Microbial production of cellulosic isobutanol: integrating ecology and evolutionary approaches with engineering

by

Jeremy J. Minty

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Chemical Engineering)
in the University of Michigan
2013

Doctoral Committee:

Professor Nina Lin, Chair
Professor Richard Lenski, Michigan State University
Professor Michael Mayer
Professor Henry Yee-Neen Wang
Professor George Zhang

2013

For Jennifer

A C K N O W L E D G M E N T S

I would first like to express my deepest and most sincere gratitude to my advisor, Professor Nina Lin, for all of her support, guidance, and patience. Nina's advocacy was instrumental in helping me gain admission to graduate school, and I owe her a tremendous debt for believing in me and taking me as a student. Working with Nina has been a transformative experience, and through her mentoring I have developed as a researcher, teacher, and leader. Nina gave me much freedom and ownership over my research projects, permitting me to creatively explore many different ideas and forge my own path. Beyond her dedication to visionary research, I am also very appreciative of Nina's efforts to provide an inclusive and nurturing environment in her lab. Overall, Nina has been an outstanding advisor and source of inspiration to me, and it has been and honor and pleasure to work with her.

In addition to Nina, I also want to give thanks to all the members of her lab with whom I have worked. It was a wonderful experience to work with such an interesting, bright, and diverse group of people, and I am grateful for the efforts many of them made to keep our day to day operations running smoothly. My colleagues and I engaged in numerous productive collaborations, and I would like to thank them for devoting their time and energy to working with me and for making substantial contributions to the research presented in this dissertation. In particular, I worked closely with Alissa Kerner and Scott Scholz on our synthetic microbial consortium project (Part I of this dissertation), and Fengming Lin and Yu Chen on our evolutionary-genomics approach for improving microbial stress tolerance (Part II of this dissertation). Jihyang Park, Artur Veloso, Steven Wang, Mathieu Rossion, Tatyana Saleski, Ping Jin, and Trinh Pham also made important contributions to our stress tolerance project. Furthermore, I would also like to thank Michael Nelson and Marjan Varedi for providing technical assistance with HPLC and Western blots, respectively, as well as many interesting discussions in the office.

Over the course of my graduate studies, I had the privilege to mentor and work with 15 different undergraduate and Master's students. Without their help, much of the research presented in this dissertation would not have been possible, and through them I also learned much about mentoring and teaching. I would like to thank Marc Singer, Ian Faulkner, LJ Bae, Jungho Ahn, Chris Moline, Shreyas Hirekhan, and Angelo Solari for their contributions to the consortium project, and Ann Lesnefsky, Ted Zaroff, Lawrence Lai, Lian Zhu, Lauren Kennedy, David Boyer, Ted Farrand, and Matthew Gross for contributing to the stress tolerance project. Many of these students went well above their call of duty in working with me. I would particularly like to recognize the efforts of Ann and Marc, who were both exceptionally dedicated and talented, and who

made great efforts and sacrifices to help me. I owe all of these students an immense debt for the help they gave me, and I feel privileged to have worked with such motivated and intelligent people.

Outside the Lin lab, many other members of the academic community have helped and supported me. I met many exceptional students through serving as an advisor to the University of Michigan 2009, 2010, 2011, and 2012 iGEM teams, and I am honored to have worked with people of such caliber. I would like to thank the Erdogan Gulari lab (especially Saadet Albayrak and Yusuf Murgha) and Mark Burns lab for being friendly and generous neighbors and collaborators. Greg Boggy (Peter Woolf lab) kindly helped me with qPCR experiments and data analysis, for which I am very grateful. Over my time at the Department of Chemical Engineering, I have gotten to know many outstanding faculty and staff members. I am especially grateful to our undergraduate advisor, Susan Montgomery, for her heartfelt advice, kindness, and dedication to education, to Phil Savage for serving as the graduate advisor, and to Robert Ziff for proof-reading and checking the mathematical models we developed in the consortium project. I would also like to thank the staff, particularly Mike Africa, Pablo Lavalle, Susan Hamlin, Pamela Bogdanski, Kelly Raickovich, and Shelly Fellers for their technical and administrative assistance. Finally, I would like to thank my committee members: Professor Richard Lenski (Michigan State University), Professor Michael Mayer, Professor Henry Wang, and Professor George Zhang. I am especially grateful to Richard for taking time out of his very busy schedule to review and provide feedback on my research, and for making the hour-long trip from Lansing to Ann Arbor for committee meetings and my defense.

Beyond the University of Michigan, I was fortunate to have many external collaborators who generously donated resources and time to our research efforts. Yuan Gao's lab (Biomedical Engineering, Johns Hopkins University) provided advice and devoted considerable resources to perform whole genome resequencing for our stress tolerance project. MYcroarray LLC was extremely generous in providing us with custom designed DNA microarrays, as well as performing gene expression studies and genome resequencing as part of our stress tolerance project. James Liao's lab (Chemical and Biomolecular Engineering, UCLA) kindly provided us with isobutanol producing *E. coli* strains. George Church's lab (Harvard Medical School) and Harris Wang's lab (Columbia University, Pathology and Cell Biology) devoted much time and resources in helping us design and perform multiplex automated genome engineering (MAGE) experiments, for which I am most grateful. Finally, I am thankful to the Biomass Conversion research Laboratory (Michigan State University; Professors Bruce Dale and Venkatesh Balan) for providing us with AFEX pre-treated biomass, and to the Great Lakes Bioenergy Research Center (especially Cliff Foster) for generously devoting their time and resources to assisting us with carbohydrate analysis

of our consortium samples.

Graduate school was a challenging experience, and I would not have made it without the love, support, and encouragement of my wife, sisters, parents, close friends, and extended family. Research and academic work consumed much of my time, and I often worked through the weekends and late into the night. I am very grateful to everyone for being so understanding and supportive of my efforts and for the good times that we shared together. I would especially like to thank my friend Paul Kuhlman for the many interesting discussions we had, as our conversations about game theory ultimately inspired the research presented in Chapter 3 of this dissertation. Last but not least, I am profoundly grateful to my wife, Jennifer. Her love, consistent support, and patience helped give me the courage and motivation to weather the many personal and professional challenges I encountered over the past five years. Words cannot express how grateful and fortunate I am to have her in my life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

ABSTRACT

Microbial production of cellulosic isobutanol: integrating ecology and evolutionary approaches with engineering

by

Jeremy J. Minty

Chair: Xiaoxia Nina Lin

Biofuels derived from lignocellulosic feedstocks are widely considered to be among the most promising renewable fuels that can be produced at a large scale and in a sustainable manner. However, many challenges exist. In this work, we aim to address two of them, which are interconnected under an overall goal of achieving efficient microbial conversion of lignocellulosic feedstocks to isobutanol, an advanced biofuel: i) enabling consolidated bioprocessing of lignocellulosic feedstocks to biofuels, through engineering synthetic microbial consortia; and ii) improving microbial stress tolerance, through genome evolution and engineering.

Inspired by the versatility and robustness of ubiquitous natural microbial ecosystems, the first part of our work explores engineering synthetic multispecies microbial communities for cellulosic biofuel production. The required biochemical functions are divided between two specialist organisms: the fungus *Trichoderma reesei*, which secretes cellulases to hydrolyze lignocellulose into soluble saccharides, and the bacterium *Escherichia coli*, which metabolizes soluble saccharides into isobutanol. We developed and experimentally validated a comprehensive modeling framework, allowing us to elucidate key ecological interactions and develop mechanisms for stabilizing and tuning population composition. To illustrate bioprocessing applications, we demonstrate direct conversion of cellulosic feedstocks to isobutanol, achieving titers up to 1.86 g/L and 62% of theoretical yield.

In the second part, we leverage recent advances in DNA sequencing and genome engineering technologies to decode and refactor microbial tolerance to isobutanol, a complex phenotype

with a poorly understood genetic basis. We experimentally evolved isobutanol tolerant *E. coli* strains, and then used genome re-sequencing and functional dissection studies to reverse engineer mechanisms and genetic bases of tolerance. Next, we exploited our initial results to select genetic loci for targeted mutagenesis using Multiplex Automated Genome Engineering (MAGE), allowing us to refactor isobutanol tolerance and explore large genotype spaces for hyper-tolerant variants.

