperski, T., and Sauer, U. (2003). Metabolic-flux profiling of the yeasts *Saccharomyces cerevisiae* and *Pichia stipitis*. *Eukaryot. Cell* 2, 170–180. doi: 10.1128/EC.2.1.170-180.2003
- Fitzpatrick, J., Kricka, W., James, T., and Bond, U. (2014). Expression of three *Trichoderma reesei* cellulase genes in *Saccharomyces pastorianus* for the development of a two-step process of hydrolysis and fermentation of cellulose. *J. App. Microbiol.* doi: 10.1111/jam.12494. [Epub ahead of print].
- Fonseca, G. G., Heinzle, E., Wittmann, C., and Gombert, A. K. (2008). The yeast *Kluyveromyces marxianus* and its biotechnological potential. *Appl. Microbiol. Biotechnol.* 79, 339–354. doi: 10.1007/s00253-008-1458-6
- Fontes, C. M., and Gilbert, H. J. (2010). Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. *Annu. Rev. Biochem.* 79, 655–681. doi: 10.1146/annurev-biochem-091208-085603
- Foreman, P. K., Brown, D., Dankmeyer, L., Dean, R., Diener, S., Dunn-Coleman, N. S., et al. (2003). Transcriptional regulation of biomass-degrading enzymes in the filamentous fungus *Trichoderma reesei*. *J. Biol. Chem.* 278, 31988–31997. doi: 10.1074/jbc.M304750200
- Fujita, Y., Ito, J., Ueda, M., Fukuda, H., and Kondo, A. (2004). Synergistic saccharification, and direct fermentation to ethanol, of amorphous cellulose by use of an engineered yeast strain codisplaying three types of cellulolytic enzyme. *Appl. Environ. Microbiol.* 70, 1207–1212. doi: 10.1128/AEM.70.2.1207-1212.2004
- Fujitomi, K., Sanda, T., Hasunuma, T., and Kondo, A. (2012). Deletion of the PHO13 gene in *Saccharomyces cerevisiae* improves ethanol production from lignocellulosic hydrolysate in the presence of acetic and formic acids, and furfural. *Bioresour. Technol.* 111, 161–166. doi: 10.1016/j.biortech.2012.01.161
- Gefen, G., Anbar, M., Morag, E., Lamed, R., and Bayer, E. A. (2012). Enhanced cellulose degradation by targeted integration of a cohesin-fused beta-glucosidase into the *Clostridium thermocellum* cellulosome. *Proc. Natl. Acad. Sci. U.S.A.* 109, 10298–10303. doi: 10.1073/pnas.1202747109
- Guedon, E., Desvaux, M., and Petitdemange, H. (2002). Improvement of cellulolytic properties of *Clostridium cellulolyticum* by metabolic engineering. *Appl. Environ. Microbiol.* 68, 53–58. doi: 10.1128/AEM.68.1.53-58.2002
- Ha, S.-J., Galazka, J., Kim, S., Choi, J.-H., Yang, X., Seo, J.-H., et al. (2011). Engineered *Saccharomyces cerevisiae* capable of simultaneous cellobiose and xylose fermentation. *Proc. Natl. Acad. Sci. U.S.A.* 108, 504–509. doi: 10.1073/pnas.1010456108
- Hasunuma, T., Ismail, K., Nambu, Y., and Kondo, A. (2014). Co-expression of TAL1 and ADH1 in recombinant xylose-fermenting *Saccharomyces cerevisiae* improves ethanol production from lignocellulosic hydrolysates in the presence of furfural. *J. Biosci. Bioeng.* 117, 165–169. doi: 10.1016/j.jbiosc.2013.07.007
- Hasunuma, T., and Kondo, A. (2012). Development of yeast cell factories for consolidated bioprocessing of lignocellulose to bioethanol through cell surface engineering. *Biotechnol. Adv.* 30, 1207–1218. doi: 10.1016/j.biotechadv.2011.10.011
- Hasunuma, T., Sung, K., Sanda, T., Yoshimura, K., Matsuda, F., and Kondo, A. (2011). Efficient fermentation of xylose to ethanol at high formic acid concentrations by metabolically engineered *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* 90, 997–1004. doi: 10.1007/s00253-011-3085-x
- Hector, R., Dien, B., Cotta, M., and Mertens, J. (2013). Growth and fermentation of D-xylose by *Saccharomyces cerevisiae* expressing a novel D-xylose isomerase originating from the bacterium *Prevotella ruminicola* TC2-24. *Biotechnol. Biofuels* 6:84. doi: 10.1186/1754-6834-6-84
- Hector, R., Dien, B., Cotta, M., and Qureshi, N. (2011). Engineering industrial *Saccharomyces cerevisiae* strains for xylose fermentation and comparison for switchgrass conversion. *J. Ind. Microbiol. Biotechnol.* 38, 1193–1202. doi: 10.1007/s10295-010-0896-1
- Hong, J., Wang, Y., Kumagai, H., and Tamaki, H. (2007). Construction of thermotolerant yeast expressing thermostable cellulase genes. *J. Biotechnol.* 130, 114–123. doi: 10.1016/j.jbiotec.2007.03.008
- Hou, X. (2012). Anaerobic xylose fermentation by *Spathaspora passalidarum*. *Appl. Microbiol. Biotechnol.* 94, 205–214. doi: 10.1007/s00253-011-3694-4
- Ismail, K. S., Sakamoto, T., Hatanaka, H., Hasunuma, T., and Kondo, A. (2013). Gene expression cross-profiling in genetically modified industrial *Saccharomyces cerevisiae* strains during high-temperature ethanol production from xylose. *J. Biotechnol.* 163, 50–60. doi: 10.1016/j.jbiotec.2012.10.017
- Jaewoong, M., Liu, Z. L., Menggen, M., and Patricia, J. S. (2013). New genotypes of industrial yeast *Saccharomyces cerevisiae* engineered with YXI and heterologous xylose transporters improve xylose utilization and ethanol production. *Biocatal. Agric. Biotechnol.* 2, 247–254. doi: 10.1016/j.bcab.2013.03.005
- Karhumaa, K., Fromanger, R., Hahn-Hägerdal, B., and Gorwa-Grauslund, M.-F. (2007). High activity of xylose reductase and xylitol dehydrogenase improves xylose fermentation by recombinant *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* 73, 1039–1046. doi: 10.1007/s00253-006-0575-3
- Karhumaa, K., Hahn-Hägerdal, B., and Gorwa-Grauslund, M.-F. (2005). Investigation of limiting metabolic steps in the utilization of xylose by recombinant *Saccharomyces cerevisiae* using metabolic engineering. *Yeast* 22, 359–368. doi: 10.1002/yea.1216
- Katahira, S., Mizuike, A., Fukuda, H., and Kondo, A. (2006). Ethanol fermentation from lignocellulosic hydrolysate by a recombinant xylose- and cellobiosaccharide-assimilating yeast strain. *Appl. Microbiol. Biotechnol.* 72, 1136–1143. doi: 10.1007/s00253-006-0402-x
- Kato, H., Matsuda, F., Yamada, R., Nagata, K., Shirai, T., Hasunuma, T., et al. (2013). Cocktail  $\delta$ -integration of xylose assimilation genes for efficient ethanol production from xylose in *Saccharomyces cerevisiae*. *J. Biosci. Bioeng.* 116, 333–336. doi: 10.1016/j.jbiosc.2013.03.020
- Kim, B., Du, J., Eriksen, D. T., and Zhao, H. (2013). Combinatorial design of a highly efficient xylose-utilizing pathway in *Saccharomyces cerevisiae* for the production of cellulosic biofuels. *Appl. Environ. Microbiol.* 79, 931–941. doi: 10.1128/AEM.02736-12

- Kim, S. R., Ha, S. J., Kong, I. I., and Jin, Y. S. (2012). High expression of XYL2 coding for xylitol dehydrogenase is necessary for efficient xylose fermentation by engineered *Saccharomyces cerevisiae*. *Metab. Eng.* 14, 336–343. doi: 10.1016/j.ymben.2012.04.001
- Kubicek, C. P. (2013). Systems biological approaches towards understanding cellulase production by *Trichoderma reesei*. *J. Biotechnol.* 163, 133–142. doi: 10.1016/j.jbiotec.2012.05.020
- Leggio, L. L., Welner, D., and De Maria, L. (2012). A structural overview of GH61 proteins – fungal cellulose degrading polysaccharide monooxygenases. *Comput. Struct. Biotechnol. J.* 2, 1–8. doi: 10.5936/csbj.201209019
- Liang, C., Fioroni, M., Rodriguez-Ropero, F., Xue, Y., Schwaneberg, U., and Ma, Y. (2011). Directed evolution of a thermophilic endoglucanase (Cel5A) into highly active Cel5A variants with an expanded temperature profile. *J. Biotechnol.* 154, 46–53. doi: 10.1016/j.jbiotec.2011.03.025
- Lynd, L. R., Weimer, P. J., Van Zyl, W. H., and Pretorius, I. S. (2002). Microbial cellulose utilization: fundamentals and biotechnology. *Microbiol. Mol. Biol. Rev.* 66, 506–577. doi: 10.1128/MMBR.66.3.506-577.2002
- Martinez, D., Berka, R. M., Henrissat, B., Saloheimo, M., Arvas, M., Baker, S. E., et al. (2008). Genome sequencing and analysis of the biomass-degrading fungus *Trichoderma reesei* (syn. *Hypocrea jecorina*). *Nat. Biotechnol.* 26, 553–560. doi: 10.1038/nbt1403
- Matsushika, A., Inoue, H., Kodaki, T., and Sawayama, S. (2009a). Ethanol production from xylose in engineered *Saccharomyces cerevisiae* strains: current state and perspectives. *Appl. Microbiol. Biotechnol.* 84, 37–53. doi: 10.1007/s00253-009-2101-x
- Matsushika, A., Inoue, H., Murakami, K., Takimura, O., and Sawayama, S. (2009b). Bioethanol production performance of five recombinant strains of laboratory and industrial xylose-fermenting *Saccharomyces cerevisiae*. *Bioresour. Technol.* 100, 2392–2398. doi: 10.1016/j.biortech.2008.11.047
- Nakatani, Y., Yamada, R., Ogino, C., and Kondo, A. (2013). Synergetic effect of yeast cell-surface expression of cellulase and expansin-like protein on direct ethanol production from cellulose. *Microb. Cell Fact.* 12:66. doi: 10.1186/1475-2859-12-66
- Partow, S., Siewers, V., Bjorn, S., Nielsen, J., and Maury, J. (2010). Characterization of different promoters for designing a new expression vector in *Saccharomyces cerevisiae*. *Yeast* 27, 955–964. doi: 10.1002/yea.1806
- Peng, F., Peng, P., Xu, F., and Sun, R. C. (2012). Fractional purification and bioconversion of hemicelluloses. *Biotechnol. Adv.* 30, 879–903. doi: 10.1016/j.biotechadv.2012.01.018
- Ruel, K., Nishiyama, Y., and Joseleau, J. P. (2012). Crystalline and amorphous cellulose in the secondary walls of Arabidopsis. *Plant Sci.* 193–194, 48–61. doi: 10.1016/j.plantsci.2012.05.008
- Runquist, D., Hahn-Hägerdal, B., and Rådström, P. (2010). Comparison of heterologous xylose transporters in recombinant *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* 3:5. doi: 10.1186/1754-6834-3-5
- Ryu, S., and Karim, M. (2011). A whole cell biocatalyst for cellulosic ethanol production from dilute acid-pretreated corn stover hydrolyzates. *Appl. Microbiol. Biotechnol.* 91, 529–542. doi: 10.1007/s00253-011-3261-z
- Schwartz, K., Wenger, J. W., Dunn, B., and Sherlock, G. (2012). APJ1 and GRE3 homologs work in concert to allow growth in xylose in a natural *Saccharomyces sensu stricto* hybrid yeast. *Genetics* 191, 621–632. doi: 10.1534/genetics.112.140053
- Skory, C. D., Freer, S. N., and Bothast, R. J. (1997). Screening for ethanol-producing filamentous fungi. *Biotechnol. Lett.* 19, 203–206. doi: 10.1023/A:1018337003433
- Takahashi, S., Iikura, H., Nakamura, A., Hidaka, M., Masaki, H., and Uozumi, T. (1998). Overproduction of recombinant *Trichoderma reesei* cellulases by *Aspergillus oryzae* and their enzymatic properties. *J. Biotechnol.* 65, 163–171. doi: 10.1016/S0168-1656(98)00084-4
- Tanino, T., Ito, T., Ogino, C., Ohmura, N., Ohshima, T., and Kondo, A. (2012). Sugar consumption and ethanol fermentation by transporter-overexpressed xylose-metabolizing *Saccharomyces cerevisiae* harboring a xyloseisomerase pathway. *J. Biosci. Bioeng.* 114, 209–211. doi: 10.1016/j.jbiosc.2012.03.004
- Tien-Yang, M., Ting-Hsiang, L., Teng-Chieh, H., Chiung-Fang, H., Gia-Luen, G., and Wen-Song, W. (2012). An improved method of xylose utilization by recombinant *Saccharomyces cerevisiae*. *J. Ind. Microbiol. Biotechnol.* 39, 1477–1486. doi: 10.1007/s10295-012-1153-6
- Tsai, S. L., Goyal, G., and Chen, W. (2010). Surface display of a functional minicellulosome by intracellular complementation using a synthetic yeast consortium and its application to cellulose hydrolysis and ethanol production. *Appl. Environ. Microbiol.* 76, 7514–7520. doi: 10.1128/AEM.01777-10
- Urbina, H., Frank, R., and Blackwell, M. (2013). *Scheffersomyces cryptocercus*: a new xylose-fermenting yeast associated with the gut of wood roaches and new combinations in the Sugiyamaella yeast clade. *Mycologia* 105, 650–660. doi: 10.3852/12-094
- Usher, J., Balderas-Hernandez, V., Quon, P., Gold, N., Martin, V. J., Mahadevan, R., et al. (2011). Chemical and synthetic genetic array analysis identifies genes that suppress xylose utilization and fermentation in *Saccharomyces cerevisiae*. *G3* 1, 247–258. doi: 10.1534/g3.111.000695
- Van Rooyen, R., Hahn-Hägerdal, B., La Grange, D., and Van Zyl, W. (2005). Construction of cellobiose-growing and fermenting *Saccharomyces cerevisiae* strains. *J. Biotechnol.* 120, 284–295. doi: 10.1016/j.jbiotec.2005.06.013
- Van Vleet, J., Jeffries, T., and Olsson, L. (2008). Deleting the para-nitrophenyl phosphatase (pNPPase), PHO13, in recombinant *Saccharomyces cerevisiae* improves growth and ethanol production on D-xylose. *Metab. Eng.* 10, 360–369. doi: 10.1016/j.ymben.2007.12.002
- Voutilainen, S. P., Murray, P. G., Tuohy, M. G., and Koivula, A. (2010). Expression of *Talaromyces emersonii* cellobiohydrolase Cel7A in *Saccharomyces cerevisiae* and rational mutagenesis to improve its thermostability and activity. *Protein Eng. Des. Sel.* 23, 69–79. doi: 10.1093/protein/gzp072
- Wahlbom, C. F., Otero, R. R. C., Zyl, W. H. V., Hahn-Hägerdal, B., and Jonsson, L. J. (2003). Molecular analysis of a *Saccharomyces cerevisiae* mutant with improved ability to utilize xylose shows enhanced expression of proteins involved in transport, initial xylose metabolism, and the pentose phosphate pathway. *Appl. Environ. Microbiol.* 69, 740–746. doi: 10.1128/AEM.69.2.740-746.2003
- Wang, X. J., Peng, Y. J., Zhang, L. Q., Li, A. N., and Li, D. C. (2012). Directed evolution and structural prediction of cellobiohydrolase II from the thermophilic fungus *Chaetomium thermophilum*. *Appl. Microbiol. Biotechnol.* 95, 1469–1478. doi: 10.1007/s00253-011-3799-9
- Wanlapa, L., Ancharida, A., Motofumi, S., Moriya, O., and Somboon, T. (2013). Diversity and fermentation products of xylose-utilizing yeasts isolated from buffalo feces in Thailand. *Microbes Environ.* 28, 354–360. doi: 10.1264/jsme2.ME13023
- Wen, F., Sun, J., and Zhao, H. (2010). Yeast surface display of trifunctional minicellulosomes for simultaneous saccharification and fermentation of cellulose to ethanol. *Appl. Environ. Microbiol.* 76, 1251–1260. doi: 10.1128/AEM.01687-09
- Wenger, J., Schwartz, K., and Sherlock, G. (2010). Bulk segregant analysis by high-throughput sequencing reveals a novel xylose utilization gene from *Saccharomyces cerevisiae*. *PLoS Genet.* 6:e1000942. doi: 10.1371/journal.pgen.1000942
- Xiong, M., Chen, G., and Barford, J. (2011). Alteration of xylose reductase coenzyme preference to improve ethanol production by *Saccharomyces cerevisiae* from high xylose concentrations. *Bioresour. Technol.* 102, 9206–9215. doi: 10.1016/j.biortech.2011.06.058
- Yamada, R., Hasunuma, T., and Kondo, A. (2013). Endowing non-cellulolytic microorganisms with cellulolytic activity aiming for consolidated bioprocessing. *Biotechnol. Adv.* 31, 754–763. doi: 10.1016/j.biotechadv.2013.02.007
- Yamada, R., Taniguchi, N., Tanaka, T., Ogino, C., Fukuda, H., and Kondo, A. (2010). Cocktail delta-integration: a novel method to construct cellulolytic enzyme expression ratio-optimized yeast strains. *Microb. Cell Fact.* 9:32. doi: 10.1186/1475-2859-9-32
- Yamada, R., Taniguchi, N., Tanaka, T., Ogino, C., Fukuda, H., and Kondo, A. (2011). Direct ethanol production from cellulosic materials using a diploid strain of *Saccharomyces cerevisiae* with optimized cellulase expression. *Biotechnol. Biofuels* 4:8. doi: 10.1186/1754-6834-4-8
- Yanase, S., Yamada, R., Kaneko, S., Noda, H., Hasunuma, T., Tanaka, T., et al. (2010). Ethanol production from cellulosic materials using cellulase-expressing yeast. *Biotechnol. J.* 5, 449–455. doi: 10.1002/biot.200900291
- Young, E., Comer, A., Huang, H., and Alper, H. (2012). A molecular transporter engineering approach to improving xylose catabolism in *Saccharomyces cerevisiae*. *Metab. Eng.* 14, 401–411. doi: 10.1016/j.ymben.2012.03.004
- Young, E., Poucher, A., Comer, A., and Bailey, A. (2011). Functional survey for heterologous sugar transport proteins, using *Saccharomyces cerevisiae* as a host. *Appl. Environ. Microbiol.* 77, 3311–3319. doi: 10.1128/AEM.02651-10
- Žižňáková, L., and Baldrian, P. (2012). Fungal polysaccharide monooxygenases: new players in the decomposition of cellulose. *Fungal Ecol.* 5, 481–489. doi: 10.1016/j.funeco.2012.05.001

Zhang, G., Lin, Y., He, P., Li, L., Wang, Q., and Ma, Y. (2014). Characterization of the sugar alcohol-producing yeast *Pichia anomala*. *J. Ind. Microbiol. Biotechnol.* 41, 41–48. doi: 10.1007/s10295-013-1364-5

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 30 January 2014; paper pending published: 13 March 2014; accepted: 31 March 2014; published online: 22 April 2014.

*Citation:* Kricka W, Fitzpatrick J and Bond U (2014) Metabolic engineering of yeasts by heterologous enzyme production for degradation of cellulose and hemicellulose from biomass: a perspective. *Front. Microbiol.* 5:174. doi: 10.3389/fmicb.2014.00174

This article was submitted to *Microbial Physiology and Metabolism*, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Kricka, Fitzpatrick and Bond. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Comparison of transcriptional profiles of *Clostridium thermocellum* grown on cellobiose and pretreated yellow poplar using RNA-Seq

Hui Wei<sup>1\*†</sup>, Yan Fu<sup>2†</sup>, Lauren Magnusson<sup>1</sup>, John O. Baker<sup>1</sup>, Pin-Ching Maness<sup>1</sup>, Qi Xu<sup>1</sup>, Shihui Yang<sup>3</sup>, Andrew Bowersox<sup>1,3</sup>, Igor Bogorad<sup>1</sup>, Wei Wang<sup>1</sup>, Melvin P. Tucker<sup>3</sup>, Michael E. Himmel<sup>1</sup> and Shi-You Ding<sup>1\*</sup>

<sup>1</sup> Biosciences Center, National Renewable Energy Laboratory, Golden, CO, USA

<sup>2</sup> Center for Plant Genomics, Iowa State University, Ames, IA, USA

<sup>3</sup> National Bioenergy Center, National Renewable Energy Laboratory, Golden, CO, USA

## Edited by:

Nigel Peter Minton, University of Nottingham, UK

## Reviewed by:

Patricia Coutinho Dos Santos, Wake Forest University, USA  
Marianne Guiral, Centre National de la Recherche Scientifique, France

## \*Correspondence:

Hui Wei and Shi-You Ding, National Renewable Energy Laboratory, Biosciences Center, 15013 Denver W Pkwy, Golden, CO 80401, USA  
e-mail: hui.wei@nrel.gov;  
shi.you.ding@nrel.gov

## † Present address:

Yan Fu, Monsanto Company, St. Louis, USA

† These authors have contributed equally to this work.

The anaerobic, thermophilic bacterium, *Clostridium thermocellum*, secretes multi-protein enzyme complexes, termed cellulosomes, which synergistically interact with the microbial cell surface and efficiently disassemble plant cell wall biomass. *C. thermocellum* has also been considered a potential consolidated bioprocessing (CBP) organism due to its ability to produce the biofuel products, hydrogen, and ethanol. We found that *C. thermocellum* fermentation of pretreated yellow poplar (PYP) produced 30 and 39% of ethanol and hydrogen product concentrations, respectively, compared to fermentation of cellobiose. RNA-seq was used to analyze the transcriptional profiles of these cells. The PYP-grown cells taken for analysis at the late stationary phase showed 1211 genes up-regulated and 314 down-regulated by more than two-fold compared to the cellobiose-grown cells. These affected genes cover a broad spectrum of specific functional categories. The transcriptional analysis was further validated by sub-proteomics data taken from the literature; as well as by quantitative reverse transcription-PCR (qRT-PCR) analyses of selected genes. Specifically, 47 cellulosomal protein-encoding genes, genes for 4 pairs of SigI-RsgI for polysaccharide sensing, 7 cellodextrin ABC transporter genes, and a set of NAD(P)H hydrogenase and alcohol dehydrogenase genes were up-regulated for cells growing on PYP compared to cellobiose. These genes could be potential candidates for future studies aimed at gaining insight into the regulatory mechanism of this organism as well as for improvement of *C. thermocellum* in its role as a CBP organism.

**Keywords:** *Clostridium thermocellum*, transcriptomics, RNA-Seq, pretreated yellow poplar (PYP), cellobiose, cellulosome, ethanol, hydrogen

## INTRODUCTION

Microbial conversion of biomass to biofuels is an attractive route for biofuel development, but an essential challenge is to increase the microbial capacity both for overcoming the biomass recalcitrance and for converting the biomass-derived sugars to biofuels (Himmel et al., 2007; Alper and Stephanopoulos, 2009). *Clostridium thermocellum*, a gram-positive, thermophilic, anaerobic bacterium, is one of the model consolidated bioprocessing (CBP) systems used to study the enzymatic hydrolysis of cellulosic biomass to produce fuels (Islam et al., 2006; Brown et al., 2012; Yang et al., 2012). The following features of *C. thermocellum*

contribute to its suitability as a model cellulolytic, biofuel-producing bacterium: (1) It produces cellulosomes, a type of highly organized multi-protein enzyme complexes, which have been shown to be highly efficient enzyme systems in deconstructing plant cell wall, especially in degrading the recalcitrant substrate crystalline cellulose (Bayer et al., 1998, 2004). (2) It carries out mixed-product fermentation, producing ethanol, H<sub>2</sub> and numerous organic acids including acetate, formate, and lactate (Demain et al., 2005; Islam et al., 2006). (3) It is suitable for both submerged and solid-state fermentation (Bayer et al., 2007), the latter having similarity to compost in the set-up of feedstock (Wei et al., 2012). (4) Its genome sequence is available (<http://genome.jgi-psf.org/cloth/cloth.info.html>) and many of the cellulolytic enzymes are identified and biochemically characterized. The knowledge gained from studies of this species will benefit work on other clostridial species of industrial interest, such as *C. acetobutylicum*, known to produce the potential fuels acetone, ethanol, and butanol (Cooksey et al., 2012).

Fulfilling this potential will require a more in-depth understanding of the metabolic and genetic mechanisms by which

**Abbreviations:** Abf,  $\alpha$ -N-arabinofuranosidase; ADH, alcohol dehydrogenase; CBO, cellobiose-grown cells only; Cbp, carbohydrate-binding protein; CBP, consolidated bioprocessing; CDP, cellodextrin phosphorylase; CEP, cellobiose phosphorylase; COG, clusters of orthologous groups; *emPAL*, exponentially modified protein abundance index; FC, fold changes; Fdred, reduced ferredoxin; GH, glycoside hydrolases; NSAF, normalized spectral abundance factor; *PF*, passing filter; PTA, phosphotransacetylase; PYP, pretreated yellow poplar; PYP0, PYP-grown cells only; qRT-PCR, quantitative reverse transcription-PCR; RPKM, reads per kilobase of exon model per million mapped reads; RsgI, RsgI-like anti-sigma factors; SigI, sigma factor.



*C. thermocellum* utilizes recalcitrant biomass substrate. So far, a number of studies have been carried out on *C. thermocellum*, such as transcriptomic analysis of stress responses to ethanol, furfural, and heat during growth on pure sugars (i.e., cellobiose) (Yang et al., 2012), time course studies of cell growth on pure crystalline cellulose (Avicel) (Raman et al., 2011), and comparisons of growth on pure crystalline cellulose (Sigmacel 50) and cellobiose (Riederer et al., 2011). However, despite reports on the sub-proteomic analyses of cellulosomes from *C. thermocellum* grown on Avicel (Gold and Martin, 2007) and on combinations of Avicel with pectin and/or xylan, or on pretreated switchgrass (Raman et al., 2009), there were no genome-wide transcriptomic studies reported for growth on pretreated woody plant biomass until most recently a transcriptomic analysis comparing *C. thermocellum* cells grown on pretreated switchgrass and woody biomass, cottonwood (*Populus trichocarpa* × *Populus deltoides*; black cottonwood × eastern cottonwood in common names), was published on December 2, 2013 (Wilson et al., 2013a), after the submission of this manuscript.

Woody biomass has been found to be more recalcitrant to enzyme digestion than is herbaceous biomass. For example, whereas the Recalcitrance Index for switchgrass is ~0.35, the index for hardwoods, such as yellow poplar, is 0.56, indicating that hardwoods are more difficult to be degraded (Wei et al., 2009). Previous studies have shown that cellulolytic bacteria grown on different lignocellulosic substrates have different levels of glycoside hydrolase (GH) activities (Irwin et al., 2003). As such *C. thermocellum* grown on pretreated woody plant biomass is likely to have distinctive responsive genes, as well as different composition of cellulolytic enzymes different from those grown on herbaceous biomass. Recently, we found that species of the genus *Clostridium* (including *C. thermocellum*) were among the dominant species, comprising 6.3% of the total, in an anaerobic community decomposing yellow poplar wood chips (van der Lelie et al., 2012). In parallel, we studied the yellow poplar compost system (Wei et al., 2012), in which *Clostridium* was also found to be one of the dominant bacteria (data not shown). These results prompted us to explore *C. thermocellum* grown on dilute acid-pretreated yellow poplar (PYP) as a single-species model for the plant cell wall-degradation. PYP is a widely-utilized feedstock in process-development for conversion of biomass to fuels and chemicals, and it is important to identify specific enzymes that this organism calls into action to attack this form of the feedstock/substrate.

PYP has been previously demonstrated that 60% conversion to simple sugars can be achieved with a loading of ~8.4 mg per g biomass cellulose of a mixture of *Trichoderma reesei* CBHI and *Acidothermus cellulolyticus* E1 (95%: 5% on molar basis) (Vinzant et al., 1994; Baker et al., 1997). In this study, *C. thermocellum* was bench fermented using PYP and cellobiose as sole carbon sources, respectively, and transcriptional profiles were analyzed. RNA-seq is a recently developed technology for transcriptome profiling, which uses next-generation sequencing to reveal the presence and quantity of RNA in biological samples. The goals of this study are two-fold: first, we identified genes responsive for degradation of recalcitrant biomass substrates in PYP- vs. cellobiose-grown *C. thermocellum* cells. Secondly, we specifically

focused on candidate genes related to cellulosome, cellodextrin transport, polysaccharide signal transduction and end-product synthesis related genes. These candidate genes are likely to be valuable for mechanism study; as well as for protein-engineering to further improve the abilities of this already potent organism to degrade the cell walls of recalcitrant biomass feedstocks.

## MATERIALS AND METHODS

**Figure 1A** shows the overall experimental approach we designed to investigate the *C. thermocellum* utilization of pretreated biomass substrates as reflected by changes in the cell's transcriptome. Details of the experimental approach are described in the following sections.

### CARBON AND LIGNOCELLULOSIC SUBSTRATES

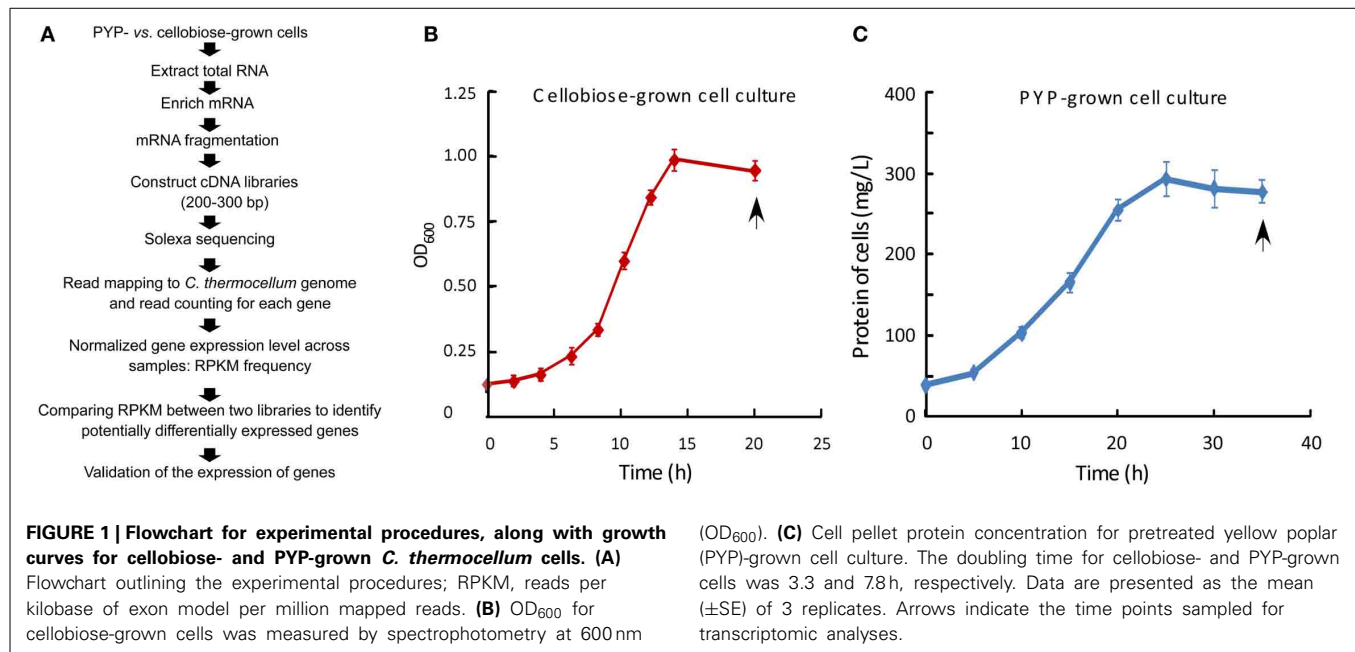
Cellobiose and other chemical compounds were purchased from Sigma (St. Louis, MO). PYP was prepared as described previously (Tucker et al., 1998). Briefly, the milled yellow biomass (20% solids loading) was pretreated in 0.21% w/w H<sub>2</sub>SO<sub>4</sub> at 200°C for 4 min. The resultant PYP contained ~65% cellulose, 4% xylan, and 31% lignin (dry weight basis). PYP was exhaustively washed with deionized water until the pH reached that of deionized water, prior to being used in medium preparation as described below.

### MICROORGANISMS AND CULTURE CONDITIONS

*C. thermocellum* ATCC 27405 was routinely cultured at 60°C anaerobically in 30-mL serum bottles containing 10 mL of ATCC medium 1191 containing cellobiose, and was subcultured with 2% inoculum taken at exponential growth phase. For this study, similarly to practices in the literature (Gold and Martin, 2007; Rydzak et al., 2009), 100-mL batch culture in 250-mL serum bottles was used for the growth of the strain anaerobically at 58°C in ATCC medium 1191, containing 0.30% (w/v) cellobiose as the control carbohydrate substrate, or 0.44% PYP as biomass substrate (sugar-content equivalent to 0.30% cellobiose) with an agitation of 130 rpm. Cell growth was monitored based on either the measurement of optical density (OD) by spectrophotometry at 600 nm (OD<sub>600</sub>) for cellobiose-grown culture, or the increase in pellet protein amount for PYP-grown culture using the method described in literature (Raman et al., 2009). Three independent sets of PYP- vs. cellobiose-grown cell cultures, harvested at the late stationary phase (20 h for cellobiose-grown cells and 36 h for PYP-grown cells), were carried through the downstream RNA extraction preparations. Two sets of RNA preparations were used for cDNA library construction and RNA-Seq, whereas all three sets of RNA preparations were used for qRT-PCR analysis in verifying RNA-Seq data of selected genes.

### MEASUREMENT OF H<sub>2</sub> AND ETHANOL, AND ISOLATION OF BACTERIAL TOTAL RNA

Cells were harvested for RNA isolation in late stationary phase after the hydrogen concentration at the headspace was analyzed using GC method, wherein an Agilent 7890A GC equipped with a Supelco 60/80 molsieve 5A column was used for the measurement of H<sub>2</sub>. The cell culture was centrifuged at high speed (8000 × g) for 5 min to pellet the bacterial cells. Cell-free culture supernatants were used to measure ethanol concentration by



using HPLC (Agilent) with a refractive index detector (Shimadzu, Kyoto, Japan). All samples were filtered through a 0.45- $\mu$ m filter before HPLC analysis. The organic acids were separated in an Aminex HPX-87H column (Bio-Rad) running at a flow rate of 0.6 ml/min at 50°C, with 4 mM H<sub>2</sub>SO<sub>4</sub> as eluent.

The cell pellets collected from 10 mL culture were used for RNA extraction, using a combined protocol (Yang et al., 2012), in which the TRIzol (Invitrogen, Carlsbad, CA) extraction aqueous phase was mixed with an equal volume of 70% ethanol, and then applied to the column of Qiagen RNeasy Mini Kit (Qiagen, Valencia, CA) for purification according to the manufacturer's instructions to obtain total RNAs. 7.5  $\mu$ g total RNAs were mixed with 3  $\mu$ L DNase I (Invitrogen; 1 U/ $\mu$ L) in 3  $\mu$ L 10x DNase I reaction buffer, with adding RNase-free water added up to a total volume of 30  $\mu$ L, and incubated at room temperature for 15 min. The DNase I was inactivated by adding 3  $\mu$ L of 25 mM EDTA and heating for 10 min at 65°C. The DNase I treated total RNA was precipitated by the ethanol-glycogen method and re-dissolved in 15  $\mu$ L of 1 mM EDTA. Since the total RNAs contains 75% rRNA (Chen and Duan, 2011), mRNA was enriched by using the MICROBExpress Bacterial mRNA Enrichment Kit (cat # AM1905, Ambion, Life Technologies, Austin, TX) to remove 16 and 23 s rRNAs. The resultant RNAs were quantified and analyzed for integrity on the Agilent 2100 Bioanalyzer, and used for cDNA library construction as described below.

#### cDNA LIBRARY CONSTRUCTION, SEQUENCING, AND READ MAPPING

cDNA libraries were constructed with the RNA-Seq sample preparation kit (RS-100-0801, Illumina, San Diego, CA), using the procedure provided by the manufacturer. Briefly, the above enriched mRNA samples were fragmented, and then annealed to random hexamers and reverse transcribed. After second strand synthesis, end repair, and A-tailing, cDNA fragment ends were ligated to adapters that were complementary to

sequencing primers. Resultant cDNA libraries were size separated on agarose gels with ~200 bp fragments being excised, and amplified by 15 cycles of PCR. The prepared cDNA libraries were sequenced on an Illumina Genome Analyzer II by the Michigan State University DNA facility, using the standard protocols and running for 75 cycles of data acquisition. Solexa reads were aligned to the reference genome sequence of *C. thermocellum* strain (ATCC 27405) deposited at NCBI (Accession: NC\_009012.1) using Bowtie (Langmead et al., 2009). The reference genome is 3.8-Mb long with 3305 predicted genes, including 71 structural RNA genes, according to the NCBI website for the genome (<http://www.ncbi.nlm.nih.gov/genome/?term=Clostridium+thermocellum>) accessed on April 28, 2010. The best end-to-end alignment (ties broken by read quality) was used with no more than two mismatches. The reads that cannot be unambiguously mapped were not included for further analysis. The above alignment approach and criteria were similar to those used in literature (Jia et al., 2009; Lin et al., 2011).

#### GENE EXPRESSION NORMALIZATION AND IDENTIFICATION OF DIFFERENTIALLY EXPRESSED GENES

After read mapping, the midpoint of the read-reference alignment was used to determine which gene that read belongs to (or was derived from), and the RNA-seq read counts for each gene can be then calculated. To facilitate comparison of gene expression levels, both within and between two samples, we quantified and normalized gene expression levels by calculating the reads per kilobase of exon model per million mapped reads (RPKM), i.e., calculating the number of reads mapped per kilobase of exon model divided by the total number of mapped reads in the whole dataset (Mortazavi et al., 2008). As an indicator of gene expression level, the RPKM is widely recognized as a measure of Solexa read density that reflects the molar concentration of a transcript in the starting RNA sample, thus making the

normalized gene expression levels comparable within and among samples.

The fold-change for the expression of an individual gene was calculated by taking the ratio of RPKM in PYP-grown cells to that in cellobiose-grown cells ( $\text{RPKM}_{\text{PYP}}/\text{RPKM}_{\text{cellobiose}}$ ). The cutoff value for defining a gene as “differentially expressed” is either a two-fold increase (with the value of  $\text{RPKM}_{\text{PYP}}/\text{RPKM}_{\text{cellobiose}}$  larger than 2.0), or a two-fold decrease (with the value of  $\text{RPKM}_{\text{PYP}}/\text{RPKM}_{\text{cellobiose}}$  less than 0.5).

### STATISTICAL ANALYSIS

The log<sub>2</sub>-transformed raw RPKM dataset was imported into the statistical analysis software JMP Genomics 6.0 software (SAS Institute, NC), and the data of PYP- and cellobiose-grown cell samples were normalized together using the LOWESS normalization algorithm within JMP Genomics. To determine if the differential expression levels between PYP- and cellobiose-grown cell samples were statistically significant, the normalized Log<sub>2</sub>(RPKM) data were subjected to One-Way analysis of variance (ANOVA) as described in literature (Yang et al., 2012, 2013; Wilson et al., 2013b). The False Discovery Rate (FDR) testing method was used with a significance threshold of  $p < 0.05$  being considered statistically significant.

### BIOLOGICAL INTERPRETATION OF DIFFERENTIALLY EXPRESSED GENES

The identified, differentially expressed genes were interpreted and discussed in the context of biological processes and functions, using clusters of orthologous groups (COG) and Carbohydrate-Active enZymes (CAZy) analyses of proteins (www.cazy.org/).

### QUANTITATIVE REVERSE TRANSCRIPTION-PCR (qRT-PCR)

Based on their potential functional importance, 12 genes were selected for validating the results of the RNA-Seq analysis. The primers for these genes were either based on literature or designed with Primer Express 3.0 software (Applied Biosystems), and are described along with the PCR results in the Results and Discussion section. Total RNA was extracted from three sets of independent cultures grown on PYP vs. cellobiose as described above, and then converted to cDNA by random priming, using the SuperScript II kit (Invitrogen, San Diego, CA). PCR reactions were run in triplicate using procedure as previously described (Wei et al., 2012). The transcription level of genes was determined according to the  $2^{-\Delta\Delta CT}$  method, using RecA as a reference gene for the normalization of gene expression levels (Stevenson and Weimer, 2005).

### CORRELATION ANALYSIS FOR CELLULOSOME-RELATED GENE EXPRESSION AND PROTEIN ABUNDANCE

Two sets of quantitative cellulosomal protein data of cellobiose-grown *C. thermocellum* cells at late stationary phase were retrieved from literature (Gold and Martin, 2007; Raman et al., 2009), and plotted against our sub-dataset of log<sub>2</sub>(RPKM) values of cellulosomal genes in cells grown on the same substrate at the same culture phase in this study. Pearson correlation coefficient values were calculated using Microsoft Excel (Microsoft Corporation, Redmond, WA, USA), and used as an indicator for the degree of correlation for the compared pairs.

### COMPOSITIONAL ANALYSIS OF PRETREATED YELLOW POPLAR RESIDUES

The PYP residues from PYP-grown *C. thermocellum* culture were collected by centrifugation at low speed ( $100 \times g$ ) for 2 min to precipitate the insoluble substrate not consumed by the bacterial cells in the culture. Such centrifugation speed has been used in the literature to remove any insoluble substrate in *C. thermocellum* culture (Dror et al., 2005). Compositional analysis of the collected PYP residue was performed by the National Bioenergy Center, National Renewable Energy Laboratory, using method described in literature (Templeton et al., 2010).

## RESULTS AND DISCUSSION

### C. THERMOCCELLUM GROWTH AND THE PRODUCTION OF HYDROGEN AND ETHANOL

The first step of this study, as illustrated in **Figures 1A–C**, is the growth of *C. thermocellum* ATCC 27405 with two types of carbohydrate substrates, PYP being compared with cellobiose. GC analysis of gas composition in the headspace of batch cultures at the late stationary phase revealed a production yield of 1.22 vs. 0.92 mole H<sub>2</sub>/mole glucose consumed in cellobiose- and PYP-grown *C. thermocellum*, respectively (**Table 1**). Furthermore, HPLC analysis of metabolite production in supernatants of harvested cell culture revealed a production yield of 0.51 vs. 0.30 mole ethanol /mole glucose equivalent consumed in cellobiose- and PYP-grown *C. thermocellum*, respectively. Overall, the production of hydrogen and ethanol by PYP-grown *C. thermocellum* are 75 and 58% of those by cellobiose-grown *C. thermocellum*, respectively (**Table 1**). Wet chemistry analysis of spent PYP solids revealed that 52% of the glucan in the starting PYP substrate was consumed during the fermentations, which can partially explain the lower absolute H<sub>2</sub> and ethanol productions compared to the culture on cellobiose substrate (in which nearly all cellobiose was depleted at late stationary phase).

### RNA-Seq RESULTS

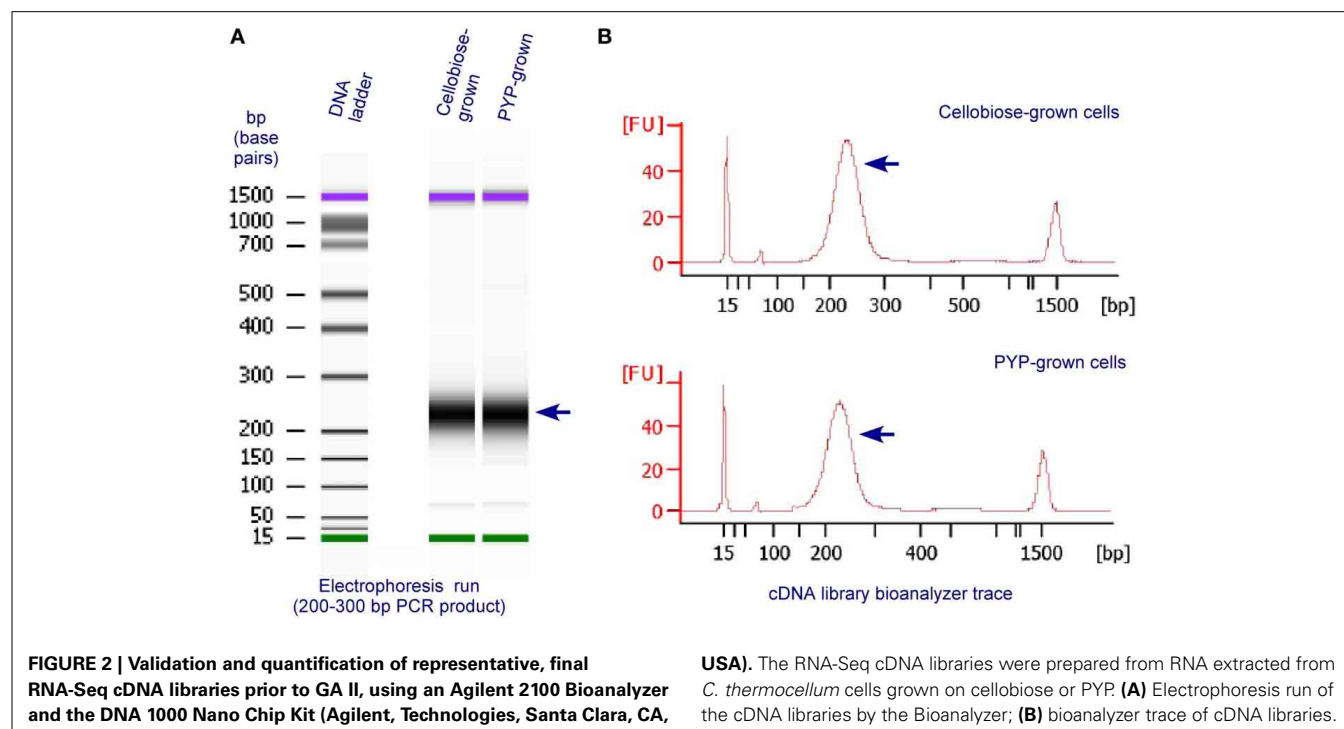
The quality of RNA-Seq cDNA libraries was assessed on a Bioanalyzer prior to GA II (**Figure 2**). The results showed the size distribution of cDNA library was between 200 and 300 bp, which met the requirement for Solexa sequencing.

For the two PYP-grown cell cDNA library samples, Solexa sequencing from the first cDNA library generated 20.8 million total raw reads, with 14.5 million passing filter reads (70% PF reads); from the second cDNA library generated 19.6 million total raw reads, with 14.1 million passing filter reads (72% PF reads). Out of these PYP-grown cell derived reads, 2.3 and 2.1 millions reads were unambiguously mapped to *C. thermocellum* genome with the criteria set in the Materials and Methods section, respectively. For the two cellobiose-grown cell cDNA library samples, Solexa sequencing from their first cDNA library generated 21.1 million total raw reads, with 15.7 million passing filter reads (74% PF reads); from the second cDNA library generated 19.7 million total raw reads, with 14.0 million passing filter reads (71% PF reads). Out of these cellobiose-grown cell-derived reads, 2.6 and 2.2 millions reads were unambiguously mapped to *C. thermocellum* genome, respectively. The RPKM value for each gene in each condition was calculated as  $\text{RPKM} = \text{Reads number}$

**Table 1 | Hydrogen, ethanol and acetate production at the late stationary phase of *C. thermocellum* culture.**

Cell culture	Yields			Acetate/Ethanol
	Hydrogen	Ethanol	Acetate	
	(Mole product/Mole glucose equivalent consumed)			
Cellobiose-grown cells	1.22 ± 0.11	0.51 ± 0.05	0.55 ± 0.07	1.08
PYP-grown cells	0.92 ± 0.09	0.30 ± 0.04	0.27 ± 0.04	0.92
Yield in PYP-grown cells/yield in cellobiose-grown cells (%)	75	58	49	

The time point for measurements was 20 and 36 h for cellobiose- and PYP-grown cells, respectively. Values for the yields of hydrogen, ethanol and acetate are means ± s.e.m. of three independent experiments. PYP, acid-pretreated yellow poplar.