In summary, we have integrated ecology and evolutionary approaches with engineering to develop novel microbial systems for biofuel production. Our synthetic microbial consortium approach provides key advantages over the conventional paradigm of engineering a single microbe ("super-bug"); in parallel, our genome evolution and engineering work has generated new insights into genetic and biochemical mechanisms underlying microbial tolerance to toxic chemicals.

CHAPTER 1

Background and motivation

1.1 Microbial production of cellulosic biofuels

With shrinking fossil fuel supplies, accelerating climate change, and intensifying geopolitical concerns, the need for renewable energy sources is becoming evermore apparent. Renewable transportation fuels are crucial for developing a sustainable energy portfolio; in the United States, transportation accounts for 28% of total energy consumption and 71% of total petroleum consumption [4]. Biofuels produced through microbial conversion of biomass feedstocks could sustainably replace a substantial fraction of the petroleum presently consumed by the transportation sector, and are considered to be among the most promising alternative fuels [5]. Numerous microbial species produce biofuel molecules as part of their natural metabolism. Many anaerobic microbes produce short-chain alcohols as terminal metabolites (such as ethanol and butanol, which can serve as gasoline substitutes), while other species are capable of accumulating lipids (possible precursors for biodiesel and other fuels) [6]. To date, most commercial efforts in biofuels have focused on microbial ethanol production from sugarcane or starch-based feedstocks such as corn. However, ethanol production has recently come under scrutiny for various reasons. Criticisms include the current reliance on food crop feedstocks (creating competition between food supplies and energy production), low ethanol yields per unit area of crop, and concerns about sustainability and environmental impacts (e.g. potentially unfavorable overall energy balance, excessive fresh water usage in the production process, etc.) [7]. Additionally, some of the physicochemical properties of ethanol are suboptimal for motor fuel use. The energy density of ethanol is approximately $2/3$ that of gasoline, leading to reduced fuel efficiency [7]. Ethanol is also hygroscopic, limiting it's use in existing fuel infrastructure [7].

1.1.1 Advanced non-ethanol biofuels

Motivated in part by the drawbacks of ethanol, recent metabolic engineering efforts have dramatically expanded the portfolio of fuel molecules that be produced biologically (Figure 1.1). Examples of next-generation biofuel molecules include isoprenoid derived hydrocarbons, fatty acid derived hydrocarbons, and non-fermentative medium chain-length alcohols (Figure 1.1). Isoprenoids are derived from an isomeric five-carbon unit called IPP (isopentenyl pyrophosphate) or DMAP (dimethyl-allyl pyrophosphate) using either the mevalonate or deoxyxylulose-5-phosphate pathways (Figure 1.1) [8]. The molecular weight ranges and branch/ring structures of isoprenoids make them promising substitutes for gasoline, as well as diesel and jet fuels [8]. Fatty acid derived biofuels include fatty alcohols, fatty acid alkyl esters, fatty acid-derived alkanes, and alkenes, which represent favorable substitutes for diesel or gasoline (Figure 1.1) [8]. Isoprenoid and fatty acid derived biofuels have many favorable fuel properties, including high energy density, low aqueous solubility (facilitating separation processes and making them compatible with existing fuel infrastructure), and combustion properties similar to petroleum based fuels. However, low titers and yields for isoprenoid and fatty acid derived biofuels have hindered commercialization [8]. To date, the highest titers and yields reported for next-generation biofuels have been achieved with non-fermentative medium chain-length alcohols, which are produced via decarboxylation and subsequent reduction of 2-keto acids (Figure 1.2). Most notably, *E. coli* has been successfully engineered to produce isobutanol in high yield (86% of theoretical maximum) and titer (22 g/L) from glucose [1]. Isobutanol has higher energy density, lower hygroscopicity, and better combustion properties than ethanol, making it a favorable gasoline substitute. Isobutanol production is presently being commercialized by Gevo Inc. and Butamax LLC.

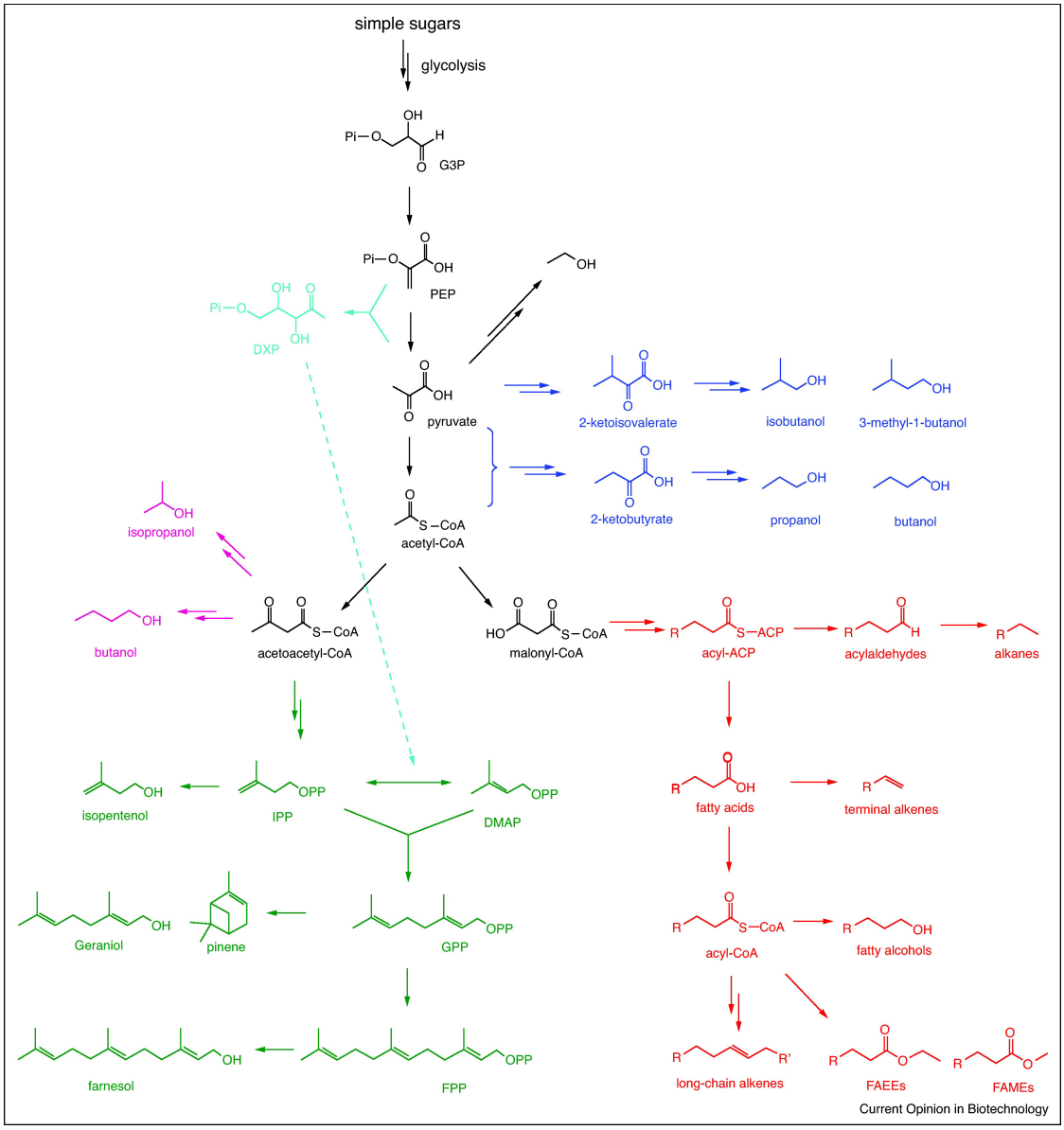

Figure 1.1: Metabolic pathways to microbial biofuels. Central metabolism is colored black. Short-chain alcohols produced by fermentative pathways are colored purple. 2-Keto acid pathways and the corresponding alcohol fuels are colored blue. Isoprenoid pathways and terpene-based fuels are colored green. Fatty acid pathway and corresponding fuels are colored red. Adapted from [8].

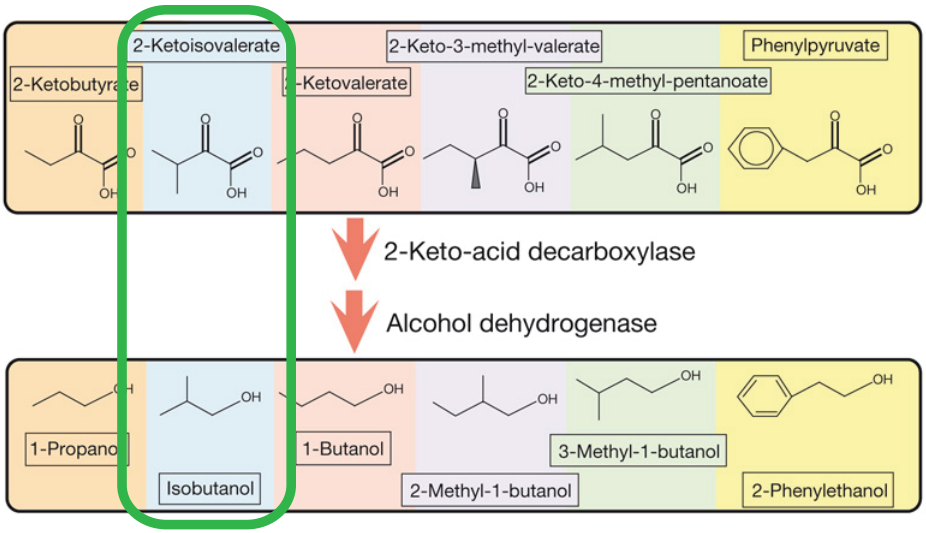

Figure 1.2: Production of alcohol biofuels from 2-keto acids (endogenous amino acid precursors) via decarboxylation and subsequent reduction. Adapted from [1].

1.1.2 Cellulosic biofuels

In parallel with metabolic engineering work to produce higher quality fuel molecules, efforts are underway to switch from sugarcane/starch based biofuel feedstocks to more sustainable lignocellulosic feedstocks. As a product of photosynthesis, lignocellulosic biomass is an abundant and renewable resource. Over half of the carbon in the biosphere is present in the form of cellulose, with approximately one trillion tons of cellulose synthesized and degraded each year globally [9]. Unlike sugarcane or starch based feedstocks, it is possible to produce and utilize lignocellulosic biomass without affecting food supplies. For example, marginal lands unsuitable for food production can provide lignocellulosic biomass via harvesting wild vegetation [10], or through intentional cultivation of robust bioenergy crops such as switchgrass (*Miscanthus sp*). Land use for producing cellulosic biofuels could be dramatically reduced by switching to carbohydrate rich microalgae as a feedstock [11] [12], though challenges remain in developing scalable methods for algae culture. In addition to bioenergy crops, many underutilized waste streams are rich in cellulose (such as crop residues and municipal solid waste) and could serve as potential feedstocks for biofuel production. It is estimated that the United States is capable of sustainably producing 1.4 billion tons of lignocellulosic biomass annually, enough to replace 30% or more of our current petroleum consumption [13].

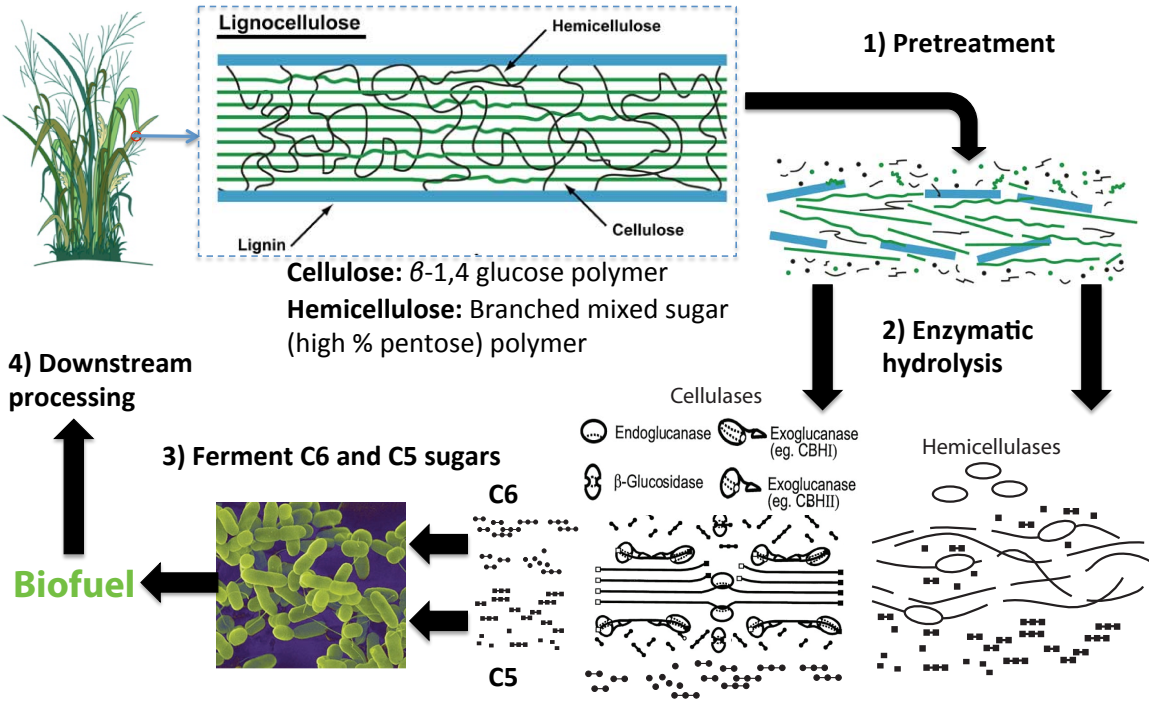

Figure 1.3: Bioprocessing lignocellulosic feedstocks into biofuels. Portions of figure adapted from http://www.biomassmagazine.com and [14].

Cellulose and related compound hemicellulose are polysaccharides which, along with lignin, are primary structural components of plant cell walls [5]. Cellulose is a polysaccharide of glucose monomers linked by β-1,4 glucosidic bonds, while hemicellulose is a polysaccharide of mixed composition and structure, containing a large proportion of pentose sugars linked by β-1,4 bonds [5]. Lignin has a complex structure, and tends to be hydrophobic with a high proportion of aromatic groups [5]. Bioprocessing of lignocellulose into fuels typically comprises four main steps: pre-treatment, saccharification (usually via enzymatic hydrolysis), fermentation of soluble hexose (C6) and pentose (C5) saccharides, and downstream processing (e.g. product separation) (Figure 1.3). In the first step, lignocellulosic biomass is subjected to mechanical and/or thermochemical treatments to improve digestibility. Pre-treatment alters the microstructure of lignocellulosic biomass, often via redistribution or removal of lignin and reduction of cellulose crystallinity, leading to improved hydrolysis rates [14]. Pre-treatment contributes substantially to overall processing costs. Cellulolytic microbes produce sophisticated and synergistic enzyme systems called cellulases and hemicellulases that effectively hydrolyze cellulose and hemicellulose, respectively [14]. In the saccharification step, cellulases and hemicellulases are used to hydrolyze insoluble cellulose and hemicellulose into soluble C6 and C5 saccharides, respectively [15]. Cellulose is a structurally complex material that is highly recalcitrant to degradation, making enzymatic hydrolysis a limiting step in microbial biofuel production [14]. In the fermentation

step, mono and oligosaccharide hydrolysis products are metabolized by microbes into biofuel molecules, such as ethanol [15]. Many microbes are unable to metabolize pentose sugars, and microbes that are capable of pentose metabolism generally consume these sugars diauxically, with hexoses consumed in preference to pentoses [16]. Additionally, the toxicity of biofuel products and inhibitory compounds present in the feedstock often limits fermentation productivity and titers. Achieving cost-effective pre-treatment and lignin removal, efficient cellulose hydrolysis, co-utilization of hexose/pentose saccharides, and mitigating the the toxic effects of biofuel products on microbes are all vital requirements for economically viable cellulosic biofuel production [16] [14].