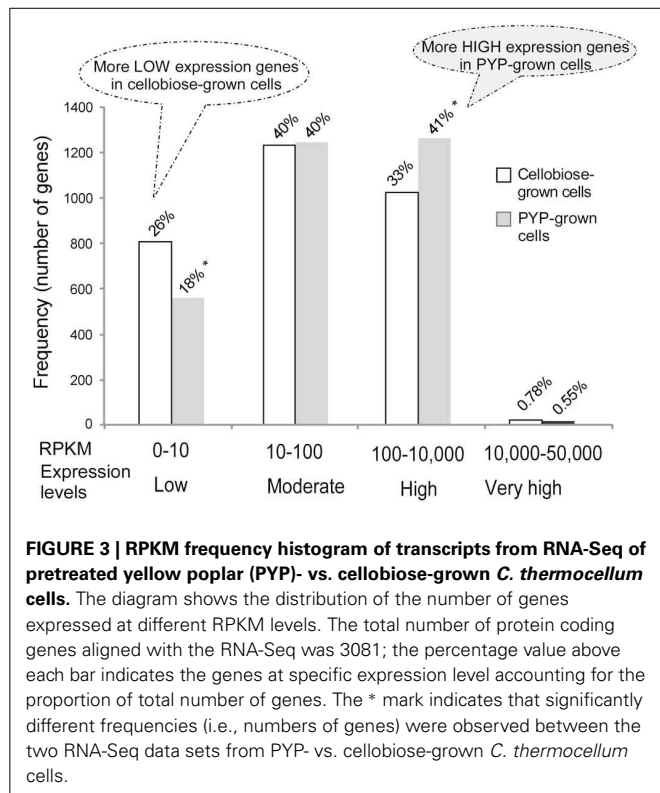
mapped to gene/Length of the gene (kb)/Total reads number (million reads), as described in the Materials and Methods section.

Overall, sequence analysis successfully aligned the Solexa reads to 3081 protein-coding genes and 61 structural RNA genes in cellobiose- and/or PYP-grown *C. thermocellum* mRNA samples, accounting for 95% [i.e., (3081 + 61)/3305] of all *C. thermocellum* genes in the genome, indicating that RNA-Seq analysis in this study achieved comprehensive coverage of the *C. thermocellum* transcriptome. In addition, the above result is consistent with the reports that in the genomes of other microorganisms such as *Saccharomyces cerevisiae* and *S. pombe*, more than 90% of the genes are transcriptionally active and expressed (Nagalakshmi et al., 2008; Wilhelm et al., 2008; Wang et al., 2009). The detailed lists of genes for the RNA-Seq sequences identified in the cellobiose- and PYP-grown cells are presented in **Supplementary Data Sheet 1**. Each of the RPKM values of cellobiose-grown cells and RPKM of PYP-grown cells was the average of two replicates.

Our data from Solexa-read transcriptome measurement of PYP- vs. cellobiose-grown *C. thermocellum* illustrate some key characteristics of the results. First, the obtained RPKM values for most of the transcripts of the active protein-coding genes in PYP- and cellobiose-grown *C. thermocellum*, are in the range of 1–50,000 (which can be revealed by re-sorting the **Supplementary Data Sheet 1**). Such range of RPKM values is comparable to the reported range of RPKM values in RNA-Seq whole-transcriptome analysis of other organism samples (Tang et al., 2009).

The gene expression levels can be classified into four levels: low, moderate, high, and very high, as illustrated in the X axis in **Figure 3**. While the RNA-Seq data sets of PYP- vs. cellobiose-grown cells had comparable numbers of genes that fall in the moderate expression levels (i.e., both 40% of the active protein-encoding genes), the PYP-grown cells had more highly expressed genes (41 vs. 33% in that of cellobiose-grown culture; **Figure 3**). In contrast, cellobiose-grown cells had more lowly expressed genes (26 vs. 18% in that PYP-grown culture).





### CHANGES IN GLOBAL GENE EXPRESSION AND CLUSTERS OF ORTHOLOGOUS GROUPS (COG) ANALYSIS

To compare the differential expression of genes between PYP- and control cellobiose-grown cells, fold changes (FC) were computed as the ratio of the RPKM values obtained for individual genes in PYP- against cellobiose-grown cells. Analysis of changes in global gene expression had identified 1211 genes that show a two-fold or greater increase in expression (i.e.,  $FC \geq 2.0$ ) or detected in PYP-grown cells only (referred as PYPO genes), and 314 genes with a two-fold or greater decrease (i.e.,  $FC \leq 0.5$ ) in transcript abundance for PYP- against cellobiose-grown cells or detected in cellobiose-grown cells only (referred as CBO genes), as listed in **Supplementary Data Sheet 2** with related statistic analyses showing that they were statistically significant.

The COG distribution for these up- and down-regulated genes in the transcriptome was determined and the result is shown in **Figure 4**. The top two categories for up- and down-regulated genes belong to the categories “general function prediction, [R]” (equivalent to unclassified) and “inorganic ion transport and metabolism, [P].” A closer examination of the distribution chart reveals that for the up-regulated genes, the categories “signal transduction mechanisms, [T],” “amino acid transport and metabolism, [E],” and “energy production and conversion, [C]” (which is important for the energy-consuming process for biofuel end-product production), were also the most well-represented categories in the transcriptome, with the number of induced genes above 200 for each category (**Figure 4**).

An important category of COG for the degradation and utilization of lignocellulosic substrate is “carbohydrate transport and

metabolism, [G],” which includes primarily cellulosome-related genes. In this category, the number of up- and down-regulated genes is 175 and 46, respectively (**Figure 4**). These differentially expressed genes are described in detail in later section.

### VALIDATION OF RNA-Seq RESULTS WITH QUANTITATIVE REVERSE TRANSCRIPTION-PCR

Quantitative reverse transcription-PCR (qRT-PCR) is a well accepted method for verifying microarray (Ferreira et al., 2010; Yang et al., 2012) and RNA-Seq data (Cusick et al., 2012; Huang et al., 2012; Ji et al., 2013). We used this method to validate the expression patterns of 12 selected genes in independent biological replicates. The selected genes mainly represented different functional categories involving in cellulosome, hydrogen and ethanol production and had a range of FC values based on RNA-Seq. In addition to these 12 genes, the gene *RecA* was chosen as a reference gene to normalize the real time RT-PCR data because it has been commonly used as reference gene for *C. thermocellum* (Stevenson and Weimer, 2005). As a further confirmation for the use of this gene as reference gene, *RecA* did not differ in expression in our cell samples grown on two substrates, the values being RPKM 1166 for PYP-grown cells vs. RPKM 1237 for cellobiose-grown cells. The primers for all 13 genes, along with one-by-one comparisons of the fold-changes in expression of each gene as measured by RNA-Seq and qRT-PCR, are listed in **Supplementary Data Sheet 3, section 1**. Most of the qRT-PCR data matched the RNA-Seq based FC values with a correlation coefficient of 0.95 for the set of 12 selected genes, which indicated that our RNA-Seq result is accurate and the conclusion from RNA-Seq should be reliable.

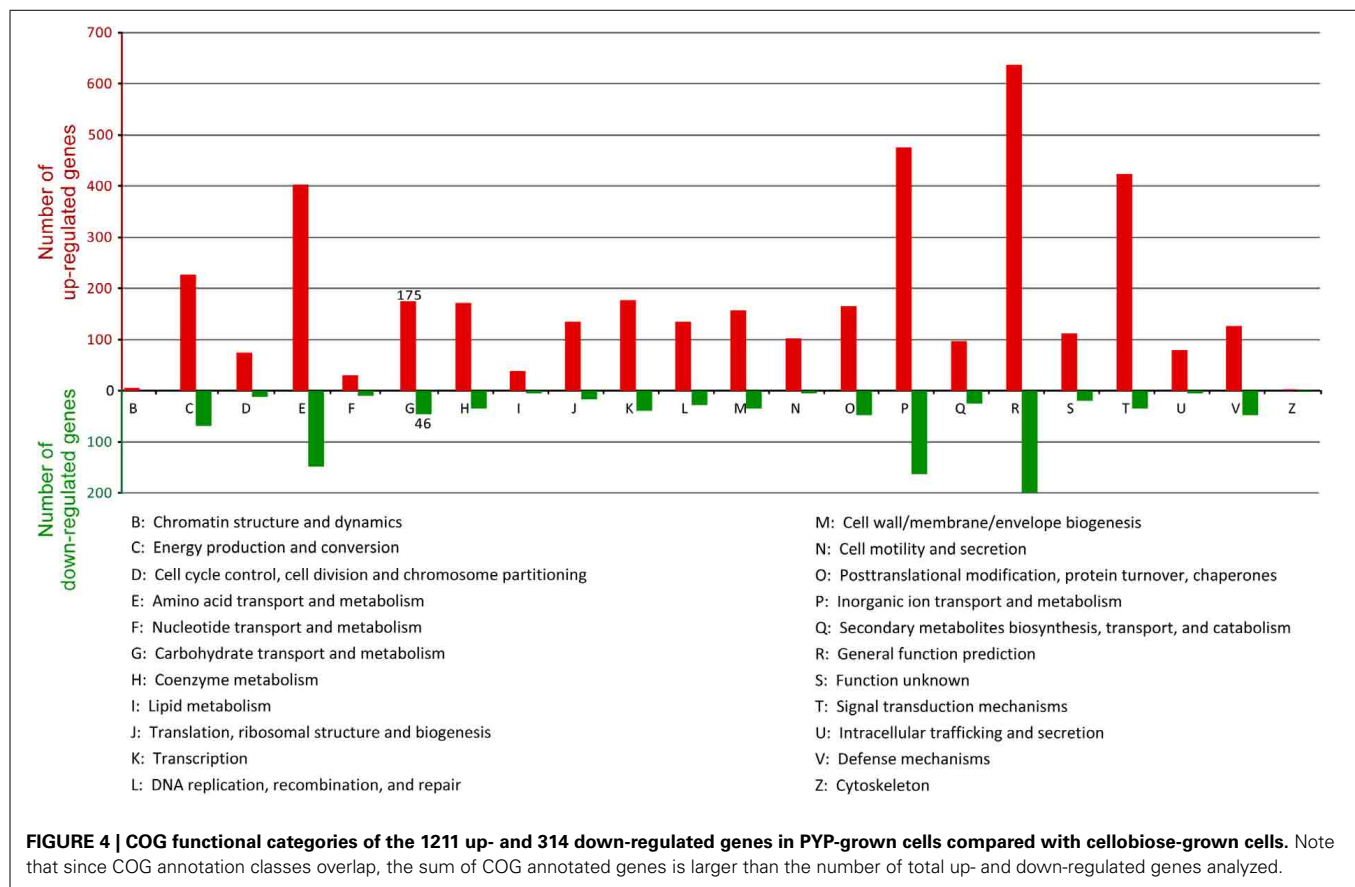
### TRANSCRIPTIONAL CHANGES OF CELLULOSOMAL COMPONENTS

Cellulosome is an extracellular supramolecular machine that can efficiently degrade crystalline cellulosic substrates and associated plant cell wall materials. We are especially interested in the genes encoding cellulosomal component proteins in response to cellulosic substrates, which can lead to the identification of regulatory or rate-limiting components regarding cellulolysis.

The number of reported cellulosomal genes in the genome of ATCC 27405 strain has been increasing from the initial numbers of 71 (Zverlov et al., 2005) and 72 (Zverlov and Schwarz, 2006), to more recently, 81 genes (Raman et al., 2011). Note that the latter number is consistent with the updated genome annotation. Of these 81 genes, all were detected transcriptionally in this study. The list of 81 cellulosomal genes with their FC values is described in **Table 2**.

Analysis of this sub-transcriptome showed the following features: first, the overall cellulosome-associated genes were up-regulated significantly in PYP- vs. cellobiose-grown cells, reflected by the facts that out of the 81 cellulosomal genes, 47 (i.e., 58%) were up-regulated with  $FC \geq 2.0$ . In contrast, only 4 out of 81 (i.e., about 5%) cellulosomal genes, including a GH5 gene (*Cthe\_2193*), *CelU* (*Cthe\_2360*), *XghA* (*Cthe\_1398*), and *Cthe\_0438*, were down-regulated by two-fold or detected in cellobiose-grown cells only. The overall average FC value for all these 81 cellulosomal genes is 3.5 (see the bottom row in **Table 2**),





suggesting the whole cellulosome machinery was “geared up” at the late stationary phase on PYP substrate.

Secondly, the primary scaffoldin, CipA, and the main secondary scaffoldins, OlpB, Orf2P, and SdbA, have shown significant up-regulation and were found to have the highest abundance on the transcriptional level. This implies that the cellulosomal system is crucial for the efficient degradation of pretreated PYP. Our data further verifies the notion that the *C. thermocellum* cellulosome is the main contributor to the extremely high activity observed in cellulose degradation. Additionally, one putative scaffoldin gene, Cthe\_0736 (cellulosome anchoring protein) has been up-regulated as much as 2.8 times, i.e., FC 2.8 (Table 2, row 31). OlpC, which has been recently identified as an important outer layer protein - cellulosome anchoring protein cohesin subunit (Pinheiro et al., 2009), was also found to be up-regulated in this study (Cthe\_0452, FC 4.1; Table 2, row 19).

Thirdly, the major cellulosomal cellulases, such as CelS (Cthe\_2089, exo-, FC 4.5; RPKM 3848—the most abundant cellulosomal transcripts in PYP-grown cells), CelA (Cthe\_0269, endo-, FC 1.7, RPKM 1885—the third most abundant transcript in PYP-grown cells), CbhA (Cthe\_0413, exo/endo, FC 7.5, RPKM 367), and Cel124A (Cthe\_0435, endo-; FC 2.0; RPKM 401) were remarkably up-regulated, as shown in Table 2. Exo- and endo- refer to the substrate site upon which the GHs act. While exo-cellulases remove one or more sugar units from the ends of

polysaccharide chain, endoglucanases randomly hydrolyze the internal glycosidic bonds of polysaccharides. Among the above listed enzymes, CelS was reported to display high synergy with the endo-Cel124A (Brás et al., 2011). Our observation that the CelS had the highest RPKM value in PYP-grown cells is consistent with a report that a knockout mutant of this gene showed a ~60% reduction in cell cellulolytic performance (Olson et al., 2010), and is also consistent with a most recent finding that CelS gene was found to be highly expressed in *C. thermocellum* cells grown on both pretreated switchgrass and cottonwood substrates (Wilson et al., 2013a). This study furthered such transcriptional analysis by showing that the transcript of CelS is not only highly abundant, but also significantly responsive to biomass substrate, as its RPKM level in cellobiose-grown cell was 4.5 times lower (Table 2).

In addition, there was one cellulosome dockerin type I gene (Cthe\_0438) for which the transcript was detected only in cellobiose-grown cells, with a low RPKM value of 4 (Table 2, row 81). This observation is consistent with a previous report regarding its absence in the sub-proteome of cellulosomes from cells grown on pretreated switchgrass (Raman et al., 2009). The domain architecture of Cthe\_0438 is “DUF843-type I dockerin” (DUF indicates domain of unknown function). It is difficult to classify the association of this gene with any of the specific catalytic enzyme types, cellulases or hemicellulases, and thus this issue may warrant further studies.

**Table 2 | List of 81 cellulosomal genes with their RPKM and fold change (FC) values determined in PYP- vs. cellobiose-grown *C. thermocellum* cells.**

Row no.	Gene ID	Protein	GH, CBM family	RPKM of PYP cells	RPKM of cellobiose cells	FC
<b>47 UP-REGULATED GENES (FC <math>\geq</math> 2.0); ALL STATISTICALLY SIGNIFICANT (P &lt; 0.05)</b>						
1	Cthe_0798	Acetyl xylan esterase	CE3	483	20	24.3
2	Cthe_0211	LicB	GH16	1624	86	18.8
3	Cthe_0912	XynY	GH10, CE1, CBM22	261	19	13.8
4	Cthe_1890	Dockerin type I cellulosome prot	n/a	45	3	14.0
5	Cthe_1472	CelH	GH5, GH26, CBM11	334	26	12.7
6	Cthe_3080	Cellulosome anchoring prot, cohesin region	n/a	1490	142	10.5
7	Cthe_0412	CelK	GH9, CBM4	1618	202	8.0
8	Cthe_0413	CbhA	GH9, CBM4	367	49	7.5
9	Cthe_0660	$\beta$ -1,3-glucanase	GH81	4	1	6.7
10	Cthe_3077	CipA	CBM3	2135	363	5.9
11	Cthe_3078	OlpB	n/a	333	59	5.6
12	Cthe_0239	Cellulosome prot, dockerin type I	n/a	50	9	5.4
13	Cthe_2872	CelG	GH5	1038	197	5.3
14	Cthe_0825	CelD	GH9	63	13	4.9
15	Cthe_3079	Orf2p	n/a	1515	338	4.5
16	Cthe_2089	CelS	GH48	3848	864	4.5
17	Cthe_0043	CelN	GH9, CBM3	122	28	4.4
18	Cthe_2147	CelO	GH5, CBM3	63	15	4.3
19	Cthe_0452	OlpC	n/a	140	34	4.1
20	Cthe_0661	Ricin B lectin	GH43, CBM13	62	17	3.8
21	Cthe_2549	Dockerin type I cellulosome prot	n/a	27	7	3.7
22	Cthe_2811	ManA	GH26, CBM	517	153	3.4
23	Cthe_1806	Cellulosome prot, dockerin type I	Dockerin type I	95	29	3.3
24	Cthe_0405	CelL	GH5	96	29	3.3
25	Cthe_0246	PL11	PL11, CBM6	119	36	3.3
26	Cthe_2950	PL1	PL1, CBM35	49	15	3.2
27	Cthe_2271	Dockerin type I cellulosome prot	n/a	60	19	3.2
28	Cthe_0745	CelW	GH9, CBM3	124	40	3.1
29	Cthe_3141	Lipolytic enzyme, G-D-S-L	CE12, CBM6	23	8	2.9
30	Cthe_1400	GH53	GH53	31	11	2.8
31	Cthe_0736	Cellulosome anchoring prot, cohesin region	n/a	122	43	2.8
32	Cthe_0536	CelB	GH5	972	352	2.8
33	Cthe_0640	Cellulosome prot, dockerin type I	n/a	50	18	2.7
34	Cthe_2761	GH9	GH9, CBM3	44	16	2.7
35	Cthe_0032	GH26	GH26	423	159	2.7
36	Cthe_2195	CBM6	CBM6	4	2	2.5
37	Cthe_2138	GH43	GH43, CBM42	13	5	2.4
38	Cthe_0258	Cellulosome prot, dockerin type I	n/a	73	31	2.3
39	Cthe_2196	GH43	GH43, CBM6	2	1	2.3
40	Cthe_3012	GH30	GH30, CBM6	106	47	2.3
41	Cthe_2139	GH30, GH43, GH54	GH30, GH43, GH54	13	6	2.3
42	Cthe_2590	XynD	GH10, CBM22	148	67	2.2
43	Cthe_0015	Alpha-L-arabinofuranosidase B	GH43, GH54	203	92	2.2
44	Cthe_2038	Cellulosome prot, dockerin type I	n/a	48	22	2.1
45	Cthe_1307	SdbA	n/a	207	101	2.1
46	Cthe_2197	GH2	GH2, CBM6	2	1	2.0
47	Cthe_0435	Cel124A	GH124	401	205	2.0
<b>30 GENES THAT DEFINED AS UNCHANGED (2 &gt; FC &gt; 0.5)</b>						
48	Cthe_1838	XynC	GH10, CBM22	746	405	1.8
49	Cthe_0191	Proteinase inhibitor I4, serpin	n/a	16	9	1.8
50	Cthe_2194	CE1	CE1, CBM6	12	7	1.8

(Continued)

Table 2 | Continued

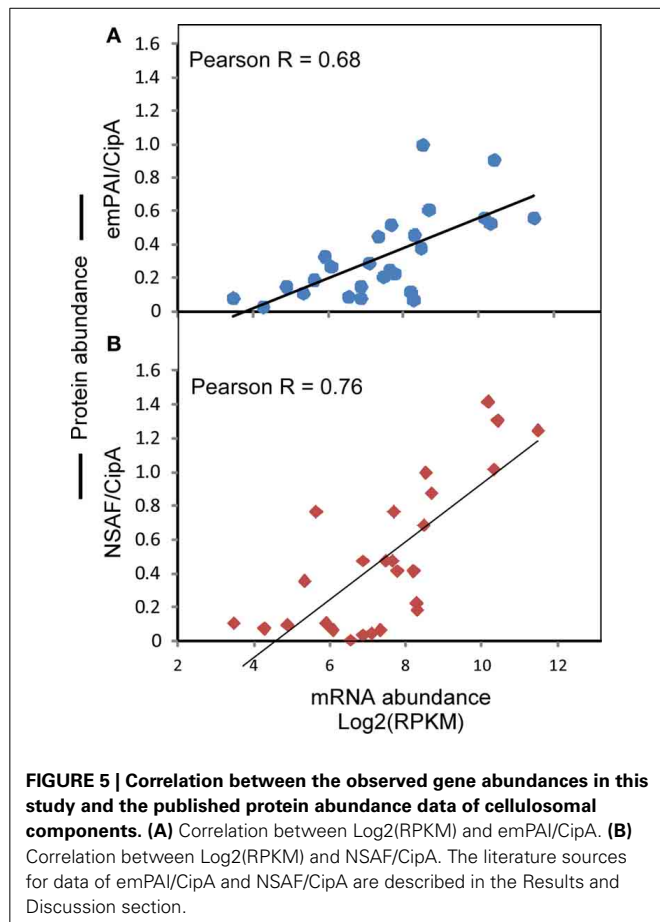
Row no.	Gene ID	Protein	GH, CBM family	RPKM of PYP cells	RPKM of cellobiose cells	FC
51	Cthe_2760	CelV	GH9, CBM3	162	92	1.8
52	Cthe_0190	Proteinase inhibitor I4, serpin	n/a	20	12	1.8
53	Cthe_3136	Peptidase S8/S53 subtilisin kexin sedolisin	n/a	36	22	1.7
54	Cthe_0269	CelA	GH8	1885	1136	1.7
55	Cthe_0274	CelP	GH9	116	73	1.6
56	Cthe_0543	CelF	GH9, CBM3	270	175	1.5
57	Cthe_1271	GH43	GH43, CBM6	168	116	1.5
58	Cthe_1963	XynZ	GH10, CE1, CBM6	532	390	1.4
59	Cthe_0109	Cellulosome prot, dockerin type I	n/a	24	18	1.3
60	Cthe_0821	Coagulation factor 5/8 type-like	GH5	1694	1266	1.3
61	Cthe_2949	Pectinesterase	CE8	10	8	1.3
62	Cthe_2812	CelT	GH9	342	291	1.2
63	Cthe_0729	Cellulosome prot, dockerin type I	n/a	66	59	1.1
64	Cthe_0735	Cellulosome anchoring prot, cohesin region	n/a	141	127	1.1
65	Cthe_0578	CelR	GH9, CBM3	238	216	1.1
66	Cthe_2179	Pectate lyase/Amb allergen	PL1, PL9, CBM35	23	22	1.1
67	Cthe_0433	GH9	GH9, CBM3	308	306	1.0
68	Cthe_0918	Cellulosome prot, dockerin type I	n/a	50	50	1.0
69	Cthe_0270	ChiA	GH18	85	92	0.9
70	Cthe_2972	XynA/U	GH11, CE4, CBM4	1212	1357	0.9
71	Cthe_2137	Cellulosome prot, dockerin type I	GH39, CBM35	7	8	0.9
72	Cthe_0624	CelJ	GH9, GH44, CBM30	93	116	0.8
73	Cthe_3132	Cellulosome prot, dockerin type I	n/a	63	79	0.8
74	Cthe_0625	CelQ	GH9, CBM3	190	248	0.8
75	Cthe_2879	Cellulosome prot, dockerin type I	CE6	13	18	0.7
76	Cthe_0044	Cellulosome prot, dockerin type I	n/a	17	24	0.7
77	Cthe_0797	CelE	GH5, CE2	180	311	0.6
<b>3 DOWN-REGULATED GENES (FC ≤ 0.5); ALL STATISTICALLY SIGNIFICANT (P &lt; 0.05)</b>						
78	Cthe_2193	GH5	GH5, CBM6	67	135	0.5
79	Cthe_2360	CelU	GH9, CBM3	13	32	0.4
80	Cthe_1398	XghA	GH74	989	2817	0.4
<b>1 GENE DETECTED IN CELLOBIOSE-GROWN CELLS ONLY (CBO)</b>						
81	Cthe_0438	Cellulosome prot, dockerin type I	n/a	0	4	
<b>Average FC value for all genes (except the CBO)</b>						<b>3.5</b>

The table is sorted in the order of FC values. The glycoside hydrolase (GH) and carbohydrate-binding module (CBM) families were determined according to the annotations in GenBank and the CAZY database ([www.CAZY.org](http://www.CAZY.org)). Other abbreviations: Cbh, cellobiohydrolase; CBO, transcript that detected only in cellobiose-grown cells; CE, carbohydrate esterases; Cel, cellulase; Chi, chitinase; CipA, cellulose-integrating protein A; Lic, lichenase; Man, mannanase; n/a, not applicable; OlpB, outer layer protein; ORF2p, open reading frame 2p; PL, polysaccharide lyase; Prot, protein; SdbA, scaffoldin-dockerin binding component A; XghA, xyloglucanase Xgh74A; Xyn, xylanase.

### EXPLORING THE CORRELATION BETWEEN RNA-Seq AND PUBLISHED PROTEOMIC DATA FOR CELLULOSOMAL GENES

Literature reports have described the quantitative sub-proteomic data of cellulosomes extracted from cell-free culture filtrates of *C. thermocellum* grown to late stationary phase on cellobiose, Avicel, and other cellulosic substrates (Gold and Martin, 2007; Raman et al., 2009). In one study (Gold and Martin, 2007), an “emPAI,” defined as the exponentially modified protein abundance index, showed a linear relationship with protein concentration and was normalized to the value obtained for CipA. Similarly, in another study (Raman et al., 2009), the normalized spectral abundance factor (NSAF) represented the number

of spectral counts divided by the number of amino acid residues in the protein and was also normalized to the value obtained for CipA. To determine whether or not a correlative relationship exists between these published cellulosomal sub-proteomic data and the RNA-Seq RPKM data in the present study, we first retrieved the emPAI/CipA and NSAF/CipA data from the literature (**Supplementary Data Sheet 3, section 2**), and then plotted it against the Log2(RPKM) data for the genes encoding the same proteins in this study. The results showed that despite the fact that the plotted data were obtained from three different research groups, there remain strong correlations between the RNA-Seq RPKM and the protein abundance indicators of emPAI/CipA and



NSAF/CipA, with Pearson correlation coefficient values being 0.68 and 0.76, respectively (Figures 5A,B).

### NON-CELLULOSOMAL GLYCOSIDE HYDROLASE PROTEINS AND “FREE” CELLULASES

In addition to the GHs existing in cellulosomes, the *C. thermocellum* genome includes non-cellulosomal genes coding for GHs, among which 12 were up-regulated by more than two-fold (Table 3). Notably, both CelC (Cthe\_2807) and Lic16A (Cthe\_2809), the two members in the putative CelC- GlyR3- Lic16A operon (Newcomb et al., 2011), were up-regulated and predicted to be extracellular and/or bacterial cell wall associated (Table 3). This was consistent with a recent report that genes in this operon were mainly expressed in the stationary phase (very little during exponential phase) of pure cellulose (Avicel) fermentation (Raman et al., 2011). In addition, Lic16A contains CBM54 and a tandem of four CBM4s, in which both types of CBMs are able to bind xylan; as well as cellulose (Dvortsov et al., 2010). Remarkably, the grouping of four CBM4s in this protein has an ~100-fold higher binding constant for xylan and cellulose than that of a protein with a single CBM4 module (Dvortsov et al., 2012).

It is surprising that Cel9I (Cthe\_0040), an important non-cellulosomal processive endoglucanase that could digest crystalline cellulose with high efficiency, showed low abundance

and FC value (RPKM 62 in PYP-grown cells, FC 1.3; see **Supplementary Data Sheet 1**, row no. 1875). Similarly, the only non-cellulosomal exo-cellulase CelY (Cthe\_0071, exo-) also showed both low abundance and FC value (RPKM 61, FC 1.4, see **Supplementary Data Sheet 1**, row no. 1741), which is consistent with a recent report that knocking out the CelY gene had no significant impact on the cellulolytic capacity of the strain (Olson et al., 2010). Based on the evidence above, it would seem that the free-enzyme system plays a less important role than does the cellulosome system in the degradation of PYP by *C. thermocellum*.

### RESPONSIVE SIGMA FACTORS COUPLED WITH MEMBRANE-ASSOCIATED ANTI-SIGMA FACTORS AS A MECHANISM FOR BIOSENSING BIOMASS SUBSTRATES

Because COG analysis had revealed that the category of “signal transduction mechanisms, [T]” was among the most represented categories in the transcriptome of PYP-grown cells, we checked the genes related to the signal transmission attributed to sensing cellulose and other polysaccharide substrates. Recently, it was proposed the possible role of seven membrane-associated RsgI-like anti-sigma factors (referred to as RsgI) and their sigma factor partners (referred as SigI) in extracellular carbohydrate-sensing and glycosidase gene regulation in *C. thermocellum* (Kahel-Raifer et al., 2010; Nataf et al., 2010). In their proposed model, RsgI senses the presence of cellulose and other biomass components in the extracellular medium via its CBM domains, whereas SigI mediates the intracellular activation of different glycosidase genes (Kahel-Raifer et al., 2010; Nataf et al., 2010; Bahari et al., 2011). These predicted gene pairs include: SigI1- RsgI1 (Cthe\_0058- Cthe\_0059),  $\sigma$ I2-RsgI2 (Cthe\_0268- Cthe\_0267),  $\sigma$ I3-RsgI3 (Cthe\_0315- Cthe\_0316),  $\sigma$ I4-RsgI4 (Cthe\_0403- Cthe\_0404),  $\sigma$ I5-RsgI5 (Cthe\_1272- Cthe\_1273),  $\sigma$ I6-RsgI6 (Cthe\_2120- Cthe\_2119), and  $\sigma$ 24C-Rsi24C (Cthe\_1470- Cthe\_1471). Among them, three SigI genes (Cthe\_0058, Cthe\_0268, Cthe\_0403) were up-regulated during later stages of pure cellulose (Avicel) fermentation (Raman et al., 2011).

In this study, while the transcripts of all seven pairs of genes had been detected (see **Supplementary Data Sheet 1**), three pairs of genes and an extra SigI-RsgI pair (which was not included in the initial literature prediction shown above) were up-regulated with FC values above 3.0 in the PYP- against cellobiose-grown cells at the late stationary phase, as described below:

- (1)  $\sigma$ I3-RsgI3 (Cthe\_0315- Cthe\_0316), with FC 4.2, and 5.9, respectively. According to literature, RsgI3 contains a PA14 dyad domain that target pectin (Nataf et al., 2010). The differential expression of this pair was not previously reported in cellulose-grown *C. thermocellum*.
- (2)  $\sigma$ I4-RsgI4 (Cthe\_0403- Cthe\_0404), with FC 3.1, and 3.9, respectively. RsgI4 contains CBM3 that binds cellulose. This observation supports previous observation of this pairs being involved in the time course of pure cellulose fermentation by *C. thermocellum* (Raman et al., 2011).
- (3)  $\sigma$ 24C-Rsi24C (Cthe\_1470- Cthe\_1470), with FC 4.0 and 5.9, respectively. Note that Rsi24C contains a module that resembles GH5 that target cellulose based on literature (Nataf et al.,

**Table 3 | Twelve non-cellulosomal GH genes that were up-regulated by two or more folds in pretreated yellow poplar (PYP)- compared with cellobiose-grown *C. thermocellum* cells.**

Gene ID	Protein and GH family <sup>a</sup>	FC	Cellular location prediction <sup>b</sup>
Cthe_1256	GH3; bgIB; $\beta$ -glucosidase B	28.6	n/a
Cthe_1428	GH1; bgI	9.1	Cytoplasmic (Psortb)
Cthe_2744	GH23; lytic transglycosylase	7.5	Membrane (UniProt)
Cthe_2809	GH16, CBM4, CBM54; $\beta$ -1,3-1,4-glucanase, Lic16A	6.0	Cell wall (Psortb) Extracellular (Psortb)
Cthe_1471	GH5; Rsi24C	5.9	Membrane (Bahari et al., 2011)
Cthe_2807	GH5; CelC, endoglucanase C	5.0	Extracellular (Psortb)
Cthe_0043	GH9, CBM3; CelN, Cel9N	4.4	Membrane (Psortb)
Cthe_0322	GH3	4.3	Extracellular (UniProt)
Cthe_2191	$\alpha$ -glucan branching enzyme; GH13	3.3	Membrane (Burton, 2010)
Cthe_2548	GH51, $\alpha$ -N-arabinofuranosidase	3.3	Cytoplasmic (UniProt)
Cthe_1221	GH94, GT84; glycosyltransferase	3.0	Membrane (Psortb)
Cthe_0795	GH13, CBM34; $\alpha$ -amylase, catalytic region	2.2	Cytoplasmic (Psortb)

<sup>a</sup>GH family characterization was based on CAZy database (<http://www.cazy.org/>).

<sup>b</sup>The cellular location was based on three sources: literature, UniProt database (<http://www.uniprot.org/>), or prediction by Psortb program (<http://www.psort.org/psortb/index.html>). All listed genes were statistically significantly up-regulated ( $p < 0.05$ ). n/a, not available.

2010). Similar to  $\sigma$ I3-RsgI3, the differential expression of this pair was also not previously reported in cellulose-grown *C. thermocellum*.

- (4) Equally interesting, we found that another SigI-RsgI pair (Cthe\_2521-Cthe\_2522) was detected on transcriptome, with FC values of 3.9 and 4.8, respectively (Supplementary Data Sheet 1). The protein of sigma factor Cthe\_2521 had been detected in the proteome of *C. thermocellum* cultured on cellobiose (Rydzak et al., 2012), but was not in the initial 7 pairs of SigI-RsgI proposed for polysaccharide signal transmission in the literature (Nataf et al., 2010). The possible role of this pair warrants further investigation.

## GENES RELATED TO CELLODEXTRIN TRANSPORT AND PHOSPHORYLATION

*C. thermocellum* has been reported to use ABC-type transporters for uptake of oligosaccharides derived from cellulose hydrolysis (Strobel et al., 1995), which is an important energy-conserving mechanism by which importing long cellooligosaccharides can reduce the cost of transport as one-ATP molecule is consumed per transport event. So far, four cellooligosaccharide ABC transporters (carbohydrate-binding protein CbpA, B, C, and D) have been characterized for their substrate binding features (Nataf et al., 2009). Whereas CbpA (Cthe\_0393) binds only to cellobiose (G3), CbpB (Cthe\_1020) binds to G2-G5 cellooligosaccharides, and CbpC (Cthe\_2128) and D (Cthe\_2446) bind to G3-G5 cellooligosaccharides (Nataf et al., 2009; Rydzak et al., 2012). Several transcripts of these annotated genes have been detected in the transcriptome of *C. thermocellum* (Raman et al., 2011; Riederer et al., 2011). Previously, six cellooligosaccharide ABC transporter genes (Cthe\_1862, Cthe\_0391-0393, and 1019-1020, including CbpA and CbpB) were found to be expressed at high levels throughout the course of Avicel alone fermentation (Raman et al., 2011). Most recently, Cthe\_0391-0393 were found to be highly expressed on both pretreated switchgrass and cottonwood substrates (Wilson

et al., 2013a). This study furthered such transcriptional analysis by comparing the gene expression between PYP- and cellobiose-grown cells at late stationary phase. A total of 12 transcripts of cellooligosaccharide ABC transporters have been detected (Supplementary Data Sheet 3, section 3; also Figure 6), among which, seven were up-regulated in PYP- against cellobiose-grown cells, with FC values in the range of 3.1–10.3, suggesting that the PYP-grown cells had a mechanism for enhancing the uptake and utilization of polysaccharides derived from biomass substrates.

For the subsequent phosphorylation of the imported oligosaccharides and cellobiose, this study identified the transcripts of one cellooligosaccharide phosphorylase (CDP, Cthe\_2989) and two cellobiose phosphorylases (CEP, Cthe\_0275 and Cthe\_1221). Among these, the CEP Cthe\_1221 was up-regulated with a FC value of 3.0 (Supplementary Data Sheet 3, section 3; also Figure 6), suggesting this gene may warrant further studies.

## GENES RELATED TO GLYCOLYSIS, PYRUVATE CATABOLISM, AND END-PRODUCT SYNTHESIS

The deduced pathway for cellulolysis, glycolysis, ethanol, and  $H_2$  production in *C. thermocellum* ATCC 27405 is illustrated in Figure 6, which is in accordance to accumulated literature findings (Demain et al., 2005; Rydzak et al., 2009, 2011; Riederer et al., 2011; Carere, 2013). The set of gene IDs for the enzymes involved in above pathway were retrieved from the KEGG PATHWAY database (<http://www.genome.jp/kegg>) (Kanehisa et al., 2008), and were shown in Supplementary Data Sheet 3, section 3 with their FC values in PYP- vs. cellobiose-grown cells. Out of the listed genes, 18 and 12 were significantly up- and down-regulated, as shown in red and green text in Supplementary Data Sheet 3, section 3, respectively.

For pyruvate catabolism and end-product synthesis, *C. thermocellum* may convert pyruvate into (1) formate via pyruvate formate lyase (PFL) and pyruvate formate lyase activating enzyme (PFL-AE), (2) lactate via lactate dehydrogenase (LDH), (3)  $CO_2$ ,





**Table 4 | *C. thermocellum* genes encoding putative hydrogenases, sensory hydrogenases, and NADH:Fd oxidoreductases using ferredoxin and NAD(P)H as electron carriers.**

Locus ID	Gene name	FC
<b>TYPE 1. [NiFe] H<sub>2</sub>ase</b>		
<b>Fd-dependent ECH: Cthe_3013-Cthe_3024</b>		
Cthe_3013	Hydrogenase expression/formation protein HypE	1.6
Cthe_3014	Hydrogenase formation HypD protein	1.4
Cthe_3015	Hydrogenase assembly chaperone hypC/HupF	0.8
Cthe_3016	[NiFe] hydrogenase maturation protein HypF	0.8
Cthe_3017	Hydrogenase accessory protein HypB	0.9
Cthe_3018	Hydrogenase expression/synthesis HypA	0.6
Cthe_3019	4Fe-4S ferredoxin, iron-sulfur binding	0.5
Cthe_3020	Ech hydrogenase subunit E (Fd H <sub>2</sub> ase)	0.3
Cthe_3021	Ech hydrogenase subunit D	0.4
Cthe_3022	Ech hydrogenase subunit C	1.5
Cthe_3023	Ech hydrogenase subunit B	0.8
Cthe_3024	Ech hydrogenase subunit A	0.5
<b>TYPE 2. [FeFe] H<sub>2</sub>ases</b>		
<b>Type 2.1. Bifurcating: Cthe_0338-Cthe_0342; Cthe_0428-Cthe_0430</b>		
Cthe_0338	NADH-dependent hydrogenases	0.9
Cthe_0340	Ferredoxin	0.9
Cthe_0341	NADH dehydrogenase (ubiquinone)	1.2
Cthe_0342	[FeFe] H <sub>2</sub> ase (NADH H <sub>2</sub> ase)	1.3
Cthe_0428	NADH dehydrogenase (ubiquinone)	6.4
Cthe_0429	NADH dehydrogenase (ubiquinone)	4.4
Cthe_0430	[FeFe] H <sub>2</sub> ase	7.9
<b>Type 2.2. Sensory: Cthe_0425-Cthe_0426</b>		
Cthe_0425	Putative hydrogenase	9.9
Cthe_0426	[FeFe] hydrogenase	11.0
<b>Type 2.3. NAD(P)H dependent: Cthe_3003-3004</b>		
Cthe_3003	[FeFe] H <sub>2</sub> ase, putative <sup>a</sup>	1.5
Cthe_3004	Ferredoxin	0.6
<b>OTHER RELATED GENES</b>		
Cthe_0335	Hydrogenase large subunit-like protein	3.2

In addition, hydrogenase maturation proteins are also listed. The table is arranged according to the hydrogenase classification described by two literature sources, as cited in the related Results and Discussion section; within each sub-category, it was arranged in order of locus IDs. For each gene, fold change (FC) value was calculated by dividing the RPKM of PYP cells by RPKM of cellobiose cells, as listed in **Supplementary Data Sheet 1**. RPKM, reads per kilobase of exon model per million mapped reads. Text in black represents the genes with no significant transcriptional changes between the two types of cells, i.e.,  $2.0 > FC \text{ value} > 0.5$ . Text in red and green represents the genes that found to be significantly up-regulated and down-regulated, respectively, in PYP- against cellobiose-grown cells, based on statistic analysis in **Supplementary Data Sheet 2**. ECH, energy conserving hydrogenase. PYPO, transcript that detected in PYP-grown cells only.

<sup>a</sup>Cthe\_3003 ([FeFe] H<sub>2</sub>ase, putative) was characterized as Fd-linked by Schut and Adams (2009), but as NAD(P)H dependent recently by Carere et al. (2012).

Fd<sub>red</sub>, and acetyl-CoA, in which acetyl-CoA eventually leads to the production of acetate and ethanol, the Fd<sub>red</sub> leads to the formation of hydrogen (**Figure 6**). The up-regulation of two out of four alcohol dehydrogenase (ADH) genes, namely Cthe\_0423 (FC value 7.0) and Cthe\_0394 (FC value 5.3), raise the average FC value for ADHs to 3.6. In contrast, the phosphotransacetylase (PTA; Cthe\_1029), whose activity had been verified in *C. thermocellum* (Lamed and Zeikus, 1980), was significantly down-regulated in PYP- against cellobiose-grown cells, with an FC value of 0.2. These data suggest that based on mRNA profiling, carbon flux of acetyl-CoA may preferentially be channeled to ethanol production in PYP-grown cells. The observed decrease of acetate/ethanol ratio, from 1.08 in cellobiose-grown cells to 0.92 in PYP-grown cells (**Table 1**), supports this carbon shift at the late stationary phase of PYP-grown cells.

Lactate and formate were not monitored in this study as previous study has shown that in the late growth phase of *C. thermocellum* cultures grown on cellobiose substrate, lactate, and formate together represent only a small fraction of the total end products produced (Islam et al., 2006). Future studies should also monitor the production of lactate, formate and CO<sub>2</sub>, which could provide a whole picture for carbon balance for the cells utilizing biomass substrates.

#### DYNAMICS FOR H<sub>2</sub> PRODUCTION WITH PYP AS SUBSTRATE

*C. thermocellum* genes encoding putative hydrogenases and sensory hydrogenases using ferredoxin and NAD(P)H as electron carriers are listed **Table 4**. In addition, hydrogenase maturation proteins are also listed. The classification of the above hydrogenases are mainly based on literatures that systematically characterized the putative hydrogenases in *C. thermocellum* and other species (Schut and Adams, 2009; Carere et al., 2012). Briefly, there are two types of hydrogenases (H<sub>2</sub>ases) according to the metal content in the respective active sites:

The first type is [NiFe] H<sub>2</sub>ase, a putative Fd-dependent energy converting hydrogenase (ECH). Its hexameric structural subunits are encoded by Cthe\_3019-Cthe\_3024, with assembly of its active site assisted by a suite of maturation proteins encoded by Cthe\_3013-Cthe\_3018 (**Table 4**, **Figure 6**). This study showed that the FC values for 7 of the 12 detected Fd-H<sub>2</sub>ase and related genes (Cthe\_3013-Cthe\_3015, Cthe\_3017-Cthe\_3018, and Cthe\_3022-Cthe\_3023) were in the range of 0.6–1.6, suggesting an unchanged status at the transcriptional level for these genes. However, the remaining five genes (Cthe\_3019-Cthe\_3021, Cthe\_3023-Cthe\_3024) encoding the Fd-H<sub>2</sub>ase structural subunits were significantly down-regulated with FC values between 0.3 and 0.5, which might account for the observed lower H<sub>2</sub> yield in PYP-grown cells, compared to the control cellobiose-grown cells. This finding implies that the [NiFe] ECH H<sub>2</sub>ase likely functions in the H<sub>2</sub> production direction since its down-regulation led to less H<sub>2</sub> accumulation in the culture headspace.

The second type is [FeFe] H<sub>2</sub>ases that include two bifurcating H<sub>2</sub>ases: both being trimeric, and encoded by Cthe\_0338-Cthe\_0342 and Cthe\_0428-Cthe\_0430, respectively. Only seven genes are listed in **Table 4** because Cthe\_0339 is annotated as

histidine kinase in GenBank protein database. The second type also includes a sensory H<sub>2</sub>ase (Cthe\_0425-Cthe\_0426) and a NAD(P)H dependent H<sub>2</sub>ase (Cthe\_3003-Cthe\_3004), as listed in **Table 4**. The data showed that the FC values for the trimeric bifurcating H<sub>2</sub>ase genes (Cthe\_0338, Cthe\_0340-Cthe\_0342) and the NAD(P)H dependent H<sub>2</sub>ase genes (Cthe\_3003-3004) were in the range of 0.6–1.5, indicating an unchanged status for their transcription (**Figure 6**). A homolog of this trimeric H<sub>2</sub>ase has been uncovered in *Acetobacterium woodii* which functions in H<sub>2</sub> consumption yielding reduced Fd and NAD(P)H (Schut and Adams, 2009; Hess et al., 2013a,b). In contrast, genes encoding the trimeric bifurcating H<sub>2</sub>ase (Cthe\_0428-Cthe\_0430) were significantly up-regulated in PYP-grown cells, with FC values of 6.4–7.9 (**Figure 6**). Based on literature (Schut and Adams, 2009; Calusinska et al., 2010), the trimeric [FeFe] H<sub>2</sub>ase identified in *C. thermocellum* is a putative bifurcating hydrogenase. This cytoplasmic enzyme was initially characterized in *Thermotoga maritima* and uses both reduced ferredoxin and NAD(P)H as substrates, and functions in H<sub>2</sub> production (Schut and Adams, 2009; Hess et al., 2013a). Yet the up-regulation of Cthe\_0428-Cthe\_0430 (this work) led to decreased H<sub>2</sub> production when cultured in PYP- vs. cellobiose-grown cells. As such, the exact role of the *C. thermocellum* trimeric H<sub>2</sub>ase is unknown, which makes it hard to link the above transcriptional data to the observed lower H<sub>2</sub> yield in PYP-grown cells. Further studies are warranted to investigate the exact direction (production or consumption of H<sub>2</sub>) for each of the above [FeFe] H<sub>2</sub>ases.

In addition, the sensory H<sub>2</sub>ase encoded by Cthe\_0425-Cthe\_0426 had a FC value of 9.9 and 11.0, respectively (**Table 4**), and future studies are also needed to explore the implication of the significant up-regulation of these two genes in PYP-grown cells.

#### GENES FOR DEGRADING HEMICELLULOSE AND LIGNIN IN FEEDSTOCK BIOMASS

We noticed that some important hemicellulase genes, such as XynY (Cthe\_0912, FC 13.8), XynD (Cthe\_2590, FC 2.2), XynZ (Cthe\_1963, FC 1.4), ManA (Cthe\_2811, FC 3.4) were up-regulated in PYP- against cellobiose-grown cells (**Table 2**; cellosomal proteins). In addition, some auxiliary enzymes such as  $\alpha$ -N-arabinofuranosidase (Cthe\_2548, FC 3.3; **Table 3**, non-cellosomal protein) were also up-regulated. This is probably caused by the adaptation of the strain to digest the PYP, of which the chemical composition and structure are much different from those of cellobiose. In contrast, we did not identify the primary genes involved in lignin modification and degradation, such as laccases, lignin peroxidases, or manganese peroxidases.

#### GROWTH PHASE AND THE TRANSCRIPTOME OF CELLS

The main focus of this study was to investigate the effects of different carbon sources (PYP vs. cellobiose) on the transcriptome of the cells. To attain this goal, the selection of sampling point was crucial. The experimental design of this study in choosing the late stationary phase is in accordance with literature (Gold and Martin, 2007; Raman et al., 2009). One merit for choosing this sampling point is that our transcriptomic data can be cross-analyzed with the literature's protein data of

cellulosomes extracted from cellobiose-grown cells, which are used as one means of validating the transcriptomic data in this study.

Another merit in choosing this sampling point is that compared with other early growth stages (exponential or early stationary phases), the levels of simple, soluble sugars in PYP- and cellobiose-grown culture are both very low: due to either the limit of bacterial enzymes in releasing sugars from PYP or the depletion of cellobiose, respectively. In this sense, late stationary phase is relatively more suitable phase than other phases for comparing transcriptomes derived from different carbon substrate-grown cells in this study.