1.1.3 Consolidated bioprocessing

Process configurations for cellulosic biofuel production can be categorized based on the extent to which the biologically mediated steps (cellulase production, enzymatic saccharification, and fermentation of soluble saccharides) are consolidated (Figure 1.4A) [14]. Most cellulosic biofuel processes presently under commercial development use process configurations known as simultaneous saccharification and fermentation (SSF) or simultaneous saccharification and co-fermentation (SSCF) [15]. SSF features a dedicated step for enzyme production (Figure 1.4A). Cellulases and hemicellulases produced in the enzyme production step are then combined with pre-treated biomass and microbes for simultaneous saccharification and fermentation to biofuel (Figure 1.4A). Hexose and pentose saccharides are co-fermented in a single step for SSCF, while in SSF hexose and pentose conversions occur in separate bioreactors (Figure 1.4A). Having a dedicated step for enzyme production contributes substantially to total processing costs (example cost analysis for ethanol production shown in Figure 1.4B). Consolidated Bioprocessing (CBP) is a promising process configuration that integrates all biochemical transformations - cellulase/hemicellulase production, saccharification, hexose fermentation, and pentose fermentation - into a single step, and may thus significantly improve process economics (Figure 1.4B)[15]. In addition to reducing processing costs via consolidation, CBP may provide other benefits such as increased hydrolysis rates due to synergy between microbes and enzymes during fermentation [15].

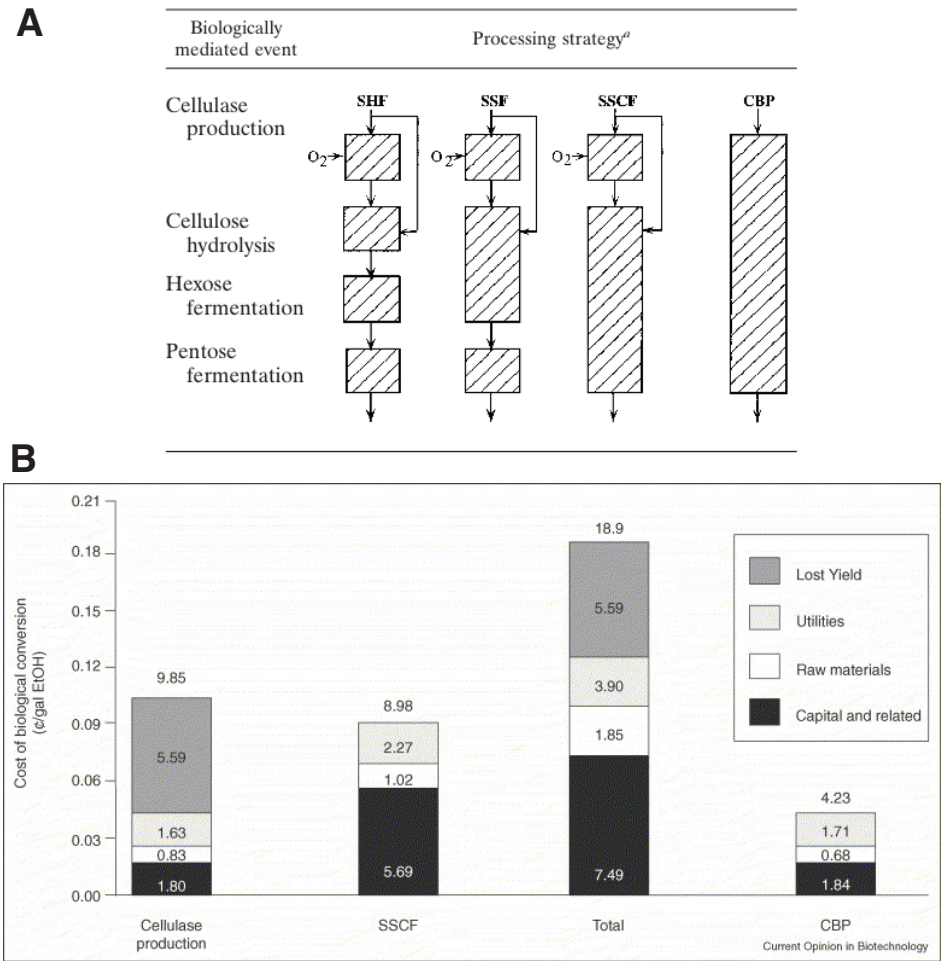

Figure 1.4: Process configurations and cost analysis for cellulosic biofuel production **(A)** Process configurations: separate hydrolysis and fermentation (SHF), simultaneous saccharification and fermentation (SSF), simultaneous saccharification and co-fermentation of hexose and pentose saccharides (SSCF), and consolidated bioprocessing (CBP). Each box represents a separate bioreactor. Adapted from [14]. **(B)** Comparative cost analysis for ethanol production by CBP vs. SSCF. Adapted from [15]; see reference for assumptions made in cost calculations.

The prevailing trend in present CBP research entails engineering microbes that incorporate all required biological functionalities into a single host (Figure 1.5). Two broad approaches are being pursued: the native cellulolytic strategy, and the recombinant cellulolytic strategy (Figure 1.5). In the native cellulolytic strategy, anaerobic cellulolytic microbes (which naturally ferment cellulose to alcohols and mixed organic acids) are engineered to improve alcohol production [15]. Most natively cellulolytic species under consideration for CBP development produce ethanol, although species producing n-butanol have recently been isolated [17]. Cellulolytic *Clostridia* species, including *C. phytofermentans*, *C. cellulolyticum*, and *C. thermocellum* are the among the best developed native hosts [18]. The native cellulolytic strategy is hindered by limited molecular biology tools for candidate microbes, substantially restricting genetic engineering

efforts [18]. Strain improvement approaches have therefore generally been limited to small-scale metabolic engineering efforts to improve ethanol yields (i.e. knocking out competing pathways to direct metabolism towards ethanol production) and developing strains with improved ethanol tolerance through adaptive evolution [18]. *C. thermocellum* was recently engineered to produce ethanol with high titer (38 g/L) and yield (82% of theoretical) from microcrystalline cellulose (in co-culture with *T. saccharolyticum*) [19]. These results represent the current state-of-the-art and are very close to projected performance targets for economically viable ethanol production (50 g/L titer and 90% theoretical yield) [18]. However, this study did not utilize real lignocellulosic biomass or industrially relevant culture conditions. Additionally, the paucity of genetic tools for *C. thermocellum* will make it difficult to use this species as a platform for production of non-ethanol biofuels. While the native cellulolytic approach has mostly focused on ethanol production, *C. cellulolyticum* was recently engineered to produce isobutanol via a heterologous pathway [20]. This study represents an important proof-of-concept for producing next-generation biofuels using the native cellulolytic strategy, however the reported isobutanol titer was low (660 mg/L) and *C. cellulolyticum* has limited potential as a CBP host due to poor cellulose utilization [21].

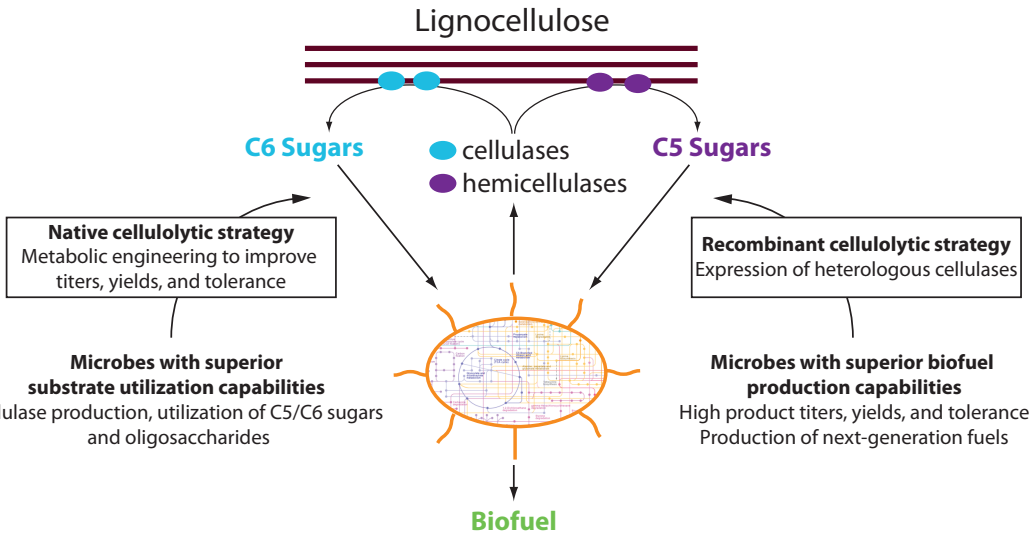

Figure 1.5: Organism development strategies for consolidated bioprocessing (CBP). Adapted from [15].

In the recombinant cellulolytic strategy, microbes that are amenable to genetic manipulation and have good biofuel production properties (e.g. biofuel production strains of *E. coli*, *S. cerevisiae*, *B. subtilis*, etc.) are engineered to produce cellulases/hemicellulases [15]. Since *E. coli* and *S. cerevisiae* have been metabolically engineered to produce a variety of biofuel molecules [16], the recombinant cellulolytic strategy could be used to produce non-ethanol fuel

molecules with desirable properties. However, engineering heterologous cellulase/hemicellulase production is extremely difficult since typical cellulase systems are comprised of dozens of different synergistically acting enzymes, which need to be co-expressed and secreted from non-native hosts. Numerous proof-of-concept demonstrations of the recombinant cellulolytic approach have been reported, however to date cellulose conversion and product titers/yields remain low [18]. The most significant advances in expressing heterologous cellulases have been with *S. cerevisiae*. Several examples of *S. cerevisiae* strains engineered to secrete cellulases and produce ethanol from cellulose have been reported, though most of these studies use amorphous cellulose (easily digestible model substrate that is much less recalcitrant than real lignocellulose) and have achieved ethanol titers in the range of only 1 to 10 g/L [18]. Beyond the work done with *S. cerevisiae*, another notable advancement is the development of a *B. subtilis* strain engineered to secrete cellulase and produce lactate as a fermentation product, with titers up to 3.1 g/L and yields 64% of the theoretical maximum. In addition to model organisms, there is interest in engineering other biofuel producing microbes to express cellulases. For example, the highly ethanologenic bacterium *Z. mobilis* has been engineered to secrete cellulases, with recent work demonstrating ethanol titers up to 4 % (v/v) using NaOH pre-treated sugarcane bagasse as a substrate [22]. There has also been progress in engineering butanologenic *Clostridium acetobutylicum* to produce extracellular cellulases, although to date none of the resulting strains are able to grow on cellulose as the sole carbon source [23].

Due to inherent challenges in both the recombinant and native cellulolytic strategies, the present CBP approach of trying to incorporate all required biological functionalities into a single super-organism has hitherto proven to be difficult. Despite intensive research efforts spanning several decades, there have been few reports of achieving commercially viable product yields and titers [15]. As an alternative approach to engineering microbes for CBP applications, it may be more tractable to compartmentalize the required biochemical functions (cellulase production and hexose/pentose fermentation) into different hosts which when co-cultured together form a synthetic microbial consortium capable of single-step conversion of cellulose to desired products. In the next section, microbial consortia are discussed from a perspective of possible bioprocessing applications.

1.2 Microbial consortia and applications in bioprocessing

In contrast to the CBP paradigm of integrating many functions into a single host, in natural environments microorganisms often live in synergistic multi-species communities in which individual

species with specialized roles cooperate to survive and thrive together [24]. Microbial consortia play many important roles in nature, ranging from participating in global biogeochemical cycles to assisting animals with food digestion [25]. Natural microbial consortia have a number of properties that are appealing in the context of bioprocessing, such the ability to perform complex tasks that would be difficult to achieve with a single species, and robustness to fluctuations in resources and environmental conditions [24]. As an example relevant to bioenergy production, natural methanogenic consortia are often utilized in anaerobic bioreactors to convert organic wastes of complex and variable composition (e.g. mixtures of polysaccharides, peptides, lipids, etc.) to CH_4 (Figure 1.6A) [26]. The conversion of complex organic substrates to CH_4 involves several distinct trophic stages, and some of the reactions are thermodynamically unfavorable and require low H_2 partial pressures [26]. As a result, methanogenic consortia often grow as highly structured aggregates in order to ensure efficient transfer of H_2 and other intermediates between species (Figure 1.6B) [26]. Methanogenic communities thus exemplify the ability of consortia to perform complex tasks, as methanogenesis requires both metabolic cooperation and spatial organization between different microbial species and would be difficult to achieve in a single unstructured microbe (Figure 1.6) [26]. In addition to their ability to perform complex tasks, long-term studies of methanogenic consortia in bioreactors have revealed that these communities are remarkably stable [27]. Community function (in terms of methane yield, substrate removal, and by-product composition in bioreactor effluent) is generally resilient to fluctuations in process temperature, pH, feed rates, and microbial populations [27]. Additionally, population composition itself tends to be highly stable, with populations rebounding to previous levels following disturbances [27].

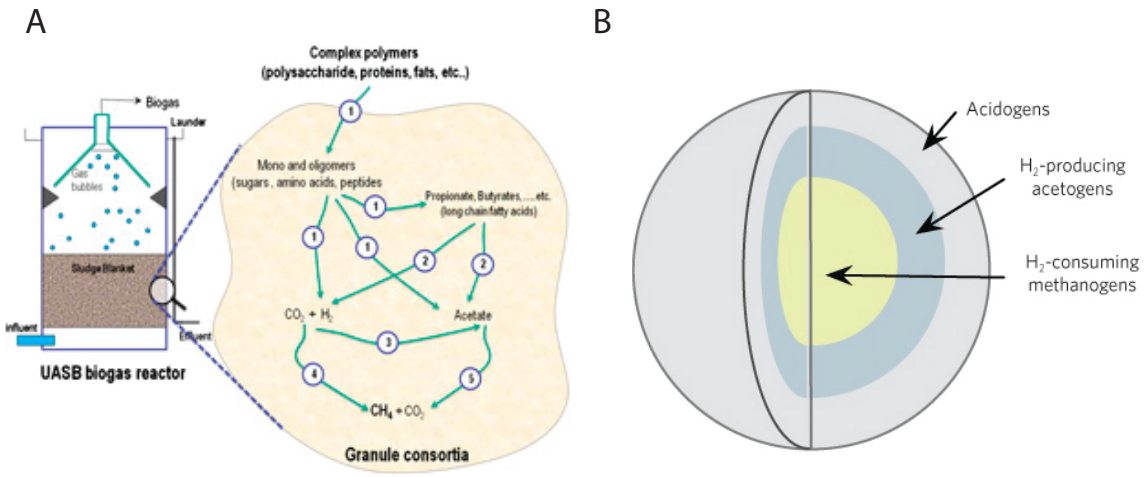

Figure 1.6: Methanogenic consortia used in anaerobic digestion. **(A)** Schematic of consortia in upflow anaerobic sludge bed (UASB) reactors. Trophic steps: 1, fermentative bacteria; 2 obligate hydrogen producing acetogenic bacteria; 3, hydrogen-oxidizing acetogens; 4, carbon dioxide reducing, hydrogen-oxidizing methanogens and 5, aceticlastic methanogens. Adapted from [26]. **(B)** Structure of typical sludge granule. Adapted from [28].

Beyond methanogenic consortia, there have been numerous efforts to isolate lignocellulose degrading consortia from natural environments with consideration towards potential CBP applications. For instance, in a recent study, enrichment of household compost on cellulose containing media yielded an aerotolerant microbial community capable of degrading cellulose to acetate and ethanol [29]. Subsequent analysis of the community revealed a diversity of species, including both facultative and strict anaerobes, with *Clostridia* species being particularly prevalent [29]. There are numerous other examples of natural cellulolytic communities in the literature; for a review, see [30]. While natural anaerobic celluloytic consortia are appealing due to their superb ability to degrade lingocellulosic biomass and metabolic capability for producing biofuel molecules (e.g. alcohols), adaptation and deployment of such consortia to industrial bioprocesses faces several key challenges. These natural consortia tend to produce mixed fermentation products, with relatively low yields of fuel molecules such as ethanol. In principle, it may be possible to engineer or otherwise manipulate these consortia to improve specificity and yield of desired products. However, these communities are often highly complex, featuring intricate interaction networks between large numbers of species, many of which are uncharacterized [30]. Lack of genetic tools and detailed knowledge of community dynamics and function thus limits our ability to engineer and improve these systems. While further study of natural microbial consortia will likely improve our understanding of the underlying interactions and design principles of these complex communities, a promising complementary approach is also possible: engineering synthetic microbial consortia.