Nevertheless, it is noteworthy that different growth phases may cause some shifts to some metabolic pathways; thus future studies comparing cells grown on cellobiose vs. cells grown on woody biomass at the exponential phase would provide another angle to investigate the effects of biomass substrate on the transcriptomes of *C. thermocellum* cells, which may lead to the identification of both growth phase- and biomass substrate-responsive genes. In addition, the transcriptional profiles of these two cell samples could also be affected by the distinct concentrations of carbohydrate sources as well as the residual concentrations of metabolites/nutrients present at time of sampling, which will remain a challenge for future studies in managing these variables—some of which are nearly uncontrollable.

#### CONCLUSIONS

We conducted a RNA-Seq analysis of the transcriptional profiles of *C. thermocellum* grown on PYP and cellobiose, which is different from previous transcriptional studies that focused on degrading cellulose alone substrates, or sub-proteomic studies that focused on cellosomal protein components. We found that nearly 60% of the genes encoding the protein components of the cellulosomes—the core machinery for cellulose degradation—were up-regulated, whereas only 5% were down-regulated. The top up-regulated cellosomal genes, along with the responsive SigI-RsgI and cellobiose transporter genes, present promising candidate genes for engineering *C. thermocellum* strains to improve their capacity in effectively converting lignocellulosic biomass substrates. Furthermore, the identified differentially expressed NAD(P)H H<sub>2</sub>ase, ADH, and PTA genes may provide insight into how the cells regulate the production of H<sub>2</sub> and ethanol under the carbon-limited condition.

#### ACKNOWLEDGMENTS

This work was funded by the Laboratory Directed Research and Development (LDRD) program at the National Renewable Energy Laboratory (NREL), and by the U.S. Department of Energy's Office of Science through the BioEnergy Science Center (BESC), Bioenergy Technology Office (DOE-BETO) and the Fuel Cell Technologies Office. This work was supported by DOE under Contract No. DE-AC36-08-GO28308 with NREL. Andrew Bowersox and Igor Bogorad were sponsored by DOE Academies Creating Teacher Scientists (ACTS) and Science Undergraduate Laboratory Internship (SULI) programs, respectively. We thank Dr. Katherine Chou of NREL for discussion.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fmicb.2014.00142/abstract>

**Supplementary Data Sheet 1 | List of 3081 genes with their RPKM and fold change values determined in PYP- vs. cellobiose-grown *C. thermocellum* cells.**

**Supplementary Data Sheet 2 | List of 1211 up-regulated and 314 down-regulated genes in PYP-grown *C. thermocellum* cells.**

**Supplementary Data Sheet 3 | Section 1.** Forward (F) and reverse (R) primer sequences for quantitative reverse transcription-PCR (qRT-PCR) analysis of selected genes in *C. thermocellum*. **Section 2.** RPKM values of cellulosomal genes based on this study and their protein abundance data retrieved from literature for cellobiose-grown *C. thermocellum* cells at the late stationary phase. **Section 3.** Fold change values for genes related to cellobiose transport and phosphorylation, glycolysis, pyruvate catabolism and end-product synthesis in *C. thermocellum*.

## REFERENCES

- Alper, H., and Stephanopoulos, G. (2009). Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential? *Nat. Rev. Microbiol.* 7, 715–723. doi: 10.1038/nrmicro2186
- Bahari, L., Gilad, Y., Borovok, I., Kahel-Raifer, H., Dassa, B., Nataf, Y., et al. (2011). Glycoside hydrolases as components of putative carbohydrate biosensor proteins in *Clostridium thermocellum*. *J. Ind. Microbiol. Biotechnol.* 38, 825–832. doi: 10.1007/s10295-010-0848-9
- Baker, J. O., Vinzant, T. B., Ehrman, C. I., Adney, W. S., and Himmel, M. E. (1997). Use of a new membrane-reactor saccharification assay to evaluate the performance of celluloses under simulated ssf conditions. *Appl. Biochem. Biotechnol.* 63, 585–595. doi: 10.1007/BF02920456
- Bayer, E. A., Belaich, J. P., Shoham, Y., and Lamed, R. (2004). The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. *Annu. Rev. Microbiol.* 58, 521–554. doi: 10.1146/annurev.micro.57.030502.091022
- Bayer, E. A., Chanzy, H., Lamed, R., and Shoham, Y. (1998). Cellulose, cellulases and cellulosomes. *Curr. Opin. Struct. Biol.* 8, 548–557. doi: 10.1016/S0959-440X(98)80143-7
- Bayer, E. A., Lamed, R., and Himmel, M. E. (2007). The potential of cellulases and cellulosomes for cellulosic waste management. *Curr. Opin. Biotechnol.* 18, 237–245. doi: 10.1016/j.copbio.2007.04.004
- Brás, J. L., Cartmell, A., Carvalho, A. L. M., Verzé, G., Bayer, E. A., Vazana, Y., et al. (2011). Structural insights into a unique cellulase fold and mechanism of cellulose hydrolysis. *Proc. Natl. Acad. Sci. U.S.A.* 108, 5237–5242. doi: 10.1073/pnas.1015006108
- Brown, S. D., Lamed, R., Morag, E., Borovok, I., Shoham, Y., Klingeman, D. M., et al. (2012). Draft genome sequences for *Clostridium thermocellum* wild-type strain YS and derived cellulose adhesion-defective mutant strain AD2. *J. Bacteriol.* 194, 3290–3291. doi: 10.1128/JB.00473-12
- Burton, E. (2010). *Proteomic Analysis of Cellulose Metabolism in Clostridium thermocellum*. Master thesis, Concordia University, Montreal.
- Calusinska, M., Happe, T., Joris, B., and Wilmotte, A. (2010). The surprising diversity of clostridial hydrogenases: a comparative genomic perspective. *Microbiology* 156, 1575–1588. doi: 10.1099/mic.0.032771-0
- Carere, C. R., Rydzak, T., Verbeke, T. J., Cicek, N., Levin, D. B., and Sparling, R. (2012). Linking genome content to biofuel production yields: a meta-analysis of major catabolic pathways among select H<sub>2</sub> and ethanol-producing bacteria. *BMC Microbiol.* 12:295. doi: 10.1186/1471-2180-12-295
- Carere, R. C. (2013). *Genomics of Cellulolytic Clostridia and Development of Rational Metabolic Engineering Strategies*. Ph.D., University of Manitoba.
- Chen, Z., and Duan, X. (2011). Ribosomal RNA depletion for massively parallel bacterial RNA-sequencing applications. *Methods Mol. Biol.* 733, 93–103. doi: 10.1007/978-1-61779-089-8\_7
- Cooksley, C. M., Zhang, Y., Wang, H., Redl, S., Winzer, K., and Minton, N. P. (2012). Targeted mutagenesis of the *Clostridium acetobutylicum* Acetone-Butanol-Ethanol fermentation pathway. *Metab. Eng.* 14, 630–641. doi: 10.1016/j.ymben.2012.09.001
- Cusick, K., Lee, Y.-Y., Youchak, B., and Belas, R. (2012). Perturbation of FlhL interferes with *Proteus mirabilis* swarmer cell gene expression and differentiation. *J. Bacteriol.* 194, 437–447. doi: 10.1128/JB.05998-11
- Demain, A. L., Newcomb, M., and Wu, J. H. D. (2005). Cellulase, clostridia, and ethanol. *Microbiol. Mol. Biol. Rev.* 69, 124–154. doi: 10.1128/MMBR.69.1.124-154.2005
- Dror, T. W., Rolider, A., Bayer, E. A., Lamed, R., and Shoham, Y. (2005). Regulation of major cellulosomal endoglucanases of *Clostridium thermocellum* differs from that of a prominent cellulosomal xylanase. *J. Bacteriol.* 187, 2261–2266. doi: 10.1128/JB.187.7.2261-2266.2005
- Dvortsov, I., Lunina, N., Zverlov, V., and Velikodvorskaya, G. (2010). Substrate-binding properties of the family 54 module of *Clostridium thermocellum* Lic16A laminarinase. *Mol. Biol.* 44, 591–595. doi: 10.1134/S002689331004014X
- Dvortsov, I., Lunina, N., Zverlov, V., and Velikodvorskaya, G. (2012). Properties of four C-terminal carbohydrate-binding modules (CBM4) of laminarinase Lic16A of *Clostridium thermocellum*. *Mol. Biol.* 46, 817–822. doi: 10.1134/S0026893312060039
- Ferreira, S. J., Senning, M., Sonnewald, S., Keßling, P.-M., Goldstein, R., and Sonnewald, U. (2010). Comparative transcriptome analysis coupled to X-ray CT reveals sucrose supply and growth velocity as major determinants of potato tuber starch biosynthesis. *BMC Genomics* 11:93. doi: 10.1186/1471-2164-11-93
- Gold, N., and Martin, V. (2007). Global view of the *Clostridium thermocellum* cellulosome revealed by quantitative proteomic analysis? *J. Bacteriol.* 189, 6787–6795. doi: 10.1128/JB.00882-07
- Hess, V., González, J. M., Parthasarathy, A., Buckel, W., and Müller, V. (2013a). Caffeate respiration in the acetogenic bacterium *Acetobacterium woodii*: a coenzyme A loop saves energy for caffeate activation. *Appl. Environ. Microbiol.* 79, 1942–1947. doi: 10.1128/AEM.03604-12
- Hess, V., Schuchmann, K., and Müller, V. (2013b). The ferredoxin: NAD<sup>+</sup> oxidoreductase (Rnf) from the acetogen *Acetobacterium woodii* requires Na<sup>+</sup> and is reversibly coupled to the membrane potential. *J. Biol. Chem.* 288, 31496–31502. doi: 10.1074/jbc.M113.510255
- Himmel, M. E., Ding, S. Y., Johnson, D. K., Adney, W. S., Nimlos, M. R., Brady, J. W., et al. (2007). Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science* 315, 804–807. doi: 10.1126/science.1137016
- Huang, W., Nadeem, A., Zhang, B., Babar, M., Soller, M., and Khatib, H. (2012). Characterization and comparison of the leukocyte transcriptomes of three cattle breeds. *PLoS ONE* 7:e30244. doi: 10.1371/journal.pone.0030244
- Irwin, D., Leathers, T., Greene, R., and Wilson, D. (2003). Corn fiber hydrolysis by *Thermobifida fusca* extracellular enzymes. *Appl. Environ. Microbiol.* 61, 352–358. doi: 10.1007/s00253-002-1210-6
- Islam, R., Cicek, N., Sparling, R., and Levin, D. (2006). Effect of substrate loading on hydrogen production during anaerobic fermentation by *Clostridium thermocellum* 27405. *Appl. Microbiol. Biotechnol.* 72, 576–583. doi: 10.1007/s00253-006-0316-7
- Ji, H., Gheysen, G., Denil, S., Lindsey, K., Topping, J. F., Nahar, K., et al. (2013). Transcriptomic analysis through RNA sequencing of giant cells induced by *Meloidogyne graminicola* in rice roots. *J. Exp. Bot.* 64, 3885–3898. doi: 10.1093/jxb/ert219
- Jia, Y., Lisch, D. R., Ohtsu, K., Scanlon, M. J., Nettleton, D., and Schnable, P. S. (2009). Loss of RNA-dependent RNA polymerase 2 (RDR2) function causes widespread and unexpected changes in the expression of transposons, genes, and 24-nt small RNAs. *PLoS Genet.* 5:e1000737. doi: 10.1371/journal.pgen.1000737
- Kahel-Raifer, H., Jindou, S., Bahari, L., Nataf, Y., Shoham, Y., Bayer, E. A., et al. (2010). The unique set of putative membrane-associated anti- $\sigma$  factors in *Clostridium thermocellum* suggests a novel extracellular carbohydrate-sensing mechanism involved in gene regulation. *FEMS Microbiol. Lett.* 308, 84–93. doi: 10.1111/j.1574-6968.2010.01997.x
- Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., et al. (2008). KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 36, D480–D484. doi: 10.1093/nar/gkm882
- Lamed, R., and Zeikus, J. (1980). Ethanol production by thermophilic bacteria: relationship between fermentation product yields of and catabolic enzyme activities in *Clostridium thermocellum* and *Thermoanaerobium brockii*. *J. Bacteriol.* 144, 569–578.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10:R25. doi: 10.1186/gb-2009-10-3-r25



- Lin, C., Garrett, A. S., De Kumar, B., Smith, E. R., Gogol, M., Seidel, C., et al. (2011). Dynamic transcriptional events in embryonic stem cells mediated by the super elongation complex (SEC). *Genes Dev.* 25, 1486–1498. doi: 10.1101/gad.205921
- Mortazavi, A., Williams, B. A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628. doi: 10.1038/nmeth.1226
- Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., et al. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320, 1344. doi: 10.1126/science.1158441
- Nataf, Y., Bahari, L., Kahel-Raifer, H., Borovok, I., Lamed, R., Bayer, E. A., et al. (2010). *Clostridium thermocellum* cellulosomal genes are regulated by extracytoplasmic polysaccharides via alternative sigma factors. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18646–18651. doi: 10.1073/pnas.1012175107
- Nataf, Y., Yaron, S., Stahl, F., Lamed, R., Bayer, E. A., Scheper, T.-H., et al. (2009). Cellodextrin and laminaribiose ABC transporters in *Clostridium thermocellum*. *J. Bacteriol.* 191, 203–209. doi: 10.1128/JB.01190-08
- Newcomb, M., Millen, J., Chen, C.-Y., and Wu, J. D. (2011). Co-transcription of the celC gene cluster in *Clostridium thermocellum*. *Appl. Microbiol. Biotechnol.* 90, 625–634. doi: 10.1007/s00253-011-3121-x
- Olson, D. G., Tripathi, S. A., Giannone, R. J., Lo, J., Caiazza, N. C., Hogsett, D. A., et al. (2010). Deletion of the Cel48S cellulase from *Clostridium thermocellum*. *Proc. Natl. Acad. Sci. U.S.A.* 107, 17727–17732. doi: 10.1073/pnas.1003584107
- Pinheiro, B., Gilbert, H., Sakka, K., Fernandes, V., Prates, J., Alves, V., et al. (2009). Functional insights into the role of novel type I cohesin and dockerin domains from *Clostridium thermocellum*. *Biochem. J.* 424, 375–384. doi: 10.1042/BJ20091152
- Raman, B., McKeown, C. K., Rodriguez, M., Brown, S. D., and Mielenz, J. R. (2011). Transcriptomic analysis of *Clostridium thermocellum* ATCC 27405 cellulose fermentation. *BMC Microbiol.* 11:134. doi: 10.1186/1471-2180-11-134
- Raman, B., Pan, C., Hurst, G. B., Rodriguez, M., McKeown, C. K., Lankford, P. K., et al. (2009). Impact of pretreated switchgrass and biomass carbohydrates on *Clostridium thermocellum* ATCC 27405 cellulosome composition: a quantitative proteomic analysis. *PLoS ONE* 4:e5271. doi: 10.1371/journal.pone.0005271
- Riederer, A., Takasuka, T. E., Makino, S.-I., Stevenson, D. M., Bukhman, Y. V., Elsen, N. L., et al. (2011). Global gene expression patterns in *Clostridium thermocellum* as determined by microarray analysis of chemostat cultures on cellulose or cellobiose. *Appl. Environ. Microbiol.* 77, 1243–1253. doi: 10.1128/AEM.02008-10
- Rydzak, T., Levin, D. B., Cicek, N., and Sparling, R. (2009). Growth phase-dependent enzyme profile of pyruvate catabolism and end-product formation in *Clostridium thermocellum* ATCC 27405. *J. Biotechnol.* 140, 169–175. doi: 10.1016/j.jbiotec.2009.01.022
- Rydzak, T., Levin, D. B., Cicek, N., and Sparling, R. (2011). End-product induced metabolic shifts in *Clostridium thermocellum* ATCC 27405. *Appl. Microbiol. Biotechnol.* 92, 199–209. doi: 10.1007/s00253-011-3511-0
- Rydzak, T., McQueen, P., Krokhin, O., Spicer, V., Ezzati, P., Dwivedi, R., et al. (2012). Proteomic analysis of *Clostridium thermocellum* core metabolism: relative protein expression profiles and growth phase-dependent changes in protein expression. *BMC Microbiol.* 12:214. doi: 10.1186/1471-2180-12-214
- Schut, G. J., and Adams, M. W. (2009). The iron-hydrogenase of *Thermotoga maritima* utilizes ferredoxin and NADH synergistically: a new perspective on anaerobic hydrogen production. *J. Bacteriol.* 191, 4451–4457. doi: 10.1128/JB.01582-08
- Stevenson, D., and Weimer, P. (2005). Expression of 17 genes in *Clostridium thermocellum* ATCC 27405 during fermentation of cellulose or cellobiose in continuous culture. *Appl. Environ. Microbiol.* 71:4672. doi: 10.1128/AEM.71.8.4672-4678.2005
- Strobel, H., Caldwell, F., and Dawson, K. (1995). Carbohydrate transport by the anaerobic thermophile *Clostridium thermocellum* LQRI. *Appl. Environ. Microbiol.* 61, 4012–4015.
- Tang, F. C., Barbacioru, C., Wang, Y. Z., Nordman, E., Lee, C., Xu, N. L., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. *Nat. Methods* 6, U377–U386. doi: 10.1038/nmeth.1315
- Tatusov, R. L., Galperin, M. Y., Natale, D. A., and Koonin, E. V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28, 33–36. doi: 10.1093/nar/28.1.33
- Templeton, D. W., Scarlata, C. J., Sluiter, J. B., and Wolfrum, E. J. (2010). Compositional analysis of lignocellulosic feedstocks. 2. Method uncertainties. *J. Agric. Food Chem.* 58, 9054–9062. doi: 10.1021/jf100807b
- Tucker, M. P., Farmer, J. D., Keller, F. A., Schell, D. J., and Nguyen, Q. A. (1998). Comparison of yellow poplar pretreatment between NREL digester and suns hydrolyzer. *Appl. Biochem. Biotechnol.* 70–72, 25–35. doi: 10.1007/BF02920121
- van der Lelie, D., Taghavi, S., McCorkle, S. M., Li, L.-L., Malfatti, S. A., Monteleone, D., et al. (2012). The metagenome of an anaerobic microbial community decomposing poplar wood chips. *PLoS ONE* 7:e36740. doi: 10.1371/journal.pone.0036740
- Vinzant, T. B., Ponfick, L., Nagle, N. J., Ehrman, C. I., Reynolds, J. B., and Himmel, M. E. (1994). SSF comparison of selected woods from southern sawmills. *Appl. Biochem. Biotechnol.* 45, 611–626. doi: 10.1007/BF02941834
- Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* 10, 57–63. doi: 10.1038/nrg2484
- Wei, H., Tucker, M. P., Baker, J. O., Harris, M., Luo, Y., Xu, Q., et al. (2012). Tracking dynamics of plant biomass composting by changes in substrate structure, microbial community, and enzyme activity. *Biotechnol. Biofuels* 5:20. doi: 10.1186/1754-6834-5-20
- Wei, H., Xu, Q., Taylor, L., Baker, J., Tucker, M., and Ding, S. (2009). Natural paradigms of plant cell wall degradation. *Curr. Opin. Biotechnol.* 20, 330–338. doi: 10.1016/j.copbio.2009.05.008
- Wilhelm, B. T., Marguerat, S., Watt, S., Schubert, F., Wood, V., Goodhead, I., et al. (2008). Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. *Nature* 453, 1239–1243. doi: 10.1038/nature07002
- Wilson, C. M., Rodriguez, M. Jr., Johnson, C. M., Martin, S. L., Chu, T. M., Wolfinger, R. D., et al. (2013a). Global transcriptome analysis of *Clostridium thermocellum* ATCC 27405 during growth on dilute acid pretreated Populus and switchgrass. *Biotechnol. Biofuels* 6:179. doi: 10.1186/1754-6834-6-179
- Wilson, C. M., Yang, S., Rodriguez, M. Jr., Ma, Q., Johnson, C. M., Dice, L., et al. (2013b). *Clostridium thermocellum* transcriptomic profiles after exposure to furfural or heat stress. *Biotechnol. Biofuels* 6:131. doi: 10.1186/1754-6834-6-131
- Yang, S., Giannone, R., Dice, L., Yang, Z., Engle, N., Tschaplinski, T., et al. (2012). *Clostridium thermocellum* ATCC27405 transcriptomic, metabolomic and proteomic profiles after ethanol stress. *BMC Genomics* 13:336. doi: 10.1186/1471-2164-13-336
- Yang, S., Guarnieri, M. T., Smolinski, S., Ghirardi, M., and Pienkos, P. T. (2013). De novo transcriptomic analysis of hydrogen production in the green alga *Chlamydomonas moewusii* through RNA-Seq. *Biotechnol. Biofuels* 6:118. doi: 10.1186/1754-6834-6-118
- Zverlov, V. V., Kellermann, J., and Schwarz, W. H. (2005). Functional subgenomics of *Clostridium thermocellum* cellulosomal genes: identification of the major catalytic components in the extracellular complex and detection of three new enzymes. *Proteomics* 5, 3646–3653. doi: 10.1002/pmic.200401199
- Zverlov, V. V., and Schwarz, W. H. (2006). “The *Clostridium thermocellum* cellulosome: novel components and insights from the genomic sequence,” in *Cellulosome*, eds I. Kataeva and L. Ljungdahl (New York, NY: Nova Science Publishers), 119–151.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 21 November 2013; accepted: 19 March 2014; published online: 11 April 2014.

Citation: Wei H, Fu Y, Magnusson L, Baker JO, Maness P-C, Xu Q, Yang S, Bowersox A, Bogorad I, Wang W, Tucker MP, Himmel ME and Ding S-Y (2014) Comparison of transcriptional profiles of *Clostridium thermocellum* grown on cellobiose and pretreated yellow poplar using RNA-Seq. *Front. Microbiol.* 5:142. doi: 10.3389/fmicb.2014.00142

This article was submitted to *Microbial Physiology and Metabolism*, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Wei, Fu, Magnusson, Baker, Maness, Xu, Yang, Bowersox, Bogorad, Wang, Tucker, Himmel and Ding. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# A mathematical model of metabolism and regulation provides a systems-level view of how *Escherichia coli* responds to oxygen

Michael Ederer<sup>1\*</sup>, Sonja Steinsiek<sup>2</sup>, Stefan Stagge<sup>2</sup>, Matthew D. Rolfe<sup>3</sup>, Alexander Ter Beek<sup>4</sup>, David Knies<sup>1</sup>, M. Joost Teixeira de Mattos<sup>4</sup>, Thomas Sauter<sup>5</sup>, Jeffrey Green<sup>3</sup>, Robert K. Poole<sup>3</sup>, Katja Bettenbrock<sup>2</sup> and Oliver Sawodny<sup>1</sup>

<sup>1</sup> Institute for System Dynamics, University of Stuttgart, Stuttgart, Germany

<sup>2</sup> Max Planck Institute for Dynamics of Complex Technical Systems, Magdeburg, Germany

<sup>3</sup> Department of Molecular Biology and Biotechnology, The University of Sheffield, Sheffield, UK

<sup>4</sup> Molecular Microbial Physiology, Swammerdam Institute for Life Sciences, University of Amsterdam, Amsterdam, Netherlands

<sup>5</sup> Life Sciences Research Unit, Université du Luxembourg, Luxembourg, Luxembourg

## Edited by:

Katherine M. Pappas, University of Athens, Greece

## Reviewed by:

Dong-Woo Lee, Kyungpook National University, Korea (South)

Armen Trchounian, Yerevan State University, Armenia

## \*Correspondence:

Michael Ederer, Institute for System Dynamics, University of Stuttgart, Pfaffenwaldring 9, Stuttgart D-70569, Germany  
e-mail: michael.ederer@isys.uni-stuttgart.de

The efficient redesign of bacteria for biotechnological purposes, such as biofuel production, waste disposal or specific biocatalytic functions, requires a quantitative systems-level understanding of energy supply, carbon, and redox metabolism. The measurement of transcript levels, metabolite concentrations and metabolic fluxes *per se* gives an incomplete picture. An appreciation of the interdependencies between the different measurement values is essential for systems-level understanding. Mathematical modeling has the potential to provide a coherent and quantitative description of the interplay between gene expression, metabolite concentrations, and metabolic fluxes. *Escherichia coli* undergoes major adaptations in central metabolism when the availability of oxygen changes. Thus, an integrated description of the oxygen response provides a benchmark of our understanding of carbon, energy, and redox metabolism. We present the first comprehensive model of the central metabolism of *E. coli* that describes steady-state metabolism at different levels of oxygen availability. Variables of the model are metabolite concentrations, gene expression levels, transcription factor activities, metabolic fluxes, and biomass concentration. We analyze the model with respect to the production capabilities of central metabolism of *E. coli*. In particular, we predict how precursor and biomass concentration are affected by product formation.

**Keywords:** *Escherichia coli*, mathematical modeling, metabolism, regulation, respiration, fermentation, thermokinetic modeling

## 1. INTRODUCTION

*Escherichia coli* is able to utilize a variety of electron and carbon donors, such as glucose or glycerol, and electron acceptors, such as oxygen or nitrate. Energy currencies in form of the proton motive force (pmf) and the ATP/ADP ratio are supplied by either substrate level phosphorylation or by proton translocation against the pmf during membrane-associated electron transport (Lengeler et al., 1998). The membrane-associated electron transport chain transfers electrons from cytoplasmatic metabolites, mostly the electron carriers NADH and FADH<sub>2</sub>, to quinones and from the quinones to the external electron acceptor (Inglelew and Poole, 1984). The thermodynamic force of these redox reactions can be used to translocate protons and thus contribute to the maintenance of the pmf. Dependent on the extracellular medium, *E. coli* uses different strategies to match the balance of carbon, electrons, and energy. We focus on the use of glucose as electron and carbon donor and oxygen as the electron acceptor. If oxygen is available, the redox balance is maintained by transferring electrons to the external acceptor oxygen. Membrane-associated electron transfer is coupled to proton translocation. Glucose is

oxidized to carbon dioxide or partially oxidized products such as acetate and succinate, often referred to as overflow products, or it is converted into precursors for biosynthesis. If no oxygen and no other external electron acceptors are available, ATP is gained mainly by substrate level phosphorylation in glycolysis. Redox balance is maintained by the excretion of different metabolites (e.g., acetate, formate, ethanol, and succinate). Even in the absence of extracellular electron acceptors, the electron transport chain can be active, since intracellular electron acceptors, such as fumarate, are available (Lengeler et al., 1998).

The metabolism of bacteria needs to match the requirements of growth and maintenance for carbon, electrons and energy with the supply from the medium. Complex regulatory networks control this process, such as those that operate following a change in oxygen availability. The redesign of bacteria for biotechnological purposes, such as biofuel production, puts additional loads on the metabolism. A high level expression of a production pathway is often not sufficient for satisfactory production capabilities. For optimal production, metabolic regulation needs to be adapted accordingly (Shimizu, 2009).

The adaption to a change of oxygen availability is controlled by transcriptional regulation centered around the transcription factors FNR and ArcA (Sawers, 1999). The activity of FNR depends directly on oxygen (Jordan et al., 1997) and the activity of ArcA depends on the quinols and quinones of the electron transport chain (Georgellis et al., 1997; Bekker et al., 2010; Alvarez et al., 2013; Sharma et al., 2013).

Alexeeva et al. (2000, 2002, 2003) introduced the aerobiosis scale that allows the reproducible adjustment of different microaerobic steady states in a continuous, glucose-limited chemostat culture. Anaerobic growth corresponds to an aerobiosis value of 0%. An aerobiosis value of 100% is defined as the steady state with the minimal oxygen input into the reactor where no fermentation products are excreted, i.e., where all carbon is either incorporated into biomass or respired to carbon dioxide. The aerobiosis scale is linear with the oxygen input into the reactor. The dependency of the biomass-specific oxygen uptake flux on the aerobiosis level is concave because at higher oxygen availability the steady state biomass concentration is higher. For the *E. coli* wild type, biomass-specific acetate production decreases linearly with aerobiosis until it vanishes at 100% aerobiosis units. The aerobiosis scale allows the reproducible analysis of microaerobic states on the physiological (Alexeeva et al., 2000, 2002, 2003; Bekker et al., 2010; Steinsiek et al., 2011) and transcriptional level (Partridge et al., 2006, 2007; Rolfe et al., 2011, 2012; Trotter et al., 2011). Using the aerobiosis scale, results of different laboratories using different reactors can be compared. The aerobiosis scale thus provides an ideal basis for mathematical modeling.

The analysis of measurement data of transcript levels, protein abundances, metabolite concentrations, and fluxes is a valuable tool to reveal bottlenecks of production pathways. In simple cases, such as a feedback inhibition in a linear path, the repressing conditions can be identified by the measurement of metabolite concentrations and reaction fluxes, and countermeasures can be implemented genetically. In more complex cases, the inherent correlation between different measured quantities may not always be so apparent. The metabolic pathways of the central metabolism form a strongly interconnected network with complex interdependencies. A thorough analysis of manipulations of this network requires integration of measurement data of different types, in particular transcript and protein levels, metabolite concentrations, fluxes, and transcription factor activities, by analyzing their dependencies within the network structure. Mathematical modeling of the central metabolism can provide tools for analyzing and predicting the effect of genetic intervention and thus provide guidance when redesigning organisms for biofuel production. The model-based integration of signal transduction, regulation, and metabolism is still not standard and most models are restricted to describing either metabolic processes, signal transduction or regulation (Gonçalves et al., 2013). The response of cellular metabolism to oxygen was studied previously using mathematical models. For example, Varma et al. (1993) used flux balance analysis to analyze the yield optimal behavior of *E. coli* for different oxygen availabilities. Peercy et al. (2006) presented a kinetic model of the respiratory chain of *E. coli* and its regulation via FNR

and ArcA. They demonstrated that the model is able to show complex dynamic behavior such as oscillations and hysteresis. Beard (2005) described the electron transport chain of mitochondria by choosing a force for each reaction that is consistent with the requirements of thermodynamics. Similarly, Klamt et al. (2008) used linear relationships of affinity and flux to describe the kinetics of the electron transport chain of purple non-sulfur bacteria.

Here, we present a modeling approach that integrates several levels of information. The goal of the modeling approach was to provide a physically consistent systems-level view of the central carbon and energy metabolism of *E. coli* and its regulation. We show how the model is able to explain the steady state response of *Escherichia coli* to oxygen by comparing model simulation and measurement data for different values of aerobiosis. We demonstrate the utility of the model by making predictions on the effect of biofuel production pathways on bacterial metabolism.

## 2. MATERIALS AND METHODS

### 2.1. MODELING

The complete model information is available in the supplementary files. In particular, **Supplementary Data Sheet 1** shows an overview of all model elements in alphabetical order and **Supplementary Data Sheet 2** shows the model definition file. The model definition file together with the Mathematica package TKMOD (Thermo-Kinetic Modeling) can be downloaded from <https://seek.sysmo-db.org/models/23>. Together they provide a runnable version of the model.

#### 2.1.1. Model of the metabolism

We use the thermodynamic-kinetic modeling formalism (Ederer and Gilles, 2007; Ederer, 2010) to describe the metabolic reaction network. We assume an ideal, aqueous solution with the chemical potentials  $\mu_i'' = \mu_i''^o(T, p, a_{\text{H}_2\text{O}}, \text{pH}, I) + R^* \cdot T \cdot \log(c_i/c^o) + z_i \cdot F \cdot \phi_i$ . For  $\mu_i''^o$  we use the transformed Gibbs formation energy of metabolite *i*. A Legendre transformation is conducted to adapt the Gibbs energy to constant pH and water activity  $a_{\text{H}_2\text{O}}$  (Alberty, 2003). The Debye-Hückel equation is used to correct for the effect of ionic strength *I* (Alberty, 2003). Temperature *T* and pressure *p* are constant. The symbols  $R^*$  and *F* denote the ideal gas constant and the Faraday constant, respectively. The charge number  $z_i$  of biochemical species *i* is approximated by the charge of the dominant species at pH 7 and taken from Reed et al. (2003). The electrical potential of the compartment of metabolite *i* is denoted by  $\phi_i$ . According to Ederer (2010), we get that the relationship between thermokinetic potential  $\xi_i$  and concentration *c<sub>i</sub>* is  $c_i = C_i \cdot \xi_i$  with  $C_i = c^o \cdot \exp(-(\mu_i''^o + z_i \cdot F \cdot \phi_i)/(R^* \cdot T))$  where  $C_i$  is the thermokinetic capacity.

Fluxes  $J_k$  of metabolic reactions are modeled according to Ederer (2010) as  $(R_j(\xi)/c_{E,j}) \cdot J_j = F_j(\xi)$  where  $F_j(\xi)$  is the thermokinetic force,  $R_j(\xi)$  is the enzyme-specific thermokinetic resistance and  $c_{E,j}$  is the enzyme concentration. For example, the thermokinetic force of reaction  $A + B \rightleftharpoons C$  is  $F = \xi_A \cdot \xi_B - \xi_C$ . The above expression reflects three major effects that control metabolic fluxes: the thermokinetic force *F* describes the

influence of reactants and products, the resistance  $R_j(\xi)$  describes the specific enzyme activity that may depend on further activators and inhibitors, and the enzyme concentration  $c_{E,j}$  describes the influence of the enzyme concentration on the metabolic reaction. For most reactions the thermokinetic resistance  $R_j(\xi)$  is assumed to be independent of the metabolite potentials  $\xi$ . For some reactions, where enzymatic regulation proved to be important for describing the experimental data according non-constant terms were included in  $R_j(\xi)$  (see **Supplementary Data Sheet 2**).

### 2.1.2. De novo synthesis of conserved moieties

The *de novo* synthesis of AMP, NAD, NADP, and CoA was modeled such that the concentrations of these conserved moieties are constant despite dilution due to growth. The *de novo* synthesis of quinones is modeled as a function of the aerobiosis value in order to reproduce the observed changes in the total quinone concentrations. In the model, the pool concentration of ubiquinone and ubiquinol increases linearly with aerobiosis and the pool concentration of menaquinone and menaquinol decreases linearly with aerobiosis. At oxygenation levels higher than 100% aerobiosis these concentrations are constant. In order to reproduce the observation that even in the complete anaerobic case a substantial part of the quinone pool is oxidized, we introduced a constant pool of oxidized quinones that does not participate in any reaction.

### 2.1.3. Transcription factor activity

Transcription factors control gene expression by activating or repressing the expression of many genes. The observed transcription factor activities are the result of a complex interplay of the amount of the transcription factor that in turn may be controlled by other transcription factors and its activation that is often controlled by a metabolic signal, for example oxygen in the case of FNR. As a simplification, we introduce the activity  $a_{TF,i}$  of transcription factor  $i$ . If the activity of transcription factor  $i$  is minimal, we write that  $a_{TF,i} = 0$  and if it is maximal, we write that  $a_{TF,i} = 1$ . We assume a phenomenological Hill type equation  $a_{TF,i} = x_{TF,i}^{n_{TF,i}} / (x_{TF,i}^{n_{TF,i}} + k_{TF,i}^{n_{TF,i}})$  that describes  $a_{TF,i}$  in dependence on the respective metabolic signal  $x_{TF,i}$ . For  $n_{TF,i} > 0$  transcription factor  $i$  is activated by its metabolic signal  $x_{TF,i}$ . For  $n_{TF,i} < 0$  it is inhibited. **Supplementary Data Sheet 3** lists the transcription factors with the respective metabolic signals.

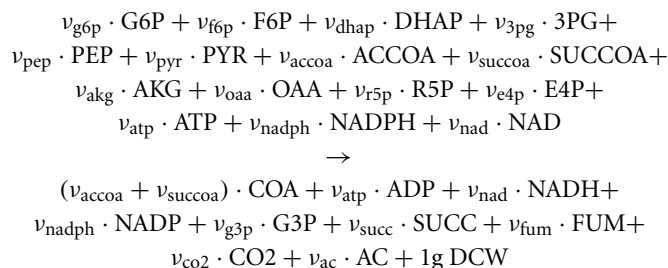
### 2.1.4. Gene expression

Gene expression is described by the equation  $\dot{c}_{E,i} = J_{syn,i}(a_{TF}) - \mu \cdot c_{E,i}$  where  $c_{E,i}$  is the concentration of enzyme  $i$ ,  $\mu$  is the dilution rate due to growth and  $J_{syn,i}$  is the synthesis rate of enzyme  $i$  that depends on the activities of the transcription factors  $a_{TF}$ . The complex dependency of the synthesis rate on the transcription factors is approximated by a phenomenological relationship. For example, the expression of a gene that is activated by transcription factors 1 and 2 and inhibited by transcription factor 3 is modeled by  $J_{syn,i} = s(k_1, a_{TF,1}) \cdot s(k_2, a_{TF,2}) \cdot s(k_3, 1 - a_{TF,3})$  where  $s(k, a_{TF}) = 2^{-k} + (1 - 2^{-k}) \cdot a_{TF}$ . This means that each transcription factor is able to change the expression of a gene by a factor of  $2^k$  and the interaction of different transcription factors is multiplicative. As will be seen, this model allows the description

of the measurement data and therefore the use of more complex models allowing for example for additive interactions is not necessary at this stage.

### 2.1.5. Growth and maintenance

Central metabolism provides biosynthesis with precursor molecules and with redox and energy equivalents. In order to describe the growth of the biomass we assume a stoichiometric relation for the reaction of precursors into biomass:



where the stoichiometric coefficients  $\nu_i$  define how much precursor is needed to produce 1 gram of dry cell mass (see Neidhardt et al., 1990). We assume that the rate of this reaction depends on the concentrations of the reactants with linlog kinetics. Growth of *E. coli* can only occur if the adenylate energy charge is above a threshold (Chapman et al., 1971). To model this fact, the linlog kinetics are extended by a factor realizing a ramp function with saturation dependent on the ATP/ADP ratio:

$$\mu = \begin{cases} 0 & \text{if } k_{lo} < c_{atp}/c_{adp} \\ k_a \cdot (k_b + \sum_i \nu_i \cdot \log(c_i)) \cdot (c_{atp}/c_{adp} - k_{lo}) / (k_{hi} - k_{lo}) & \text{if } k_{lo} \leq c_{atp}/c_{adp} \leq k_{hi} \\ k_a \cdot (k_b + \sum_i \nu_i \cdot \log(c_i)) & \text{if } k_{hi} < c_{atp}/c_{adp} \end{cases}$$

where  $i$  runs over the reactants. Cellular maintenance is modeled by assuming a hydrolysis rate of ATP to ADP that is not coupled to processes of the central metabolism or growth. It follows a ramp function with saturation dependent on the ATP/ADP ratio. The use of ramp functions assures that below a certain threshold no growth or maintenance occurs and that the rate of growth or maintenance saturates above a certain threshold.

### 2.1.6. Chemostat environment

With the specific growth rate we get according to the chemostat equations (Smith and Waltman, 2008) for the biomass concentration  $\dot{c}_X = \mu \cdot c_X - D \cdot c_X$ . The extracellular metabolites are described by  $\dot{c}_i = J_i \cdot c_X + D \cdot c_{in,i} - D \cdot c_i$ , where  $c_{in,i}$  is the concentration of  $i$  in the inflow,  $J_i$  is the specific production (positive) or consumption (negative) rate of metabolite  $i$  by the biomass and  $D$  is the dilution rate. For the gaseous compounds oxygen and carbon dioxide the equation is modified to  $\dot{c}_i = J_i \cdot c_X + k_{in,i} - k_{out,i} \cdot c_i$  because these compounds are mainly exchanged via the gas phase, but not the liquid phase. The parameter  $k_{in,i}$  describes the supply of the medium by the aeration flow. For oxygen this parameter depends on the aerobiosis value. The parameter  $k_{out,i}$  describes the outgassing.

### 2.1.7. Model reduction

The resulting model spans the time scales from fast metabolic reactions up to steady state growth. The stiffness of the model calls for model reduction. For several reactions of the central metabolism it is known that they proceed near thermodynamic equilibrium (Kümmel et al., 2006a). We assume quasi-equilibrium for several reactions. For example the reaction  $G6P \rightleftharpoons F6P$  is usually rapid such that we can assume that the concentrations of G6P and F6P are in equilibrium with each other. This allows the reduction of the order and the stiffness of the model. In thermokinetic modeling this can be achieved by assuming a vanishing resistance ( $R_j = 0$ ). The resulting differential-algebraic equation system has index 2 but can be simplified to index 1 (Ederer, 2010).

### 2.1.8. TKMOD

Modeling is done using the tool TKMOD (Ederer, 2010). TKMOD reads a model description (see **Supplementary Data Sheet 2**) where stoichiometry, thermokinetic capacities and resistances can be defined. Then it derives the model equations by using the computer algebra system Mathematica (Wolfram Research, 2010). TKMOD automatically performs reduction steps for fast reactions with  $R_j = 0$  and writes a FORTRAN code for the simulation equations. TKMOD uses DASKR that is a solver for differential-algebraic equation systems for simulation (Brown et al., 1994, 1998, 2007). A version of TKMOD including DASKR is packaged together with the model files and available from <https://seek.sysmo-db.org/models/23>.

### 2.1.9. Parameters

The model described above has different types of parameters. The stoichiometric parameters of the reactions are taken mainly from Reed et al. (2003). The amount of translocated protons during electron transport follows Borisov et al. (2011). The thermokinetic capacities are computed from the standard Gibbs energies of formation. Standard Gibbs formation energies for most metabolites are taken from Alberty (2003). Thermodynamic data for the ubiquinone/quinol and menaquinone/quinol pairs are from Alvarez et al. (2013). Data for metabolites in the pentose phosphate pathway are taken from Kümmel et al. (2006b). The parameters of gene expression, gene regulation and the thermokinetic resistances are manually adjusted in order to fit the experimental data. The thermokinetic resistance  $R_j$  of reaction  $j$  is related to the thermokinetic capacities of the reactants  $C_i$  and the specific forward rate constant  $k_{+j}$  by  $R_j = k_{+j}^{-1} \cdot \prod_{i \in E_j} C_i^{-|v_{ij}|}$ . In order to restrict the search space, we assume that  $k_{+j}$  can take only values of the type  $10^x$  with an integer  $x$ . Similarly, we assume that the parameters  $k_i$  of the gene expression model can take only integer values. The quality of the fit we achieve with this high restriction suggests that the order of magnitude of the resistances determines most of the behavior of the model and that many features of the model are robust against uncertainties in the exact values.

### 2.1.10. Comparison of simulation and experimental data

Measurement data on biomass concentration, yield factors and fluxes can be compared directly to the simulation results. Valgepea

et al. (2010) observe a high correlation between transcript levels and enzyme concentrations for similar chemostat conditions. Also for the gene expression data used in this study a high correlation was observed (Rolfe et al., 2011; Trotter et al., 2011). For this reason, we are able to compare measured transcript levels with the simulated enzyme concentrations. The direct use of measured metabolite concentrations in the mathematical model is subject to several uncertainties. The Gibbs formation energies used to parametrize the model are measured for dilute aqueous solution different from the crowded cytoplasm (Cossins et al., 2011). Systematic losses during quenching and probe preparation may prevent an absolute quantification. For these reasons, we treat the measured metabolite concentrations as relative values.

Relative measurement values  $x_m$  (transcripts, metabolites) are scaled with a factor  $f$  before comparing them (in a plot) with the respective simulated variables  $x_s$ . The factor  $f$  is calculated such that the quadratic difference  $\|1/f \cdot x_s - x_m\|_2^2$  between measured and simulated variables is minimal and can be computed as  $f = \frac{\vec{x}_s^T \cdot \vec{x}_s}{\vec{x}_m^T \cdot \vec{x}_s}$ . Here,  $\vec{x}_s$  and  $\vec{x}_m$  denote the vectors with corresponding pairs of simulated and measured data points. This means that  $x_s$  is plotted together not with  $x_m$  but with  $f \cdot x_m$ .

## 2.2. EXPERIMENTAL DATA

Experimental conditions and strain *E. coli* MG1655 were as described in (Rolfe et al., 2012). Biomass and extracellular metabolite concentrations were determined as in (Steinsiek et al., 2011). Transcript data were measured via DNA microarray technology and are taken from (Rolfe et al., 2011). Measurements were complemented with RT-PCR data as described in (Steinsiek et al., 2011).

For determination of intracellular metabolite concentrations cells were first quenched following the method of Link et al. (2008) and afterwards a modified extraction procedure published by Ritter et al. (2008) was used. Following the method of Link et al. (2008) 10 ml of cell containing medium from continuous cultivations were immediately quenched in 20 ml methanol-glycerol solution (60/40% v/v) at  $-60^\circ\text{C}$  thereby holding the temperature below  $-20^\circ\text{C}$ . Samples were thoroughly mixed and immediately transferred to dry ice and cooled to  $-50^\circ\text{C}$ . After centrifugation for 30 min at 10,000g and  $-20^\circ\text{C}$  the cell pellet was washed with methanol-glycerol solution (60%/40% v/v) at  $-20^\circ\text{C}$ . After a second centrifugation step all the supernatant was removed and the pellet was kept at  $-80^\circ\text{C}$  until extraction. The cell pellet was extracted with 1 ml methanol and immediately after resuspension 500  $\mu\text{l}$  of trichloromethane were added and the solution was mixed vigorously. The sample was split into three aliquots and 450  $\mu\text{l}$  trichloromethane pre-chilled on ice were added to each aliquot. Samples were thoroughly vortexed. Afterwards 900  $\mu\text{l}$  of methanol/tricine buffer (9:10 parts; final concentration of tricine 1 mM, pH = 7.4) were added, the sample was vortexed again and centrifuged for 10 min at 16,000 g at  $4^\circ\text{C}$ . 800  $\mu\text{l}$  of the upper (hydrophilic) phase were collected and stored. This step was repeated and the supernatant was collected, combined with the first sample and boiled for 4 min at  $90^\circ\text{C}$ . The sample was again centrifuged at 16000g at  $4^\circ\text{C}$  and the supernatant was evaporated to dryness under nitrogen stream.



Samples were afterwards analyzed by anion exchange chromatography using a BioLC type DX320 (Dionex) as described by Ritter et al. (2006).

Quinone/ol concentrations were determined as described in (Bekker et al., 2007).

### 3. RESULTS AND DISCUSSION

We present a mathematical model of the oxygen response of an *E. coli* population in a glucose-limited chemostat. Modeling is facilitated by the restriction to steady state conditions and the use of the aerobiosis scale. The use of the aerobiosis scale allows the integration of the experiments in different reactors with one unique parameter set. Due to the restriction to steady state conditions, differences in initial conditions (e.g., initial pH, cell density, gene expression levels) do not need to be considered. The overall model contains a thermokinetic model of the metabolic network. The model reproduces the effect of reactants, products, important activators and inhibitors, as well as the enzyme concentrations on the metabolic reaction rate by simplified kinetic laws. The metabolic model is complemented by a gene expression model. The synthesis rate of enzymes depends on the activities of transcription factors. The activities of transcription factors depends in turn on their respective metabolic signals. Information about the network structure is based on the EcoCyc database (Keseler et al., 2013). The metabolism and regulation model are embedded into a model describing the growth of the bacterial population and the chemostat environment. The final model is able to provide an integrated description of metabolic fluxes and concentrations, gene expression levels and genetic regulation. A further hallmark of the model is that the balances of the metabolites ATP, ADP, and AMP, as well as NADH, NAD, NADPH, and NADP are explicitly considered. In many models of smaller sub-networks the concentrations of these ubiquitous metabolites are assumed to be constant because only a small subset of all producing and consuming reactions is modeled. The present model seeks a complete description of the balance of these metabolites and thus reflects the constraints that arise from energy and redox requirements.

The parameters of the model fall into two classes: (1) The stoichiometric parameters and the Gibbs formation energies are largely organism-independent and can be taken from available databases. (2) The thermokinetic resistances and the parameters of gene expression and gene regulation are free and (within bounds) not subject to physical constraints. Their values depend on properties of enzymes, transcription factors, and consensus sequences that may vary between strains. By adjusting the different parameter values for the latter class of parameters, the model can describe different physically feasible behaviors of the cell population. In order to test the model, the free parameters are adapted to a data set describing a steady state chemostat at different levels of aerobiosis. The data set includes metabolite concentrations, gene expression data and uptake and excretion fluxes. The model reproduces the steady state values of most measured variables and predicts the values of many others at several values of aerobiosis. This means that the model is able to describe the steady state response of *E. coli* to oxygen.

Despite the complexity and size of the considered system, the use of simplifying assumptions keeps the model tractable. For most reactions a constant thermokinetic resistance is assumed. The resistance is allowed to vary only by integer factors of 10 to restrict search space. Resistances of rapid reactions are assumed to be zero allowing for a reduction of model size and stiffness. Gene expression and gene regulation are described with phenomenological equations. Parameters describing the influence of transcription factors on gene expression are allowed to take only integer values. Growth and maintenance are described by reactions with phenomenological kinetics. The rationale behind these simplifications is that they facilitate modeling and parameter adaptation while still preserving the basic physical and regulatory constraints on the cellular behavior.