1.2.1 Engineered microbial consortia for bioprocessing

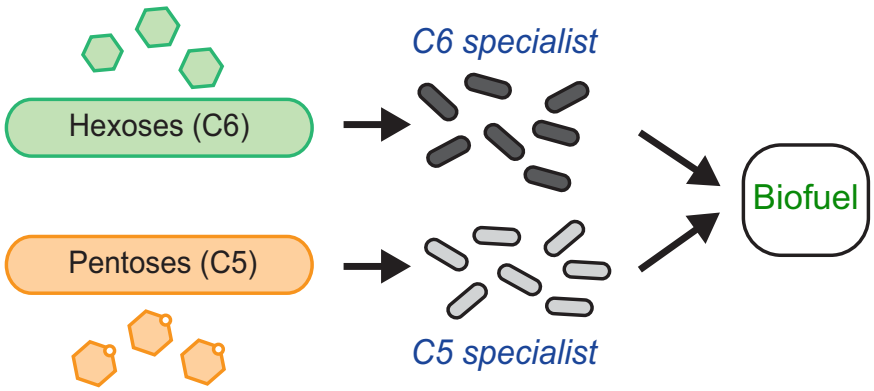

Figure 1.7: General scheme of consortia for co-fermentation of pentose (C5) and hexose (C6) sugars. Adapted from [31].

Inspired in part by the powerful and interesting features of natural consortia, there is rapidly growing interest in engineering synthetic microbial consortia for both fundamental research and biotechnology applications. Synthetic microbial consortia for consolidated bioprocessing of cellulosic feedstocks to fuels have been investigated in numerous studies. In a notable recent study, Bayer *et al*. engineered *S. cerevisiae* to aerobically produce CH_3I (which can be converted to hydrocarbons with zeolite catalysts) from acetate and ethanol [32]. By combining the engineered *S. cerevisiae* with *Actinotalea fermentans*, which aerobically ferments cellulose to acetate and ethanol, a co-culture capable of converting cellulose to CH_3I was created [32]. Engineering hexose and pentose specialized microbes, each exclusively metabolizing it's respective carbon source, is another emerging trend in engineering microbial consortia for biofuel applications (Figure 1.7). Co-cultures of hexose and pentose specialists would be expected to utilize both types of sugar simultaneously (in contrast to sequential utilization, observed in most natural species), thus improving the conversion rate of mixed sugars derived from lignocellulosic biomass (Figure 1.7). This approach was reported by Trinh *et al*. 2008 [33], Eiteman *et al*. 2008 [34], Eiteman *et al*. 2009 [35], Xia *et al*. 2012 [36], and was also pursued in our research independently during the same time period (results not yet published; [31]), all using *E. coli* as a base species. Eiteman *et al*. demonstrated that co-cultures of hexose and pentose specialized *E. coli* gave higher rates of fermentation and more complete utilization of glucose/xylose mixtures compared to fermentation by a single diauxic *E. coli* strain [35]. Furthermore, the hexose/pentose specialist co-culture was able to adapt to fluctuations in feed composition [35].

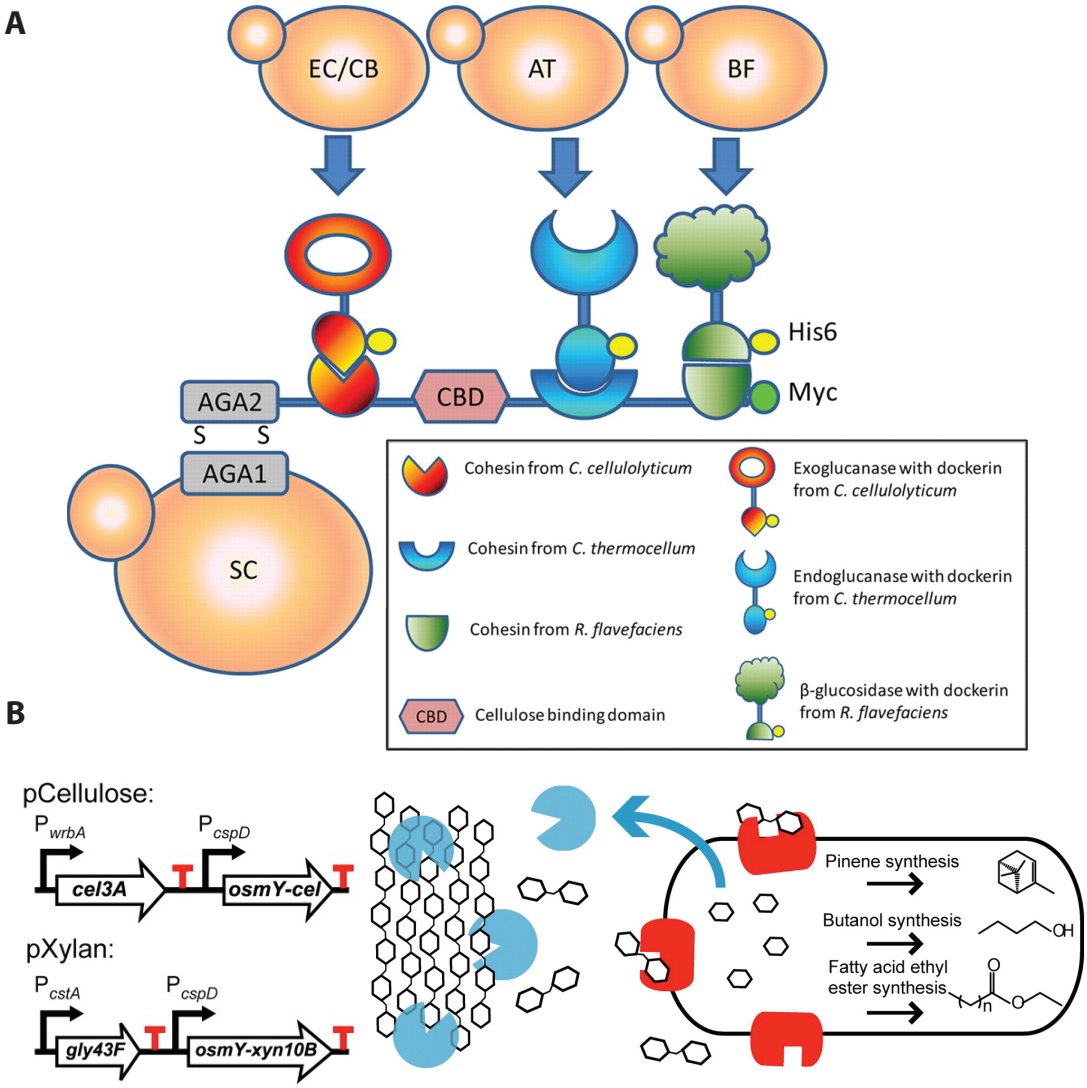

Figure 1.8: Cellulase producing consortia for CBP. **(A)** *S. cerevisiae* consortium for assembly of synthetic extracellular cellulosome complexes. SC cells surface display scaffold proteins and EC/CB, AT, and BF cells secrete cellulases. Adapted from [37]. **(B)** *E. coli* consortium for simultaneous conversion of cellulose and hemicellulose into pinene, n-butanol, or fatty acid ethyl esters (FAEE). Cells harbor plasmids containing oligosaccharide transporters (*cel3A* or *gly43F*) and secretable hydrolases (*osmY-cel* or *osmY-xyn10B*) for either cellulose (pCellulose) or xylan (pXylan). By integrating pCellulose or pXylan into hosts containing heterologous pathways for biofuel production, direct conversion of lingocellulosic biomass into biofuels can be achieved. Adapted from [38].