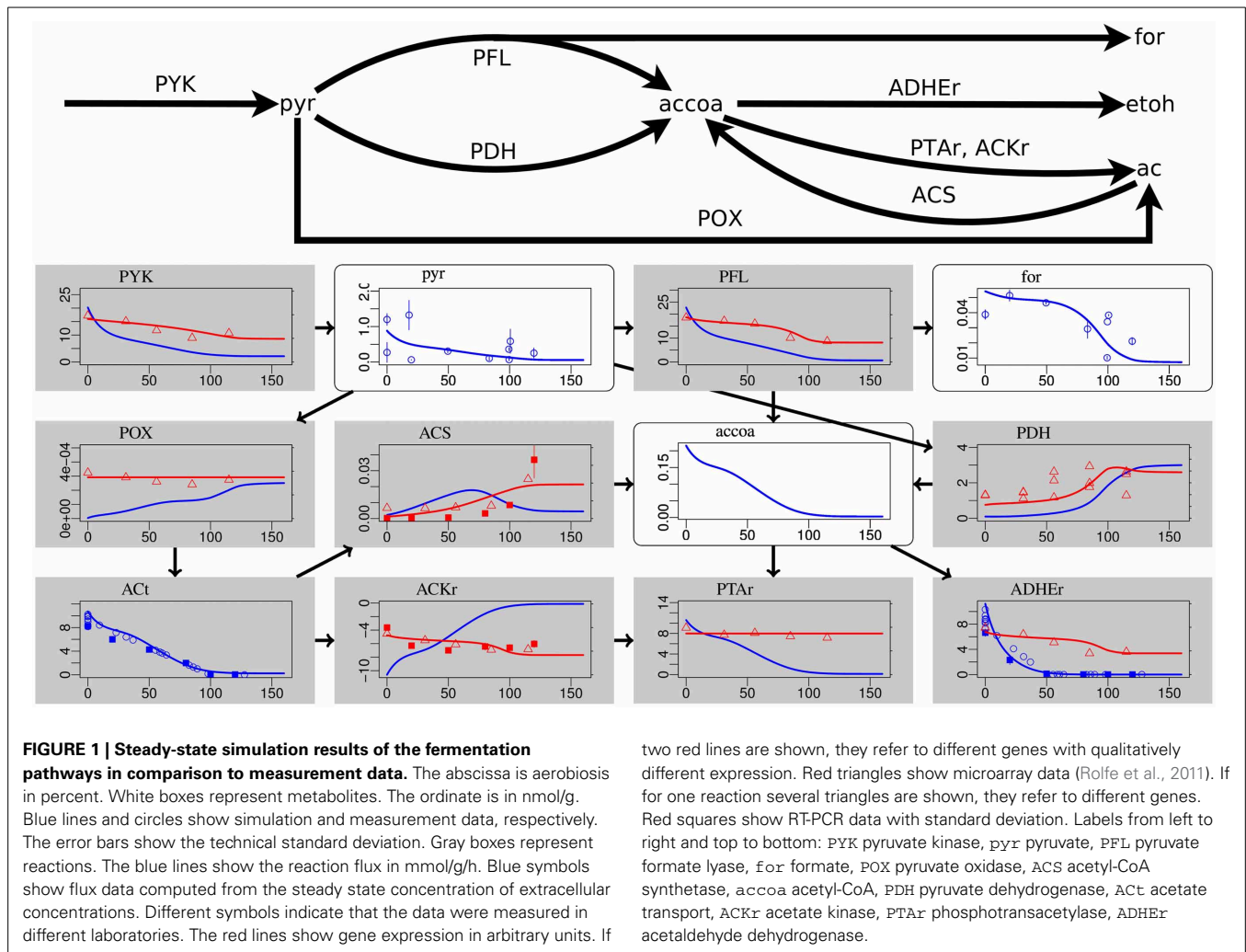
The following section presents the comparison of model results and experimental data for several aerobiosis values. The subsequent section demonstrates the use of the model by providing predictions of the effects of production pathways on the central metabolism.

#### 3.1. BEHAVIOR ACROSS THE AEROBIOSIS SCALE

The model is compared to measurement data of transcripts, metabolites and uptake/excretion fluxes for several values of aerobiosis. Under identical (Rolfe et al., 2011; Trotter et al., 2011) and similar experimental conditions (Valgepea et al., 2010), it was shown that transcript levels correlate well with protein levels. Thus, here we compare modeled enzyme levels with measured transcript levels. **Figures 1–4** show important parts of the model's results. **Supplementary Data Sheet 4** shows a more complete overview of the simulation results.

Most measurement data at the flux, metabolite, transcript, and gene regulatory level are reproduced in an integrated, thermodynamically consistent way by our model. The relative tendencies of metabolite concentrations are described well for glycolysis (**Supplementary Data Sheet 4**) and fermentation pathways (**Figure 1**). For low oxygen availability the biomass yield is low and glucose uptake is high to allow for a growth rate equal to the dilution rate. Consequently, the metabolite concentrations in the first steps of glycolysis (glucose 6-phosphate *g6p*, fructose 6-phosphate *f6p*, and fructose 1,6-bisphosphate *fdp*) before the enzymatic reaction catalyzed by glyceraldehyde-3-phosphate dehydrogenase GAPD decrease with oxygen availability. GAPD catalyzes a rapid reaction with NADH as a product. Because the concentration of NADH decreases strongly at high aerobiosis values, the thermodynamic pull of NADH reverses the pattern for 3-phospho-glycerophosphate *13dpg*, glycerate 2-phosphate *2pg*, 3-phospho-glycerate *3pg*, and phosphoenolpyruvate *pep*. The pattern is again inverted after the essentially irreversible pyruvate kinase PYK such that pyruvate *pyr* and the metabolites of the fermentation pathways are high for low oxygen availability and low for high oxygen availability. This is consistent with the observed inverse correlation of fermentation product excretion and aerobiosis. As a test to check if the model could predict trends of metabolite concentrations, we computed the steady state solutions for different dilution rates that in a steady-state chemostat are equal to the growth rates. The results are qualitatively consistent with the experimental



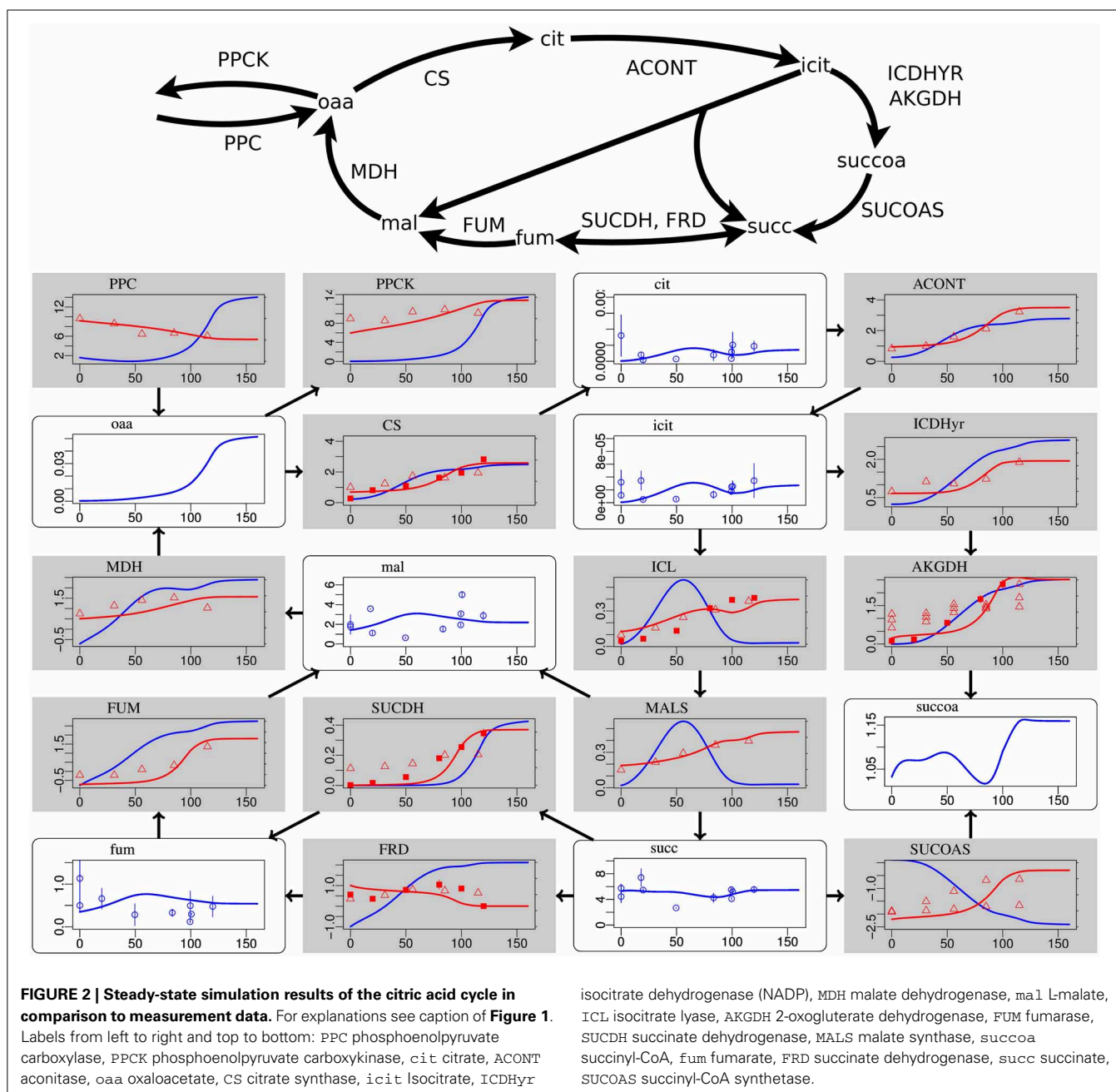


observation from Schaub and Reuss (2008) in that the concentrations of fructose 1,6-bisphosphate, dihydroxyacetone phosphate, and glyceraldehyde 3-phosphate increase with dilution rate, whereas the concentrations of phosphoenolpyruvate, glyceralate 2-phosphate and 3-phospho-glycerate decrease with dilution rate.

The concentrations of the different quinone species in the electron transport chains are well matched, in particular the non-monotone behavior of the ubiquinone redox state (Figure 3). The NADH redox state follows the expected pattern from high reduction potential at low aerobiosis values to low reduction potential at high aerobiosis values. The same holds for the menaquinone redox state that however could not be experimentally measured. The level of oxidized ubiquinone *q8* correlates to the concentration of its oxidant, oxygen, with low but almost constant levels for microaerobic levels and high level in the fully aerobic state. The strongly differentially regulated *de novo* synthesis of ubiquinones leads to strong increase of the total concentration (oxidized plus reduced) of ubiquinones. In the microaerobic range this leads to a seemingly paradoxical increase of the concentration of reduced ubiquinolins such that maximal reduction is

reached for intermediate aerobiosis levels and not for anaerobic growth.

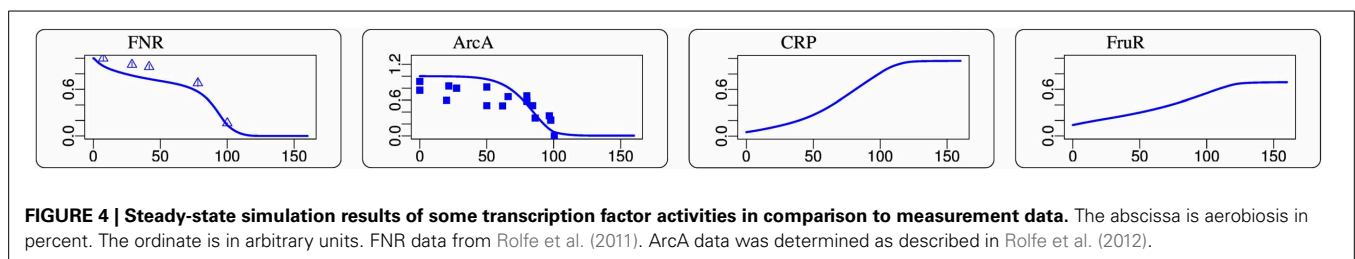
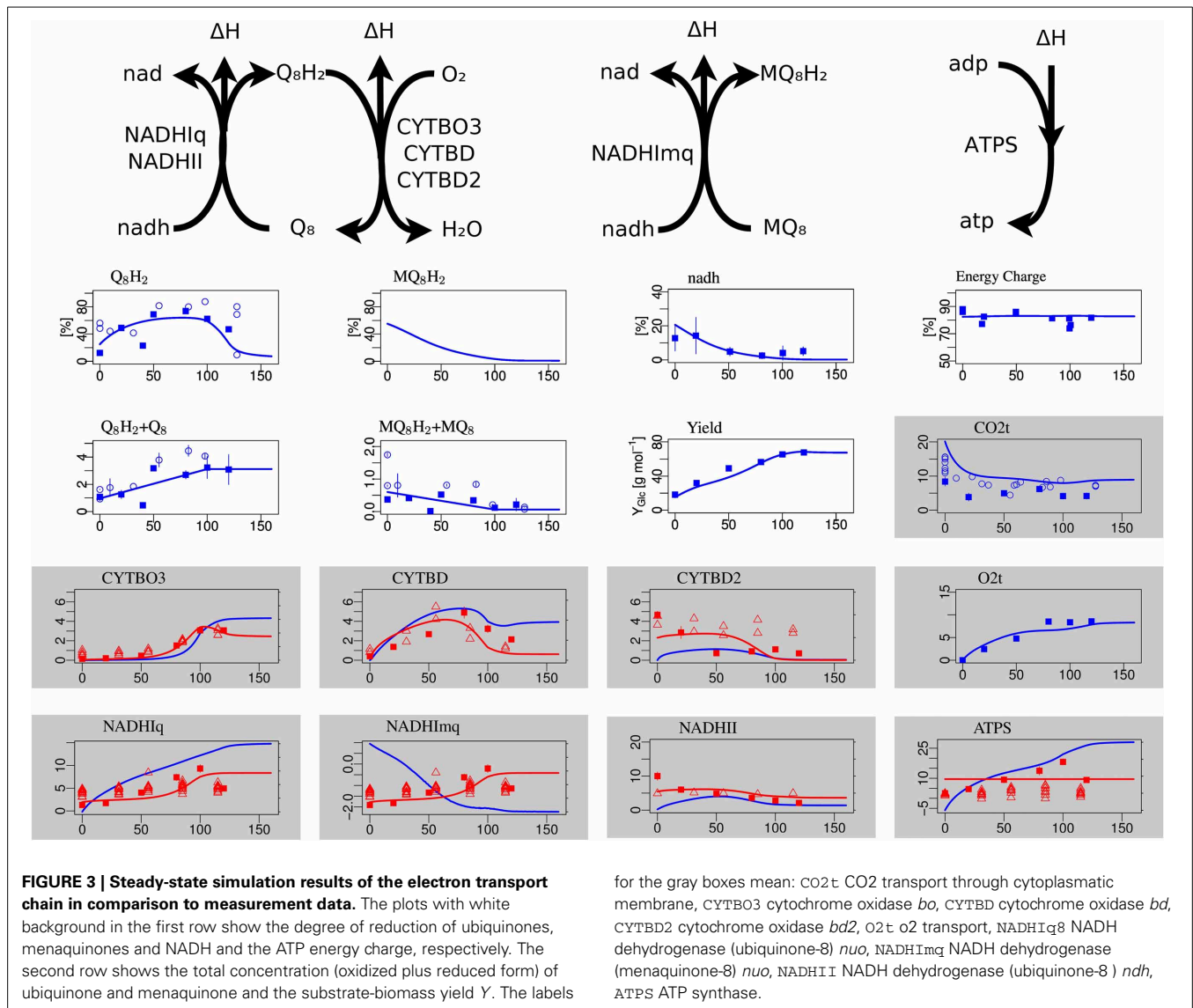
Deviations occur for metabolites in the citric acid cycle (Figure 2) and the pentose phosphate pathway (Supplementary Data Sheet 4). Since the dynamic ranges of these metabolites are small in both, measurement and simulation, and since transcript levels and uptake and production fluxes are described well by the model, these deviations are considered to be minor. Driven by the redox state of NADH, the citric acid cycle shows the switch from its branched form at low aerobiosis values to the cyclic form at high aerobiosis values. The branched mode is characterized by reaction fluxes directed from oxaloacetate (*oaa*) and citrate (*cit*) to succinate (*succ*) and 2-oxoglutarate *akg*, whereas in the cyclic mode the 2-oxoglutarate dehydrogenase *AKGDH* couples both branches and flux occurs in the direction from succinate *succ* to oxaloacetate (*oaa*). Succinate dehydrogenase *SUCDH* and fumarate reductase *FRD* participate in the cyclic and branched modes, respectively. This selectivity is caused by differential gene expression and the different use of quinone species (menaquinone for *FRD* and ubiquinone for *SUCDH*). Thus, the model reflects the expected behavior of the citric acid



cycle for different oxygen availabilities. The model comprises the anaplerotic reaction phosphoenolpyruvate carboxylase PPC and the reverse phosphoenolpyruvate carboxykinase PPCK as well as the anaplerotic glyoxylate shunt realized by isocitrate lyase ICL and malate synthase MALS. The model is able to fit the data for different distributions of anaplerosis over both possible pathways (PPC/PPCK and ICL/MALS), and the simulation results present only one possibility.

The activities of the two measured transcription factors in the simulation and measurement are in agreement (Figure 4). FNR and ArcA activities follow the expected pattern of high activity at low aerobiosis values and low activity at high aerobiosis

values (Sawers, 1999). The model provides predictions for the behavior of several other transcription factors that could not be measured. An indirect validation of these predictions is provided by the good fits of the transcript data. Some glycolytic enzymes are repressed by FruR and the slight increase of FruR activity across the aerobiosis scale is likely. The increase of CRP can be seen for example in the increased expression of the *mgl* operon that is activated by CRP. The *mgl* operon encodes a methyl- $\beta$ -D-galactoside and galactose ABC transporter. Since this transporter is also able to transport glucose and is active under glucose-limited conditions (Death and Ferenci, 1993; Ferenci, 1996; Hua et al., 2004; Steinsiek and Bettenbrock, 2012) it is



modeled as a glucose transporter (GLCabc). The simulated CRP activity is also consistent with the inferred CRP activity from Rolfe et al. (2012). Since FruR activity is controlled by the concentration of fructose 1,6-bisphosphate ( $\text{f}_\text{dp}$ ) and CRP activity is controlled by the ratio of phosphoenolpyruvate ( $\text{p}_\text{ep}$ ) to pyruvate ( $\text{p}_\text{yr}$ ), these predictions follow directly from the distribution of concentrations in glycolysis. The observation that PdhR is more active under aerobic conditions than under

anaerobic conditions (Supplementary Data Sheet 4) is consistent with the experimental observation of Ogasawara et al. (2007). AppY is known to be a regulator active under anaerobic conditions (Brøndsted and Atlung, 1996; Atlung et al., 1997) as it is in the model (Supplementary Data Sheet 4). The transcript levels of both target genes of IclR (MAL5 and ICL in Figure 2) match the predicted course of IclR activity (Supplementary Data Sheet 4).

### 3.2. ASSESSMENT OF PRODUCTION CAPABILITIES

We use the above described model to assess the production capabilities of *E. coli* in a chemostat. For this purpose, we introduce a reaction into the model that represents the effect of a production pathway. For example, if a pathway uses the precursors A and B and 2 ATP to produce a product (possibly via some intermediates), we introduce the reaction  $A + B + 2\text{ATP} \rightarrow 2\text{ADP}$ . For every production pathway, we show steady state values of several quantities depending on the enforced biomass-specific production flux  $J_{\text{prod}}$  given in mmol/g/h. These quantities are the biomass concentration, concentrations of precursors of the production pathway and the overall productivity per reactor volume  $q_{\text{prod}}$  in mmol/l/h. The productivity is given by  $q_{\text{prod}} = J_{\text{prod}} \cdot c_X$  where  $c_X$  is the biomass concentration in g/l. In this way, we can assess the production capabilities independently of the kinetics of the production pathway.

**Figure 5** shows results for hypothetical production pathways where each pathway requires only a single precursor molecule. These results demonstrate the abilities and limitations of the central metabolism to provide biosynthesis with precursors. The aerobic and the anaerobic case are shown. For the aerobic case aeration is fixed to a level that results in 160% aerobiosis of the undisturbed case ( $J_{\text{prod}} = 0$ ). Plots of all twelve precursor molecules are shown in the **Supplementary Data Sheet 5**.

The concentration of glucose 6-phosphate (g6p) increases with production flux, i.e., the consumption rate of glucose 6-phosphate by the production pathway, under aerobic and anaerobic conditions (**Figure 5**). This seemingly paradoxical behavior is explained by the autocatalytic structure of glycolysis. Glucose is phosphorylated to glucose 6-phosphate using ATP (or phosphoenolpyruvate). ATP is regained by metabolism of glucose 6-phosphate. An additional consumption of glucose 6-phosphate leads to an increase of the glucose uptake flux. An increased glucose uptake requires more ATP for phosphorylation of glucose. Since a fixed flux of ATP is required to form biomass in the steady state chemostat, the flux away from glucose 6-phosphate increases and provides a higher ATP production flux. Because the ATP concentration is kept nearly constant and

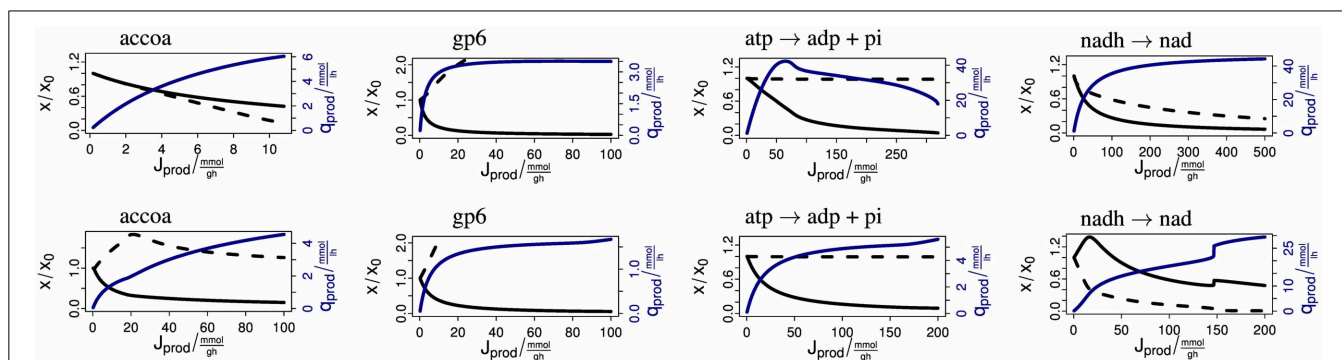
since glycolytic enzymes are nearly constitutively expressed, an increase of the glucose 6-phosphate concentration is needed to drive this flux in order to achieve a steady state. Depending on the kinetics of the production pathway, this behavior may cause dynamic instabilities such as oscillations.

Aerobically, the concentration of acetyl-CoA (accoa) and the biomass concentration drop with an increasing consumption of acetyl-CoA (**Figure 5** upper row). Anaerobically, the concentration of acetyl-CoA increases initially and then decreases slightly (**Figure 5** lower row). However, the biomass concentration decreases rapidly. In the model, up to 6 mmol/l/h acetyl-CoA can be produced aerobically and up to 5 mmol/l/h anaerobically. Due to the much smaller biomass concentration in the anaerobic case, this corresponds to a much higher biomass-specific flux in the anaerobic case.

Aerobically and anaerobically, the concentration of ATP (atp) decreases only slightly with a consumption of ATP (**Figure 5**). This is in line with the observation that the energy charge is kept constant over a wide range of conditions (Chapman et al., 1971). In the model this behavior emerges from the activation of the phosphofructokinase by ADP and by the high sensitivity of growth and maintenance with respect to the ATP/ADP ratio. Aerobically, we obtain the maximal volume-specific ATP production flux for intermediate values of the biomass-specific ATP production flux, whereas for the anaerobic case volume-specific and biomass-specific flux are related monotonously.

Aerobically, the oxidation of NADH (nadh) from the cell leads to a decrease of the (already low) NADH concentration and the biomass (**Figure 5** upper row). But anaerobically, the oxidation of NADH has initially a positive effect on the population size (**Figure 5** lower row), because it removes the requirement for the excretion of carbon in the form of ethanol to maintain the electron balance.

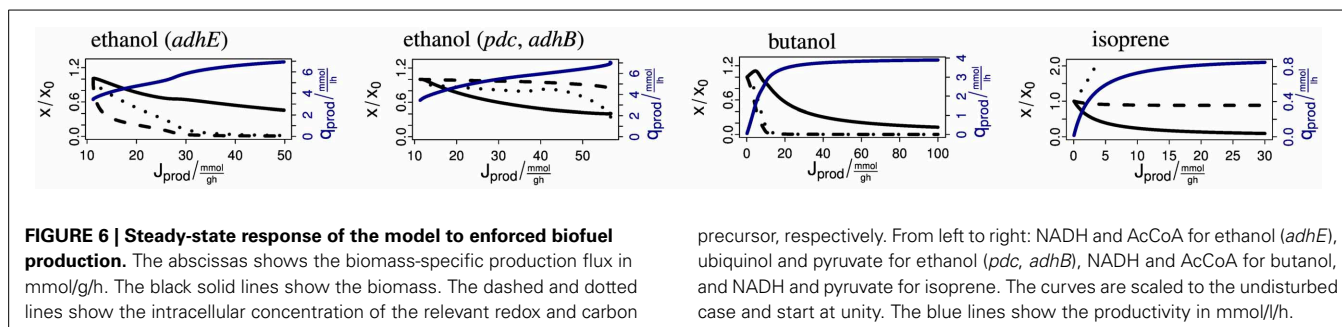
The above examples show that the model provides detailed predictions about the complex effects of production pathways on central metabolism. The synthesis of any biofuel requires a combination of precursor molecules. **Figure 6** shows four anaerobic example cases. (1) *E. coli* naturally produces ethanol via



**FIGURE 5 | Steady-state response of the model to enforced precursor consumption rates.** The upper and lower rows show the results for the aerobic and anaerobic case, respectively. The abscissas shows the biomass-specific production flux in mmol/g/h. The black solid lines show the biomass. The dashed lines show the intracellular

concentration of the respective metabolite. The curves are scaled to the undisturbed case and start at unity. The blue lines show the productivity in mmol/l/h. Labels mean accoa acetyl-CoA, g6p glucose 6-phosphate, atp → adp + pi dephosphorylation of ATP, nadh → nad oxidation of NADH.





the aldehyde-alcohol dehydrogenase AdhE with stoichiometry  $\text{AcCoA} + 2\text{NADH} \rightarrow \text{CoA} + 2\text{NAD} + \text{ethanol}$ . If the flux is increased (for example by overexpression of AdhE), the AcCoA and NADH concentrations drop rapidly. (2) Ohta et al. (1991) expressed the ethanol production pathway from *Zymomonas mobilis* in *E. coli*. The pathway uses the pyruvate decarboxylase Pdc and the alcohol dehydrogenase II AdhB with the stoichiometry  $\text{pyruvate} + \text{ubiquinol} \rightarrow \text{ubiquinone} + \text{ethanol}$ . When enforcing an increasing flux via this pathway, the concentrations of the precursors ubiquinol and pyruvate drop initially only slightly. Only immediately before reaching the maximum productivity do the concentrations decrease. This demonstrates the superior properties of this pathway over the native pathway. **Figure 6** shows that pyruvate supply becomes limiting. (3) Inui et al. (2008) uses a pathway from *Clostridium acetobutylicum* for the production of butanol in *E. coli*. Its total stoichiometry is  $2\text{AcCoA} + 4\text{NADH} \rightarrow 2\text{CoA} + 4\text{NAD} + \text{butanol}$ . The concentrations of the precursors AcCoA and NADH decrease strongly with the production flux. The biomass concentration initially rises because the removal of reducing power is advantageous under anaerobic conditions. (4) The last example is the production of isoprene via the methylerythritol phosphate (MEP) pathway (Zhao et al., 2011). The overall stoichiometry is  $\text{glyceraldehyde 3-phosphate} + \text{pyruvate} + 2\text{NADPH} + \text{NADH} + 1\text{ATP} \rightarrow \text{isoprene} + 2\text{NADP} + \text{NAD} + \text{AMP} + \text{ADP} + \text{diphosphate} + \text{CO}_2$ . When enforcing an increasing flux over this pathway, the biomass concentration decreases strongly until it approaches zero. The ATP concentration stays almost constant, the concentrations of NADH and NADPH (not shown) drop slightly and the concentrations of pyruvate and glyceraldehyde 3-phosphate (not shown) increase. Since biomass is clearly the limiting resource in this example, supporting biomass production by providing a complex medium may increase productivity.

The above examples demonstrate the use of the model for assessing the production capabilities of *E. coli*. We enforce the consumption rate of precursor molecules in stoichiometric relations corresponding to the requirements for the synthesis of biofuel. The analysis of the steady state response of the precursor concentrations, the biomass concentration and the volume-specific productivity gives a detailed picture of the capacity of the metabolism in chemostat conditions. In an ideal situation a high production flux is possible and biomass and precursor concentrations are stable over a wide range of production rates. In real examples, either the decrease in biomass concentration or the decrease in concentration of a precursor limits

the productivity of the pathway. The model gives detailed predictions about the responses of metabolism to specific production pathways. For example, the model makes predictions for which biomass-specific production flux the highest volume-specific productivity occurs. The model predicts how precursor concentrations change with varying production rate. For pathways with several precursors, the model shows which precursor concentrations decrease strongly and become limiting. In addition, complex phenomena can be observed. For example, it is possible that, due to autocatalytic effects, the concentration of a precursor increases despite an enforced consumption rate of the precursor. Depending on the kinetics of the production pathway, this may lead to dynamic instabilities that could be manifest as oscillations.

These results are predictions of the model. The approach of dissecting the parameters into an experiment- and organism-independent class and a class that is variable within bounds, assures that the resulting model is generic but can be adapted to specific experimental situations. The validity of model predictions decreases when one departs from the experimental conditions that were used to determine the values of the free parameters. Under changed experimental conditions, other, currently unmodeled, regulatory systems may play a relevant role. Changed experimental conditions may be the introduction of production pathways, as discussed here, the use of different substrates, dilution rates or strains. However, since the model by construction cannot violate the basic mass balance and thermodynamic constraints, the predictions will share the features enforced by these constraints.

#### 4. CONCLUSION

To our knowledge, this model presents the first comprehensive, integrated model of steady state metabolism and regulation of the oxygen response of *E. coli*. A model of metabolism describing metabolite concentrations and reaction fluxes is complemented with a model of genetic regulation describing gene expression and transcription factor activity. This model of cellular metabolism is embedded into a model of growth describing the requirement of precursor for biomass formation, a model of maintenance describing the non-growth-associated needs for ATP, and a model of the chemostat describing the concentrations of extracellular compounds and biomass. The model provides a thermodynamically consistent view that integrates experimental data at the metabolite, flux, transcript, and regulator levels.

The combined model reflects the major constraints that restrict the behavior of *E. coli*. The use of the thermokinetic



modeling formalism and the availability of the Gibbs formation energies of the metabolites assures that the metabolic model is thermodynamically consistent. This means that the energetic constraints of the cellular metabolism are properly addressed. The mass balances of all relevant metabolites of central metabolism are considered. In particular the model includes also the balances of the biosynthetic precursors and of important energy and redox carriers, as ATP/ADP/AMP, NADH/NAD, and NADPH/NADP. Thus, the model integrates major mass balance constraints with the respective energetic constraints and the respective cellular regulation. The model is able to describe the competition of different pathways for energy and carbon and provides predictions of the effect of the introduction of production pathways on central metabolism.

## FUNDING

We thank the ERASysbio SysMO (Systems Biology of Microorganisms) initiative for funding the SUMO and SUMO2 consortia. The research was funded by the Biotechnology and Biological Sciences Research Council (BBSRC), the Bundesministerium für Bildung und Forschung (BMBF) and the Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO). ME further acknowledges support by the Ministerium für Wissenschaft, Forschung und Kunst Baden-Württemberg within the Ideenwettbewerb Biotechnologie und Medizintechnik Baden-Württemberg.

## ACKNOWLEDGMENTS

We would like to thank all members of the SysMO SUMO consortium for fruitful discussions and collaborations especially M. Bekker, F. Bruggemann, B. Cseke, K. J. Hellingwerf, M. Holcombe, E. D. Gilles, A. Graham, S. Henkel, W. Jia, A. Maleki-Dizaji, T. Nye, G. Sanguinetti.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fmicb.2014.00124/abstract>

**Supplementary Data Sheet 1 | Index of Model Elements.**

**Supplementary Data Sheet 2 | Model Definition File.**

**Supplementary Data Sheet 3 | Transcription Factors and Their Metabolic Signals.**

**Supplementary Data Sheet 4 | Map of Measurement Data and Model Results.**

**Supplementary Data Sheet 5 | Steady State Answer of the Model to enforced precursor effluxes.**

## REFERENCES

- Alberty, R. A. (2003). *Thermodynamics of Biochemical Reactions*. Hoboken, NJ: Wiley-Interscience. doi: 10.1002/0471332607
- Alexeeva, S., Hellingwerf, K. J., and Teixeira de Mattos, M. J. (2002). Quantitative assessment of oxygen availability: perceived aerobiosis and its effect on flux distribution in the respiratory chain of *Escherichia coli*. *J. Bacteriol.* 184, 1402–1406. doi: 10.1128/JB.184.5.1402-1406.2002
- Alexeeva, S., Hellingwerf, K. J., and Teixeira de Mattos, M. J. (2003). Requirement of ArcA for redox regulation in *Escherichia coli* under microaerobic but not anaerobic or aerobic conditions. *J. Bacteriol.* 185, 204–209. doi: 10.1128/JB.185.1.204-209.2003
- Alexeeva, S., Kort, B. D. E., Sawers, G., Hellingwerf, K. J., and Teixeira de Mattos, M. J. (2000). Effects of limited aeration and of the ArcAB system on intermediary pyruvate catabolism in *Escherichia coli*. *J. Bacteriol.* 182, 4934–4940. doi: 10.1128/JB.182.17.4934-4940.2000
- Alvarez, A. F., Rodriguez, C., and Georgellis, D. (2013). Ubiquinone and menaquinone electron carriers represent the yin and yang in the redox regulation of the ArcB sensor kinase. *J. Bacteriol.* 195, 3054–3061. doi: 10.1128/JB.00406-13
- Atlung, T., Knudsen, K., Heerfordt, L., and Brøndsted, L. (1997). Effects of sigmaS and the transcriptional activator AppY on induction of the *Escherichia coli* *hya* and *cbdAB-appA* operons in response to carbon and phosphate starvation. *J. Bacteriol.* 179, 2141–2146.
- Beard, D. A. (2005). A biophysical model of the mitochondrial respiratory system and oxidative phosphorylation. *PLoS Comput. Biol.* 1:e36. doi: 10.1371/journal.pcbi.0010036
- Bekker, M., Alexeeva, S., Laan, W., Sawers, G., Teixeira de Mattos, M. J., and Hellingwerf, K. (2010). The ArcBA two-component system of *Escherichia coli* is regulated by the redox state of both the ubiquinone and the menaquinone pool. *J. Bacteriol.* 192, 746–754. doi: 10.1128/JB.01156-09
- Bekker, M., Kramer, G., Hartog, A. F., Wagner, M. J., de Koster, C. G., Hellingwerf, K. J., et al. (2007). Changes in the redox state and composition of the quinone pool of *Escherichia coli* during aerobic batch-culture growth. *Microbiology (Reading, Engl.)* 153, 1974–1980. doi: 10.1099/mic.0.2007/006098-0
- Borisov, V. B., Murali, R., Verkhovskaya, M. L., Bloch, D. A., Han, H., Gennis, R. B., et al. (2011). Aerobic respiratory chain of *Escherichia coli* is not allowed to work in fully uncoupled mode. *Proc. Natl. Acad. Sci. U.S.A.* 108, 17320–17324. doi: 10.1073/pnas.1108217108
- Brøndsted, L., and Atlung, T. (1996). Effect of growth conditions on expression of the acid phosphatase (*cyx-appA*) operon and the *appY* gene, which encodes a transcriptional activator of *Escherichia coli*. *J. Bacteriol.* 178, 1556–1564.
- Brown, P., Hindmarsh, A., and Petzold, L. (1994). Using Krylov methods in the solution of large-scale differential-algebraic systems. *SIAM J. Sci. Comput.* 15, 1467–1488. doi: 10.1137/0915088
- Brown, P., Hindmarsh, A., and Petzold, L. (1998). Consistent initial condition calculation for differential-algebraic systems. *SIAM J. Sci. Comput.* 19, 1495–1512. doi: 10.1137/S1064827595289996
- Brown, P., Hindmarsh, A., and Petzold, L. (2007). *DASKR Package: DAE Solver with Krylov Methods and Rootfinding*. Available online at: <http://www.netlib.org/ode/>
- Chapman, A. G., Fall, L., and Atkinson, D. E. (1971). Adenylate energy charge in *Escherichia coli* during growth and starvation. *J. Bacteriol.* 108, 1072–1086.
- Cossins, B., Jacobson, M., and Guallar, V. (2011). A new view of the bacterial cytosol environment. *PLoS Comput. Biol.* 7:e1002066. doi: 10.1371/journal.pcbi.1002066
- Death, A., and Ferenci, T. (1993). The importance of the binding-protein-dependent Mgl system to the transport of glucose in *Escherichia coli* growing on low sugar concentrations. *Res. Microbiol.* 144, 529–537. doi: 10.1016/0923-2508(93)90002-J
- Ederer, M. (2010). *Thermokinetic Modeling and Model Reduction of Reaction Networks*. Dissertation, University of Stuttgart, Stuttgart.
- Ederer, M., and Gilles, E. D. (2007). Thermodynamically feasible kinetic models of reaction networks. *Biophys. J.* 92, 1846–1857. doi: 10.1529/biophysj.106.094094
- Ferenci, T. (1996). Adaptation to life at micromolar nutrient levels: the regulation of *Escherichia coli* glucose transport by endoinduction and cAMP. *FEMS Microbiol. Rev.* 18, 301–317. doi: 10.1111/j.1574-6976.1996.tb00246.x
- Georgellis, D., Lynch, A., and Lin, E. (1997). *In vitro* phosphorylation study of the arc two-component signal transduction system of *Escherichia coli*. *J. Bacteriol.* 179, 5429–5435.
- Gonçalves, E., Bucher, J., Ryll, A., Niklas, J., Mauch, K., Klamt, S., et al. (2013). Bridging the layers: towards integration of signal transduction, regulation and metabolism into mathematical models. *Mol. Biosyst.* 9, 1576–1583. doi: 10.1039/c3mb25489e
- Hua, Q., Yang, C., Oshima, T., Mori, H., and Shimizu, K. (2004). Analysis of gene expression in *Escherichia coli* in response to changes of growth-limiting nutrient in chemostat cultures. *Appl. Environ. Microbiol.* 70, 2354–2366. doi: 10.1128/AEM.70.4.2354-2366.2004
- Inglede, W. J., and Poole, R. K. (1984). The respiratory chains of *Escherichia coli*. *Microbiol. Rev.* 48, 222–271.

- Inui, M., Suda, M., Kimura, S., Yasuda, K., Suzuki, H., Toda, H., et al. (2008). Expression of *Clostridium acetobutylicum* butanol synthetic genes in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 77, 1305–1316. doi: 10.1007/s00253-007-1257-5
- Jordan, P. A., Thomson, A. J., Ralph, E. T., Guest, J. R., and Green, J. (1997). FNR is a direct oxygen sensor having a biphasic response curve. *FEBS Lett.* 416, 349–352. doi: 10.1016/S0014-5793(97)01219-2
- Keseler, I. M., Mackie, A., Peralta-Gil, M., Santos-Zavaleta, A., Gama-Castro, S., Bonnavides-Martínez, C., et al. (2013). EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Res.* 41, D605–D612. doi: 10.1093/nar/gks1027
- Klamt, S., Grammel, H., Straube, R., Ghosh, R., and Gilles, E. D. (2008). Modeling the electron transport chain of purple non-sulfur bacteria. *Mol. Syst. Biol.* 4:156. doi: 10.1038/msb4100191
- Kümmel, A., Panke, S., and Heinemann, M. (2006a). Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. *Mol. Syst. Biol.* 2, 20060034. doi: 10.1038/msb4100074
- Kümmel, A., Panke, S., and Heinemann, M. (2006b). Systematic assignment of thermodynamic constraints in metabolic network models. *BMC Bioinformatics* 7:512. doi: 10.1186/1471-2105-7-512
- Lengeler, J. W., Drews, G., and Schlegel, H. G. (1998). *Biology of the Prokaryotes*. Stuttgart: Thieme.
- Link, H., Anselment, B., and Weuster-Botz, D. (2008). Leakage of adenylates during cold methanol/glycerol quenching of *Escherichia coli*. *Metabolomics* 4, 240–247. doi: 10.1007/S11306-008-0114-6
- Neidhardt, F. C., Ingraham, J. L., and Schaechter, M. (1990). *Physiology of the Bacterial Cell: A Molecular Approach*. Sunderland, MA: Sinauer Associates Inc., U.S.
- Ogasawara, H., Ishida, Y., Yamada, K., Yamamoto, K., and Ishihama, A. (2007). PdhR (pyruvate dehydrogenase complex regulator) controls the respiratory electron transport system in *Escherichia coli*. *J. Bacteriol.* 189, 5534–5541. doi: 10.1128/JB.00229-07
- Ohta, K., Beall, D. S., Mejia, J. P., Shanmugam, K. T., and Ingram, L. O. (1991). Genetic improvement of *Escherichia coli* for ethanol production: chromosomal integration of *Zymomonas mobilis* genes encoding pyruvate decarboxylase and alcohol dehydrogenase II. *Appl. Environ. Microbiol.* 57, 893–900.
- Partridge, J. D., Sanguinetti, G., Dibden, D. P., Roberts, R. E., Poole, R. K., and Green, J. (2007). Transition of *Escherichia coli* from aerobic to micro-aerobic conditions involves fast and slow reacting regulatory components. *J. Biol. Chem.* 282, 11230–11237. doi: 10.1074/jbc.M700728200
- Partridge, J. D., Scott, C., Tang, Y., Poole, R. K., and Green, J. (2006). *Escherichia coli* transcriptome dynamics during the transition from anaerobic to aerobic conditions. *J. Biol. Chem.* 281, 27806–27815. doi: 10.1074/jbc.M603450200
- Peercy, B. E., Cox, S. J., Shalel-Levanon, S., San, K.-Y., and Bennett, G. (2006). A kinetic model of oxygen regulation of cytochrome production in *Escherichia coli*. *J. Theor. Biol.* 242, 547–563. doi: 10.1016/j.jtbi.2006.04.006
- Reed, J., Vo, T., Schilling, C., and Palsson, B. (2003). An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biol.* 4:R54. doi: 10.1186/gb-2003-4-9-r54
- Ritter, J. B., Genzel, Y., and Reichl, U. (2006). High-performance anion-exchange chromatography using on-line electrolytic eluent generation for the determination of more than 25 intermediates from energy metabolism of mammalian cells in culture. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* 843, 216–226. doi: 10.1016/j.jchromb.2006.06.004
- Ritter, J. B., Genzel, Y., and Reichl, U. (2008). Simultaneous extraction of several metabolites of energy metabolism and related substances in mammalian cells: optimization using experimental design. *Anal. Biochem.* 373, 349–369. doi: 10.1016/j.ab.2007.10.037
- Rolfe, M. D., Ocone, A., Stapleton, M. R., Hall, S., Trotter, E. W., Robert, K., et al. (2012). Systems analysis of transcription factor activities in environments with stable and dynamic oxygen concentrations. *Open Biol.* 2:120091. doi: 10.1098/rsob.120091
- Rolfe, M. D., Ter Beek, A., Graham, A. I., Trotter, E. W., Asif, H. M. S., Sanguinetti, G., et al. (2011). Transcript profiling and inference of *Escherichia coli* K-12 ArcA activity across the range of physiologically relevant oxygen concentrations. *J. Biol. Chem.* 286, 10147–10154. doi: 10.1074/jbc.M110.211144
- Sawers, G. (1999). The aerobic/anaerobic interface. *Curr. Opin. Microbiol.* 2, 181–187. doi: 10.1016/S1369-5274(99)80032-0
- Schaub, J., and Reuss, M. (2008). *In vivo* dynamics of glycolysis in *Escherichia coli* shows need for growth rate dependent metabolome analysis. *Biotechnol. Prog.* 24, 1402–1407. doi: 10.1002/btpr.59
- Sharma, P., Stagge, S., Bekker, M., Bettenbrock, K., and Hellingwerf, K. J. (2013). Kinase activity of ArcB from *Escherichia coli* is subject to regulation by both ubiquinone and demethylmenaquinone. *PLoS ONE* 8:e75412. doi: 10.1371/journal.pone.0075412
- Shimizu, K. (2009). Toward systematic metabolic engineering based on the analysis of metabolic regulation by the integration of different levels of information. *Biochem. Eng. J.* 46, 235–251. doi: 10.1016/j.bej.2009.06.006
- Smith, H. L., and Waltman, P. (2008). *The Theory of the Chemostat: Dynamics of Microbial Competition (Cambridge Studies in Mathematical Biology)*. 2nd Edn. Cambridge: Cambridge University Press.
- Steinsiek, S., and Bettenbrock, K. (2012). Glucose transport in *Escherichia coli* mutant strains with defects in sugar transport systems. *J. Bacteriol.* 194, 5897–5908. doi: 10.1128/JB.01502-12
- Steinsiek, S., Frixel, S., Stagge, S., and Bettenbrock, K. (2011). Characterization of *E. coli* MG1655 and *frdA* and *sdhC* mutants at various aerobiosis levels. *J. Biotechnol.* 154, 35–45. doi: 10.1016/j.jbiotec.2011.03.015
- Trotter, E. W., Rolfe, M. D., Hounslow, A. M., Craven, C. J., Williamson, M. P., Sanguinetti, G., et al. (2011). Reprogramming of *Escherichia coli* K-12 metabolism during the initial phase of transition from an anaerobic to a micro-aerobic environment. *PLoS ONE* 6:e25501. doi: 10.1371/journal.pone.0025501
- Valgepea, K., Adamberg, K., Nahku, R., Lahtvee, P.-J., Arike, L., and Vilu, R. (2010). Systems biology approach reveals that overflow metabolism of acetate in *Escherichia coli* is triggered by carbon catabolite repression of acetyl-CoA synthetase. *BMC Syst. Biol.* 4:166. doi: 10.1186/1752-0509-4-166
- Varma, A., Boesch, B. W., and Palsson, B. O. (1993). Stoichiometric interpretation of *Escherichia coli* glucose catabolism under various oxygenation rates. *Appl. Environ. Microbiol.* 59, 2465–2473.
- Wolfram Research, I. (2010). *Mathematica Edition: Version 8.0*. Available online at: <http://www.wolfram.com/mathematica/>
- Zhao, Y., Yang, J., Qin, B., Li, Y., Sun, Y., Su, S., et al. (2011). Biosynthesis of isoprene in *Escherichia coli* via methylerythritol phosphate (MEP) pathway. *Appl. Microbiol. Biotechnol.* 90, 1915–1922. doi: 10.1007/s00253-011-3199-1

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 15 January 2014; accepted: 11 March 2014; published online: 27 March 2014.

Citation: Ederer M, Steinsiek S, Stagge S, Rolfe MD, Ter Beek A, Knies D, Teixeira de Mattos MJ, Sauter T, Green J, Poole RK, Bettenbrock K and Sawodny O (2014) A mathematical model of metabolism and regulation provides a systems-level view of how *Escherichia coli* responds to oxygen. *Front. Microbiol.* 5:124. doi: 10.3389/fmicb.2014.00124

This article was submitted to Microbial Physiology and Metabolism, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Ederer, Steinsiek, Stagge, Rolfe, Ter Beek, Knies, Teixeira de Mattos, Sauter, Green, Poole, Bettenbrock and Sawodny. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Death by a thousand cuts: the challenges and diverse landscape of lignocellulosic hydrolysate inhibitors

Jeff S. Piotrowski\*, Yaoping Zhang, Donna M. Bates, David H. Keating, Trey K. Sato, Irene M. Ong and Robert Landick

DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, Madison, WI, USA

## Edited by:

Shihui (Shane) Yang, National Renewable Energy Laboratory, USA

## Reviewed by:

Shiyong Peng, National Institute of Health, USA

Qiang J. Fei, National Renewable Energy Lab, USA

## \*Correspondence:

Jeff S. Piotrowski, DOE Great Lakes Bioenergy Research Center, University of Wisconsin-Madison, 1552 University Ave., WE14152, Madison, WI 53726, USA  
e-mail: jpiotrowski@wisc.edu

Lignocellulosic hydrolysate (LCH) inhibitors are a large class of bioactive molecules that arise from pretreatment, hydrolysis, and fermentation of plant biomass. These diverse compounds reduce lignocellulosic biofuel yields by inhibiting cellular processes and diverting energy into cellular responses. LCH inhibitors present one of the most significant challenges to efficient biofuel production by microbes. Development of new strains that lessen the effects of LCH inhibitors is an economically favorable strategy relative to expensive detoxification methods that also can reduce sugar content in deconstructed biomass. Systems biology analyses and metabolic modeling combined with directed evolution and synthetic biology are successful strategies for biocatalyst development, and methods that leverage state-of-the-art tools are needed to overcome inhibitors more completely. This perspective considers the energetic costs of LCH inhibitors and technologies that can be used to overcome their drain on conversion efficiency. We suggest academic and commercial research groups could benefit by sharing data on LCH inhibitors and implementing “translational biofuel research.”

**Keywords:** cellulosic biofuels, lignocellulosic hydrolysate inhibitors, systems biology, chemical genomics, metabolic modeling, ethanologens

## INTRODUCTION

Lignocellulosic biofuels offer the promise of sustainable, domestically produced fuels with favorable carbon balances. Fast-growing grasses like *Miscanthus* and agricultural residues provide fermentable sugars at lower energy and fertilizer costs than grains (Schmer et al., 2008), making them preferable feedstocks for advanced biofuels. Cellulosic ethanol is an obvious next-generation biofuel to implement given its production and delivery infrastructures are compatible with existing fuels.

Central to the success of cellulosic ethanol is efficient conversion of biomass-derived sugars to ethanol by microbes such as *Saccharomyces cerevisiae*, *Escherichia coli*, and *Zymomonas mobilis* (Alper and Stephanopoulos, 2009; Lau et al., 2010; Yang et al., 2010a). Under optimal conditions, these microbes are powerful ethanologens; however, lignocellulosic hydrolysates (LCH) and industrial scale fermentation tanks are not optimal conditions. Thermal, osmotic, and ethanol stresses are just some of the environmental factors that inhibit fermentation and reduce yield (Attfield, 1997; Gibson et al., 2007; Jin et al., 2013). Industrial microbes are pushed to the limits of stress tolerance to make biofuel production energetically favorable.

Although environmental stressors limit yields in present day ethanol facilities, cellulosic biomass conversion comes with new challenges. Specifically, LCH inhibitors, a group of small, bioactive molecules can significantly reduce conversion efficiency. LCH inhibitors such as aliphatic acids, furans, and phenolics are released or condensed from cellulose and hemicellulose during pretreatment and hydrolysis (Larsson et al., 1999, 2000; Yang et al., 2010a); however, chemical residues from newer hydrolysis strategies and synergies with biofuel end products (ethanol,

isobutanol) are less well studied. Removal of these inhibitors can be expensive and may reduce titers of fermentable sugars; some estimates suggest that detoxification can remove up to 26% of total fermentable sugars (Larsson et al., 1999). Thus, a preferred strategy is to develop microbial strains with properties that minimize the effects of LCH inhibitors on biofuel yields.

With commercially available industrial stains that are robust to thermal and ethanol stress (e.g., Ethanol Red, Fermentis, Milwaukee, WI, USA), recent attention has been directed to overcoming the challenge of LCH inhibitors. These compounds are ubiquitous in hydrolysates, and their abundance and composition depends on pretreatment (Chundawat et al., 2010), feedstock (Klinke et al., 2004; Almeida et al., 2007), and seasonality (Bunnell et al., 2013; Greenhalf et al., 2013). Given their chemical diversity, these compounds can target many cellular processes. LCH inhibitors can also generate a substantial cellular energy drain. Cells have evolved to detoxify, excrete inhibitors, or repair the resultant cellular damage fast enough to reproduce. However, evolved coping mechanisms may also negatively affect the efficiency of conversion by competing for cellular resources (Bellissimi et al., 2009; Miller et al., 2009). Although it is in the microbe's best interest to use its resources to limit the effects of LCH inhibitors and maintain cellular viability, this may be reducing biofuel production. In this perspective, we consider the diversity and cellular costs of LCH inhibitors from traditional and novel pretreatment and hydrolysis strategies, describe new technologies and their application to strain development, and finally identify key needs of the cellulosic biofuel community that will empower “translational biofuel research” to take discoveries quickly to industrial scale.

## DIVERSITY OF FERMENTATION INHIBITORS

Prior to microbial conversion of lignocellulosic sugars into bio-fuel, biomass must be deconstructed into monomeric sugars by enzymatic or chemical hydrolysis. This hydrolysis step is often preceded by a pretreatment step that expands the plant fibers and allows cellulolytic enzymes access to the polysaccharide matrices. The resulting hydrolysates are complex, ill-defined mixtures that include sugars and a diversity of bioactive molecules (**Table 1**).

Small acids and phenolic compounds are released from cellulose and hemicelluloses during hydrolysis and furans arise from the dehydration of pentose and hexose monomers (Klinke et al., 2004). Pretreatments such as acid hydrolysis, steam explosion or  $\text{NH}_3$  expansion each impart their own “profile” of LCH inhibitors. For example, AFEX (Ammonia Fiber EXpansion) uses high-pressure/temperature ammonia to alter the cellulose matrix to allow hydrolysis by cellulases (Lau and Dale, 2009), and

**Table 1 | Classes of lignocellulosic hydrolysate (LCH) inhibitors and their described modes of toxicity.**

Inhibitors	Mode of action	References
<b>LIGNOCELLULOSE DERIVED</b>		
<b>Small acids</b>		
Acetic acid, formic acid, levulinic acid	Decreases cellular pH, Decreases cellular ATP, Inhibits macromolecule biosynthesis, Inhibits DNA synthesis/repair, Inhibits glycolytic enzymes	Sinha, 1986; Cherrington et al., 1990; Holyoak et al., 1996; Stratford and Anslow, 1998; Bellissimi et al., 2009; Ullah et al., 2012; Ding et al., 2013
<b>Furans</b>		
Furfural, HMF, 2-furoic acid	Damages membranes, Oxidative damage, Damages nucleic acids, Damages proteins, Limits sulfur assimilation, Reduces NADH/NADPH pools, Inhibits enzymes	Ingram, 1976; Hadi et al., 1989; Khan et al., 1995; Zaldivar and Ingram, 1999; Modig et al., 2002; Miller et al., 2009; Allen et al., 2010; Wang et al., 2013
<b>Phenolics</b>		
Ferulic acid, coumaric acid Vanillin, Syringaldehyde Coniferyl alcohol, Eugenol Acetovanillin, Feruloyl amide, Coumaryl amide	Damages membranes, Decreases cellular pH, Decrease cellular ATP, Inhibits translation, Oxidative damage, Denatures proteins, Damages cytoskeleton, DNA mutagenesis, Induces apoptosis	Krebs et al., 1983; Mikulášová et al., 1990; Verduyn et al., 1992; Chambel et al., 1999; Fitzgerald et al., 2004; Iwaki et al., 2013; Nguyen et al., 2014
<b>PROCESS DERIVED INHIBITORS</b>		
<b>Ionic liquids</b>		
1-Ethyl-3-methylimidazolium-Ac	Unknown	Docherty and Kulpa, 2005
<b>Surfactants</b>		
Triton-X, Tween	Damages membranes	King et al., 1991; Laour et al., 1996
<b>Metal ions</b>		
Copper, Sodium, Zirconium	Damages membranes, nucleic acids, and enzymes	Shapiro and Ling, 1998; Schmitt and Tampé, 2002
$\gamma$ -valerolactone	Unknown	Luterbacher et al., 2014
<b>End product inhibitors</b>		
Ethanol Isobutanol	Damages membranes Damages DNA Inhibits enzymes	Nagodawithana and Steinkraus, 1976; Dombek and Ingram, 1984; Alexandre et al., 1994; Ibeas and Jimenez, 1997; Huffer et al., 2011

this produces amide versions of inhibitors (e.g., feruloyl amide from ferulic acid) with potentially new biological properties (Chundawat et al., 2010).

Besides these common inhibitors, residual pretreatment chemicals may complicate fermentation. Ionic liquids are pretreatment and hydrolysis solvents, but are toxic to many microorganisms (Docherty and Kulpa, 2005; Ouellet et al., 2011). Alkaline hydrogen peroxide (AHP) pretreatment limits the production of furans; however this method introduces significant amounts of Na<sup>+</sup> from NaOH, which can cause osmotic stress (Sato et al., 2014). Copper(II) 2,2'-bipyridine is a catalyst that enhances AHP pretreatment by reducing the H<sub>2</sub>O<sub>2</sub> requirement (Li et al., 2013), but copper is toxic to most microbes. Next-generation pre-treatments and hydrolysis methods like  $\gamma$ -valerolactone (Luterbacher et al., 2014), surfactants (Sindhu et al., 2013), zirconium phosphate catalysts (Gliozzi et al., 2014), and other incipient hydrolysis technologies may imbue hydrolysates with novel toxicities and synergisms with common inhibitors.

Biofuel end-products themselves are inhibitory. Ethanol can directly damage cellular membranes, DNA, as well as inhibit enzymes (Nagodawithana and Steinkraus, 1976; Dombek and Ingram, 1984; Alexandre et al., 1994; Ibeas and Jimenez, 1997; Huffer et al., 2011). The ethanologens *S. cerevisiae* and *Z. mobilis* are not immune to ethanol toxicity at high concentrations (Carmona-Gutierrez et al., 2012; Yang et al., 2013). Advanced biofuels like isobutanol are toxic at significantly lower concentrations than ethanol (Brynildsen and Liao, 2009; Atsumi et al., 2010; Huffer et al., 2011; Minty et al., 2011). Inhibition by end products has been an area of research interest (Baez et al., 2011; McEwen and Atsumi, 2012; Zingaro et al., 2013), and ethanol tolerance is a pre-requisite for all industrial yeast.

Effects of inhibitors at a minimum can be additive, but an even greater concern is the potential for synergy between LCH inhibitors and fermentation condition, including high osmolarity and absence of O<sub>2</sub>. Some studies have described synergies between acetic acid, furfural, and phenolics in yeast (Oliva et al., 2006; Ding et al., 2011), but a comprehensive evaluation of synergisms between compounds and conditions on both growth rate and fermentation will be essential. Such assessment will be a massive undertaking that will also require defined synthetic hydrolysate media to permit meaningful definition of minimum inhibitory concentrations (MICs) of the individual inhibitors, and how this value will change in combination with other LCH inhibitors (synergy/antagonism) and fermentation conditions. Nevertheless, documenting interactions between inhibitors on sugar conversion is crucial to prioritizing future research for improved biofuel microbes.

## SMALL MOLECULE INHIBITORS DEplete CELLULAR RESOURCES

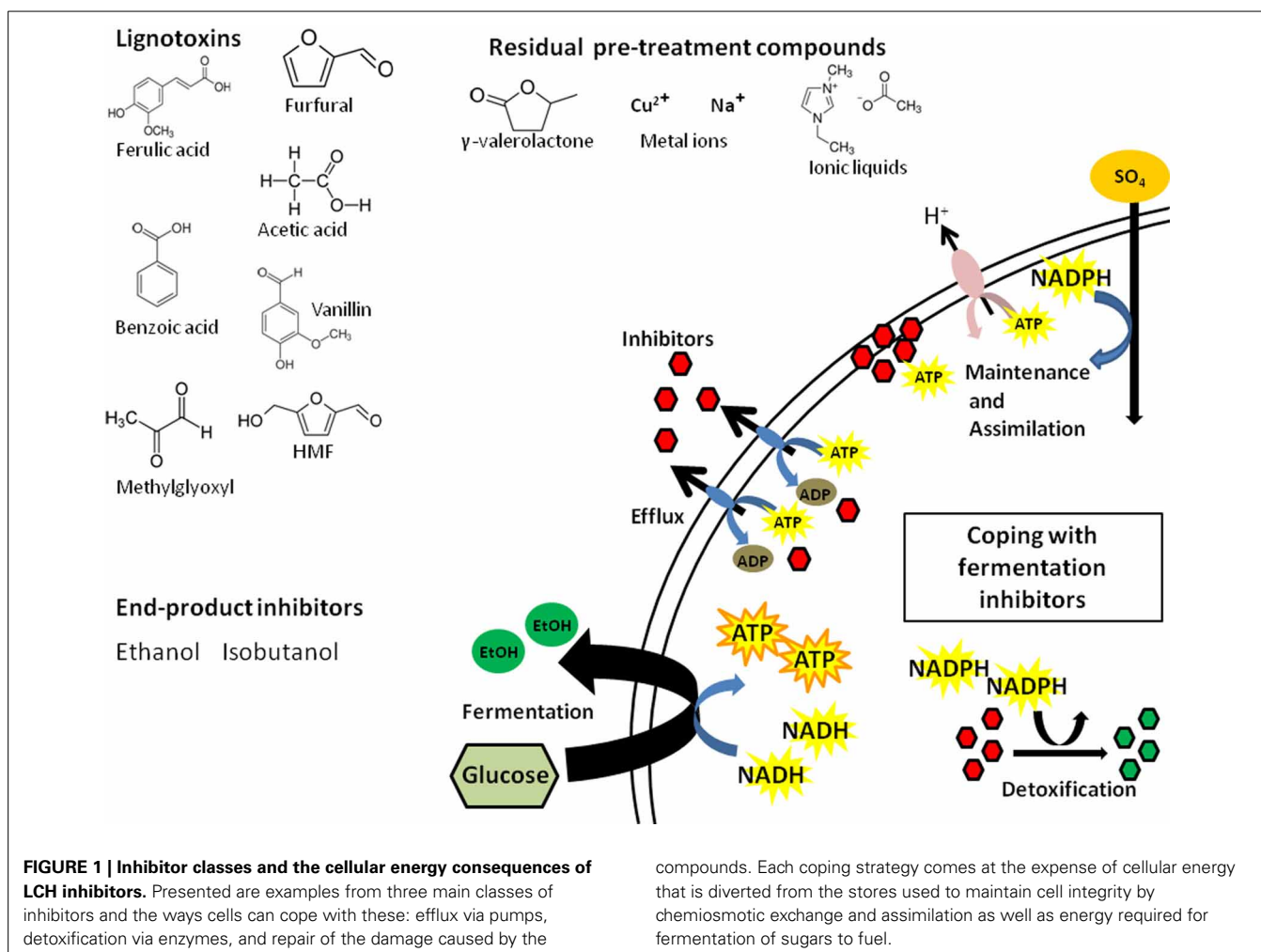
LCH inhibitors directly affect biofuel yield as well as the production rate, which can extend fermentation time and result in higher operating costs. In simplest terms, these inhibitors affect conversion efficiency by depleting cellular energy resources (e.g., ATP, NADH, NADPH; **Figure 1**). Some inhibitors can act broadly and damage key enzymes of fermentation pathways

(Modig et al., 2002). The coping mechanisms available to the biofuel microbes fall into 4 main categories: (i) detoxification, (ii) efflux, (iii) repair, or (iv) tolerance. The first three are of most concern given their effects on cellular energy and resources.

Detoxification and efflux are the most well characterized mechanisms of inhibitor tolerance in microbes, with deep literature not only from biofuel research but also in the medical literature from a wealth of antibiotic and pharmaceutical research. Detoxification is the major route of tolerance for aldehydes in both bacteria and yeast. Reduction of compounds like furfural and vanillin to less toxic alcohols by NADH/NADPH dependent reductases occurs in ethanol fermentation using *S. cerevisiae* and *E. coli* (Jarboe, 2011). Oxidoreductase expression significantly increases in yeast and *E. coli* in the presence of aldehydes (Liu et al., 2008; Wang et al., 2011), depleting cellular NADH/NADPH. This results in inhibition of NADPH-dependent processes (e.g., assimilation of sulfur) leading ultimately to slower conversion of sugars (Miller et al., 2009). Interestingly, yields can be increased by disabling the detoxification pathway, suggesting that tolerance may be more energetically efficient than detoxification (Wang et al., 2013). Alternatively, changing the source of reducing equivalent for aldehyde detox from NADPH to NADH also can improve biofuel yield (Wang et al., 2013).

Efflux is mediated by ATP-dependent trans-membrane pumps that selectively or non-selectively pump out toxic compounds usually at the cost of 1 ATP per molecule (Shapiro and Ling, 1998; Schmitt and Tampé, 2002). The yeast *S. cerevisiae* has 29 different ATP-binding cassette (ABC) efflux transporters (Decottignies and Goffeau, 1997) and 5% of the *E. coli* genome is composed of genes with ABC-transporter domains, many involved in efflux (Linton and Higgins, 1998). In both yeast and *E. coli*, expression of transporters increases in response to LCH inhibitors (Schüller et al., 2004; Lee et al., 2012; Schwalbach et al., 2012). The yeast weak acid response is mediated by increased expression of ABC-transporters (e.g., Pdr1p and Pdr5p) (Schüller et al., 2004; Pereira Rangel et al., 2010). As long as cells are exposed to LCH inhibitors, a significant fraction of cellular ATP will be diverted to efflux pumps. Of particular importance for overall efficiency of LCH conversion, ATP-depletion may have a disproportionate effect on xylose conversion compared to glucose conversion because xylose produces less cellular energy per molecule transported (Matsushika et al., 2013). Bellissimi et al. (2009) found that acetic acid could specifically inhibit xylose fermentation, but that this effect could be reversed with glucose addition. The authors posit that ATP generated from xylose fermentation cannot match ATP depletion from the weak acid response and efflux/proton pumps used to maintain cellular pH. Inhibitors that require ATP-dependent coping mechanisms can directly reduce xylose conversion. Some inhibitors can be particularly draining, affecting both NADPH and ATP pools. Ask et al. (2013) found that furfural and HMF not only reduced cellular NADPH in *S. cerevisiae*, but also elicited increased expression of ATP-dependent efflux pumps Pdr5p and Yor1p. This suggests that coping with furans requires both NADPH dependent detoxification and ATP-dependent efflux.





compounds. Each coping strategy comes at the expense of cellular energy that is diverted from the stores used to maintain cell integrity by chemiosmotic exchange and assimilation as well as energy required for fermentation of sugars to fuel.

An unanswered question is whether the ATP-dependent action of efflux pumps has a net positive (by inhibitor removal) or net negative (by energy consumption) effect on biofuel production. Although the answer may vary by inhibitor, because some inhibitors like aldehydes are more damaging to cells than others, a general test of the positive or negative consequences of efflux pumps for biofuel yield will help advance strategies for biofuel microbe design.

The cellular energy costs of maintenance and repair are more difficult to quantify but could account for significant energy loss. If cells can repair the damage caused by fermentation inhibitors quickly, then fermentation may proceed. Inhibitors can acidify cells (Verduyn et al., 1992), damage cellular membranes (Russell, 1992), DNA (Allen et al., 2010), and individual proteins (Modig et al., 2002). Repairing structures requires biogenesis; this comes at the expense of ATP, NADPH, carbon, and nitrogen. Maintaining pH is mediated by ATP-dependent proton pump Pma1p; and ATP cost under acidic conditions is the primary cause of reduced cellular growth (Verduyn et al., 1992; Ullah et al., 2012). Biogenesis requires NADPH-dependent assimilation of nutrients like sulfur, which is drained by repair enzymes. Phenolics and furans can damage membranes, requiring more

energy to maintain the proton gradient required for basic metabolism (Ding et al., 2012; Schwalbach et al., 2012; Stratford et al., 2013). Growth and sugar conversion will be slowed as resources are diverted to maintenance and repair. Given that hydrolysates contain a mixture of inhibitors with diverse modes of action requiring all of these coping mechanisms simultaneously, the energy drain from fermentation inhibitors is truly death by a thousand cuts.

## ADAPTING TO THE CHANGING LANDSCAPE OF LCH INHIBITORS

Strain development is an economical route to deal with LCH inhibitors. Resistance to the suite of inhibitors requires complex response of many cellular systems, and as such is not easily conferred by engineering of individual genes. Moreover, inhibitor pools can vary between hydrolysate preparations, thus even the most robust strains in ammonia-pretreated hydrolysate may wither in dilute acid pretreated hydrolysates. It is unlikely that one strain will be optimal for all conditions. The reality is that microbial strains will need to be tailored to specific hydrolysates through engineering and directed evolution. Accelerating this process is

crucial to making new cellulosic technologies industrially viable.

The tools of systems biology can give a detailed view of the microbial stress response (Jozefczuk et al., 2010; Lee et al., 2011). Transcriptomic, proteomic, and metabolomic responses to inhibitors can be tracked in detail, and this “multiomic” approach can give high-resolution insight to the global cellular consequences (Miller et al., 2009; Yang et al., 2010b; Schwabach et al., 2012; Skerker et al., 2013; Yang et al., 2013). Advanced techniques such as ribosome profiling can give a view into relationships between transcription and protein abundance in the presence of LCH inhibitors (Ingolia et al., 2009; Brar et al., 2012). Metabolic and flux-balance models are valuable in determining energy balances within cells (Varma and Palsson, 1994; Fiaux et al., 2003; Jin and Jeffries, 2004; Herrgård et al., 2008; Taymaz-Nikerel et al., 2010). These models combined with a systems biology view of protein and gene expression can be used to identify key energetic bottlenecks as targets for engineering. Recently, Wei et al. (2013) demonstrated an elegant way to overcome the redox cofactor imbalance in yeast designed to ferment xylose by engineering acetate metabolism from *E. coli* into *S. cerevisiae*. The authors combined an acetate utilization pathway that consumes NADH with a xylose utilization pathway that produces NADH to overcome the redox imbalance of engineered xylose fermentation. The resultant strain has both better xylose conversion and the ability to detoxify acetate (Wei et al., 2013). Detailed accounting of ATP and NAD(P)H in the presence of LCH inhibitors and industrial conditions will be necessary to disentangle and understand points for rational engineering of microbial catalysts.

Model biofuel microbes like *S. cerevisiae* and *E. coli* benefit from well-developed suites of functional genomics resources, such as deletion mutant or overexpression collections (Giaever et al., 2002; Baba et al., 2006; Kitagawa et al., 2006). These tools have revealed effects of some inhibitors such as furfural (Gorsich et al., 2006), vanillin (Endo et al., 2008; Iwaki et al., 2013), and acetic acid (Mira et al., 2010). Genome-wide mutant collections also allow powerful studies of inhibitors via “chemical genomics” (Giaever et al., 2004; Parsons et al., 2006; Ho et al., 2011). This new tool in the multiomic arsenal, when combined with the information in genetic interaction networks (Butland et al., 2008; Costanzo et al., 2010), can allow precise predictions of the cellular targets of fermentation inhibitors. Recently, Skerker et al. (2013) used a chemical genomics approach to discover a previously undescribed inhibitor in acid pretreated hydrolysate, methyl glyoxal (MG), and identified mutations that confer MG resistance. Chemical genomics can be used to identify the chemical biological signatures within hydrolysates and mutations conferring resistance, but more broadly can serve as a “biological fingerprinting” technique for hydrolysate to identify variation in production, and as a method to benchmark the biological properties of novel hydrolysates. Resources such as the MoBY-ORF collections (Ho et al., 2009; Magtanong et al., 2011), which are barcoded plasmids collections carrying nearly all *S. cerevisiae* used to assess the effects of increased gene dose and gene complementation, could be used with industrial, wild, and engineered yeast to identify genetic interactions within diverse

yeast strains. Combined with traditional selection for resistance and directed evolution, systems biology tools offer great potential for strain development.

Further, new tools are now available that can accelerate strain development. Genome editing and optimization techniques such as CRISPR/Cas9 and MAGE allows rapid, detailed genome editing in both bacteria and eukaryotes (Wang et al., 2012; Cong et al., 2013; DiCarlo et al., 2013; Gilbert et al., 2013). Next steps in strain development will be tuning genomes to inhibitor tolerance by parallel disruption or activation of inhibitor responsive genes using CRISPR/Cas9-based systems. Additionally, large-scale gene synthesis can be used to identify genes that not only aid xylose utilization, but also confer inhibitor tolerance. Systems biology tools for biocatalyst development are coalescing to a pipeline that can keep pace with the changing landscape of fermentation inhibitors, allowing for the rapid tailoring of ethanologens with robust and efficient sugar to biofuel conversion from any new hydrolysate.

## NECESSARY TOOLS TO MEET A COMMON GOAL

Much like the interface between commercial and academic drug discovery communities, applying next-generation biofuel technologies will require cooperative translational research. Academic biofuel communities are developing advanced system biology techniques whereas the commercial community excels at scale-up commercialization. However, similar to drug discovery, these two groups are often isolated in their research. In both cases, public-private partnerships such as the NIH translational medicine initiatives (Zerhouni, 2003) and NCERC industrial partnerships (<http://www.siue.edu/ethanolresearch/>), as well as shared computational resources like Kbase (<http://www.kbase.us/>) can help bridge the divide.

How can researchers and funding agencies best enhance collaboration? A key advance would be greater data sharing about LCH inhibitors. Diverse hydrolysates and their respective inhibitors are major variables in the field of cellulosic biofuel production, but detailed information on specific hydrolysate compositions is not broadly available. Comprehensive efforts to identify all major LCH components and inhibitors across feedstocks and hydrolysis treatments are needed and will require a central repository of LCH data that includes data on feedstocks, pretreatment, hydrolysis, nutrients, and inhibitor. Chemical genomic profiling could be used to generate “biological fingerprints” of hydrolysates to allow comparisons among hydrolysates and identify the biological effects of LCH components by standard analytical methods. A central resource would allow researchers to compare the composition and biological fingerprints of new hydrolysates with existing knowledge about tolerant microbes for further strain development. The DOE’s Systems Biology Portal, KBase, offers the best outlet for community resources, and could serve as the authority of the response of biocatalysts to LCH inhibitors with open-source, community-developed analytical tools for chemical genomics datasets.

## CONCLUSIONS

LCH inhibitors are major barriers for cellulosic biofuels. The cellular energy costs of coping with these compounds are a

significant drain on the already thin margins of biofuel production. However, the increasingly powerful tools of systems biology can be used to gain a detailed understanding of the cellular consequences of individual and mixtures of fermentation inhibitors, which will serve as a basis for rational engineering of customizable microbes. The biofuel research community would benefit from shared computational and database resources that can improve communication between the academic and commercial sides of biofuels.

## ACKNOWLEDGMENTS

All authors are funded by the DOE Great Lakes Bioenergy Research Center (DOE BER Office of Science DE-FC02-07ER64494).

## REFERENCES

- Alexandre, H., Berlot, J. P., and Charpentier, C. (1994). Effect of ethanol on membrane fluidity of protoplasts from *Saccharomyces cerevisiae* and *Kloeckera apiculata* grown with or without ethanol, measured by fluorescence anisotropy. *Biotechnol. Tech.* 8, 295–300. doi: 10.1007/BF02428970
- Allen, S. A., Clark, W., McCaffery, J. M., Cai, Z., Lancot, A., Slininger, P. J., et al. (2010). Furfural induces reactive oxygen species accumulation and cellular damage in *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* 3, 2. doi: 10.1186/1754-6834-3-2
- Almeida, J. R., Modig, T., Petersson, A., Hahn-Hägerdal, B., Lidén, G., and Gorwa-Grauslund, M. F. (2007). Increased tolerance and conversion of inhibitors in lignocellulosic hydrolysates by *Saccharomyces cerevisiae*. *J. Chem. Technol. Biotechnol.* 82, 340–349. doi: 10.1002/jctb.1676
- Alper, H., and Stephanopoulos, G. (2009). Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential? *Nat. Rev. Microbiol.* 7, 715–723. doi: 10.1038/nrmicro2186
- Ask, M., Bettiga, M., Mapelli, V., and Olsson, L. (2013). The influence of HMF and furfural on redox-balance and energy-state of xylose-utilizing *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* 6, 22. doi: 10.1186/1754-6834-6-22
- Atsumi, S., Wu, T.-Y., Machado, I. M. P., Huang, W.-C., Chen, P.-Y., Pellegrini, M., et al. (2010). Evolution, genomic analysis, and reconstruction of isobutanol tolerance in *Escherichia coli*. *Mol. Syst. Biol.* 6, 449. doi: 10.1038/msb.2010.98
- Attfield, P. V. (1997). Stress tolerance: the key to effective strains of industrial baker's yeast. *Nat. Biotechnol.* 15, 1351–1357. doi: 10.1038/nbt1297-1351
- Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., et al. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* 2. doi: 10.1038/msb4100050
- Baez, A., Cho, K.-M., and Liao, J. C. (2011). High-flux isobutanol production using engineered *Escherichia coli*: a bioreactor study with *in situ* product removal. *Appl. Microbiol. Biotechnol.* 90, 1681–1690. doi: 10.1007/s00253-011-3173-y
- Bellissimi, E., Van Dijken, J. P., Pronk, J. T., and Van Maris, A. J. A. (2009). Effects of acetic acid on the kinetics of xylose fermentation by an engineered, xylose-isomerase-based *Saccharomyces cerevisiae* strain. *FEMS Yeast Res.* 9, 358–364. doi: 10.1111/j.1567-1364.2009.00487.x
- Brar, G. A., Yassour, M., Friedman, N., Regev, A., Ingolia, N. T., and Weissman, J. S. (2012). High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science* 335, 552–557. doi: 10.1126/science.1215110
- Brynildsen, M. P., and Liao, J. C. (2009). An integrated network approach identifies the isobutanol response network of *Escherichia coli*. *Mol. Syst. Biol.* 5, 277. doi: 10.1038/msb.2009.34
- Bunnell, K., Rich, A., Luckett, C., Wang, Y.-J., Martin, E., and Carrier, D. J. (2013). Plant maturity effects on the physicochemical properties and dilute acid hydrolysis of switchgrass (*Panicum virgatum*, L.) hemicelluloses. *ACS Sustain. Chem. Eng.* 1, 649–654. doi: 10.1021/sc4000175
- Butland, G., Babu, M., Diaz-Mejia, J. J., Bohdana, F., Phanse, S., Gold, B., et al. (2008). eSGA: *E. coli* synthetic genetic array analysis. *Nat. Methods* 5, 789–795. doi: 10.1038/nmeth.1239
- Carmona-Gutierrez, D., Sommer, C., Andryushkova, A., Kroemer, G., and Madeo, F. (2012). A higher spirit: avoiding yeast suicide during alcoholic fermentation. *Cell Death Differ.* 19, 913–914. doi: 10.1038/cdd.2012.31
- Chambel, A., Viegas, C. A., and Sá-Correia, I. (1999). Effect of cinnamic acid on the growth and on plasma membrane H<sup>+</sup>-ATPase activity of *Saccharomyces cerevisiae*. *Int. J. Food Microbiol.* 50, 173–179. doi: 10.1016/S0168-1605(99)00100-2
- Cherrington, C. A., Hinton, M., and Chopra, I. (1990). Effect of short-chain organic acids on macromolecular synthesis in *Escherichia coli*. *J. Appl. Bacteriol.* 68, 69–74. doi: 10.1111/j.1365-2672.1990.tb02550.x
- Chundawat, S. P. S., Vismeh, R., Sharma, L. N., Humpala, J. F., da Costa Sousa, L., Chambliss, C. K., et al. (2010). Multifaceted characterization of cell wall decomposition products formed during ammonia fiber expansion (AFEX) and dilute acid based pretreatments. *Bioresour. Technol.* 101, 8429–8438. doi: 10.1016/j.biortech.2010.06.027
- Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823. doi: 10.1126/science.1231143
- Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E. D., Sevier, C. S., et al. (2010). The genetic landscape of a cell. *Science* 327, 425–431. doi: 10.1126/science.1180823
- Decottignies, A., and Goffeau, A. (1997). Complete inventory of the yeast ABC proteins. *Nat. Genet.* 15, 137–145. doi: 10.1038/ng0297-137
- DiCarlo, J. E., Norville, J. E., Mali, P., Rios, X., Aach, J., and Church, G. M. (2013). Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.* 41, 4336–4343. doi: 10.1093/nar/gkt135
- Ding, J., Bierma, J., Smith, M. R., Poliner, E., Wolfe, C., Hadduck, A. N., et al. (2013). Acetic acid inhibits nutrient uptake in *Saccharomyces cerevisiae*: auxotrophy confounds the use of yeast deletion libraries for strain improvement. *Appl. Microbiol. Biotechnol.* 97, 7405–7416. doi: 10.1007/s00253-013-5071-y
- Ding, M.-Z., Wang, X., Yang, Y., and Yuan, Y.-J. (2011). Metabolomic study of interactive effects of phenol, furfural, and acetic acid on *Saccharomyces cerevisiae*. *OMICS J. Integr. Biol.* 15, 647–653. doi: 10.1089/omi.2011.0003
- Ding, M.-Z., Wang, X., Yang, Y., and Yuan, Y.-J. (2012). Comparative metabolic profiling of parental and inhibitors-tolerant yeasts during lignocellulosic ethanol fermentation. *Metabolomics* 8, 232–243. doi: 10.1007/s11306-011-0303-6
- Docherty, K. M., and Kulpa, C. F. Jr. (2005). Toxicity and antimicrobial activity of imidazolium and pyridinium ionic liquids. *Green Chem.* 7, 185–189. doi: 10.1039/B419172B
- Dombek, K. M., and Ingram, L. O. (1984). Effects of ethanol on the *Escherichia coli* plasma membrane. *J. Bacteriol.* 157, 233–239.
- Endo, A., Nakamura, T., Ando, A., Tokuyasu, K., and Shima, J. (2008). Genome-wide screening of the genes required for tolerance to vanillin, which is a potential inhibitor of bioethanol fermentation, in *Saccharomyces cerevisiae*. *Biotechnol. Biofuels* 1, 3. doi: 10.1186/1754-6834-1-3
- Fiaux, J., Çakar, Z. P., Sonderegger, M., Wüthrich, K., Szyperski, T., and Sauer, U. (2003). Metabolic-flux profiling of the yeasts *Saccharomyces cerevisiae* and *Pichia stipitis*. *Eukaryot. Cell* 2, 170–180. doi: 10.1128/EC.2.1.170-180.2003
- Fitzgerald, D. J., Stratford, M., Gasson, M. J., Ueckert, J., Bos, A., and Narbad, A. (2004). Mode of antimicrobial action of vanillin against *Escherichia coli*, *Lactobacillus plantarum* and *Listeria innocua*. *J. Appl. Microbiol.* 97, 104–113. doi: 10.1111/j.1365-2672.2004.02275.x
- Giaever, G., Chu, A. M., Ni, L., Connelly, C., Riles, L., Véronneau, S., et al. (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418, 387–391. doi: 10.1038/nature00935
- Giaever, G., Flaherty, P., Kumm, J., Proctor, M., Nislow, C., Jaramillo, D. F., et al. (2004). Chemogenomic profiling: identifying the functional interactions of small molecules in yeast. *Proc. Natl. Acad. Sci. U.S.A.* 101, 793–798. doi: 10.1073/pnas.0307490100
- Gibson, B. R., Lawrence, S. J., Leclaire, J. P. R., Powell, C. D., and Smart, K. A. (2007). Yeast responses to stresses associated with industrial brewery handling. *FEMS Microbiol. Rev.* 31, 535–569. doi: 10.1111/j.1574-6976.2007.00076.x
- Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154, 442–451. doi: 10.1016/j.cell.2013.06.044
- Giozzi, G., Innorta, A., Mancini, A., Bortolo, R., Perego, C., Ricci, M., et al. (2014). Zr/P/O catalyst for the direct acid chemo-hydrolysis of non-pretreated microcrystalline cellulose and softwood sawdust. *Appl. Catal. B Environ.* 145, 24–33. doi: 10.1016/j.apcatb.2012.12.035
- Gorsich, S. W., Dien, B. S., Nichols, N. N., Slininger, P. J., Liu, Z. L., and Skory, C. D. (2006). Tolerance to furfural-induced stress is associated with pentose

- phosphate pathway genes ZWF1, GND1, RPE1, and TKL1 in *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* 71, 339–349. doi: 10.1007/s00253-005-0142-3
- Greenhalf, C. E., Nowakowski, D. J., Yates, N., Shield, I., and Bridgwater, A. V. (2013). The influence of harvest and storage on the properties of and fast pyrolysis products from *Miscanthus x giganteus*. *Biomass Bioenergy* 56, 247–259. doi: 10.1016/j.biombioe.2013.05.007
- Hadi, S. M., Shahabuddin, and Rehman, A. (1989). Specificity of the interaction of furfural with DNA. *Mutat. Res. Lett.* 225, 101–106. doi: 10.1016/0165-7992(89)90125-5
- Herrgård, M. J., Swainston, N., Dobson, P., Dunn, W. B., Arga, K. Y., Arvas, M., et al. (2008). A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat. Biotechnol.* 26, 1155–1160. doi: 10.1038/nbt1492
- Ho, C. H., Magtanong, L., Barker, S. L., Gresham, D., Nishimura, S., Natarajan, P., et al. (2009). A molecular barcoded yeast ORF library enables mode-of-action analysis of bioactive compounds. *Nat. Biotechnol.* 27, 369–377. doi: 10.1038/nbt1534
- Ho, C. H., Piotrowski, J., Dixon, S. J., Baryshnikova, A., Costanzo, M., and Boone, C. (2011). Combining functional genomics and chemical biology to identify targets of bioactive compounds. *Curr. Opin. Chem. Biol.* 15, 66–78. doi: 10.1016/j.cbpa.2010.10.023
- Holyoak, C. D., Stratford, M., McMullin, Z., Cole, M. B., Crimmins, K., Brown, A. J., et al. (1996). Activity of the plasma membrane H(+)-ATPase and optimal glycolytic flux are required for rapid adaptation and growth of *Saccharomyces cerevisiae* in the presence of the weak-acid preservative sorbic acid. *Appl. Environ. Microbiol.* 62, 3158–3164.
- Huffer, S., Clark, M. E., Ning, J. C., Blanch, H. W., and Clark, D. S. (2011). Role of alcohols in growth, lipid composition, and membrane fluidity of yeasts, bacteria, and archaea. *Appl. Environ. Microbiol.* 77, 6400–6408. doi: 10.1128/AEM.00694-11
- Ibeas, J. I., and Jimenez, J. (1997). Mitochondrial DNA loss caused by ethanol in *Saccharomyces* flor yeasts. *Appl. Environ. Microbiol.* 63, 7–12.
- Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S., and Weissman, J. S. (2009). Genome-wide analysis *in vivo* of translation with nucleotide resolution using ribosome profiling. *Science* 324, 218–223. doi: 10.1126/science.1168978
- Ingram, L. O. (1976). Adaptation of membrane lipids to alcohols. *J. Bacteriol.* 125, 670–678.
- Iwaki, A., Ohnuki, S., Suga, Y., Izawa, S., and Ohya, Y. (2013). Vanillin inhibits translation and induces messenger ribonucleoprotein (mRNP) granule formation in *Saccharomyces cerevisiae*: application and validation of high-content, image-based profiling. *PLoS ONE* 8:e61748. doi: 10.1371/journal.pone.0061748
- Jarboe, L. R. (2011). YqhD: a broad-substrate range aldehyde reductase with various applications in production of biorenewable fuels and chemicals. *Appl. Microbiol. Biotechnol.* 89, 249–257. doi: 10.1007/s00253-010-2912-9
- Jin, M., Sarkis, C., Gunawan, C., Bice, B. D., Simonett, S. P., Avanas Narasimhan, R., et al. (2013). Phenotypic selection of a wild *Saccharomyces cerevisiae* strain for simultaneous saccharification and co-fermentation of AFEX™ pretreated corn stover. *Biotechnol. Biofuels* 6, 108. doi: 10.1186/1754-6834-6-108
- Jin, Y.-S., and Jeffries, T. W. (2004). Stoichiometric network constraints on xylose metabolism by recombinant *Saccharomyces cerevisiae*. *Metab. Eng.* 6, 229–238. doi: 10.1016/j.ymben.2003.11.006
- Jozefczuk, S., Klie, S., Catchpole, G., Szymanski, J., Cuadros-Inostroza, A., Steinhäuser, D., et al. (2010). Metabolomic and transcriptomic stress response of *Escherichia coli*. *Mol. Syst. Biol.* 6:364. doi: 10.1038/msb.2010.18
- Khan, Q. A., Shamsi, F. A., and Hadi, S. M. (1995). Mutagenicity of furfural in plasmid DNA. *Cancer Lett.* 89, 95–99. doi: 10.1016/0304-3835(95)90163-9
- King, A. T., Davey, M. R., Mellor, I. R., Mulligan, B. J., and Lowe, K. C. (1991). Surfactant effects on yeast cells. *Enzyme Microb. Technol.* 13, 148–153. doi: 10.1016/0141-0229(91)90171-6
- Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., et al. (2006). Complete set of ORF clones of *Escherichia coli* ASKA library (A Complete Set of *E. coli* K-12 ORF Archive): unique resources for biological research. *DNA Res.* 12, 291–299. doi: 10.1093/dnares/dsi012
- Klinke, H. B., Thomsen, A. B., and Ahring, B. K. (2004). Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass. *Appl. Microbiol. Biotechnol.* 66, 10–26. doi: 10.1007/s00253-004-1642-2
- Krebs, H. A., Wiggins, D., Stubbs, M., Sols, A., and Bedoya, F. (1983). Studies on the mechanism of the antifungal action of benzoate. *Biochem. J.* 214, 657–663.
- Laouar, L., Lowe, K. C., and Mulligan, B. J. (1996). Yeast responses to non-ionic surfactants. *Enzyme Microb. Technol.* 18, 433–438. doi: 10.1016/0141-0229(95)00122-0
- Larsson, S., Quintana-Sáinz, A., Reimann, A., Nilvebrant, N. O., and Jönsson, L. J. (2000). Influence of lignocellulose-derived aromatic compounds on oxygen-limited growth and ethanolic fermentation by *Saccharomyces cerevisiae*. *Appl. Biochem. Biotechnol.* 84–86, 617–632. doi: 10.1385/ABAB:84-86:1-9:617
- Larsson, S., Reimann, A., Nilvebrant, N.-O., and Jönsson, L. J. (1999). Comparison of different methods for the detoxification of lignocellulose hydrolyzates of spruce. *Appl. Biochem. Biotechnol.* 77, 91–103. doi: 10.1385/ABAB:77:1-3:91
- Lau, M. W., and Dale, B. E. (2009). Cellulosic ethanol production from AFEX-treated corn stover using *Saccharomyces cerevisiae* 424A(LNH-ST). *Proc. Natl. Acad. Sci. U.S.A.* 106, 1368–1373. doi: 10.1073/pnas.0812364106
- Lau, M. W., Gunawan, C., Balan, V., and Dale, B. E. (2010). Comparing the fermentation performance of *Escherichia coli* KO11, *Saccharomyces cerevisiae* 424A(LNH-ST) and *Zymomonas mobilis* AX101 for cellulosic ethanol production. *Biotechnol. Biofuels* 3, 11. doi: 10.1186/1754-6834-3-11
- Lee, M. V., Topper, S. E., Hubler, S. L., Hose, J., Wenger, C. D., Coon, J. J., et al. (2011). A dynamic model of proteome changes reveals new roles for transcript alteration in yeast. *Mol. Syst. Biol.* 7:514. doi: 10.1038/msb.2011.48
- Lee, S., Nam, D., Jung, J. Y., Oh, M.-K., Sang, B.-I., and Mitchell, R. J. (2012). Identification of *Escherichia coli* biomarkers responsive to various lignin-hydrolysate compounds. *Bioresour. Technol.* 114, 450–456. doi: 10.1016/j.biortech.2012.02.085
- Li, Z., Chen, C. H., Liu, T., Mathrubootham, V., Hegg, E. L., and Hodge, D. B. (2013). Catalysis with Cu(II) (bpy) improves alkaline hydrogen peroxide pretreatment. *Biotechnol. Bioeng.* 110, 1078–1086. doi: 10.1002/bit.24793
- Linton, K. J., and Higgins, C. F. (1998). The *Escherichia coli* ATP-binding cassette (ABC) proteins. *Mol. Microbiol.* 28, 5–13. doi: 10.1046/j.1365-2958.1998.00764.x
- Liu, Z. L., Moon, J., Andersh, B. J., Slininger, P. J., and Weber, S. (2008). Multiple gene-mediated NAD(P)H-dependent aldehyde reduction is a mechanism of *in situ* detoxification of furfural and 5-hydroxymethylfurfural by *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* 81, 743–753. doi: 10.1007/s00253-008-1702-0
- Luterbacher, J. S., Rand, J. M., Alonso, D. M., Han, J., Youngquist, J. T., Marvelas, C. T., et al. (2014). Nonenzymatic sugar production from biomass using biomass-derived  $\gamma$ -valerolactone. *Science* 343, 277–280. doi: 10.1126/science.1246748
- Magtanong, L., Ho, C. H., Barker, S. L., Jiao, W., Baryshnikova, A., Bahr, S., et al. (2011). Dosage suppression genetic interaction networks enhance functional wiring diagrams of the cell. *Nat. Biotechnol.* 29, 505–511. doi: 10.1038/nbt.1855
- Matsushika, A., Nagashima, A., Goshima, T., and Hoshino, T. (2013). Fermentation of xylose causes inefficient metabolic state due to carbon/energy starvation and reduced glycolytic flux in recombinant industrial *Saccharomyces cerevisiae*. *PLoS ONE* 8:e69005. doi: 10.1371/journal.pone.0069005
- McEwen, J. T., and Atsumi, S. (2012). Alternative biofuel production in non-natural hosts. *Curr. Opin. Biotechnol.* 23, 744–750. doi: 10.1016/j.copbio.2011.12.019
- Mikulášová, M., Vodní, Š., and Pekarovičová, A. (1990). Influence of phenolics on biomass production by *Candida utilis* and *Candida albicans*. *Biomass* 23, 149–154. doi: 10.1016/0144-4565(90)90032-F
- Miller, E. N., Jarboe, L. R., Turner, P. C., Pharkya, P., Yomano, L. P., York, S. W., et al. (2009). Furfural inhibits growth by limiting sulfur assimilation in ethanologenic *Escherichia coli* strain LY180. *Appl. Environ. Microbiol.* 75, 6132–6141. doi: 10.1128/AEM.01187-09
- Minty, J. J., Lesnfsky, A. A., Lin, F., Chen, Y., Zaroff, T. A., Veloso, A. B., et al. (2011). Evolution combined with genomic study elucidates genetic bases of isobutanol tolerance in *Escherichia coli*. *Microb. Cell Fact.* 10, 18. doi: 10.1186/1475-2859-10-18
- Mira, N. P., Palma, M., Guerreiro, J. F., and Sa-Correia, I. (2010). Genome-wide identification of *Saccharomyces cerevisiae* genes required for tolerance to acetic acid. *Microb. Cell Fact.* 9, 79. doi: 10.1186/1475-2859-9-79
- Modig, T., Lidén, G., and Taherzadeh, M. J. (2002). Inhibition effects of furfural on alcohol dehydrogenase, aldehyde dehydrogenase and pyruvate dehydrogenase. *Biochem. J.* 363, 769–776. doi: 10.1042/0264-6021:3630769