Another emerging approach in developing consortia for biofuel production entails engineering specialist strains to secrete different synergistically acting cellulases and/or hemicellulases (examples shown in Figure 1.8). Such an approach was pursued by Arai *et al*. 2007 [39], Tsai *et al*. 2010 [37], Goyal *et al*. 2011 [40], and Bokinsky *et al*. 2011 [38]. Tsai *et al*. engineered *S. cerevisiae* strains to secrete three different cellulases and a scaffold protein, respectively, which were then combined to form a synthetic consortium (Figure 1.8A) [37]. The secreted cellulases assembled extracellularly on the scaffold protein, forming a synergistic cellulosome complex

(Figure 1.8A) [37]. Through the production of extracellular cellulosomes, the consortium was able to convert phosphoric acid swollen cellulose (PASC) to ethanol at yield of up to 0.475 g/g (93% theoretical) and titer of 1.87 g/L, using an optimized strain ratio (SC:AT:CB:BF) of 7:2:4:2 (Figure 1.8A) [37]. In contrast, a monoculture employing a similar enzyme system achieved much lower ethanol yields and titers, possibly due to the burden of expressing and secreting multiple heterologous proteins in a single host [37]. As another example, Bokinsky *et al*. engineered *E. coli* strains to produce cellulases or xyalanases, respectively, which could then be co-cultured to accomplish simultaneous conversion of both hemicellulose and cellulose [38] (Figure 1.8B). By integrating heterologous pathways for butanol, pinene, or fatty acid ethyl esters (FAEE), consortia of xyalanase/cellulase producing *E. coli* strains were able to directly convert ionic liquid (IL) pre-treated switchgrass to these respective biofuels (Figure 1.8B) [38]. While the titers and yields achieved in this study are very low, it serves as an important proof-of-concept of using microbial consortia for CBP production of next-generation biofuels.

1.2.2 Synthetic microbial consortia: design and population coordination

The above examples illustrate a broader trend of utilizing synthetic microbial consortia to compartmentalize pathways into different hosts for individual optimization, and/or demonstrate that consortia may be engineered for division of labor to accomplish complex tasks proven difficult to achieve with monocultures. Another broad approach to engineering microbial consortia involves using a bottom-up strategy of programming specific interactions between microbes [24]. These approaches have been used to construct several canonical ecological and logic systems for proof-of-concept and fundamental study. For example, Balagaddé *et al*. constructed a synthetic predator-prey system in *E. coli*, in which the predator and prey strains communicate bi-directionally through quorum sensing and regulate each other's gene expression and survival via synthetic genetic circuits (Figure 1.9A) [41]. Predator cells kill the prey by inducing transcription of a toxic protein in the prey, while the prey rescue predator cells by inducing transcription of an anti-toxin protein in the predator (Figure 1.9A) [41]. The synthetic ecosystem mimics natural predator-prey systems in terms of logic and long-range dynamics [41]. As another noteworthy example, Shou *et al*. engineered a synthetic symbiosis between two *S. cerevisiae* auxotrophs (Figure 1.9B) [42]. In minimal media, the auxotrophs cross-feed each other and are mutually dependent on one another for growth (Figure 1.9B) [42]. The populations stably coexist over long time periods, and exhibit complex oscillatory dynamics that occur due to delay in nutrient release until cell death, among other factors (Figure 1.9B) [42].

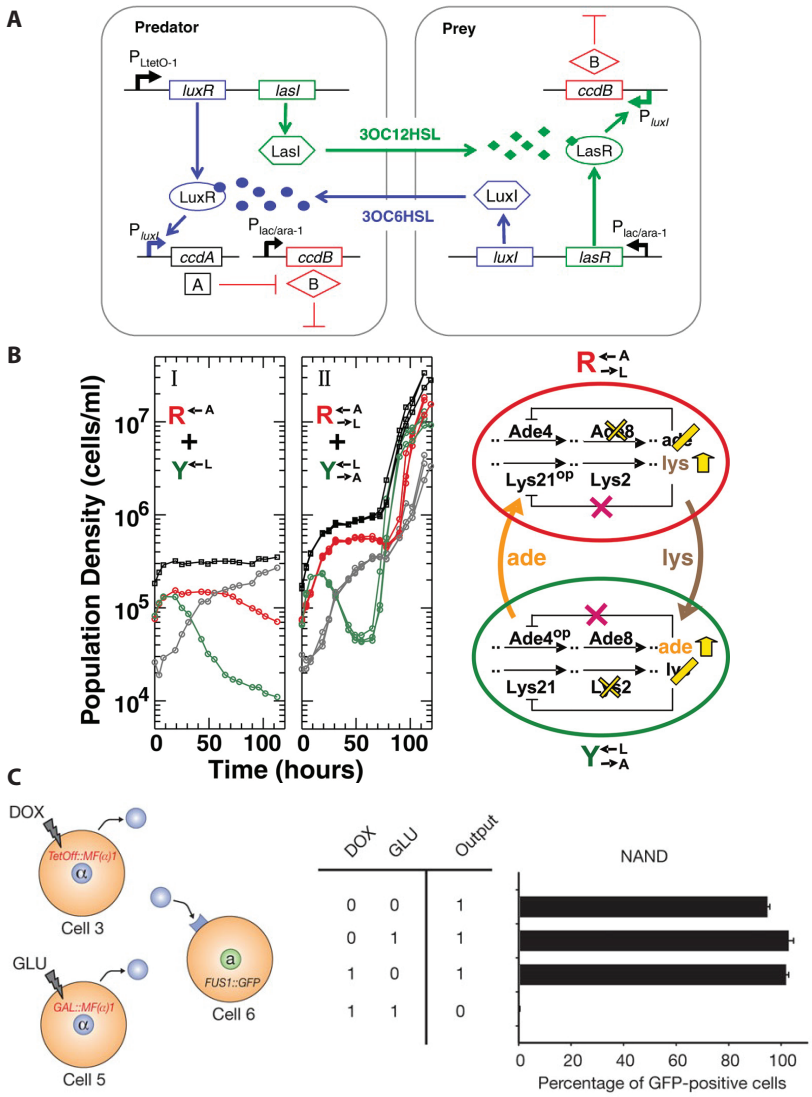

Figure 1.9: Programming synthetic consortia with genetic circuits and intercellular communication **(A)** Synthetic predator-prey system. Predator cells produce quorum sensing (QS) molecule 30C12HSL, activating transcription of toxic protein *ccdB* in prey cells. Prey cells rescue predator cells by producing QS 3OC6HSL, which activates transcription of anti-toxin protein *ccdA*, thus rescuing cells from endogenously produced *ccdB*. Adapted from [41]. **(B)** Synthetic cooperation. Engineered *S. cerevisiae* strains auxotrophic for Lys (Y) and Ade (R) cross-feed each other, forming a stable mutualism. Population dynamics of live R (red), live Y (green), dead (gray), and total (black) cells are shown to the left. Adapted from [42]. **(C)** Example of an engineered multicellular NAND gate using three *S. cerevisae* strains and two extracellular inputs. Doxycycline and 2% glucose are added to the culture as shown in the table, and the resulting output GFP expression is shown to the right. Adapted from [43].

In a final example, Regot *et al*. constructed Boolean logic gates through the use of combinations of different engineered *S. cerevisiae* strains [43]. The cells can either respond to an exogenous signal (i.e. an added inducer) and/or an output from another *S. cerevisae* strain (i.e. α factor, a *S. cerevisiae* pheromone) and then release either another signal or the final output, in this case green fluorescent protein (GFP) (Figure 1.9C) [43]. Different strains can be combined

to yield Boolean operators such as AND, NAND, OR, NOR, NIMPLIES, IDENTITY, and NOT gates, and the overall computation is determined by the number of cells involved and the types of each engineered cell. An example of a three-strain combination yielding a NAND gate is shown in Figure 1.9C [43]. Using only three inputs and three cell types, the authors predict that over 100 different functions can be created [43].

All three of the above studies follow a bottom-up synthetic biology approach of combining relatively simple modular parts to build systems with complex behaviors and functions. These examples highlight the use of synthetic genetic circuits and intercellular communication to coordinate populations in microbial consortia. Population coordination is of key importance in both natural and synthetic microbial consortia, and will be further discussed in Chapter 3. While bottom-up design and programmed interactions have been used to construct a number of interesting proof-of-concept consortia, there have been few demonstrations of this approach for biotechnology applications, despite much tantalizing potential [24]. As a notable exception, Prindle *et al*. recently developed a microbial Arsenic detection system by engineering an oscillatory circuit synchronized across an entire cell population in which the oscillatory period is modulated as a function of Arsenic concentration [44].