- Nagodawithana, T. W., and Steinkraus, K. H. (1976). Influence of the rate of ethanol production and accumulation on the viability of *Saccharomyces cerevisiae* in "rapid fermentation". *Appl. Environ. Microbiol.* 31, 158–162.
- Nguyen, T. T. M., Iwaki, A., Ohya, Y., and Izawa, S. (2014). Vanillin causes the activation of Yap1 and mitochondrial fragmentation in *Saccharomyces cerevisiae*. *J. Biosci. Bioeng.* 117, 33–38. doi: 10.1016/j.jbiosc.2013.06.008
- Oliva, J. M., Negro, M. J., Sáez, F., Ballesteros, I., Manzanera, P., González, A., et al. (2006). Effects of acetic acid, furfural and catechol combinations on ethanol fermentation of *Kluyveromyces marxianus*. *Process Biochem.* 41, 1223–1228. doi: 10.1016/j.procbio.2005.12.003
- Ouellet, M., Datta, S., Dibble, D. C., Tamrakar, P. R., Benke, P. I., Li, C., et al. (2011). Impact of ionic liquid pretreated plant biomass on *Saccharomyces cerevisiae* growth and biofuel production. *Green Chem.* 13, 2743. doi: 10.1039/c1gc15327g
- Parsons, A., Lopez, A., Givoni, I., Williams, D., Gray, C., Porter, J., et al. (2006). Exploring the mode-of-action of bioactive compounds by chemical-genetic profiling in yeast. *Cell* 126, 611–625. doi: 10.1016/j.cell.2006.06.040
- Pereira Rangel, L., Fritzen, M., Yunes, R. A., Leal, P. C., Creczynski-Pasa, T. B., and Ferreira-Pereira, A. (2010). Inhibitory effects of gallic acid ester derivatives on *Saccharomyces cerevisiae* multidrug resistance protein Pdr5p. *FEMS Yeast Res.* 10, 244–251. doi: 10.1111/j.1567-1364.2010.00603.x
- Russell, J. B. (1992). Another explanation for the toxicity of fermentation acids at low pH: anion accumulation versus uncoupling. *J. Appl. Bacteriol.* 73, 363–370. doi: 10.1111/j.1365-2672.1992.tb04990.x
- Sato, T. K., Liu, T., Parreiras, L. S., Williams, D. L., Wohlbach, D. J., Bice, B. D., et al. (2014). Harnessing genetic diversity in *Saccharomyces cerevisiae* for improved fermentation of xylose in hydrolysates of alkaline hydrogen peroxide pretreated biomass. *Appl. Environ. Microbiol.* 80, 540–554. doi: 10.1128/AEM.01885-13
- Schmer, M. R., Vogel, K. P., Mitchell, R. B., and Perrin, R. K. (2008). Net energy of cellulosic ethanol from switchgrass. *Proc. Natl. Acad. Sci. U.S.A.* 105, 464–469. doi: 10.1073/pnas.0704767105
- Schmitt, L., and Tampé, R. (2002). Structure and mechanism of ABC transporters. *Curr. Opin. Struct. Biol.* 12, 754–760. doi: 10.1016/S0959-440X(02)00399-8
- Schüller, C., Mamnun, Y. M., Mollapour, M., Krapf, G., Schuster, M., Bauer, B. E., et al. (2004). Global phenotypic analysis and transcriptional profiling defines the weak acid stress response regulon in *Saccharomyces cerevisiae*. *Mol. Biol. Cell* 15, 706–720. doi: 10.1091/mbc.E03-05-0322
- Schwalbach, M. S., Keating, D. H., Tremaine, M., Marner, W. D., Zhang, Y., Bothfeld, W., et al. (2012). Complex physiology and compound stress responses during fermentation of alkali-pretreated corn stover hydrolysate by an *Escherichia coli* ethanologen. *Appl. Environ. Microbiol.* 78, 3442–3457. doi: 10.1128/AEM.07329-11
- Shapiro, A. B., and Ling, V. (1998). Transport of LDS-751 from the cytoplasmic leaflet of the plasma membrane by the rhodamine-123-selective site of P-glycoprotein. *Eur. J. Biochem. FEBS* 254, 181–188. doi: 10.1046/j.1432-1327.1998.2540181.x
- Sindhu, R., Kuttiraja, M., Elizabeth Preeti, V., Vani, S., Sukumaran, R. K., and Binod, P. (2013). A novel surfactant-assisted ultrasound pretreatment of sugarcane tops for improved enzymatic release of sugars. *Bioresour. Technol.* 135, 67–72. doi: 10.1016/j.biortech.2012.09.050
- Sinha, R. P. (1986). Toxicity of organic acids for repair-deficient strains of *Escherichia coli*. *Appl. Environ. Microbiol.* 51, 1364–1366.
- Skerker, J. M., Leon, D., Price, M. N., Mar, J. S., Tarjan, D. R., Wetmore, K. M., et al. (2013). Dissecting a complex chemical stress: chemogenomic profiling of plant hydrolysates. *Mol. Syst. Biol.* 9, 674. doi: 10.1038/msb.2013.30
- Stratford, M., and Anslow, P. A. (1998). Evidence that sorbic acid does not inhibit yeast as a classic "weak acid preservative." *Lett. Appl. Microbiol.* 27, 203–206. doi: 10.1046/j.1472-765X.1998.00424.x
- Stratford, M., Nebe-von-Caron, G., Steels, H., Novodvorska, M., Ueckert, J., and Archer, D. B. (2013). Weak-acid preservatives: pH and proton movements in the yeast *Saccharomyces cerevisiae*. *Int. J. Food Microbiol.* 161, 164–171. doi: 10.1016/j.ijfoodmicro.2012.12.013
- Taymaz-Nikerel, H., Borujeni, A. E., Verheijen, P. J. T., Heijnen, J. J., and van Gulik, W. M. (2010). Genome-derived minimal metabolic models for *Escherichia coli* MG1655 with estimated *in vivo* respiratory ATP stoichiometry. *Biotechnol. Bioeng.* 107, 369–381. doi: 10.1002/bit.22802
- Ullah, A., Orij, R., Brul, S., and Smits, G. J. (2012). Quantitative analysis of the modes of growth inhibition by weak organic acids in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* 78, 8377–8387. doi: 10.1128/AEM.02126-12
- Varma, A., and Palsson, B. O. (1994). Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. *Appl. Environ. Microbiol.* 60, 3724–3731.
- Verduyn, C., Postma, E., Scheffers, W. A., and Van Dijken, J. P. (1992). Effect of benzoic acid on metabolic fluxes in yeasts: a continuous-culture study on the regulation of respiration and alcoholic fermentation. *Yeast* 8, 501–517. doi: 10.1002/yea.320080703
- Wang, H. H., Kim, H., Cong, L., Jeong, J., Bang, D., and Church, G. M. (2012). Genome-scale promoter engineering by coselection MAGE. *Nat. Methods* 9, 591–593. doi: 10.1038/nmeth.1971
- Wang, X., Miller, E. N., Yomano, L. P., Zhang, X., Shanmugam, K. T., and Ingram, L. O. (2011). Increased furfural tolerance due to overexpression of NADH-Dependent oxidoreductase FucO in *Escherichia coli* strains engineered for the production of ethanol and lactate. *Appl. Environ. Microbiol.* 77, 5132–5140. doi: 10.1128/AEM.05008-11
- Wang, X., Yomano, L. P., Lee, J. Y., York, S. W., Zheng, H., Mullinnix, M. T., et al. (2013). Engineering furfural tolerance in *Escherichia coli* improves the fermentation of lignocellulosic sugars into renewable chemicals. *Proc. Natl. Acad. Sci. U.S.A.* 110, 4021–4026. doi: 10.1073/pnas.1217958110
- Wei, N., Quarterman, J., Kim, S. R., Cate, J. H. D., and Jin, Y.-S. (2013). Enhanced biofuel production through coupled acetic acid and xylose consumption by engineered yeast. *Nat. Commun.* 4:2580. doi: 10.1038/ncomms3580
- Yang, S., Land, M. L., Klingeman, D. M., Pelletier, D. A., Lu, T.-Y. S., Martin, S. L., et al. (2010a). Paradigm for industrial strain improvement identifies sodium acetate tolerance loci in *Zymomonas mobilis* and *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.* 107, 10395–10400. doi: 10.1073/pnas.0914506107
- Yang, S., Pan, C., Tschaplinski, T. J., Hurst, G. B., Engle, N. L., Zhou, W., et al. (2013). Systems biology analysis of *Zymomonas mobilis* ZM4 ethanol stress responses. *PLoS ONE* 8:e68886. doi: 10.1371/journal.pone.0068886
- Yang, S., Pelletier, D. A., Lu, T.-Y. S., and Brown, S. D. (2010b). The *Zymomonas mobilis* regulator hfq contributes to tolerance against multiple lignocellulosic pretreatment inhibitors. *BMC Microbiol.* 10:135. doi: 10.1186/1471-2180-10-135
- Zaldivar, J., and Ingram, L. O. (1999). Effect of organic acids on the growth and fermentation of ethanologenic *Escherichia coli* LY01. *Biotechnol. Bioeng.* 66, 203–210. doi: 10.1002/(SICI)1097-0290(1999)66:4<203::AID-BIT1>3.0.CO;2-#
- Zerhouni, E. (2003). The NIH Roadmap. *Science* 302, 63–72. doi: 10.1126/science.1091867
- Zingaro, K. A., Nicolaou, S. A., and Papoutsakis, E. T. (2013). Dissecting the assays to assess microbial tolerance to toxic chemicals in bioprocessing. *Trends Biotechnol.* 31, 643–653. doi: 10.1016/j.tibtech.2013.08.005

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 16 November 2013; accepted: 18 February 2014; published online: 14 March 2014.

Citation: Piotrowski JS, Zhang Y, Bates DM, Keating DH, Sato TK, Ong IM and Landick R (2014) Death by a thousand cuts: the challenges and diverse landscape of lignocellulosic hydrolysate inhibitors. *Front. Microbiol.* 5:90. doi: 10.3389/fmicb.2014.00090

This article was submitted to Microbial Physiology and Metabolism, a section of the journal Frontiers in Microbiology.

Copyright © 2014 Piotrowski, Zhang, Bates, Keating, Sato, Ong and Landick. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.





# Modeling of *Zymomonas mobilis* central metabolism for novel metabolic engineering strategies

Uldis Kalnenieks<sup>1\*</sup>, Agris Pentjuss<sup>2</sup>, Reinis Rutkis<sup>1</sup>, Egils Stalidzans<sup>1,2,3</sup> and David A. Fell<sup>4</sup>

<sup>1</sup> Institute of Microbiology and Biotechnology, University of Latvia, Riga, Latvia

<sup>2</sup> Department of Computer Systems, Latvia University of Agriculture, Jelgava, Latvia

<sup>3</sup> SIA TIBIT, Jelgava, Latvia

<sup>4</sup> Department of Biological and Medical Sciences, Oxford Brookes University, Oxford, UK

## Edited by:

Katherine M. Pappas, University of Athens, Greece

## Reviewed by:

Jason Warren Cooley, University of Missouri, USA

Patrick Hallenbeck, University of Montreal, Canada

## \*Correspondence:

Uldis Kalnenieks, Institute of Microbiology and Biotechnology, University of Latvia, Kronvalda boulevard 4, Riga, LV-1586, Latvia  
e-mail: kalnen@lanet.lv

Mathematical modeling of metabolism is essential for rational metabolic engineering. The present work focuses on several types of modeling approach to quantitative understanding of central metabolic network and energetics in the bioethanol-producing bacterium *Zymomonas mobilis*. Combined use of Flux Balance, Elementary Flux Mode, and thermodynamic analysis of its central metabolism, together with dynamic modeling of the core catabolic pathways, can help to design novel substrate and product pathways by systematically analyzing the solution space for metabolic engineering, and yields insights into the function of metabolic network, hardly achievable without applying modeling tools.

**Keywords:** stoichiometric modeling, elementary flux modes, kinetic modeling, systems biology, metabolic engineering, Entner–Doudoroff pathway, central metabolism, *Zymomonas mobilis*

## INTRODUCTION

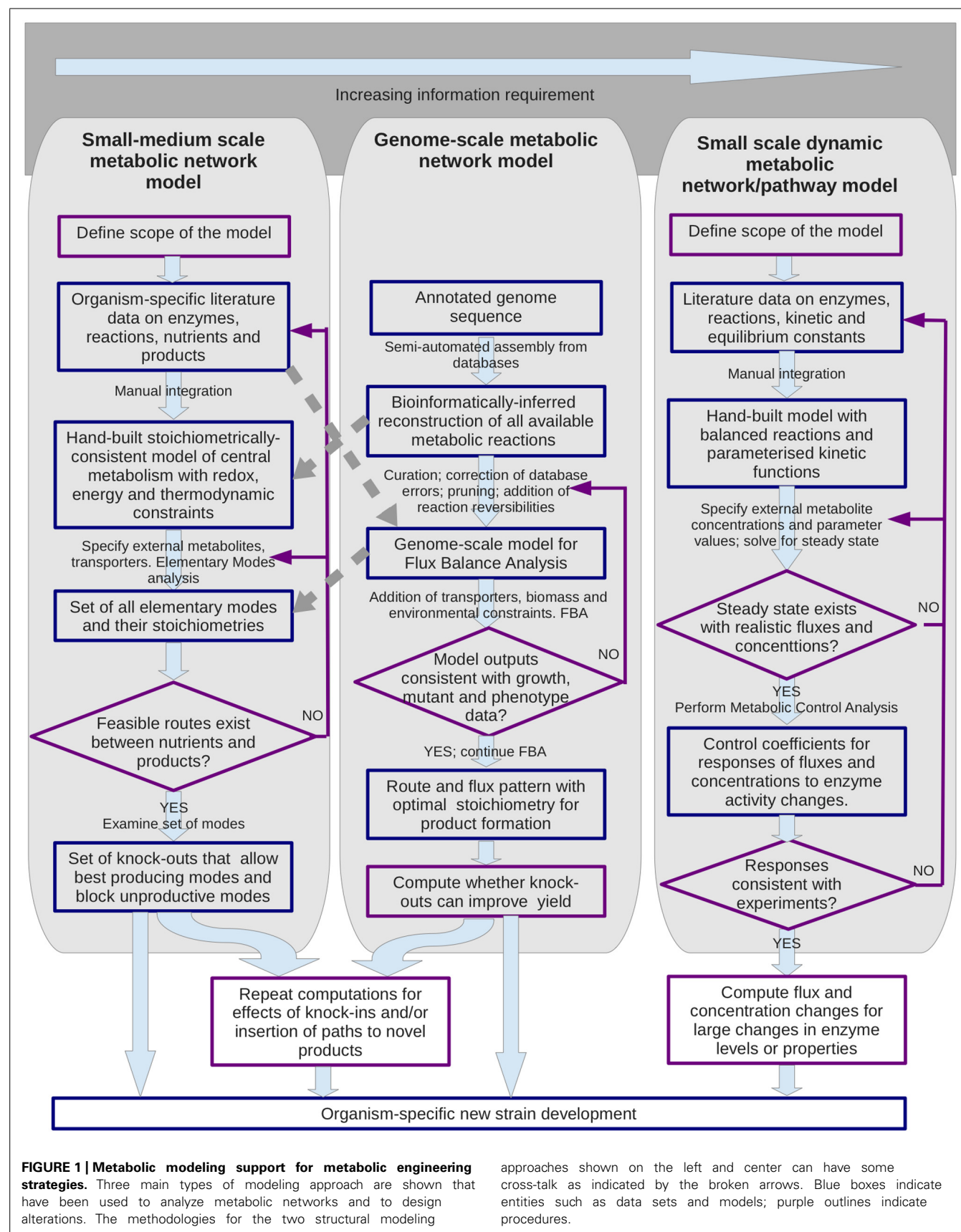
*Zymomonas mobilis*, a member of the family of *Sphingomonadaceae*, is an unusual facultatively anaerobic Gram-negative bacterium, which has a very efficient homoethanol fermentation pathway. High ethanol yields, outstanding ethanol productivity (exceeding by 3–5 fold that of yeast; see Rogers et al., 1982), and tolerance to high ethanol and sugar concentrations, keep *Z. mobilis* in the focus of biotechnological research over four decades. The complete genome sequence of *Z. mobilis* ZM4, consisting of a single circular chromosome of 2,056,416 bp, was reported by Seo et al. (2005), followed by the genomes of several other strains (Kouvelis et al., 2009; Pappas et al., 2011; Desiniotis et al., 2012). Its small genome size, together with high specific rate of sugar catabolism via the Entner–Doudoroff (ED) pathway, and a relatively simple central metabolic network, make *Z. mobilis* a promising candidate for metabolic engineering (Sprenger, 1996; Rogers et al., 2007). Currently, recombinant *Z. mobilis* capable of fermenting pentose sugars is regarded as a potential alternative to yeast and recombinant *Escherichia coli* for ethanol biofuel synthesis from agricultural and forestry waste (Dien et al., 2003; Panesar et al., 2006; Rogers et al., 2007; Lau et al., 2010).

In spite of the seeming simplicity of its metabolism, *Z. mobilis* is a bacterium with an interesting physiology (Kalnenieks, 2006), posing researchers some long-standing challenges. Its extremely rapid glucose catabolism, far exceeding the biosynthetic demands of the cell, and the presence of an active respiratory chain with a low apparent P/O ratio (Bringer et al., 1984; Strothdeicher et al., 1990; Kalnenieks et al., 1993) are major manifestations of its so-called uncoupled growth. There are serious gaps in our understanding of the mechanistic basis of uncoupled growth, and in particular, the reason for the low degree of coupling in the respiratory chain of *Z. mobilis*.

Mathematical modeling and *in silico* simulations are the most powerful tools of systems biology for understanding of complex metabolic phenomena, and often lead to novel, counterintuitive conclusions. A quantitative picture of physiology and metabolism is a key for rational, model-driven metabolic engineering. Some of the different metabolic modeling approaches that can support the design of novel metabolic engineering strategies are summarized in **Figure 1** (for reviews see: Schuster et al., 2000; Liu et al., 2010; Santos et al., 2011; Schellenberger et al., 2011; Rohwer, 2012). Compared with qualitative, pathway-oriented approaches, computational network analyses can enforce strict mass, energy and redox balancing and give an overall stoichiometric equation for predicted conversions (c.f. de Figueiredo et al., 2009). Here we outline recent advances and perspectives from applying such systems biology approaches to the physiology of *Z. mobilis*. We discuss some recent results gained by stoichiometric and kinetic modeling of its central metabolism, and their potential application to the design of novel substrate pathways, synthesis of novel products, and to the study of the uncoupled growth phenomenon *per se*.

## RECONSTRUCTION OF *Z. mobilis* CENTRAL METABOLIC NETWORK

Two medium-scale (Tsantili et al., 2007; Pentjuss et al., 2013) and two genome-scale stoichiometric reconstructions of *Z. mobilis* (Lee et al., 2010; Widiastuti et al., 2010) have been reported so far, representing instances of the left and center panels of **Figure 1** respectively. These reconstructions were based on the available genome annotation (Seo et al., 2005; Kouvelis et al., 2009) and provided an overall picture of *Z. mobilis* metabolism. The recent reconstruction made by Pentjuss et al. (2013) was focussed solely on the reactions of central metabolism and for the first time for *Z. mobilis* provided simulation-ready model files. That



decreased the scale, yet allowed an improvement in the accuracy of reconstruction, by combining the genome-derived information with the preexisting biochemical evidence on *Z. mobilis*, available mostly for the reactions of catabolism and central metabolism.

Notably, several key reactions of central metabolism, common for the majority of the chemoheterotrophic, facultatively anaerobic bacteria, are absent in *Z. mobilis*. The Embden–Meyerhof–Parnas (EMP) glycolytic pathway is not operating in this bacterium. Absence of the EMP pathway has been confirmed by [ $1\text{-}^{13}\text{C}$ ]glucose experiments (Fuhrer et al., 2005), and furthermore, the gene for phosphofructokinase is lacking in the genome (Seo et al., 2005). *Z. mobilis* is the only known microorganism that uses the ED pathway anaerobically in place of the EMP glycolysis. Since the EMP pathway produces two ATP per glucose while the ED produces only one, it might seem that *Z. mobilis* suffers from ATP deficiency. However, it has been recently shown by means of thermodynamic analysis that, for a given glycolytic flux, the ED pathway requires significantly less enzymatic protein than the EMP pathway (Flamholz et al., 2013). On the other hand, the amount of the ED pathway enzymes in *Z. mobilis* cell is reported to be very high, reaching 50% of the cell's soluble protein (Algar and Scopes, 1985; An et al., 1991). The high level of expression of the pathway together with its inherent speed, therefore, makes ATP production by the *Z. mobilis* ED pathway very rapid and, in fact, excessive for the needs of cell. Energy dissipation in order to regenerate ADP is thus essential for its balanced operation (Kalnenieks, 2006).

The TCA cycle is truncated, and consists of two branches, leading to 2-oxoglutarate and fumarate as the end products (Bringer-Meyer and Sahm, 1989). The genes for the 2-oxoglutarate dehydrogenase complex and malate dehydrogenase are absent (Seo et al., 2005), and accordingly,  $^{13}\text{C}$ -labeling patterns of 2-oxoglutarate and oxaloacetate do not support cyclic function of this pathway in *Z. mobilis* (de Graaf et al., 1999). Also, the pentose phosphate pathway is incomplete: transaldolase activity is lacking (Feldmann et al., 1992; de Graaf et al., 1999). The activity of 6-phosphogluconate dehydrogenase, the first reaction of the oxidative part of the pentose phosphate pathway, was reported to be very low (Feldmann et al., 1992). Subsequently, the corresponding gene (*gnd*) could not be identified in the sequenced genomes.

The aerobic redox cofactor balance and the function of electron transport chain represent yet another part of *Z. mobilis* metabolism that differs from that typically found in other bacteria. *Z. mobilis* is one of the few known bacteria in which both NADH and NADPH can serve as electron donors for the respiratory type II NADH dehydrogenase (ZMO1113; Bringer et al., 1984; Strohmeyer et al., 1990; Kalnenieks et al., 2008). Because of the truncated Krebs cycle, the ED pathway is the only source of reducing equivalents in catabolism, and therefore the electron transport chain competes for the limited NADH with the highly active alcohol dehydrogenases (Kalnenieks et al., 2006). Withdrawal of NADH from the alcohol dehydrogenase reaction would cause accumulation of acetaldehyde, which inhibits growth of aerobic *Z. mobilis* culture (Wecker and Zall, 1987). Nevertheless, this bacterium possesses a respiratory chain with

high rates of oxygen consumption. The apparent P/O ratio of its respiratory chain is low (Bringer et al., 1984; Kalnenieks et al., 1993) though the mechanistic basis for that is not clear. However, for metabolic engineering purposes, an active, yet energetically inefficient electron transport has advantages for the needs of redox balancing during synthesis of novel products via metabolic pathways for which regeneration of  $\text{NAD(P)}^+$  is essential, whereas the aerobic increase of biomass yield is unwanted.

## QUEST FOR NOVEL SUBSTRATES AND PRODUCTS: STOICHIOMETRIC AND THERMODYNAMIC ANALYSIS

Much of the metabolic engineering in *Z. mobilis* has been devoted to broadening of its substrate spectrum and expanding its product range beyond bioethanol with a particular focus on the pathway of pentose sugar utilization for synthesis of bioethanol (Sprenger, 1996; Rogers et al., 2007). Advanced pentose-assimilating strains of *Z. mobilis* have been developed during the last couple of decades that can, in several respects, compete with the analogous recombinant strains of *E. coli* and *S. cerevisiae* (Lau et al., 2010). We were interested to explore the biotechnological potential of the low-efficiency respiratory chain of this bacterium for expanding its substrate and product spectrum.

Based on a medium-scale reconstruction of central metabolism (Pentjuss et al., 2013), stoichiometric modeling was used to search the whole solution space of the model, finding maximum product yields and the byproduct spectra with glucose, xylose, or glycerol as the carbon substrates for respiring cultures (Figure 1, left hand side). This was done by Flux Balance Analysis approach, using the COBRA Toolbox (Schellenberger et al., 2011). The stoichiometric analysis suggested several metabolic engineering strategies for obtaining products, such as glycerate, succinate, and glutamate that would use the electron transport chain to oxidize the excess  $\text{NAD(P)H}$ , generated during synthesis of these metabolites. Oxidation of the excess  $\text{NAD(P)H}$  would also be needed for synthesis of ethanol from glycerol.

It is essential, however, to complement the stoichiometric analysis with estimation of the thermodynamic feasibility of the underlying reactions. Glycerol utilization can serve as an example. Being a cheap, renewable carbon source, a byproduct of biodiesel technology, glycerol represents an attractive alternative substrate for *Z. mobilis* metabolic engineering. It is not expected to have serious growth-inhibitory effects, and also, little genetic engineering seems to be needed to make it consumable by *Z. mobilis*, and to channel it into the rapid ED pathway. Conversion of glycerol to ethanol by *Z. mobilis* would require expression of a heterologous transmembrane glycerol transporter and a glycerol kinase. Its genome contains genes for the two subsequent conversion steps, glycerolphosphate dehydrogenase and triose phosphate isomerase, leading to the ED intermediate glyceraldehyde-3-phosphate although, their overexpression might be needed. The further reactions from the glyceraldehyde-3-phosphate to ethanol represent a part of *Z. mobilis* natural ethanologenic pathway, and should be both rapid and redox-balanced. The extra  $\text{NAD(P)H}$ , generated by the glycerolphosphate dehydrogenase reaction could be oxidized by the respiratory chain. If succinate is the desired product



(Pentjuss et al., 2013), the extra reducing equivalents could be used for reduction of fumarate by the respiratory fumarate reductase.

The pathway from glycerol to glyceraldehyde-3-phosphate via phosphorylation and following oxidation and isomerization steps is presented in biochemistry textbooks as the pathway of glycerol catabolism after breakdown of triacylglycerols in higher animals and humans (see e.g., Lehninger Principles of Biochemistry, 6th edition, Fig. 17-4). Though feasible from the stoichiometric point of view, it reveals problems when subjected to thermodynamic analysis. The equilibrium of the glycerolphosphate dehydrogenase reaction appears to be shifted very much toward formation of glycerol phosphate. *In silico* kinetic simulations of glycerol uptake for a putative engineered *Z. mobilis* demonstrate a dramatic accumulation of glycerol-3-phosphate, reaching concentrations of several molar even at a high rate of NAD(P)H withdrawal by the respiratory chain (Rutkis et al., unpublished). Apparently, while the estimated overall stoichiometry of aerobic glycerol conversions is correct, thermodynamic analysis suggests the need to search for alternative reaction sequences to avoid excessive intracellular accumulation of metabolites.

### AEROBIC ELEMENTARY FLUX MODES OF THE PENTOSE PHOSPHATE PATHWAY

A metabolic network can function according to many different pathway options. Elementary flux mode (EFM) analysis has emerged as a systems biological tool that dissects a metabolic network into its basic building blocks, the EFMs (Schuster et al., 2000, 2002). All metabolic capabilities in steady states represent a weighted average of the EFMs, which are the minimal sets of enzymes that can each generate a valid steady state. The EFM approach has proved to be efficient for designing sets of knock-out mutations in order to minimize unwanted metabolic functionality in the producer strains. For example, in engineered *E. coli*, EMF-based mutation analysis helped to eliminate catabolite repression and to increase carbon flux toward the target product ethanol (Trinh et al., 2008).

By decomposing a network of highly interconnected reactions, the EFM analysis may reveal unexpected flux options. Recently we applied EFM analysis to the interaction between the ED, pentose phosphate pathway and respiratory chain in an engineered *Z. mobilis*, which expresses heterologous *gnd* and enzymes for pentose conversion, using the metabolic modeling package ScrumPy (Poolman, 2006). We were interested in the EFMs that such non-growing engineered *Z. mobilis* might employ for aerobic catabolism of glucose and xylose. Analysis revealed several EFMs in respiring cells (Figure 2) that have considerable interest for study of aerobic energy-coupling in this bacterium. With both monosaccharides, knocking out *edd* (encoding 6-phosphogluconate dehydratase), and overexpressing heterologous *gnd* (encoding 6-phosphogluconate dehydrogenase), would lead to generation of additional NAD(P)H and CO<sub>2</sub> in the pentose phosphate pathway, while lowering the ethanol yield. Yet, most importantly, decrease of the ethanol yield would not be accompanied by accumulation of acetaldehyde and acetoin.

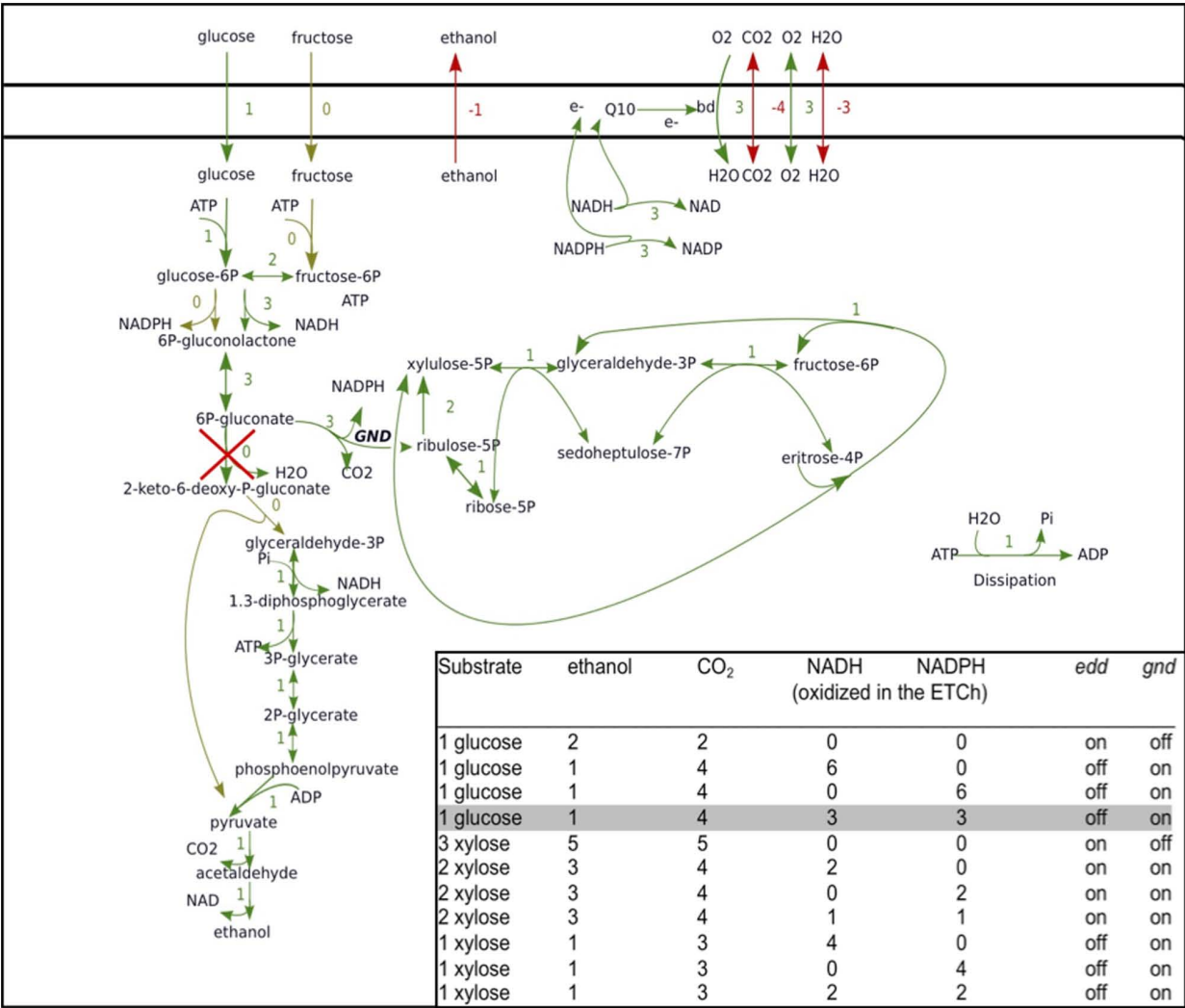
Thus, a simple EFM analysis suggests how to modify *Z. mobilis* aerobic metabolism so that its electron transport chain would

receive more reducing equivalents without accumulation of inhibitory byproducts. Strains with such metabolic modifications might be very useful for study of the mechanisms underlying the uncoupled mode of oxidative phosphorylation in this bacterium.

### KINETIC MODELING OF THE ENTNER–DOUDOROFF PATHWAY

Despite the diverse studies of *Z. mobilis* physiology and genetics, little has been done so far to combine the accumulated knowledge in a form of kinetic model of central metabolism that would be comparable to the existing models for *E. coli* and yeast, and could be used to develop efficient metabolic engineering strategies (c.f. Figure 1, right-hand panel). A kinetic model reported by Altintas et al. (2006) focussed mainly on the interaction between the heterologous enzymes of pentose phosphate pathway and the native *Z. mobilis* ED glycolysis. Providing predictions for optimization of expression levels of the heterologous genes, this study contributed to strategies for maximizing xylose conversion to ethanol. However, the authors assumed constant intracellular concentrations of all adenylate cofactors. Since the ED pathway itself is a major player in ANP and NAD(P)(H) turnover, this might lead to erroneous conclusions on the pathway kinetics and restrict the range of model application. The recent kinetic model by Rutkis et al. (2013): (i) treated the cofactor levels as variables, making the interplay between adenylate cofactor levels and the pathway kinetics explicit, and (ii) introduced equilibrium constants in the kinetic equations to account for the reversibility of reactions more correctly. Metabolic control analysis (MCA) carried out with the model pointed to the ATP turnover as a major bottleneck, showing that the ATP consumption (dissipation) exerts a high level of control over glycolytic flux under various conditions (Rutkis et al., 2013).

Indeed, experimental studies of the ED pathway flux have shown that moderate overexpression of the ED pathway and alcohol dehydrogenase genes do not affect the glycolytic flux (Arfman et al., 1992; Snoep et al., 1995). Larger increases of the expression levels even caused a decrease in flux, exerting also a negative impact on *Z. mobilis* growth rate (Snoep et al., 1995). This clearly indicated that glycolytic flux in *Z. mobilis* must be controlled at some point(s) outside the ED pathway itself. The negative effects of overexpression apparently did not result from intrinsically negative flux control coefficients of the ED enzymes, but were attributable to the protein burden effect (Snoep et al., 1995), whereby overexpression of an enzyme with a small flux control coefficient caused reductions in the expression of other enzymes that have a greater influence on the flux. These results together with MCA studies on the kinetic model suggested that, due to the negligible flux control coefficients for the majority of reactions, single enzymes of the ED pathway should not be considered as prime targets for overexpression to increase the glycolytic flux in *Z. mobilis* (Rutkis et al., 2013). The calculated effects of several glycolytic enzyme (*gap*, *pgk*, *pgm*) and both alcohol dehydrogenase isoenzyme (*adhA* and *adhB*) overexpression, in accordance with previous experimental observations, predicted little or no increase of glycolytic flux (Arfman et al., 1992; Snoep et al., 1995). The somewhat higher flux control coefficient for the pyruvate decarboxylase (*pdc*) reaction suggested



**FIGURE 2 | Elementary flux modes of aerobic glucose and xylose catabolism for a strain with engineered pentose phosphate pathway enzymes.** Elementary flux mode for catabolism of glucose in cells with knocked-out *edd* and overexpressed heterologous *gnd* via the Entner–Doudoroff and pentose phosphate pathway, involving both NADH- and NADPH-oxidizing activity of the respiratory chain, is shown. ScrumPy

modeling software EFM drawing algorithm (Pentjuss et al., unpublished) was used for visualization. Inset: complete list of elementary flux modes of glucose and xylose catabolism in *Z. mobilis*, involving the Entner–Doudoroff pathway, pentose phosphate pathway and the respiratory chain, with ethanol and carbon dioxide as the sole products. The explicitly shown elementary flux mode is shaded in gray.

that overexpression of this enzyme by more than 3-fold, might lead to an increase of glycolytic flux of almost 23% (Rutkis et al., 2013). However, quite the opposite was observed experimentally: approximately 10-fold increase of *pdh* was shown to slow down glycolysis by up to 25%, thereby implying that the protein burden might be a serious side effect of catabolic enzyme overexpression in *Z. mobilis*. Usually effects of protein burden are of minor importance in optimization of catabolic fluxes, due to relatively low concentrations of the enzymes in catabolic routes. This is not the case for *Z. mobilis* catabolism, however, since over 50% of the cell protein already is engaged in the function of the ED pathway (Algar and Scopes, 1985). Fortunately, flux control coefficient estimations still indicate a certain solution space for flux improvement: simultaneous overexpression of *pdh*, *eno*, *pgm* within the 3-fold range of initial

enzyme activities (which most probably would be below the putative protein burden threshold), has the potential to increase the glycolytic flux by up to 25% (to reach 6.6 g glucose, g dry wt<sup>-1</sup>h<sup>-1</sup>; Rutkis et al. (2013)). Obviously, another option would be to raise ATP dissipation. That could be done by overexpression of the H<sup>+</sup>-dependent F<sub>0</sub>F<sub>1</sub>-ATPase, a major ATP-dissipating activity. Reyes and Scopes (1991) have estimated the F<sub>0</sub>F<sub>1</sub>-ATPase contribution being over 20% of the total intracellular ATP turnover. It should be noted, however, that overexpression of ATP-dissipating reaction(s) might disturb the intracellular ATP homeostasis, with successive suspension of glycolysis (by slowing down the first reaction of the ED pathway, phosphorylation of glucose). Co-response analysis indicates (Rutkis et al., 2013) that, at the highest glycolytic flux considered



(4.6 g/g/h), the cellular capacity to maintain the ATP homeostasis is close to its limit, since even 1% further increase of glycolytic flux due to rise of ATP dissipation would be associated with a 4% decrease in ATP concentration.

## CONCLUSION

Although *Z. mobilis* metabolism has been subject to extensive research, and genome sequence data for several strains are now also available, it is only quite recently that modeling of its central metabolic network has started to gain momentum. These latest results of modeling *Z. mobilis* illustrate the relevance of combined stoichiometric, thermodynamic and kinetic analysis of central metabolism at different scales for microorganisms producing biorenewables. Concerted application of structural and dynamic modeling will help to identify targets for future metabolic engineering in a systematic manner, and provide novel insights into the biotechnological potential of this bacterium.

## AUTHOR CONTRIBUTIONS

All authors have equally contributed to the manuscript and have accepted the final version to be published.

## ACKNOWLEDGMENTS

This work was supported by the Latvian ESF projects 2009/027/1DP/1.1.1.2.0/09/APIA/VIAA/128 and 2009/0138/1DP/1.1.2.1.2/09/IPIA/VIAA/004, and by the Latvian Council of Science project 536/2012.

## REFERENCES

- Algar, E. M., and Scopes, R. K. (1985). Studies on cell-free metabolism: ethanol production by extracts of *Zymomonas mobilis*. *J. Biotechnol.* 2, 275–287.
- Altintas, M. M., Eddy, C. K., Zhang, M., McMillan, J. D., and Kompala, D. S. (2006). Kinetic modeling to optimize pentose fermentation in *Zymomonas mobilis*. *Biotechnol. Bioeng.* 94, 273–295. doi: 10.1002/bit.20843
- An, H., Scopes, R. K., Rodriguez, M., Keshav, K. F., and Ingram, L. O. (1991). Gel electrophoretic analysis of *Zymomonas mobilis* glycolytic and fermentative enzymes: identification of alcohol dehydrogenase II as a stress protein. *J. Bacteriol.* 173, 5975–5982.
- Arfman, N., Worell, V., and Ingram, L. O. (1992). Use of the tac promoter and lacIq for the controlled expression of *Zymomonas mobilis* fermentative genes in *Escherichia coli* and *Zymomonas mobilis*. *J. Bacteriol.* 174, 7370–7378.
- Bringer-Meyer, S., and Sahm, H. (1989). Junctions of catabolic and anabolic pathways in *Zymomonas mobilis*: phosphoenolpyruvate carboxylase and malic enzyme. *Appl. Microbiol. Biotechnol.* 31, 529–536.
- Bringer, S., Finn, R. K., and Sahm, H. (1984). Effect of oxygen on the metabolism of *Zymomonas mobilis*. *Arch. Microbiol.* 139, 376–381.
- de Figueiredo, L. F., Schuster, S., Kaleta, C., and Fell, D. A. (2009). Can sugars be produced from fatty acids? a test case for pathway analysis tools. *Bioinformatics* 25, 152–158. doi: 10.1093/bioinformatics/btn621
- de Graaf, A. A., Striegel, K., Wittig, R. M., Laufer, B., Schmitz, G., Wiechert, W., et al. (1999). Metabolic state of *Zymomonas mobilis* in glucose-, fructose-, and xylose-fed continuous cultures as analysed by <sup>13</sup>C- and <sup>31</sup>P-NMR spectroscopy. *Arch. Microbiol.* 171, 371–385.
- Desiniotis, A., Kouvelis, V. N., Davenport, K., Bruce, D., Detter, C., Tapia, R., et al. (2012). Complete genome sequence of the ethanol-producing *Zymomonas mobilis* subsp. *mobilis* centrotypic ATCC 29191. *J. Bacteriol.* 194, 5966–5967. doi: 10.1128/JB.01398–1312
- Dien, B. S., Cotta, M. A., and Jeffries, T. W. (2003). Bacteria engineered for fuel ethanol production: current status. *Appl. Microbiol. Biotechnol.* 63, 258–266. doi: 10.1007/s00253-003-1444-y
- Feldmann, S., Sahm, H., and Sprenger, G. (1992). Pentose metabolism in *Zymomonas mobilis* wild-type and recombinant strains. *Appl. Microbiol. Biotechnol.* 38, 354–361.
- Flamholz, A., Noor, E., Bar-Even, A., Liebermeister, W., and Milo, R. (2013). Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proc. Natl. Acad. Sci. U.S.A.* 110, 10039–10044. doi: 10.1073/pnas.1215283110
- Fuhrer, T., Fischer, E., and Sauer, U. (2005). Experimental identification and quantification of glucose metabolism in seven bacterial species. *J. Bacteriol.* 187, 1581–1590. doi: 10.1128/JB.187.5.1581–1590.2005
- Kalnenieks, U. (2006). “Physiology of *Zymomonas mobilis*: some unanswered questions,” in *Advances in Microbial Physiology*, Vol. 51, ed. R. K. Poole (London: Academic Press), 73–117. doi: 10.1016/S0065-2911(06)51002-1
- Kalnenieks, U., de Graaf, A. A., Bringer-Meyer, S., and Sahm, H. (1993). Oxidative phosphorylation in *Zymomonas mobilis*. *Arch. Microbiol.* 160, 74–79.
- Kalnenieks, U., Galinina, N., Strazdina, I., Kravale, Z., Pickford, J. L., Rutkis, R., et al. (2008). NADH dehydrogenase deficiency results in low respiration rate and improved aerobic growth of *Zymomonas mobilis*. *Microbiology* 154, 989–994. doi: 10.1099/mic.0.2007/012682-0
- Kalnenieks, U., Galinina, N., Toma, M. M., Pickford, J. L., Rutkis, R., and Poole, R. K. (2006). Respiratory behaviour of a *Zymomonas mobilis* adhB::kanr mutant supports the hypothesis of two alcohol dehydrogenase isoenzymes catalysing opposite reactions. *FEBS Lett.* 580, 5084–5088. doi: 10.1016/j.febslet.2006.08.034
- Kouvelis, V. N., Saunders, E., Brettin, T. S., Bruce, D., Detter, C., Han, C., et al. (2009). Complete genome sequence of the ethanol producer *Zymomonas mobilis* NCIMB 11163. *J. Bacteriol.* 191, 7140–7141. doi:10.1128/JB.01084-09
- Lau, M. W., Gunawan, C., Balan, V., and Dale, B. E. (2010). Comparing the fermentation performance of *Escherichia coli* KO11, *Saccharomyces cerevisiae* 424A(LNH-ST) and *Zymomonas mobilis* AX101 for cellulosic ethanol production. *Biotechnol. Biofuels* 3, 11. doi: 10.1186/1754-6834-3-11
- Lee, K. Y., Park, J. M., Kim, T. Y., Yun, H., and Lee, S. Y. (2010). The genome-scale metabolic network analysis of *Zymomonas mobilis* ZM4 explains physiological features and suggests ethanol and succinic acid production strategies. *Microb. Cell Fact.* 9, 94. doi: 10.1186/1475-2859-9-94
- Liu, L., Agren, R., Bordel, S., and Nielsen, J. (2010). Use of genome-scale metabolic models for understanding microbial physiology. *FEBS Lett.* 584, 2556–2564. doi: 10.1016/j.febslet.2010.04.052
- Panesar, P. S., Marwaha, S. S., and Kennedy, J. F. (2006). *Zymomonas mobilis*: an alternative ethanol producer. *J. Chem. Technol. Biotechnol.* 81, 623–635. doi: 10.1002/jctb.1448
- Pappas, K. M., Kouvelis, V. N., Saunders, E., Brettin, T. S., Bruce, D., Detter, C., et al. (2011). Genome sequence of the ethanol-producing *Zymomonas mobilis* subsp. *mobilis* lectotype strain ATCC 10988. *J. Bacteriol.* 193, 5051–5052. doi: 10.1128/JB.05395-11
- Pentjuss, A., Odzina, I., Kostromins, A., Fell, D. A., Stalidzans, E., and Kalnenieks, U. (2013). Biotechnological potential of respiring *Zymomonas mobilis*: a stoichiometric analysis of its central metabolism. *J. Biotechnol.* 165, 1–10. doi: 10.1016/j.jbiotec.2013.02.014
- Poolman, M. G. (2006). ScrumPy: metabolic modelling with python. *Syst. Biol.* 153, 375–378. doi: 10.1049/ip-syb:20060010
- Reyes, L., and Scopes, R. K. (1991). Membrane-associated ATPase from *Zymomonas mobilis*; purification and characterization. *Biochim. Biophys. Acta* 1068, 174–178.
- Rogers, P., Lee, K., Skotnicki, M., and Tribe, D. (1982). Ethanol production by *Zymomonas mobilis*. *Adv. Biochem. Eng.* 23, 37–84.
- Rogers, P. L., Jeon, Y. J., Lee, K. J., and Lawford, H. G. (2007). *Zymomonas mobilis* for fuel ethanol and higher value products. *Adv. Biochem. Eng.* 108, 263–288. doi: 10.1007/10\_2007\_060
- Rohwer, J. M. (2012). Kinetic modeling of plant metabolic pathways. *J. Exp. Bot.* 63, 2275–2292. doi:10.1093/jxb/ers080
- Rutkis, R., Kalnenieks, U., Stalidzans, E., and Fell, D. A. (2013). Kinetic modeling of *Zymomonas mobilis* Entner-Doudoroff pathway: insights into control and functionality. *Microbiology* 159, 2674–2689. doi: 10.1099/mic.0.071340-0
- Santos, F., Boele, J., and Teusink, B. (2011). A practical guide to genome-scale metabolic models and their analysis. *Methods Enzymol.* 500, 509–532. doi: 10.1016/B978-0-12-385118-5.00024-4
- Schellenberger, J., Que, R., Fleming, R. M. T., Thiele, I., Orth, J. D., Feist, A. M., et al. (2011). Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat. Protoc.* 6, 1290–1307. doi: 10.1038/nprot.2011.308
- Schuster, S., Fell, D. A., and Dandekar, T. (2000). A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat. Biotechnol.* 18, 326–332. doi: 10.1038/73786

- Schuster, S., Hilgetag, C., Woods, J. H., and Fell, D. A. (2002). Reaction routes in biochemical reaction systems: algebraic properties, validated calculation procedure and example from nucleotide metabolism. *J. Math. Biol.* 45, 153–181. doi: 10.1007/s002850200143
- Seo, J.-S., Chong, H., Park, H. S., Yoon, K.-O., Jung, C., Kim, J. J., et al. (2005). The genome sequence of the ethanologenic bacterium *Zymomonas mobilis* ZM4. *Nat. Biotechnol.* 23, 63–68. doi: 10.1038/nbt1045
- Snoep, J. L., Yomano, L. P., Westerhoff, H. V., and Ingram, L. O. (1995). Protein burden in *Zymomonas mobilis*: negative flux and growth control due to overproduction of glycolytic enzymes. *Microbiology* 141, 2329–2337.
- Sprenger, G. A. (1996). Carbohydrate metabolism in *Zymomonas mobilis*: a catabolic highway with some scenic routes. *FEMS Microbiol. Lett.* 145, 301–307.
- Strohdecker, M., Neuss, B., Bringer-Meyer, S., and Sahm, H. (1990). Electron transport chain of *Zymomonas mobilis*. Interaction with the membrane-bound glucose dehydrogenase and identification of ubiquinone 10. *Arch. Microbiol.* 154, 536–543.
- Trinh, C. T., Unrean, P., and Sreenc, F. (2008). Minimal *Escherichia coli* cell for the most efficient production of ethanol from hexoses and pentoses. *Appl. Environ. Microbiol.* 74, 3634–3643. doi: 10.1128/AEM.02708-07
- Tsantili, I. C., Karim, M. N., and Klapa, M. I. (2007). Quantifying the metabolic capabilities of engineered *Zymomonas mobilis* using linear programming analysis. *Microb. Cell Fact.* 6, 8. doi: 10.1186/1475-2859-6-8
- Wecker, M. S. A., and Zall, R. R. (1987). Production of acetaldehyde by *Zymomonas mobilis*. *Appl. Environ. Microbiol.* 53, 2815–2820.
- Widiastuti, H., Kim, J. Y., Selvarasu, S., Karimi, I. A., Kim, H., Seo, J.-S., et al. (2010). Genome-scale modeling and in silico analysis of ethanologenic bacteria *Zymomonas mobilis*. *Biotechnol. Bioeng.* 655–665. doi: 10.1002/bit.22965

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 16 November 2013; paper pending published: 29 December 2013; accepted: 21 January 2014; published online: 05 February 2014.

Citation: Kalnenieks U, Pentjuss A, Rutkis R, Stalidzans E and Fell DA (2014) Modeling of *Zymomonas mobilis* central metabolism for novel metabolic engineering strategies. *Front. Microbiol.* 5:42. doi: 10.3389/fmicb.2014.00042

This article was submitted to Microbial Physiology and Metabolism, a section of the journal *Frontiers in Microbiology*.

Copyright © 2014 Kalnenieks, Pentjuss, Rutkis, Stalidzans and Fell. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Comparative genomics and functional analysis of rhamnose catabolic pathways and regulons in bacteria

Irina A. Rodionova<sup>1</sup>, Xiaoqing Li<sup>1</sup>, Vera Thiel<sup>2</sup>, Sergey Stolyar<sup>3†</sup>, Krista Stanton<sup>4</sup>, James K. Fredrickson<sup>3</sup>, Donald A. Bryant<sup>2,5</sup>, Andrei L. Osterman<sup>1</sup>, Aaron A. Best<sup>4\*</sup> and Dmitry A. Rodionov<sup>1,6\*</sup>

<sup>1</sup> Sanford-Burnham Medical Research Institute, La Jolla, CA, USA

<sup>2</sup> Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA, USA

<sup>3</sup> Pacific Northwest National Laboratory, Biological Sciences Division, Richland, WA, USA

<sup>4</sup> Department of Biology, Hope College, Holland, MI, USA

<sup>5</sup> Department of Chemistry and Biochemistry, Montana State University, Bozeman, MT, USA

<sup>6</sup> A.A. Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow, Russia

## Edited by:

Katherine M. Pappas, University of Athens, Greece

## Reviewed by:

Akos T. Kovacs, Friedrich Schiller University of Jena, Germany  
Sacha A.F. T. Van Hijum, UMC St. Radboud, Netherlands

## \*Correspondence:

Aaron A. Best, Department of Biology, Hope College, 35 E 12th Street, Holland, MI 49423, USA  
e-mail: best@hope.edu;  
Dmitry A. Rodionov, Sanford-Burnham Medical Research Institute, 10901 North Torrey Pines Road, La Jolla, CA 92037, USA  
e-mail: rodionov@burnham.org

## † Present address:

Sergey Stolyar, Institute for Systems Biology, Seattle, USA

L-rhamnose (L-Rha) is a deoxy-hexose sugar commonly found in nature. L-Rha catabolic pathways were previously characterized in various bacteria including *Escherichia coli*. Nevertheless, homology searches failed to recognize all the genes for the complete L-Rha utilization pathways in diverse microbial species involved in biomass decomposition. Moreover, the regulatory mechanisms of L-Rha catabolism have remained unclear in most species. A comparative genomics approach was used to reconstruct the L-Rha catabolic pathways and transcriptional regulons in the phyla Actinobacteria, Bacteroidetes, Chloroflexi, Firmicutes, Proteobacteria, and Thermotogae. The reconstructed pathways include multiple novel enzymes and transporters involved in the utilization of L-Rha and L-Rha-containing polymers. Large-scale regulon inference using bioinformatics revealed remarkable variations in transcriptional regulators for L-Rha utilization genes among bacteria. A novel bifunctional enzyme, L-rhamnulose-phosphate aldolase (RhaE) fused to L-lactaldehyde dehydrogenase (RhaW), which is not homologous to previously characterized L-Rha catabolic enzymes, was identified in diverse bacteria including Chloroflexi, Bacilli, and Alphaproteobacteria. By using *in vitro* biochemical assays we validated both enzymatic activities of the purified recombinant RhaEW proteins from *Chloroflexus aurantiacus* and *Bacillus subtilis*. Another novel enzyme of the L-Rha catabolism, L-lactaldehyde reductase (RhaZ), was identified in Gammaproteobacteria and experimentally validated by *in vitro* enzymatic assays using the recombinant protein from *Salmonella typhimurium*. *C. aurantiacus* induced transcription of the predicted L-Rha utilization genes when L-Rha was present in the growth medium and consumed L-Rha from the medium. This study provided comprehensive insights to L-Rha catabolism and its regulation in diverse Bacteria.

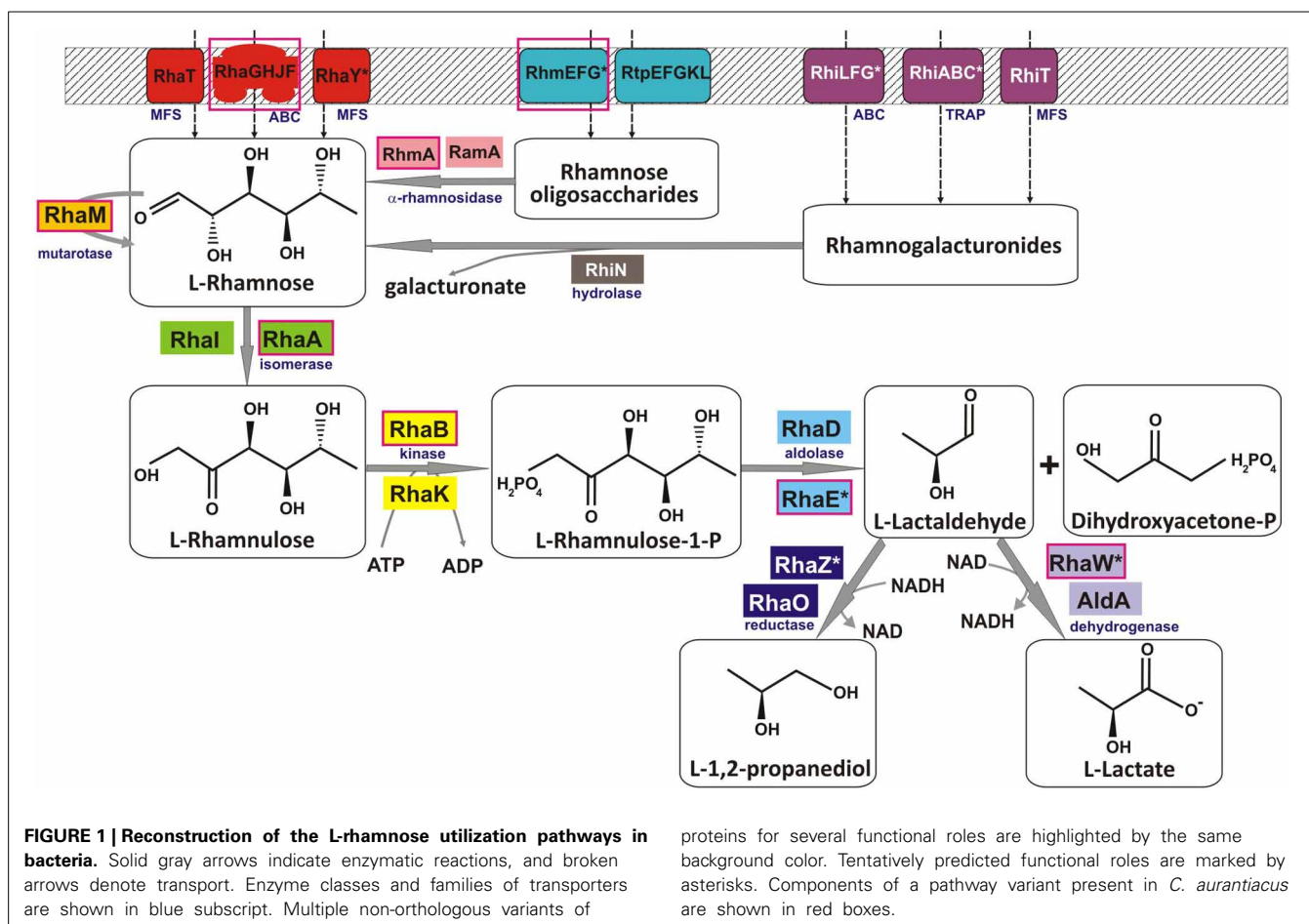
**Keywords:** L-rhamnose catabolism, metabolic reconstruction, regulon, comparative genomics, *Chloroflexus*

## INTRODUCTION

L-rhamnose (L-Rha) is a deoxy-hexose sugar commonly found in plants as a part of complex pectin polysaccharides and in many bacteria as a common component of the cell wall (Buttke and Ingram, 1975; Giraud and Naismith, 2000). Many microorganisms including the *Enterobacteriaceae* and *Rhizobiaceae* are capable of utilizing L-Rha as a carbon source (Eagon, 1961). Plant-pathogenic species (such as *Erwinia* spp.) and saprophytic species (e.g., *Bacillus subtilis*) are able to degrade rhamnogalacturonans and other L-Rha-containing polysaccharides by a set of extracellular enzymes including rhamnogalacturonate lyases (termed RhiE in *Erwinia* spp.) and  $\alpha$ -L-rhamnosidases (RhmA, RamA) (Laatu and Condemine, 2003; Ochiai et al., 2007; Avila et al., 2009). The resulting L-Rha and unsaturated rhamnogalacturonides can enter the cells by specific transport systems, the L-rhamnose permease RhaT in *Enterobacteriaceae* (Muiry

et al., 1993), and the RhiT transporter in *Erwinia chrysanthemi* (Hugouvieux-Cotte-Pattat, 2004). In the latter species, the unsaturated galacturonyl hydrolase RhiN is used to release L-Rha and unsaturated galacturonate residues to promote their further catabolism in the cytoplasm (Hugouvieux-Cotte-Pattat, 2004; Rodionov et al., 2004).

The canonical phosphorylated catabolic pathway for L-Rha described in enterobacteria is comprised of three enzymes, L-Rha isomerase (RhaA), L-rhamnulose kinase (RhaB), and L-rhamnulose-1-phosphate aldolase (RhaD) (Schwartz et al., 1974), which convert L-Rha to dihydroxyacetone phosphate (DHAP) and L-lactaldehyde (Akhy et al., 1984; Badia et al., 1989) (Figure 1). In addition, L-Rha mutarotase (RhaM) facilitates the interconversion of  $\alpha$  and  $\beta$  anomers of L-Rha, providing the stereochemically less-favored anomer for the subsequent catabolic reactions (Richardson et al., 2008). The structures



and reaction mechanisms each of these four enzymes from *Escherichia coli* have been determined (Korndorfer et al., 2000; Kroemer et al., 2003; Ryu et al., 2005; Grueninger and Schulz, 2006). Another L-Rha isomerase with broad substrate specificity (RhaI, 17% sequence identity to RhaA from *E. coli*) has been characterized in *Pseudomonas stutzeri* (Leang et al., 2004; Yoshida et al., 2007). L-lactaldehyde is a common product of both the L-rhamnose and L-fucose catabolic pathways and is further metabolized to L-lactate by the aldehyde dehydrogenase AldA or to 1,2-propanediol by the lactaldehyde reductase RhaO/FucO under certain conditions (Baldoma and Aguilar, 1988; Zhu and Lin, 1989; Patel et al., 2008). An alternative nonphosphorylated catabolic pathway for L-Rha comprising four metabolic enzymes L-rhamnose-1-dehydrogenase, L-rhamnono- $\gamma$ -lactonase, L-rhamnonate dehydratase and L-2-keto-3-deoxyrhamnonate aldolase, by which L-Rha is converted to pyruvate and L-lactaldehyde, have been identified in fungi and two bacterial species, *Azotobacter vinelandii* and *Sphingomonas* sp. (Watanabe et al., 2008; Watanabe and Makino, 2009).

Induction of the L-Rha utilization genes in *E. coli* is mediated by two rhamnose-responsive positive transcription factors (TFs) from the AraC family, RhaS, and RhaR (Tobin and Schleif, 1990; Egan and Schleif, 1993; Via et al., 1996). RhaR activates the *rhaSR* genes via binding to the inverted repeat of two 17

bp half sites separated by a 17 bp spacer. RhaS activates the *rhaBAD* and *rhaT* genes via binding to another inverted repeat of two sites whose sequence differs from the RhaR consensus binding site. In another bacterium, the plant pathogen *Erwinia chrysanthemi* from the order *Enterobacteriales*, the expanded RhaS regulon includes a similar set of genes involved in L-Rha utilization, as well as the rhamnogalacturonides utilization genes *rhiTN* (Hugouvieux-Cotte-Pattat, 2004). The L-Rha catabolic gene cluster in *Bacteroides thetaiotaomicron* is positively controlled by another AraC-family TF, which is non-orthologous to *E. coli* RhaR (16% identity) (Patel et al., 2008). In *Rhizobium leguminosarum* bv. *trifolii*, a novel negative TF of the DeoR family has been implicated in control of the L-Rha utilization regulon, which contains two divergently transcribed operons, *rhaRST-PQUK* and *rhaDI*, encoding an ABC transporter for L-Rha uptake (RhaSTPQ), an alternative kinase (RhaK, 19% identity to RhaB from *E. coli*), an isomerase (RhaI), and a mutarotase (RhaU, 41% identity to RhaM from *E. coli*) (Richardson et al., 2004, 2008; Richardson and Oresnik, 2007).

Our initial genome analysis suggested the presence of a novel variant of the L-Rha utilization pathway in anoxygenic phototrophic bacteria from the *Chloroflexi* phylum. Indeed, the existence of such pathway was implicated by the presence of *rhaA* and *rhaB* gene orthologs and the absence of *rhaD* and *rhaO*



genes in *Chloroflexus aurantiacus*. Moreover, the L-Rha catabolic pathway is not completely understood in many more bacterial species including *Bacillus subtilis*, and *Streptomyces coelicolor*. Mechanisms of transcriptional regulation of L-Rha utilization genes are also poorly understood in many species beyond the models. With the availability of hundreds of sequenced bacterial genomes, it is possible to use comparative genomics to reconstruct metabolic pathways and regulatory networks in individual taxonomic groups of Bacteria (Rodionov et al., 2010, 2011; Ravcheev et al., 2011, 2013; Leyn et al., 2013). Genome context-based techniques, including the analysis of chromosomal gene clustering, protein fusion events, phylogenetic co-occurrence profiles, and the genomic inference of metabolic regulons, are highly efficient methods for elucidation of novel sugar catabolic pathways. In our previous studies, we combined the genomic reconstruction of metabolic and regulatory networks with experimental testing of selected bioinformatic predictions to map sugar catabolic pathways systematically in two diverse taxonomic groups of bacteria, *Shewanella*, and *Thermotoga* (Rodionov et al., 2010, 2013). Furthermore, we have applied the integrated bioinformatic and experimental approaches to predict and validate novel metabolic pathways and transcriptional regulons involved in utilization of arabinose (Zhang et al., 2012), xylose (Gu et al., 2010), N-acetylglucosamine (Yang et al., 2006), N-acetylgalactosamine (Leyn et al., 2012), galacturonate (Rodionova et al., 2012a), and inositol (Rodionova et al., 2013) in diverse bacterial lineages.

In this work, we combined genomics-based reconstruction of L-Rha utilization pathways and RhaR transcriptional regulons in bacteria from diverse taxonomic lineages with the experimental validation of the L-Rha utilization system in *C. aurantiacus* and two other microorganisms. A novel bifunctional enzyme (named RhaEW) catalyzing two consecutive steps in L-Rha catabolism, L-rhamnulose-phosphate aldolase and L-lactaldehyde dehydrogenase, was identified in diverse bacterial lineages including Actinobacteria,  $\alpha$ -proteobacteria, Bacilli, Bacteroidetes, and Chloroflexi. The predicted dual function of RhaEW was validated by *in vitro* enzymatic assays with recombinant proteins from *C. aurantiacus* and *B. subtilis*. Another enzyme involved in L-lactaldehyde utilization in  $\gamma$ -proteobacteria, L-lactaldehyde reductase RhaZ, was identified and experimentally confirmed in *Salmonella* spp. Comparative analyses of upstream regions of the L-Rha utilization genes allowed identification of candidate DNA motifs for various groups of regulators from different TF families and reconstruction of putative rhamnose regulons. L-Rha-specific transcriptional induction and the predicted DNA binding motif of a novel DeoR-family regulator for of the *rha* genes were experimentally confirmed in *C. aurantiacus*.

## MATERIALS AND METHODS

### GENOMIC RECONSTRUCTION OF RHAMNOSE UTILIZATION PATHWAYS AND REGULONS

The comparative genomic analysis of L-Rha utilization subsystem was performed using the SEED genomic platform (Overbeek et al., 2005), which allowed annotation and capture of gene functional roles, their assignment to metabolic subsystems, identification of non-orthologous gene displacements, and

projection of the functional annotations across microbial genomes, as it was previously described for other sugar catabolic subsystems (Rodionov et al., 2010, 2013; Leyn et al., 2012; Rodionova et al., 2012a, 2013). The obtained functional gene annotations were captured in the SEED subsystem available online at [http://pubseed.theseed.org/SubsysEditor.cgi?page=ShowSubsystem&subsystem=L-rhamnose\\_utilization](http://pubseed.theseed.org/SubsysEditor.cgi?page=ShowSubsystem&subsystem=L-rhamnose_utilization) and are summarized in Table S1 in the Supplementary Material.

For reconstruction of RhaR regulons we used an established comparative genomics approach based on identification of candidate regulator-binding sites in closely related bacterial genomes implemented in the RegPredict Web server tool ([regpredict.lbl.gov](http://regpredict.lbl.gov)) (Novichkov et al., 2010). First, we identified potential *rhaR* transcription factor genes that are located within the conserved neighborhoods of the L-Rha catabolic genes in bacterial genomes from each studied taxonomic lineage. Identification of orthologs in closely related genomes and gene neighborhood analysis were performed in MicrobesOnline (<http://microbesonline.org/>) (Dehal et al., 2010). To find the conserved DNA-binding motifs for each group of orthologous RhaR regulators, we used initial training sets of genes that are co-localized with *rhaR* orthologs (putative operons containing at least one candidate L-Rha utilization gene and that are located in the vicinity of a maximum ten genes from a *rhaR* gene), and then we updated each set by the most likely RhaR-regulated genes confirmed by the comparative genomics tests as well as functional considerations (i.e., involvement of candidate target genes in the L-Rha utilization pathway). Using the Discover Profile procedure in RegPredict, common DNA motifs with palindromic or direct repeat symmetry were identified and their corresponding position weight matrices (PWMs) were constructed. The initial PWMs were used to scan the reference genomes and identify additional RhaR-regulated genes that share similar binding sites in their upstream regions. The conserved regulatory interactions were included in the reconstructed RhaR regulons using the clusters of co-regulated orthologous operons in RegPredict. Candidate sites associated with new members of the regulon were added to the training set, and the respective lineage-specific PWM was rebuilt to improve search accuracy. Sequence logos for the derived DNA-binding motifs were built using the Weblogo package (Crooks et al., 2004). The details of all reconstructed regulons are displayed in the RegPrecise database of regulons (Novichkov et al., 2013) available online at [http://regprecise.lbl.gov/RegPrecise/collection\\_pathway.jsp?pathway\\_id=34](http://regprecise.lbl.gov/RegPrecise/collection_pathway.jsp?pathway_id=34).

### GENE CLONING AND PROTEIN PURIFICATION

The *rhaEW* (*Caur\_2283*) and *rhaR* (*Caur\_2290*) genes from *C. aurantiacus* J-10-fl, the *rhaEW* (*yuxG*) gene from *B. subtilis*, and the *rhaZ* (*STM4044*) and *rhaD* (*STM4045*) genes from *Salmonella enterica* serovar Typhimurium LT2 were amplified by PCR from genomic DNA using specific primer pairs (see Table S2 in Supplementary Material). A pET-derived vector, pODC29 Gerdes et al. (2006), containing a T7 promoter and an N-terminal His<sub>6</sub> tag, or a similar vector, pProEX HTb (Invitrogen), with a *trc* promoter was used for cloning and protein expression. The *rhaR* gene was cloned into the pSMT3 expression vector (Mossessova and Lima, 2000) (a kind gift of Dr. Lima from Cornell University).



The obtained plasmid encodes a fusion between the RhaR protein and an N-terminal Hexa-histidine Smt3 polypeptide (a yeast SUMO ortholog), which enhances protein solubility. The resulting plasmids were transformed into *E. coli* BL21/DE3 or BL21 (Gibco-BRL, Rockville, MD). Recombinant proteins were overexpressed as fusions with an N-terminal His<sub>6</sub>tag and purified to homogeneity using Ni<sup>2+</sup>-chelation chromatography. Cells were grown in LB medium (50 ml), induced by addition of 0.2 mM isopropyl- $\beta$ -D-thiogalactopyranoside, and harvested after 4 h of additional shaking at 37°C (for Caur\_2283, and Caur\_2290) or 16 h of shaking at 25°C (for YuxG, STM4044, and STM4045). Harvested cells were resuspended in 20 mM HEPES buffer (pH 7) containing 100 mM NaCl, 0.03% Brij-35, 2 mM  $\beta$ -mercaptoethanol, and 2 mM phenylmethylsulfonyl fluoride (Sigma-Aldrich). Cells were lysed by incubation with lysozyme (1 mg/ml) for 30 min, followed by a freeze-thaw cycle and sonication. After centrifugation, Tris-HCl buffer (pH 8) was added to the supernatant (50 mM, final concentration), which was loaded onto Ni-nitrilotriacetic acid (NTA) agarose minicolumn (0.3 ml) from Qiagen Inc. (Valencia, CA). After washing with starting buffer containing 1 M NaCl and 0.3% Brij-35 bound proteins were eluted with 0.3 ml of the same buffer supplemented with 250 mM imidazole. The purified proteins were electrophoresed on a 12% (w/v) sodium dodecyl sulfate-polyacrylamide gel to monitor size and purity (>90%). Protein concentration was determined by the Quick Start Bradford Protein Assay kit from Bio-Rad.

## ENZYME ASSAYS

Aldolase/dehydrogenase activities of the purified recombinant RhaEW proteins from *C. aurantiacus* (Ca\_RhaEW) and *B. subtilis* (Bs\_RhaEW), and the St\_RhaD and St\_RhaZ proteins from *Salmonella typhimurium* were tested by a direct NADH detection assay. Because L-rhamnulose-1-P is not commercially available, we used an enzymatic coupling assay with two upstream catabolic enzymes for the conversion of L-Rha to L-rhamnulose-1-P. The L-Rha isomerase RhaA from *E. coli* (Ec\_RhaA) and the L-rhamnulose kinase RhaB from *Thermotoga maritima* (Tm\_RhaB) were expressed in *E. coli* and purified as described previously (Rodionova et al., 2012b). For Ca\_RhaEW assays, the purified recombinant enzymes, Ec\_RhaA (2  $\mu$ g) and Tm\_RhaB (2  $\mu$ g), were pre-incubated during 20 min at 37°C in 100  $\mu$ l of reaction mixture containing 150 mM Tris-HCl (pH 8), 20 mM MgCl<sub>2</sub>, 10 mM ATP, 1.4 mM NAD<sup>+</sup>, 10  $\mu$ M ZnSO<sub>4</sub>, and 8 mM L-Rha. Then Ca\_RhaEW (0.5  $\mu$ g) was added to the assay mixture and the reduction of NAD<sup>+</sup> was followed by increase in absorbance at 340 nm at different temperatures (30–70°C) in the spectrophotometer. For Bs\_RhaEW, St\_RhaD and St\_RhaZ assays, Ec\_RhaA and Tm\_RhaB were pre-incubated in a ratio of 40:1 (RhaA:RhaB) at 25°C for 40 min in a reaction mixture containing 50 mM Tris-HCl (pH 7.5), 20 mM MgCl<sub>2</sub>, 1 mM ATP, 50 mM KCl, 2.5 mM NAD<sup>+</sup> or 0.25 mM NADH, 5  $\mu$ M ZnCl<sub>2</sub>, and 2 mM L-Rha. Subsequently, either Bs\_RhaEW (2.8  $\mu$ g) or St\_RhaD (10  $\mu$ g) and St\_RhaZ (2.8  $\mu$ g) enzymes were added and the reduction of NAD<sup>+</sup> or oxidation of NADH was monitored by increase or decrease in absorbance at 340 nm, respectively, at 25°C in a final reaction volume of 200  $\mu$ l.

## GC-MS ANALYSIS

Four-step biochemical conversions of L-Rha to L-lactate and DHAP by mixtures of the three L-Rha catabolic enzymes were monitored by GC-MS. Samples from enzymatic assay mixtures (10  $\mu$ l) were dried in a vacuum centrifuge at room temperature, and derivatized at 80°C for 20 min with 75  $\mu$ l of pyridine containing 50 mg ml<sup>-1</sup> methoxyamine or ethylhydroxylamine (for lactate detection). The solution was incubated at 80°C for 60 min with 75  $\mu$ l of N,O-bis-(trimethylsilyl)trifluoroacetamide or N-tert-butyltrimethylsilyl-N-methyltrifluoroacetamide (for lactate detection). After derivatization, the samples were centrifuged for 1 min at 14,000 r.p.m. and the supernatant (1  $\mu$ l) was transferred to vials for GC-MS analysis. A QP2010 Plus GC-MS instrument was from Shimadzu (Columbia, MD). GC-MS analyses were performed as previously described in Rodionova et al. (2012a, 2013).

## BACTERIAL STRAINS AND GROWTH CONDITIONS

The *yuxG(rhaEW)* and *yceI(niaP)* disruption strains of *B. subtilis* were obtained from the joint Japanese and European *B. subtilis* consortium (Kobayashi et al., 2003). The latter strain with an insertion in the niacin transporter *niaP* was used as an isogenic negative control. Both strains were grown overnight at 37°C in chemically defined medium containing D-glucose (4 g/l), L-tryptophan (50 mg/l), L-glutamine (2 g/l), K<sub>2</sub>HPO<sub>4</sub> (10 g/l), KH<sub>2</sub>PO<sub>4</sub> (6 g/l), sodium citrate (1 g/l), MgSO<sub>4</sub> (0.2 g/l), K<sub>2</sub>SO<sub>4</sub> (2 g/l), FeCl<sub>3</sub> (4 mg/l), and MnSO<sub>4</sub> (0.2 mg/l) in the presence of erythromycin (0.5 mg/l) (pMUTIN2 marker). Overnight cultures were diluted ~10-fold to yield the same cell density (optical density at 600 nm of 0.05) in the defined medium lacking glucose and washed three times to remove residual glucose. Cells were grown in triplicate in one of two versions of the defined medium containing L-Rha (4 g/l), or no additional carbon source. *C. aurantiacus* J-10-fl was grown at 52°C in 25 ml screw capped glass tubes completely filled with BG-11 medium (Stanier et al., 1971) supplemented with 0.02% (w/v) of NH<sub>4</sub>Cl and 2 mM of NaHCO<sub>3</sub>. 0.2% of yeast extract (YE) or 35 mM of pyruvate, both with and without additional 20 mM L-Rha, were used as main carbon source and cultures grown under microaerobic starting conditions in the light. Cultures were constantly mixed on a rotation wheel during incubation. Growth of cultures was monitored at 600 nm using a ELX-808IU microplate reader from BioTek Instruments Inc. (Winooski, VT). The concentration of L-Rha in culture fluids was determined on an HPLC equipped with an HPX 78 (Bio-Rad) column.

## RT-PCR

Individual transcript levels were measured for seven genes from *C. aurantiacus*: *rhaB* (Caur\_2282), *rhaF* (Caur\_2286), *rhaR* (Caur\_2290), *rhmA* (Caur\_0361), and *Caur\_0839* (NADH-flavin oxidoreductase/NADH oxidase). The latter housekeeping gene was used as a positive control since it was found to be highly expressed under both photoheterotrophic as well as chemoheterotrophic conditions in a previous proteome study (Cao et al., 2012). Total RNA was isolated from cells grown on BG-11 medium supplied with YE, YE plus L-Rha, pyruvate, and pyruvate plus L-Rha under suboxic conditions in the light, and collected

after 3 days at optical densities at 650 nm of 1.3, 0.9, 0.4, and 0.6, respectively. RNA was isolated using a phenol-chloroform extraction method adapted from (Aiba et al., 1981; Steunou et al., 2006). Cell pellets were resuspended in 250  $\mu$ l 10 mM sodium acetate (pH 4.5) and 37.5  $\mu$ l 500 mM Na<sub>2</sub>EDTA (pH 8.0), then mixed with 375  $\mu$ l Lysis buffer (10 mM sodium acetate, 2% SDS, pH to 4.5). Hot (65°C) acidic (pH 4.5) phenol (700  $\mu$ l) was added, the sample was vortexed and incubated at 65°C for 3 min. After centrifugation (17,000  $\times$  g, 2 min), the RNA was further purified by one phenol-chloroform-isoamyl alcohol (25:24:1) and one chloroform extraction. RNA was precipitated using 0.1 volume of 10 M LiCl and 2.5 volume 100% EtOH and precipitated at -80°C for at least 30 min, washed with 80% EtOH and resuspended in DEPC treated H<sub>2</sub>O. The RNA solution was treated with DNase I (New England Biolab Inc.) and re-precipitated after an additional chloroform:isoamyl alcohol (24:1) extraction. The purified RNA was dissolved in DEPC-treated water. Semi-quantitative RT-PCR was conducted using a Biotin Tetro one-step RT-PCR kit following the manufacturer's protocol. The gene-specific primers for each gene tested are shown in Table S2 in Supplementary Material. For each reaction one control for DNA contamination was included (same template as for RT-PCR, started with inactivation of RT-Polymerase step) and a PCR positive control (using 10 ng whole genome DNA from *C. aurantiacus* as template) was used. PCR conditions were the same for each primer pair used. All started with a 30 min RT-step at 42°C followed by an RT-inactivation step at 95°C. Then a single step PCR for amplification of the genes from cDNA was conducted using 30 cycles of 30 s denaturation at 95°C, 30 s annealing at 60°C, and 90 s elongation step at 72°C before cooling down to 10°C.

## DNA BINDING ASSAYS

The interaction of the purified recombinant *C. aurantiacus* RhaR protein with its cognate DNA binding site in *C. aurantiacus* was assessed using an electrophoretic mobility-shift assay (EMSA). The His<sub>6</sub>-Smt3-tag was cleaved from the purified RhaR protein by digestion with Ulp1 protease. Complementary DNA fragments, containing the predicted 38-bp RhaR binding site from the *Caur\_2290* promoter region and flanked on each side by five guanosine residues (Table S2 in Supplementary Materials) were synthesized by Integrated DNA Technologies. One strand of oligo was 3'-labeled by a biotin label, whereas the complementary oligo was unlabeled. Double-stranded labeled DNA fragments were obtained by annealing the labeled oligonucleotides with unlabeled complementary oligonucleotides at a 1:10 ratio. The biotin-labeled 48-bp DNA fragment (0.2 nM) was incubated with increasing concentrations of the purified RhaR protein (10–1000 nM) in a total volume of 20  $\mu$ l of the binding buffer containing 50 mM Tris-HCl (pH 8.0), 150 mM NaCl, 5 mM MgCl<sub>2</sub>, 1 mM DDT, 0.05% NP-40, and 2.5% glycerol. Poly(dI-dC) (1  $\mu$ g) was added as a nonspecific competitor DNA to reduce non-specific binding. After 25 min of incubation at 50°C, the reaction mixtures were separated by electrophoresis on a 1.5% (w/v) agarose gel at room temperature. The DNA was transferred by electrophoresis onto a Hybond-N<sup>+</sup> membrane and fixed by UV-cross-linking. The biotin-labeled DNA was detected with the LightShift chemiluminiscent EMSA kit (Thermo Fisher

Scientific Inc, Rockford, IL, USA). Additional DNA fragment of the *Caur\_0003* gene upstream region (Table S2 in Supplementary Materials) was used as a negative control. The effect of D-glucose, L-Rha, and L-rhamnulose (obtained by enzymatic conversion of L-Rha by *Ec\_RhaA*) was tested by their addition to the incubation mixture.

## RESULTS

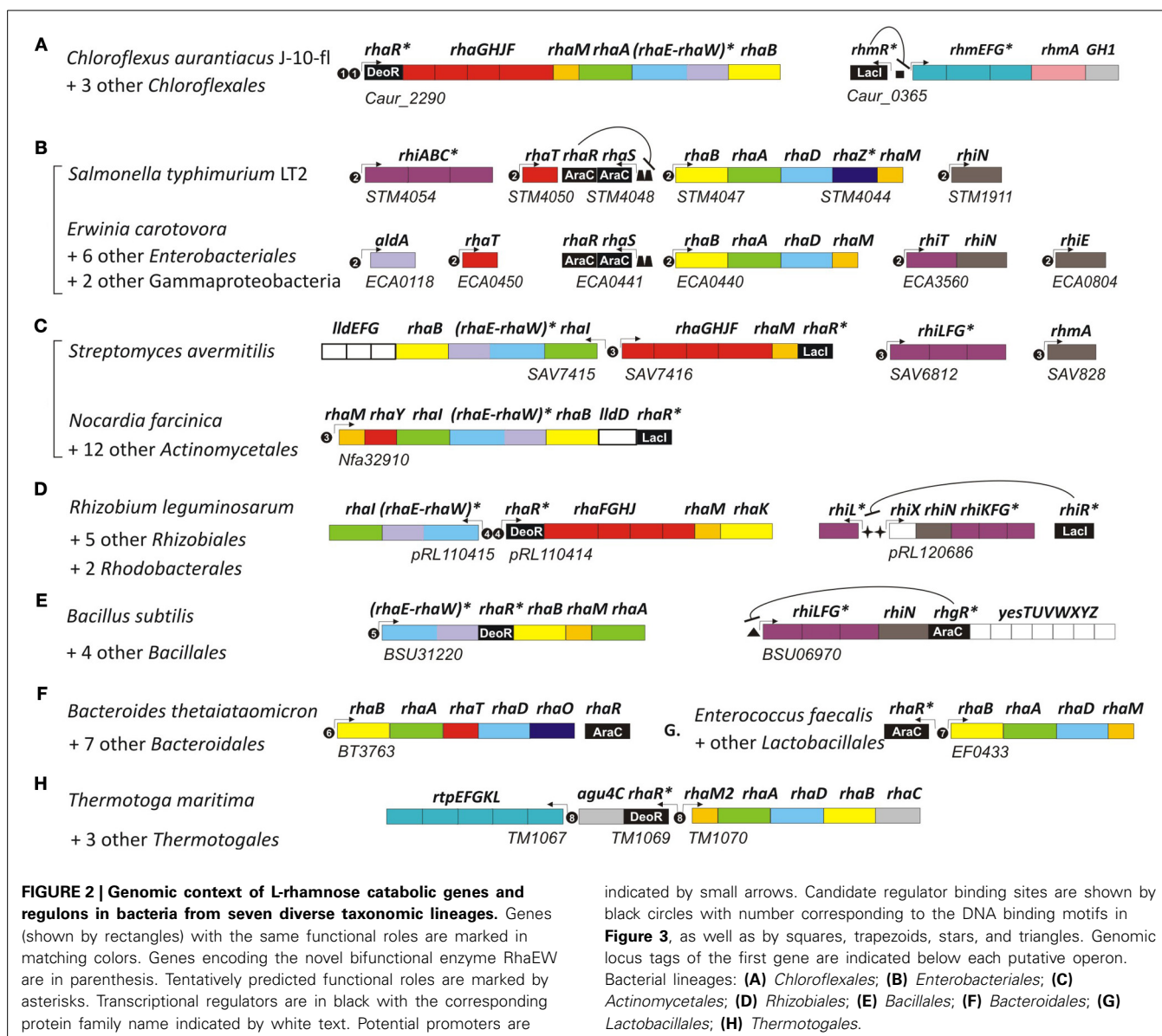
### COMPARATIVE GENOMICS OF L-RHAMNOSE UTILIZATION IN BACTERIA

To reconstruct catabolic pathways and transcriptional regulons involved in L-Rha utilization in bacteria we utilized the subsystem-based comparative genomics approach implemented in the RegPredict and the SEED Web resources (Overbeek et al., 2005; Novichkov et al., 2010). As a result, the L-Rha metabolic pathway genes and transcriptional regulons were identified in complete genomes of 55 representatives of diverse taxonomic groups of bacteria including the *Actinomycetales*, *Bacteroidales*, *Chloroflexales*, *Bacillales*, *Rhizobiales*, *Enterobacteriales*, and *Thermotogales*. The distribution of genes encoding the L-Rha catabolic enzymes and associated transporters and regulators across the studied species is summarized in Table S1 in Supplementary Material. The studied bacterial species possess many variations in key enzymes from the L-Rha catabolic pathway, as well as in mechanisms of sugar uptake and transcriptional regulation. Some of these variations are briefly described below when we describe novel functional variants of the L-Rha catabolic pathway and novel transcriptional regulons for these pathways.

#### *L-rhamnose catabolic regulons*

The transcriptional regulator RhaS in *E. coli* belongs to the AraC protein family and controls the L-Rha transporter *rhaT* and the catabolic operon *rhaBADU* (Egan and Schleif, 1993; Via et al., 1996). Orthologs of *rhaS* and these catabolic genes for L-Rha utilization are present in other *Enterobacteriales*, as well as in *Tolomonas* and *Mannheimia* spp. RhaS in *E. chrysanthemi* was additionally shown to regulate the *rhiTN* operon involved in the uptake and catabolism of rhamnogalacturonides, L-rhamnose containing oligosaccharides (Hugouvieux-Cotte-Pattat, 2004). The analysis of upstream regions of RhaS-controlled genes and their orthologs in  $\gamma$ -proteobacteria resulted in identification of the putative RhaS-binding motif, which was used for identification of additional RhaS targets in the analyzed genomes (Figures 2B, 3B).

Analysis of other taxonomic groups outside the  $\gamma$ -proteobacteria identified previously uncharacterized members of the LacI, DeoR, and AraC families as alternative transcriptional regulators of the L-rhamnose catabolic pathways (Figure 2). To infer novel L-Rha regulons in each taxonomic group, we applied the comparative genomics approach that combines identification of candidate regulator-binding sites with cross-genomic comparison of regulons. The upstream regions of L-Rha utilization genes in each group of genomes containing an orthologous TF was analyzed using a motif-recognition program to identify conserved TF-binding DNA motifs (Figure 3). The deduced palindromic DNA motifs of novel LacI-family regulators are

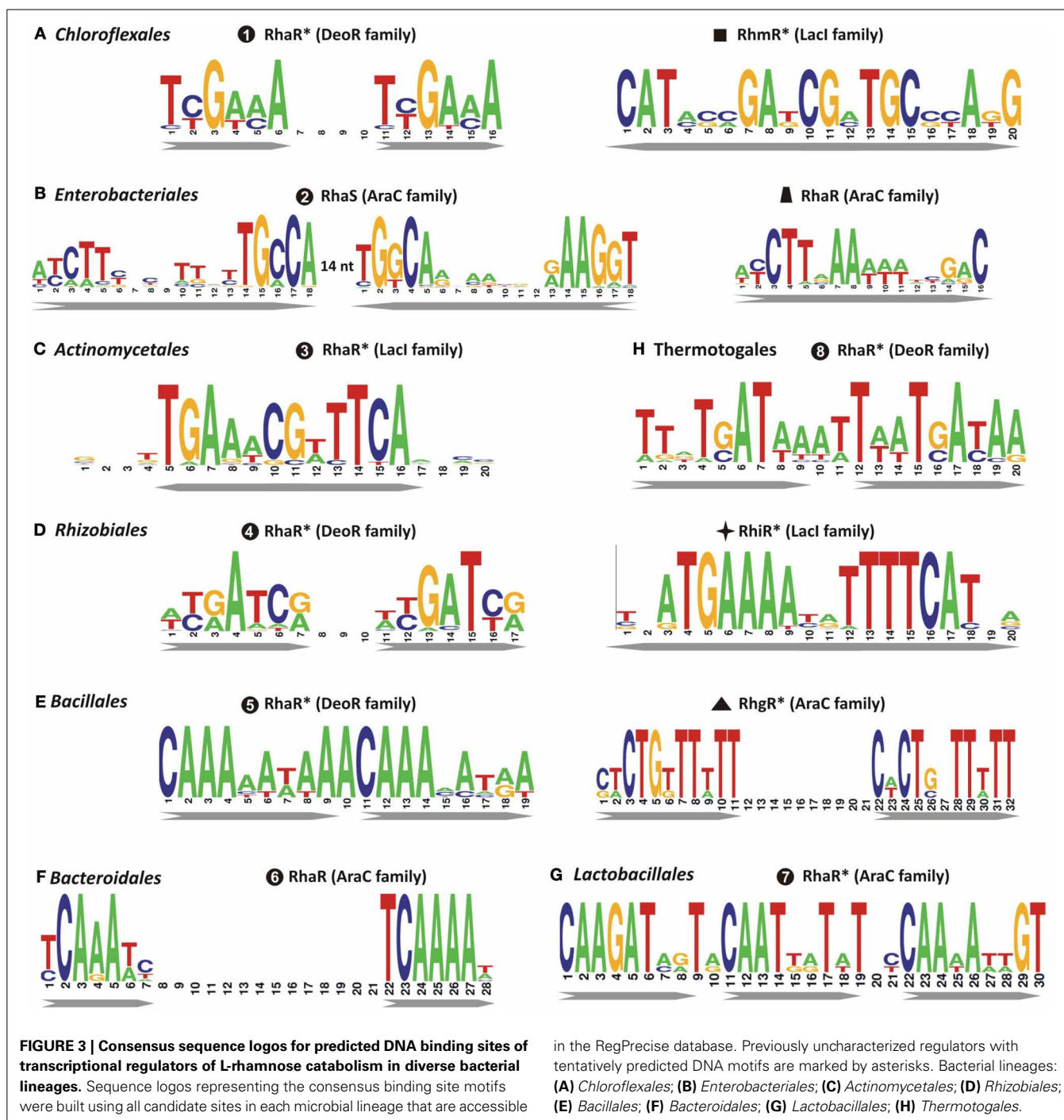


characteristic of DNA-binding sites of LacI family regulators. The predicted binding motifs of DeoR-family RhaR regulators in four distinct taxonomic groups are characterized by unique sequences; however, each of them has a similar structure that includes two imperfect direct repeats with a periodicity of 10–11 bp. Novel AraC-family regulators of L-Rha metabolism in the *Bacillales*, *Bacteroides*, and *Enterococcus* groups also are characterized by unique DNA motifs with a common structure of a direct repeat with 21-bp periodicity. Among this large set of predicted L-Rha catabolic regulators, only two transcriptional factors, an AraC-type activator in *B. thetaiotaomicron* and a DeoR-type repressor in *R. leguminosarum*, have been shown experimentally to mediate the transcriptional control of L-Rha utilization genes in the previous studies (Richardson et al., 2004; Patel et al., 2008), although specific DNA operator motifs of these two regulators were not reported before.

A detailed description of the reconstructed L-Rha catabolic regulons is available in the RegPrecise database within the

collection of regulons involved in L-Rha utilization (Novichkov et al., 2013). Overall, most of these TF regulons are local and control from one to several target operons per genome (Figure 2). In the *Bacillales*, RhaR and RhgR control genes involved in the utilization of L-rhamnose and rhamnolacturonan, respectively (Leyn et al., 2013). In the *Thermotogales*, the DeoR-family RhaR regulator co-regulates genes involved in the utilization of L-Rha mono- and oligosaccharides (Rodionov et al., 2013). In the *Rhizobiales*, RhaR from the DeoR family negatively controls the L-Rha catabolic operon (Richardson et al., 2004), whereas RhiR from the LacI family is predicted to regulate the rhamnolacturonide utilization gene cluster (named *rhi*). An orthologous LacI-family regulator controls the similar *rhi* gene locus in *Erwinia* spp. In the *Actinomycetales*, a novel LacI-type regulator (termed RhaR) co-regulates genes involved in the uptake and catabolism of L-Rha and L-Rha-containing oligosaccharides. In the *Chloroflexales*, two unique TFs control L-Rha metabolism—the DeoR-family regulator RhaR controls the





L-Rha utilization operons in both *Chloroflexus* and *Roseiflexus* spp., while the LacI-family regulator RhmR controls the *rhm* operon involved in the L-Rha oligosaccharide utilization in *C. aurantiacus*.

In summary, at least seven non-orthologous types of TFs appear to regulate the L-rhamnose utilization (*rha*) genes in diverse bacterial lineages. Uptake and catabolism of L-Rha-containing oligosaccharides is either co-regulated with *rha* genes by the same TFs (e.g., RhaRs in *Actinomycetales* and *Thermotogales*; RhaS in *Enterobacteriales*), or is under control of

other specialized TFs (RhgR in *Bacillales*, RhiR in *Rhizobiales*, and *Erwinia*, RhmR in *Chloroflexus*). In the third part of this study, we experimentally validated the predicted DNA binding sites of RhaR regulator in *C. aurantiacus*.

#### L-rhamnose catabolic pathways

Analysis of L-Rha regulons revealed various sets of genes that are presumably involved in the L-rhamnose utilization subsystem (Table S1 in Supplementary Material). By analyzing protein similarities and genomic contexts for these genes, we inferred

their potential functional roles and reconstructed the pathways (Figure 1). All four enzymatic steps of the reconstructed catabolic pathways occur in many alternative forms. The most conserved enzyme in the L-Rha subsystem is the L-rhamnulose kinase RhaB, which is substituted by a non-orthologous kinase from the same protein family in  $\gamma$ -proteobacteria (Rodionova et al., 2012b). Two alternative types of L-rhamnulose isomerase (RhaA and RhaI) are almost equally distributed among the studied genomes. All analyzed lineages except the *Bacillales* possess L-rhamnulose isomerases of a single type. Among the *Bacillales*, all studied genomes have the RhaA isomerase, whereas only *B. licheniformis* has the non-orthologous isozyme RhaI.

The canonical form of L-rhamnulose-1-P aldolase (RhaD) was found in  $\gamma$ -proteobacteria, *Bacteroidales*, *Thermotogales*, and *Lactobacillales*. Instead of RhaD, the L-Rha catabolic gene clusters in *Actinomycetales*,  $\alpha$ -proteobacteria, *Bacillales*, and *Chloroflexales* contain a chimeric gene encoding a two-domain protein (e.g., *yuxG* in *Bacillus subtilis*). The uncharacterized protein YuxG and its orthologs have an N-terminal class II aldolase domain (PF00596 protein family in PFAM) fused to a C-terminal short-chain dehydrogenase domain (PF00106). We used DELTA\_BLAST to search for distant homologues of YuxG among proteins with experimentally determined functions. The N-terminal domain of YuxG (named RhaE) is distantly homologous to three *E. coli* enzymes, L-ribulose-5-phosphate epimerase (15% identity, *E*-value  $1e^{-39}$ ), L-fucose-1-phosphate aldolase (11% identity, *E*-value  $4e^{-36}$ ), and the canonical RhaD enzyme (14% identity, *E*-value  $1e^{-25}$ ). These relationships suggest that it represents a non-orthologous substitution of aldolase RhaD. The C-terminal domain of YuxG (named RhaW) is homologous to various NADH- and NADPH-dependent sugar dehydrogenases including sorbose dehydrogenase from fungi (29% identity, *E*-value  $3e^{-14}$ ), 2,3-butanediol dehydrogenase from *Corynebacterium glutamicum* (25% identity, *E*-value  $4e^{-10}$ ), and sorbitol-6-phosphate dehydrogenase from *E. coli* (22% identity, *E*-value  $9e^{-09}$ ). The phylogenetic occurrence profile suggests that RhaW may encode the missing L-lactaldehyde dehydrogenase/reductase. Thus, the bifunctional protein RhaEW is tentatively predicted to catalyze the two final reactions in the L-Rha catabolic pathway (Figure 1).

Downstream enzymes for utilization of L-lactaldehyde varied the most among the analyzed species. Reconstruction of the RhaS regulon in  $\gamma$ -proteobacteria identified various genes that are likely involved in utilization of L-lactaldehyde. The rhamnose operons in *S. typhimurium* and five other species include an additional gene (named *rhaZ*) that encodes a hypothetical iron-containing alcohol dehydrogenase (PF00465). *E. carotovora* has a single RhaS-regulated gene *aldA* encoding alcohol dehydrogenase from another protein family (PF00171). In contrast, the RhaS regulons in *E. chrysanthemi* and *Mannheimia* spp. include the L-lactaldehyde reductase *rhaO*, whereas *aldA* and *rhaZ* are absent from their genomes. These observations suggest that  $\gamma$ -proteobacteria use three different enzymes and two different pathways for the final stage of the L-rhamnose pathway (Figure 1).

In summary, the subsystem reconstruction and genome context analyses allowed us to predict the following novel candidate genes: L-rhamnulose-1-P aldolase (RhaE) and two variants of L-lactaldehyde utilizing enzymes (RhaW and RhaZ) in diverse bacterial genomes. In the second part of this study, we experimentally validated the predicted functions of RhaEW from *B. subtilis* and *C. aurantiacus* and RhaZ from *S. typhimurium*.

### L-rhamnose transporters and upstream hydrolytic pathways

Uptake of L-Rha in *E. coli* is mediated by the L-Rha-proton symport protein, RhaT (Baldoma et al., 1990) that belongs to the Drug/Metabolite Transporter (DMT) superfamily. An orthologous L-Rha transporter was found in the genome context of L-Rha utilization genes/regulons in other  $\gamma$ -proteobacteria and in the *Bacteroidales* (Table S1 in Supplementary Material). Another L-Rha transporter belonging to the ABC superfamily, RhaSTPQ (designated RhaFGHJ here) was described in *R. leguminosarum* (Richardson et al., 2004). In this study, we identified orthologs within the L-Rha operons/regulons in all other  $\alpha$ -proteobacteria, as well as in several genomes from the *Chloroflexales*, *Actinomycetales*, and *Enterobacteriales* orders. A different putative L-Rha transporter (termed RhaY), which belongs to the Sugar Porter (SP) family of the Major Facilitator Superfamily (MFS), was identified in certain *Bacillales* and *Actinomycetales* genomes. This functional assignment is supported by the conserved co-localization on the chromosome (in *Mycobacterium/Nocardia* spp.) and by predicted co-regulation (via upstream RhaR-binding site in *Saccharopolyspora erythraea*) with other *rha* genes.

The predicted L-Rha regulons in many bacteria include several glycoside hydrolases and transport systems involved in the uptake of L-Rha-containing oligosaccharides in the cytoplasm and their consequent degradation to form L-Rha monosaccharides. The RhaS-activated operon *rhiTN* is involved in the uptake and hydrolysis of oligosaccharides produced during rhamnogalacturonan catabolism in the plant-pathogenic species from the order *Enterobacteriales* (Hugouvieux-Cotte-Pattat, 2004). Another enterobacterium, *S. typhimurium*, possesses a different RhaS-regulated transport system (named *rhiABC*), which is similar to the C4-dicarboxylate transport system Dcu (Figure 2). Based on the gene occurrence pattern and candidate co-regulation, *rhiABC* is tentatively predicted to encode an alternative transporter for rhamnogalacturonides, which replaces RhiT in *S. typhimurium*. A different transport system from the ABC family (named *rhiLFG*) and putative  $\alpha$ -L-rhamnosidases (*ramA*, *rhmA*) were detected within the RhaR regulons in several *Actinomycetales*. In the *Bacillales* and *Rhizobiales* groups, as well as in the *Ewringia* and *Chloroflexus* spp., homologous ABC transporters and rhamnohydrolases are co-regulated with several novel lineage-specific transcriptional regulons, RhgR, RhiR, and RhmR, respectively.

In summary, the comparative genomics analysis of L-Rha catabolic subsystem in bacteria revealed extensive variation for the components of transport machinery. L-Rha transport systems belong to at least three protein families. In addition to L-Rha transporters, many L-Rha-utilizing bacteria possess systems for active uptake of L-Rha containing oligosaccharides.



## EXPERIMENTAL VALIDATION OF NOVEL RHAMNOSE CATABOLIC ENZYMES

### Novel aldolase/dehydrogenase RhaEW

To provide biochemical evidence for the novel bifunctional aldolase/dehydrogenase enzyme involved in L-Rha catabolism, the recombinant protein RhaEW from *C. aurantiacus* (termed *Ca\_RhaEW*) was overexpressed in *E. coli* with the N-terminal His6 tag, purified using Ni-NTA affinity chromatography, and characterized *in vitro* by a coupled enzymatic assay using spectrophotometry and GC-MS.

Bioinformatics analysis suggested that RhaEW is a bifunctional enzyme catalyzing two sequential activities, L-rhamnulose-1-P aldolase and L-lactaldehyde dehydrogenase (Figure 1). We assayed the biochemical activity of the recombinant *Ca\_RhaEW* protein by monitoring the conversion of NAD<sup>+</sup> to NADH at 340 nm as a result of predicted L-lactaldehyde dehydrogenase reaction. The peak of *Ca\_RhaEW* activity ( $V_{max}$  2.9 U mg Major Facilitator Superfamily<sup>-1</sup>) was observed at 60–70°C (Figure 4A), which is in agreement with the optimal temperature range for the growth for *C. aurantiacus* (Hanada and Pierson, 2006). Additionally, we tested the possibility that *Ca\_RhaEW* acts as an aldolase/reductase by supplying NADH rather than NAD<sup>+</sup> in the reaction; no activity was seen under these conditions (data not shown). Thus, *Ca\_RhaEW* acts *in vitro* to convert L-rhamnulose-1-P to L-lactate and DHAP, which is consistent with the prediction made through comparative genomics analyses.

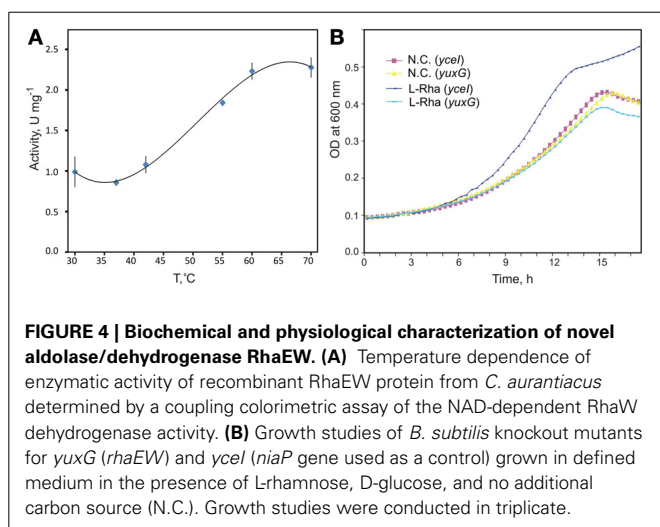
The formation of *Ca\_RhaEW* reaction products was directly confirmed by GC-MS profiling of reaction mixtures obtained by overnight incubation of L-Rha with the *Ca\_RhaEW* protein taken alone or in combination with the upstream catabolic enzymes. While incubation of L-Rha with *Ca\_RhaEW* alone did not produce any new peaks on the chromatogram, the addition to the mixture of the *Ec\_RhaA* and *Tm\_RhaB* proteins led to a decrease of two peaks corresponding to L-Rha (retention times 9.28 and 9.39 min) and the appearance of a series of novel peaks (Figure S1 in Supplementary Material). By comparison with standards and the analysis of electron ionization mass spectra ( $m/z$  299), the first two peaks with retention times 9.27 and 9.37 min were

attributed to DHAP, whereas the peak at retention time 7.75 min was assigned as lactate. Additional peaks appearing in the coupled enzymatic assay were attributed to the upstream intermediates of the L-Rha catabolic pathway, L-rhamnulose (retention times 8.85 and 8.92 min) and L-rhamnulose-1-P (13.04 min). The moderate consumption of L-Rha observed when only *Ec\_RhaA* and *Tm\_RhaB* enzymes were added increased substantially after addition of *Ca\_RhaEW* to the reaction mixture. Finally, neither DHAP nor lactate was detected in the reaction mixture after exclusion of NAD<sup>+</sup> which is an essential cofactor of L-lactaldehyde dehydrogenase. These results suggest that the activity of the L-lactaldehyde dehydrogenase domain RhaW is essential for the L-rhamnulose-1-P aldolase activity of the second domain in this bifunctional enzyme.

In order to test the hypothesis that RhaEW from *B. subtilis* functions in the catabolism of L-Rha *in vivo*, we performed growth experiments in defined medium for two mutant *B. subtilis* strains. One strain carried a knockout mutation in the gene *yuxG* (*rhaEW*), whereas the second strain carried an intact version of *yuxG* but had a knockout mutation in an unrelated gene, *ycel* (encoding a niacin transporter), to serve as an isogenic control. We expected that the growth of the *B. subtilis yuxG* mutant strain would not be stimulated by the addition of L-Rha as a carbon source when compared to the *ycel* mutant strain. The results clearly demonstrate that the *B. subtilis yuxG* knockout mutant is non-responsive to added L-Rha when compared to the *ycel* knockout strain and to both strains grown in the absence of an additional carbon source (Figure 4B). These data confirm that RhaEW is required for L-Rha utilization in *B. subtilis*. The *B. subtilis* RhaEW protein (*Bs\_RhaEW*) was cloned, purified, and tested by the same coupled enzymatic assay as described above for *Ca\_RhaEW*. The *Bs\_RhaEW* protein showed weak, but reproducible activity, measured at  $0.0127 \pm 0.001$   $\mu\text{mol mg protein}^{-1} \text{ min}^{-1}$  at 25°C. Controls removing starting substrate (L-Rha), *Bs\_RhaEW*, or *Ec\_RhaB* (effectively removing rhamnulose-1-P) from the reaction yielded no measurable activity (Figure S2A in Supplementary Materials).

### RhaZ functions as a L-lactaldehyde reductase *in vitro*

We used the reconstituted L-Rha catabolic pathway to test the prediction that *Salmonella* spp. harbor a novel L-lactaldehyde dehydrogenase, distinct from that of *E. coli* and shared among a subgroup of the  $\gamma$ -proteobacteria. We cloned, overexpressed and purified the recombinant proteins *St\_RhaD* (predicted aldolase) and *St\_RhaZ* (predicted novel dehydrogenase) from *S. typhimurium* to complete the *in vitro* pathway (Figure 1). *St\_RhaD* is 99% identical at the amino acid level to *E. coli* RhaD, for which an aldolase function has been demonstrated (Schwartz et al., 1974). To ensure that *St\_RhaD* acts as an aldolase in the L-Rha catabolism, we performed two control assays to confirm the production of DHAP and L-lactaldehyde. To test for the production of DHAP, we used purified glycerol-3-P dehydrogenase (GPDH) (Sigma) in an assay containing *Ec\_RhaA*, *Tm\_RhaB*, and *St\_RhaD*. If *St\_RhaD* acts as a L-rhamnulose-1-P aldolase, then the DHAP produced would be converted to glycerol-3-P by GPDH with the oxidation of NADH to NAD<sup>+</sup> monitored as a decrease in absorbance at 340 nm. Likewise, it was expected



that if *St\_RhaD* produced L-lactaldehyde, then the known *E. coli* L-lactaldehyde dehydrogenase, *AldA*, should be active in a reaction containing all three L-Rha catabolic enzymes, producing L-lactate, and converting  $\text{NAD}^+$  to NADH. The results of both controls confirmed the activity of *St\_RhaD* as a L-rhamnulose-1-P aldolase (data not shown), making possible to test the prediction for *St\_RhaZ*. The purified *St\_RhaZ* protein was included in an assay containing *Ec\_RhaA*, *Tm\_RhaB*, and *St\_RhaD*, using  $\text{NAD}^+$  as a cofactor. This reaction mixture should lead to the conversion of L-lactaldehyde to L-lactate (as with the *E. coli* *AldA* enzyme). Under these conditions, *St\_RhaZ* did not show activity as a L-lactaldehyde dehydrogenase. In order to assess the alternative fate for L-lactaldehyde, which is conversion to L-1,2-propanediol, we repeated the assay under identical conditions with the exception of supplying NADH as the cofactor. *St\_RhaZ* was active under these conditions (Figure 2B in Supplementary Materials), converting L-lactaldehyde to L-1,2-propanediol with a specific activity of  $0.13 \pm 0.02 \mu\text{mol mg protein}^{-1} \text{ min}^{-1}$ . This indicates that the function of *RhaZ* is a L-lactaldehyde reductase, rather than a L-lactaldehyde dehydrogenase.

#### EXPERIMENTAL VALIDATION OF RHAMNOSE UTILIZATION AND REGULON IN *CHLOROFLEXUS AURANTIACUS*

The anoxygenic phototroph *C. aurantiacus* can grow heterotrophically using various organic compounds under either oxic conditions or anoxic conditions in light (Hanada and Pierson, 2006). However, the ability of *C. aurantiacus* and other species from the order *Chloroflexales* to utilize L-Rha has not been previously investigated. In *C. aurantiacus*, the L-Rha utilization genes are organized into a nine-gene *rha* operon, which is predicted to be transcriptionally controlled by a novel DeoR-family regulator *RhaR* (Figure 2). An additional gene, termed *rhmA*, encoding a potential  $\alpha$ -L-rhamnosidase (*Caur\_0361*) is potentially involved in the utilization of L-Rha oligosaccharides by *C. aurantiacus*. A novel LacI-family transcription factor, termed *RhmR*, potentially regulates the *RhmA*-encoding operon, which also encodes a potential transport system for uptake of L-Rha-containing oligosaccharides, termed *RhmEFG* (Figure 2). In contrast to the L-Rha utilization operon, which has orthologs in all sequenced genomes of *Chloroflexus* and *Roseiflexus* spp., the *rhmA* gene locus is only conserved in the closely-related *Chloroflexus* spp. strain, Y-400-fl, but is absent in the other *Chloroflexales*. We assessed the L-Rha utilization and regulon in *C. aurantiacus* by a combination of *in vivo* and *in vitro* experimental approaches.

To validate L-Rha-specific induction of the predicted L-Rha utilization genes *in vivo*, we performed RT-PCR with specific primers designed for three *rha* operon genes, *rhaR*, *rhaF*, and *rhaB*. Total RNA was isolated from *C. aurantiacus* grown in media containing YE or pyruvate, with and without addition of L-Rha. All three genes demonstrated elevated transcript levels in the cells grown on either YE or pyruvate media supplied with L-Rha compared to that of the cells grown in the absence of L-Rha (Figure S3 in Supplementary Materials). In addition to the *rha* operon genes, *rhmA* transcription was also highly elevated in pyruvate-grown cells supplied with L-Rha. These results confirm that the *rha* and *rhmA* operons, that are predicted to be controlled by *RhaR* and *RhmR* transcription factors, respectively,

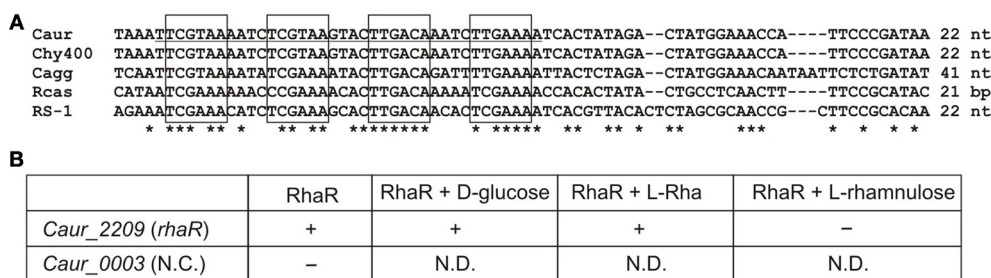
are transcriptionally induced by L-Rha. Additionally, the L-Rha grown culture samples of *C. aurantiacus* were analyzed by HPLC to monitor the L-Rha consumption from the culture fluids. The results confirm a high rate of L-Rha consumption in the samples (Figure S3 in Supplementary Materials), thus confirming that the L-Rha uptake and utilization system is functional *in vivo*.

The interaction of the predicted *RhaR* regulator with the *Caur\_2209* (*rhaR*) upstream DNA fragment containing candidate *RhaR*-binding sites in *C. aurantiacus*, and the influence of potential sugar effectors on protein-DNA interaction were assessed *in vitro* by EMSA (Figure 5). The synthetic 38-bp DNA region containing a tandem repeat of four individual *RhaR* sites (a consensus sequence TCGAAA) was incubated with increasing concentrations of the purified recombinant *RhaR* protein. The incubation was performed at 50°C, which is close to the optimal growth temperature of 55°C for *C. aurantiacus*. The EMSA results (Figure S4 in Supplementary Material) are consistent with the *in silico* predicted DNA operator region of *RhaR*. The addition of D-glucose and L-Rha had no effect on *RhaR*-DNA interaction, whereas L-rhamnulose abolished the specific DNA-binding ability of *RhaR*. The obtained results suggest that the *RhaR* repressor binds to the operator region at the *rha* operon in the absence of a sugar inducer, and that L-rhamnulose serves as a negative regulator for *RhaR* in *C. aurantiacus*.

#### DISCUSSION

L-Rha is the most common deoxy-hexose sugar in nature. In plants, it is a component of many glycosides and polysaccharides such as pectins and hemicelluloses (Peng et al., 2012). Among bacteria, L-Rha is found in the cell wall and as a part of the glycosylated carotenoids (Takaichi and Mochimaru, 2007; Takaichi et al., 2010). Utilization of L-Rha and rhamnose-containing polysaccharides has previously been studied in several free-living and plant pathogenic microbial species from the phylum *Proteobacteria*, including members of the genera *Escherichia*, *Erwinia*, *Rhizobium*, *Azotobacter*, and *Sphingomonas*. Due to significant variations in sugar catabolic pathways in bacteria, the projection of this knowledge to the genomes of more distant species, including many species important for prospective bioenergy applications, is a challenging problem (Rodionov et al., 2010, 2013). In this study, we used comparative genomics to reconstruct novel variants of catabolic pathways and novel transcriptional regulons for L-Rha utilization in the genomes of bacteria from ten taxonomic groups.

Using bioinformatics analyses of L-Rha utilization genes, we identified twelve groups of rhamnose-related transcriptional regulators from different protein families, AraC, DeoR, and LacI, and proposed binding site motifs for these regulators within tentatively reconstructed regulons (Figure S5 in Supplementary Material). Prior to this study, only four types of bacterial transcriptional regulators related to L-Rha metabolism had been identified. The AraC family includes at least five groups of non-orthologous regulators of L-Rha metabolism. These regulators have unique DNA motifs with a tandem repeat symmetry. Activators from three AraC groups have been characterized previously: *RhaR* and *RhaS* from *E. coli* and *Erwinia* spp., with previously known DNA motifs, and *RhaR* from *Bacteroides*, with



**FIGURE 5 | Experimental validation of RhaR regulon in *C. aurantiacus*.**

**(A)** Conservation or predicted RhaR binding sites (boxed) identified in the promoter regions of *rha* operons in the *C. aurantiacus* J-10-fl (*Caur*), *C. sp.* Y-400-fl (*Chy400*), *C. aggregans* DSM 9485 (*Cagg*), *Roseiflexus* sp. RS-1 (*RS-1*), and *R. castenholzii* DSM 13941 (*Rcas*). Distance to a start codon of *rhaR* is indicated. A 38-bp fragment from *C. aurantiacus* used for DNA binding assays is underlined. **(B)** Summary of the EMSA experiments assessing the potential interaction between the recombinant RhaR protein and its predicted DNA motif at the *Caur* 2209 (*rhaR*) gene. The disappearance of unbound

DNA band (shown by "+") was observed upon the addition of increasing concentrations of RhaR protein (0.25–1  $\mu$ M). Addition of 2 mM of L-rhamnose or D-glucose to the reaction mixture containing 1  $\mu$ M of RhaR did not change this pattern, whereas addition of 2 mM of L-rhamnulose led to re-appearance of the unbound DNA band (shown by "-"). As a negative control, incubation of RhaR protein (0.5  $\mu$ M) with upstream DNA fragment of *Caur\_0003* did not reveal the disappearance of unbound DNA band (shown by "-"). The EMSA gel pictures are presented in Figure S4 in Supplementary Material. Asterisks indicate the conserved nucleotides in the multiple alignment.

previously unknown DNA motif. The DeoR family includes at least four non-orthologous groups of RhaR regulators that are characterized by distinct DNA motifs with a tandem repeat symmetry. Among them, only RhaR in *Rhizobium* spp. was described previously (Richardson et al., 2004); however, its DNA binding motif was not known before this study. All LacI-family regulators of L-Rha utilization genes were analyzed for the first time in this study. They are characterized by 20-bp palindromic DNA motifs of four different consensus sequences. In summary, the results of this comparative genomics study demonstrate significant variability in the design and composition of transcriptional regulators for L-Rha metabolism in bacteria. This study has very significantly increased our knowledge about types and operator sequences for transcriptional regulators for L-Rha utilization.

Based on genomic context analyses of the reconstructed regulons, we have identified several novel enzymes and transporters involved in L-Rha utilization (**Figure 1**). A novel enzyme with two domains, termed RhaEW, encoded by the *yuxG* gene in *B. subtilis* and its orthologs in other bacterial lineages, was found to catalyze the last two steps in the catabolism of L-Rha, namely cleavage of L-rhamnulose-1-P to produce DHAP and L-lactaldehyde and oxidation of L-lactaldehyde to L-lactate. Thus, the RhaE domain functions as a non-orthologous substitute for the classical RhaD aldolase, whereas the function of the RhaW domain is analogous to the aldehyde dehydrogenase AldA from *E. coli*. A novel L-lactaldehyde reductase involved in L-Rha catabolism, termed RhaZ, that is not homologous to previously characterized RhaO/FucO, was identified in many  $\gamma$ -proteobacteria. Both functional predictions were experimentally validated *in vitro* by enzymatic assays with the purified recombinant proteins from *C. aurantiacus* and *B. subtilis* (for RhaEW), and *S. typhimurium* (for RhaZ). The function of RhaEW in L-Rha utilization *in vivo* was also confirmed by genetic techniques in *B. subtilis*. Interestingly, genes encoding L-lactate dehydrogenases (*lldD*, *lldEFG*) belong to the reconstructed RhaR regulons in certain genomes of the *Actinomycetales* and *Rhodobacterales* that encode RhaEW. Thus,

the L-Rha utilization pathways in these species are probably extended to produce pyruvate as one of the final products.

Orthologs of the novel aldolase/dehydrogenase RhaEW are broadly distributed among diverse bacterial phyla including Proteobacteria ( $\alpha$ -subdivision), Actinobacteria, Chloroflexi, Bacteroidetes, and Firmicutes (*Bacillales*), in which they are always encoded within the *rha* gene loci (Figure S6 in Supplementary Material). The L-rhamnulose-1-P aldolase domain in RhaE is distantly homologous to class II aldolases including the analogous enzyme, RhaD, and the L-fuculose-1-P aldolase, FucA, from *E. coli*. The tertiary structures and catalytic mechanisms for these enzymes have been determined (Dreyer and Schulz, 1996; Grueninger and Schulz, 2008). We aligned the amino acid sequences of all three enzymes using the multiple protein sequence and structure alignment server PROMALS3D (Pei et al., 2008) (Figure S7 in Supplementary Material). Class II aldolases are zinc-dependent enzymes, in which the metal ion is used for enolate stabilization during catalysis. In RhaD, the  $\text{Zn}^{2+}$  ion is chelated by three histidines, His<sup>141</sup>, His<sup>143</sup>, and His<sup>212</sup>, which are conserved in all RhaE proteins. An Asp residue in RhaE replaces the catalytically important Glu<sup>117</sup> in RhaD, which performs the nucleophilic attack of the C3 atom of DHAP. This conservative substitution suggests that this Asp may play the similar role in RhaE. The Gly<sup>28</sup>, Asn<sup>29</sup>, and Gly<sup>44</sup> residues that are involved in phosphate binding in FucA (Dreyer and Schulz, 1996) are conserved in both RhaD and RhaE enzymes. Conservation of the catalytically important amino acids in both types of L-rhamnulose-1-P aldolases suggests similar position of the active site and catalytic mechanism.

In summary, the phosphorylated catabolic pathway for L-Rha contains a large number of alternative enzymes including RhaI/RhaA, RhaB/RhaK, RhaD/RhaE, RhaO/RhaZ, and RhaW/AldA (**Figure 1**) and is widely-distributed among diverse bacterial phyla. An alternative pathway for the nonphosphorylated L-Rha catabolism that utilizes a unique subset of catabolic enzymes was found only in a small number of proteobacteria



(Table S1 in Supplementary Material). In addition to numerous variations among enzymes and transcriptional regulators associated with the L-Rha catabolic pathway, a similarly high level of variations and non-orthologous displacements is observed for the components of transport machinery. The L-Rha permease, RhaT, which is characteristic of members of the *Enterobacteriales* and *Bacteroidales*, appears to be functionally replaced by either a permease from a different family in some *Actinomycetales* and *Bacillales* or an ABC cassette in  $\alpha$ -proteobacteria and *Chloroflexales*. In other genomes, no candidate transporter specific for L-Rha was detected; however, the reconstructed L-Rha pathways and regulons in these species include transport systems and hydrolytic enzymes for L-Rha oligosaccharides (e.g., rhamnogalacturonides). Some of the latter species are known to grow on L-Rha, such as *B. subtilis* (this study) and *T. maritima* (Rodionov et al., 2013), thus we propose that the predicted L-Rha oligosaccharide transporters in these species are also capable of L-Rha uptake.

Previous studies of L-Rha catabolism in *E. coli* and *Salmonella*, revealed a differential fate for L-Rha under aerobic and anaerobic conditions in *E. coli*, but not in *Salmonella* (Baldoma et al., 1988; Obradors et al., 1988). *E. coli* oxidizes L-lactaldehyde to L-lactate via the activity of AldA under aerobic conditions and reduces L-lactaldehyde to L-1,2-propanediol via the activity of FucO under anaerobic conditions (Figure 1). In contrast, *Salmonella* produces L-1,2-propanediol under both aerobic and anaerobic conditions when metabolizing L-Rha. The identification of *Salmonella* RhaZ as an L-lactaldehyde reductase is consistent with these observations. *Salmonella* produce 1:1 molar equivalents of L-1,2-propanediol from the catabolism of L-Rha under both aerobic and anaerobic conditions, with growth yields higher than *E. coli* under anaerobic conditions (Baldoma et al., 1988). The production of L-1,2-propanediol through renewable, biological methods is of high importance given the current chemical based processes of production and the high use of L-1,2-propanediol in many commercial products (Cameron et al., 1998). There are several examples of recent bioengineering strategies to improve L-1,2-propanediol production in *E. coli* (Clomburg and Gonzalez, 2011), cyanobacteria (Li and Liao, 2013), and *Saccharomyces* (Jung et al., 2011) in which each strategy uses glycerol as a starting substrate. The observation of differential fates for L-Rha in *E. coli* and *Salmonella*, the identification of the activity of RhaZ, putative transport systems for rhamnogalacturonides, and predicted regulatory mechanisms in *Salmonella* raise possibilities for exploring alternative biological production strategies of the commercially important L-1,2-propanediol from L-Rha containing substrates, though L-Rha, itself, remains an expensive substrate (Cameron et al., 1998).

*C. aurantiacus* and other filamentous anoxygenic phototrophic bacteria from the *Chloroflexaceae* family were commonly found in the upper layers of microbial mats in hot springs (50–62°C), with cyanobacteria growing together with chloroflexi. Although *Chloroflexus* spp. can grow heterotrophically on various organic carbon sources, their sugar utilization pathways have remained largely unknown before this work. Here, we identified and characterized a novel variant of the L-Rha catabolic pathway in *C. aurantiacus*, which includes the L-Rha isomerase RhaA, kinase

RhaB, and a novel bifunctional enzyme, RhaEW, that catalyzes the last two steps of the pathway. *C. aurantiacus* transcribed genes for L-Rha utilization when L-Rha was present in the growth medium and consumed L-Rha from the medium. The ecophysiological importance of the L-Rha utilization pathway in members of the *Chloroflexales* is yet to be elucidated. One possibility is that cyanobacteria commonly co-occurring with chloroflexi in hot springs microbial mats may provide them L-Rha. In such microbial mats, cyanobacteria are primary producers that are thought to cross-feed low-molecular-weight organic compounds (e.g., lactate, acetate, glycolate) to members of the *Chloroflexales* (van der Meer et al., 2003, 2007). There are several potential sources of L-Rha in cyanobacteria including lipopolysaccharides in the outer membrane (Buttke and Ingram, 1975) and glycosylated carotenoids in the cytoplasmic and outer membrane that protect the cell against photooxidative damage (Takaichi and Mochimaru, 2007; Graham and Bryant, 2009). The exact source of L-Rha from a primary producer and its significance for possible metabolite exchange in the mat community requires further investigation.

## ACKNOWLEDGMENTS

We would like to thank P. Novichkov (LBNL, Berkeley, CA) for help with visualization of regulons in the RegPrecise database, Dave Kennedy (PNNL, Richland, WA) for help with HPLC analysis, David Scott (SBMRI, La Jolla, CA) for help with GC-MS analysis, and E. Dervyn (INRA, France) for *B. subtilis* knockout strains. This research was supported by the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the Pacific Northwest National Laboratory (PNNL) Foundational Scientific Focus Area. Additional funding was provided by the Russian Foundation for Basic Research (12-04-33003) and the Towsley Foundation (Midland, MI) through the Towsley Research Scholar program at Hope College.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fmicb.2013.00407/abstract>.

## REFERENCES

- Aiba, H., Adhya, S., and de Crombrughe, B. (1981). Evidence for two functional gal promoters in intact *Escherichia coli* cells. *J. Biol. Chem.* 256, 11905–11910.
- Akhy, M. T., Brown, C. M., and Old, D. C. (1984). L-Rhamnose utilisation in *Salmonella typhimurium*. *J. Appl. Bacteriol.* 56, 269–274. doi: 10.1111/j.1365-2672.1984.tb01347.x
- Avila, M., Jaquet, M., Moine, D., Requena, T., Pelaez, C., Arigoni, F., et al. (2009). Physiological and biochemical characterization of the two alpha-L-rhamnosidases of *Lactobacillus plantarum* NCC245. *Microbiology* 155, 2739–2749. doi: 10.1099/mic.0.027789-0
- Badia, J., Baldoma, L., Aguilar, J., and Boronat, A. (1989). Identification of the rhaA, rhaB and rhaD gene products from *Escherichia coli* K-12. *FEMS Microbiol. Lett.* 53, 253–257. doi: 10.1016/0378-1097(89)90226-7
- Baldoma, L., and Aguilar, J. (1988). Metabolism of L-fucose and L-rhamnose in *Escherichia coli*: aerobic-anaerobic regulation of L-lactaldehyde dissimilation. *J. Bacteriol.* 170, 416–421.
- Baldoma, L., Badia, J., Obradors, N., and Aguilar, J. (1988). Aerobic excretion of 1,2-propanediol by *Salmonella typhimurium*. *J. Bacteriol.* 170, 2884–2885.



- Baldoma, L., Badia, J., Sweet, G., and Aguilar, J. (1990). Cloning, mapping and gene product identification of rhaT from *Escherichia coli* K12. *FEMS Microbiol. Lett.* 60, 103–107. doi: 10.1111/j.1574-6968.1990.tb03870.x
- Buttke, T. M., and Ingram, L. O. (1975). Comparison of lipopolysaccharides from *Agmenellum quadruplicatum* to *Escherichia coli* and *Salmonella typhimurium* by using thin-layer chromatography. *J. Bacteriol.* 124, 1566–1573.
- Cameron, D. C., Altaras, N. E., Hoffman, M. L., and Shaw, A. J. (1998). Metabolic engineering of propanediol pathways. *Biotechnol. Prog.* 14, 116–125. doi: 10.1021/bp9701325
- Cao, L., Bryant, D. A., Schepmoes, A. A., Vogl, K., Smith, R. D., Lipton, M. S., et al. (2012). Comparison of *Chloroflexus aurantiacus* strain J-10-fl proteomes of cells grown chemoheterotrophically and photoheterotrophically. *Photosynth. Res.* 110, 153–168. doi: 10.1007/s11120-011-9711-8
- Clomburg, J. M., and Gonzalez, R. (2011). Metabolic engineering of *Escherichia coli* for the production of 1,2-propanediol from glycerol. *Biotechnol. Bioeng.* 108, 867–879. doi: 10.1002/bit.22993
- Crooks, G. E., Hon, G., Chandonia, J. M., and Brenner, S. E. (2004). WebLogo: a sequence logo generator. *Genome Res.* 14, 1188–1190. doi: 10.1101/gr.849004
- Dehal, P. S., Joachimiak, M. P., Price, M. N., Bates, J. T., Baumohl, J. K., Chivian, D., et al. (2010). MicrobesOnline: an integrated portal for comparative and functional genomics. *Nucleic Acids Res.* 38, D396–D400. doi: 10.1093/nar/gkp919
- Dreyer, M. K., and Schulz, G. E. (1996). Catalytic mechanism of the metal-dependent fucose aldolase from *Escherichia coli* as derived from the structure. *J. Mol. Biol.* 259, 458–466. doi: 10.1006/jmbi.1996.0332
- Eagon, R. G. (1961). Bacterial dissimilation of L-fucose and L-rhamnose. *J. Bacteriol.* 82, 548–550.
- Egan, S. M., and Schleif, R. F. (1993). A regulatory cascade in the induction of rhaBAD. *J. Mol. Biol.* 234, 87–98. doi: 10.1006/jmbi.1993.1565
- Gerdes, S. Y., Kurnasov, O. V., Shatalin, K., Polanuyer, B., Sloutsky, R., Vonstein, V., et al. (2006). Comparative genomics of NAD biosynthesis in cyanobacteria. *J. Bacteriol.* 188, 3012–3023. doi: 10.1128/JB.188.8.3012-3023.2006
- Giraud, M. F., and Naismith, J. H. (2000). The rhamnose pathway. *Curr. Opin. Struct. Biol.* 10, 687–696. doi: 10.1016/S0959-440X(00)00145-7
- Graham, J. E., and Bryant, D. A. (2009). The biosynthetic pathway for myxol-2' fucoside (myxoxanthophyll) in the cyanobacterium *Synechococcus* sp. strain PCC 7002. *J. Bacteriol.* 191, 3292–3300. doi: 10.1128/JB.00050-09
- Grueninger, D., and Schulz, G. E. (2006). Structure and reaction mechanism of L-rhamnulose kinase from *Escherichia coli*. *J. Mol. Biol.* 359, 787–797. doi: 10.1016/j.jmb.2006.04.013
- Grueninger, D., and Schulz, G. E. (2008). Antenna domain mobility and enzymatic reaction of L-rhamnulose-1-phosphate aldolase. *Biochemistry* 47, 607–614. doi: 10.1021/bi7012799
- Gu, Y., Ding, Y., Ren, C., Sun, Z., Rodionov, D. A., Zhang, W., et al. (2010). Reconstruction of xylose utilization pathway and regulons in Firmicutes. *BMC Genomics* 11:255. doi: 10.1186/1471-2164-11-255
- Hanada, S., and Pierson, B. K. (2006). "The family chloroflexaceae," in *The Prokaryotes*, 3rd Edn (New York, NY: Springer), 815–842.
- Hugouvieux-Cotte-Pattat, N. (2004). The RhaS activator controls the *Erwinia chrysanthemi* 3937 genes rhiN, rhiT and rhiE involved in rhamnolacturonan catabolism. *Mol. Microbiol.* 51, 1361–1374. doi: 10.1046/j.1365-2958.2003.03908.x
- Jung, J. Y., Yun, H. S., Lee, J., and Oh, M. K. (2011). Production of 1,2-propanediol from glycerol in *Saccharomyces cerevisiae*. *J. Microbiol. Biotechnol.* 21, 846–853. doi: 10.4014/jmb.1103.03009
- Kobayashi, K., Ehrlich, S. D., Albertini, A., Amati, G., Andersen, K. K., Arnaud, M., et al. (2003). Essential *Bacillus subtilis* genes. *Proc. Natl. Acad. Sci. U.S.A.* 100, 4678–4683. doi: 10.1073/pnas.0730515100
- Korndorfer, I. P., Fessner, W. D., and Matthews, B. W. (2000). The structure of rhamnose isomerase from *Escherichia coli* and its relation with xylose isomerase illustrates a change between inter and intra-subunit complementation during evolution. *J. Mol. Biol.* 300, 917–933. doi: 10.1006/jmbi.2000.3896
- Kroemer, M., Merkel, I., and Schulz, G. E. (2003). Structure and catalytic mechanism of L-rhamnulose-1-phosphate aldolase. *Biochemistry* 42, 10560–10568. doi: 10.1021/bi0349266
- Laatu, M., and Condemine, G. (2003). Rhamnolacturonate lyase RhiE is secreted by the out system in *Erwinia chrysanthemi*. *J. Bacteriol.* 185, 1642–1649. doi: 10.1128/JB.185.5.1642-1649.2003
- Leang, K., Takada, G., Ishimura, A., Okita, M., and Izumori, K. (2004). Cloning, nucleotide sequence, and overexpression of the L-rhamnose isomerase gene from *Pseudomonas stutzeri* in *Escherichia coli*. *Appl. Environ. Microbiol.* 70, 3298–3304. doi: 10.1128/AEM.70.6.3298-3304.2004
- Leyn, S. A., Gao, F., Yang, C., and Rodionov, D. A. (2012). N-acetylgalactosamine utilization pathway and regulon in proteobacteria: genomic reconstruction and experimental characterization in *Shewanella*. *J. Biol. Chem.* 287, 28047–28056. doi: 10.1074/jbc.M112.382333
- Leyn, S. A., Kazanov, M. D., Sernova, N. V., Ermakova, E. O., Novichkov, P. S., and Rodionov, D. A. (2013). Genomic reconstruction of the transcriptional regulatory network in *Bacillus subtilis*. *J. Bacteriol.* 195, 2463–2473. doi: 10.1128/JB.00140-13
- Li, H., and Liao, J. C. (2013). Engineering a cyanobacterium as the catalyst for the photosynthetic conversion of CO<sub>2</sub> to 1,2-propanediol. *Microb. Cell Fact.* 12:4. doi: 10.1186/1475-2859-12-4
- Mossessova, E., and Lima, C. D. (2000). Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol. Cell* 5, 865–876. doi: 10.1016/S1097-2765(00)80326-3
- Muiry, J. A., Gunn, T. C., McDonald, T. P., Bradley, S. A., Tate, C. G., and Henderson, P. J. (1993). Proton-linked L-rhamnose transport, and its comparison with L-fucose transport in Enterobacteriaceae. *Biochem. J.* 290(Pt 3), 833–842.
- Novichkov, P. S., Kazakov, A. E., Ravcheev, D. A., Leyn, S. A., Kovaleva, G. Y., Sutormin, R. A., et al. (2013). RegPrecise 3.0—a resource for genome-scale exploration of transcriptional regulation in Bacteria. *BMC Genomics* 14:745. doi: 10.1186/1471-2164-14-745
- Novichkov, P. S., Rodionov, D. A., Stavrovskaya, E. D., Novichkova, E. S., Kazakov, A. E., Gelfand, M. S., et al. (2010). RegPredict: an integrated system for regulon inference in prokaryotes by comparative genomics approach. *Nucleic Acids Res.* 38, W299–W307. doi: 10.1093/nar/gkq531
- Obradors, N., Badia, J., Baldoma, L., and Aguilar, J. (1988). Anaerobic metabolism of the L-rhamnose fermentation product 1,2-propanediol in *Salmonella typhimurium*. *J. Bacteriol.* 170, 2159–2162.
- Ochiai, A., Itoh, T., Kawamata, A., Hashimoto, W., and Murata, K. (2007). Plant cell wall degradation by saprophytic *Bacillus* strains: gene clusters responsible for rhamnolacturonan depolymerization. *Appl. Environ. Microbiol.* 73, 3803–3813. doi: 10.1128/AEM.00147-07
- Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H. Y., Cohoon, M., et al. (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Res.* 33, 5691–5702. doi: 10.1093/nar/gki866
- Patel, E. H., Paul, L. V., Patrick, S., and Abratt, V. R. (2008). Rhamnose catabolism in *Bacteroides thetaiotaomicron* is controlled by the positive transcriptional regulator RhaR. *Res. Microbiol.* 159, 678–684. doi: 10.1016/j.resmic.2008.09.002
- Pei, J., Tang, M., and Grishin, N. V. (2008). PROMALS3D web server for accurate multiple protein sequence and structure alignments. *Nucleic Acids Res.* 36, W30–W34. doi: 10.1093/nar/gkn072
- Peng, F., Peng, P., Xu, F., and Sun, R. C. (2012). Fractional purification and bioconversion of hemicelluloses. *Biotechnol. Adv.* 30, 879–903. doi: 10.1016/j.biotechadv.2012.01.018
- Ravcheev, D. A., Best, A. A., Sernova, N. V., Kazanov, M. D., Novichkov, P. S., and Rodionov, D. A. (2013). Genomic reconstruction of transcriptional regulatory networks in lactic acid bacteria. *BMC Genomics* 14:94. doi: 10.1186/1471-2164-14-94
- Ravcheev, D. A., Best, A. A., Tintle, N., Dejongh, M., Osterman, A. L., Novichkov, P. S., et al. (2011). Inference of the transcriptional regulatory network in *Staphylococcus aureus* by integration of experimental and genomics-based evidence. *J. Bacteriol.* 193, 3228–3240. doi: 10.1128/JB.00350-11
- Richardson, J. S., Carpena, X., Switala, J., Perez-Luque, R., Donald, L. J., Loewen, P. C., et al. (2008). RhaU of *Rhizobium leguminosarum* is a rhamnose mutarotase. *J. Bacteriol.* 190, 2903–2910. doi: 10.1128/JB.01120-07
- Richardson, J. S., Hynes, M. F., and Oresnik, I. J. (2004). A genetic locus necessary for rhamnose uptake and catabolism in *Rhizobium leguminosarum* bv. trifolii. *J. Bacteriol.* 186, 8433–8442. doi: 10.1128/JB.186.24.8433-8442.2004
- Richardson, J. S., and Oresnik, I. J. (2007). L-Rhamnose transport is sugar kinase (RhaK) dependent in *Rhizobium leguminosarum* bv. trifolii. *J. Bacteriol.* 189, 8437–8446. doi: 10.1128/JB.01032-07
- Rodionov, D. A., Gelfand, M. S., and Hugouvieux-Cotte-Pattat, N. (2004). Comparative genomics of the KdgR regulon in *Erwinia chrysanthemi* 3937 and other gamma-proteobacteria. *Microbiology* 150, 3571–3590. doi: 10.1099/mic.0.27041-0

- Rodionov, D. A., Novichkov, P. S., Stavrovskaya, E. D., Rodionova, I. A., Li, X., Kazanov, M. D., et al. (2011). Comparative genomic reconstruction of transcriptional networks controlling central metabolism in the *Shewanella* genus. *BMC Genomics* 12(Suppl. 1):S3. doi: 10.1186/1471-2164-12-S1-S3
- Rodionov, D. A., Rodionova, I. A., Li, X., Ravcheev, D. A., Tarasova, Y., Portnoy, V. A., et al. (2013). Transcriptional regulation of the carbohydrate utilization network in *Thermotoga maritima*. *Front. Microbiol.* 4:244. doi: 10.3389/fmicb.2013.00244
- Rodionov, D. A., Yang, C., Li, X., Rodionova, I. A., Wang, Y., Obratsova, A. Y., et al. (2010). Genomic encyclopedia of sugar utilization pathways in the *Shewanella* genus. *BMC Genomics* 11:494. doi: 10.1186/1471-2164-11-494
- Rodionova, I. A., Leyn, S. A., Burkart, M. D., Boucher, N., Noll, K. M., Osterman, A. L., et al. (2013). Novel inositol catabolic pathway in *Thermotoga maritima*. *Environ. Microbiol.* 15, 2254–2266. doi: 10.1111/1462-2920.12096
- Rodionova, I. A., Scott, D. A., Grishin, N. V., Osterman, A. L., and Rodionov, D. A. (2012a). Tagaturonate-fructuronate epimerase UxaE, a novel enzyme in the hexuronate catabolic network in *Thermotoga maritima*. *Environ. Microbiol.* 14, 2920–2934. doi: 10.1111/j.1462-2920.2012.02856.x
- Rodionova, I. A., Yang, C., Li, X., Kurnasov, O. V., Best, A. A., Osterman, A. L., et al. (2012b). Diversity and versatility of the *Thermotoga maritima* sugar kinome. *J. Bacteriol.* 194, 5552–5563. doi: 10.1128/JB.01136-12
- Ryu, K. S., Kim, J. I., Cho, S. J., Park, D., Park, C., Cheong, H. K., et al. (2005). Structural insights into the monosaccharide specificity of *Escherichia coli* rhamnose mutarotase. *J. Mol. Biol.* 349, 153–162. doi: 10.1016/j.jmb.2005.03.047
- Schwartz, N. B., Abram, D., and Feingold, D. S. (1974). L-Rhamnulose 1-phosphate aldolase of *Escherichia coli*. The role of metal in enzyme structure. *Biochemistry* 13, 1726–1730. doi: 10.1021/bi00705a026
- Stanier, R. Y., Kunisawa, R., Mandel, M., and Cohen-Bazire, G. (1971). Purification and properties of unicellular blue-green algae (order Chroococcales). *Bacteriol. Rev.* 35, 171–205.
- Steunou, A. S., Bhaya, D., Bateson, M. M., Melendrez, M. C., Ward, D. M., Brecht, E., et al. (2006). *In situ* analysis of nitrogen fixation and metabolic switching in unicellular thermophilic cyanobacteria inhabiting hot spring microbial mats. *Proc. Natl. Acad. Sci. U.S.A.* 103, 2398–2403. doi: 10.1073/pnas.0507513103
- Takaichi, S., Maoka, T., Takasaki, K., and Hanada, S. (2010). Carotenoids of *Gemmatimonas aurantiaca* (Gemmatimonadetes): identification of a novel carotenoid, deoxyoscollol 2-rhamnoside, and proposed biosynthetic pathway of oscollol 2,2'-dirhamnoside. *Microbiology* 156, 757–763. doi: 10.1099/mic.0.034249-0
- Takaichi, S., and Mochimaru, M. (2007). Carotenoids and carotenogenesis in cyanobacteria: unique ketocarotenoids and carotenoid glycosides. *Cell. Mol. Life Sci.* 64, 2607–2619. doi: 10.1007/s00018-007-7190-z
- Tobin, J. F., and Schleif, R. F. (1990). Purification and properties of RhaR, the positive regulator of the L-rhamnose operons of *Escherichia coli*. *J. Mol. Biol.* 211, 75–89. doi: 10.1016/0022-2836(90)90012-B
- van der Meer, M. T., Schouten, S., Damste, J. S., and Ward, D. M. (2007). Impact of carbon metabolism on 13C signatures of cyanobacteria and green non-sulfur-like bacteria inhabiting a microbial mat from an alkaline siliceous hot spring in Yellowstone National Park (USA). *Environ. Microbiol.* 9, 482–491. doi: 10.1111/j.1462-2920.2006.01165.x
- van der Meer, M. T., Schouten, S., Sinninghe Damste, J. S., de Leeuw, J. W., and Ward, D. M. (2003). Compound-specific isotopic fractionation patterns suggest different carbon metabolisms among *Chloroflexus*-like bacteria in hot-spring microbial mats. *Appl. Environ. Microbiol.* 69, 6000–6006. doi: 10.1128/AEM.69.10.6000-6006.2003
- Via, P., Badia, J., Baldoma, L., Obradors, N., and Aguilar, J. (1996). Transcriptional regulation of the *Escherichia coli* rhaT gene. *Microbiology* 142, 1833–1840. doi: 10.1099/13500872-142-7-1833
- Watanabe, S., and Makino, K. (2009). Novel modified version of nonphosphorylated sugar metabolism—an alternative L-rhamnose pathway of *Sphingomonas* sp. *FEBS J.* 276, 1554–1567. doi: 10.1111/j.1742-4658.2009.06885.x
- Watanabe, S., Saimura, M., and Makino, K. (2008). Eukaryotic and bacterial gene clusters related to an alternative pathway of nonphosphorylated L-rhamnose metabolism. *J. Biol. Chem.* 283, 20372–20382. doi: 10.1074/jbc.M801065200
- Yang, C., Rodionov, D. A., Li, X., Laikova, O. N., Gelfand, M. S., Zagnitko, O. P., et al. (2006). Comparative genomics and experimental characterization of N-acetylglucosamine utilization pathway of *Shewanella oneidensis*. *J. Biol. Chem.* 281, 29872–29885. doi: 10.1074/jbc.M605052200
- Yoshida, H., Yamada, M., Ohyama, Y., Takada, G., Izumori, K., and Kamitori, S. (2007). The structures of L-rhamnose isomerase from *Pseudomonas stutzeri* in complexes with L-rhamnose and D-allose provide insights into broad substrate specificity. *J. Mol. Biol.* 365, 1505–1516. doi: 10.1016/j.jmb.2006.11.004
- Zhang, L., Leyn, S. A., Gu, Y., Jiang, W., Rodionov, D. A., and Yang, C. (2012). Ribulokinase and transcriptional regulation of arabinose metabolism in *Clostridium acetobutylicum*. *J. Bacteriol.* 194, 1055–1064. doi: 10.1128/JB.06241-11
- Zhu, Y., and Lin, E. C. (1989). L-1,2-propanediol exits more rapidly than L-lactaldehyde from *Escherichia coli*. *J. Bacteriol.* 171, 862–867.

**Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 01 November 2013; paper pending published: 25 November 2013; accepted: 09 December 2013; published online: 23 December 2013.

Citation: Rodionova IA, Li X, Thiel V, Stolyar S, Stanton K, Fredrickson JK, Bryant DA, Osterman AL, Best AA and Rodionov DA (2013) Comparative genomics and functional analysis of rhamnose catabolic pathways and regulons in bacteria. *Front. Microbiol.* 4:407. doi: 10.3389/fmicb.2013.00407

This article was submitted to *Microbial Physiology and Metabolism*, a section of the journal *Frontiers in Microbiology*.

Copyright © 2013 Rodionova, Li, Thiel, Stolyar, Stanton, Fredrickson, Bryant, Osterman, Best and Rodionov. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