Although great progress has been made in engineering synthetic microbial consortia, many challenges remain. In stark contrast to their natural counterparts, synthetic microbial consortia are often fragile and unstable, limiting their utility in real-world applications like industrial bioprocessing. In mixed cultures created by arbitrarily combining different species, population compositions are often unstable [45]. Single competitors may dominate the population and drive other consortium members to extinction, and in other cases, the entire consortium may collapse [45]. These instabilities often arise due to mismatches between growth rates and rates of resource production/consumption between different species [45]. Beyond ecological stability, consortia featuring programmed interactions between microbes are subject to evolutionary pressures, with mutational inactivation of synthetic genetic circuits often occurring in relatively few generations of growth, depending on the size and host burden of the circuit [46]. Evolutionary and ecology theory provides a framework to understand the stability of social interactions in microbial consortia, and indeed some of these theories have been empirically tested with synthetic microbial consortia [47]. Despite this, many efforts towards engineering microbial consortia are undertaken with reductionist perspectives that fail to consider long-term stability.

In the first part of this dissertation, we seek to integrate evolutionary and ecology theory with engineering principles to design and construct stable and tunable microbial consortia for

consolidated bioprocessing of cellulosic feedstocks to biofuels, focusing on isobutanol as a proof-of-concept. In addition to achieving direct cellulose conversion, another key difficulty in microbial biofuel production is the toxicity of biofuel products and inhibitors present in pre-treated lignocellulose, which will be reviewed in the next section along with methods for engineering improved tolerance.

1.3 Toxicity issues in microbial biofuel production and genome-scale phenotype improvement methods

1.3.1 Microbial chemical toxicity

Toxicity of biofuel products and feedstock inhibitors limits final titer and volumetric productivity in fermentations, thus motivating efforts to engineer microbes with improved tolerance [48]. Numerous investigations have elucidated mechanisms of toxicity, proximal cellular responses, and possible mechanisms of tolerance to biofuel molecules and feedstock inhibitors [49] (see Figure 1.10 for overview). Biofuel molecules, like other organic solvents, intercalate into the membrane lipid bilayer, perturbing the physicochemical properties of the membrane [49]. Membrane fluidity and permeability are altered, and membrane proteins may be denatured; these changes can ultimately lead to cytosolic leakage, dissipation of membrane electrochemical potential and proton gradient, and disruption of membrane based processes such as substrate transport and respiration [49]. For many biofuel molecules, toxicity is correlated to the octanol-water partition coefficient P_{ow},; this reflects increasing tendency for more hydrophobic molecules to partition into the lipid bilayer [49]. In addition to affecting cell membranes, biofuel molecules can damage or denature other biological molecules [49]. Biofuel molecules can exert chaotropic effects, leading to protein and RNA denaturation. Biofuel molecules can also lead to increased generation of reactive oxygen species (ROS), causing oxidative damage of proteins, DNA, and lipids [49].

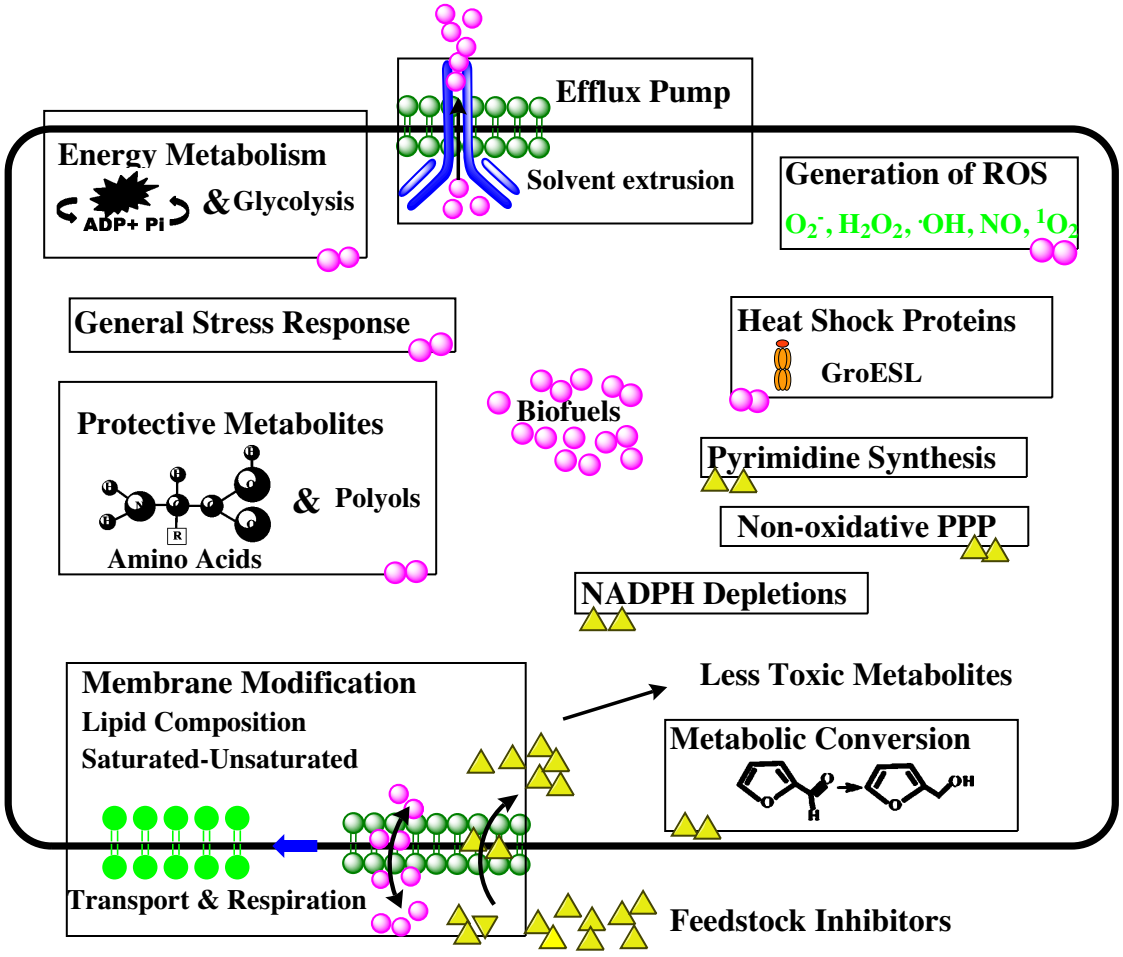

Figure 1.10: Mechanisms of toxicity and microbial tolerance to biofuels and feedstock inhibitors. Adapted from [49].

General mechanisms of biofuel molecule toxicity and tolerance have been extensively studied for short and medium chain-length alcohols (such ethanol and n-butanol) in model organisms (*E. coli* and *S. cerevisiae*), as well as natural producers (e.g. solventogenic *Clostridia* species). In addition to biofuel molecules, toxic chemical species may be generated during pre-treatment and hydrolysis of lingocellulosic feedstocks. These feedstock inhibitors comprise a wide range of different compounds, including furans, weak acids, and phenolics [50]. Furans are thought to disrupt energy metabolism through inhibition of key enzymes and depletion of cofactor pools [49]. Weak acids perturb energy metabolism as well, via disruption of membrane electrochemical potential [49]. The phenolics present in biomass hydrolysates constitute a very broad range of compounds and thus mechanisms of toxicity are difficult to generalize; however, it is believed that the cell membrane is a primary target of toxicity, similar to other organic solvents [49]. Feedstock inhibitors are known to increase generation of ROS, and may also inhibit certain metabolic enzymes [49]. Numerous cellular responses to biofuel molecule and feedstock inhibitor toxicity have been observed, including induction of general stress response (such as upregulation of heat

shock proteins), upregulation of efflux pumps, synthesis of protective metabolites, alteration of membrane and cell surface properties (e.g. by altering composition of peptidoglycan, lipopolysaccharides, membrane proteins, and lipids), adaptations in energy metabolism, changes in cellular morphology, and metabolic degradation of toxic species (Figure 1.10); some or all of these responses may be present in a given organism [49]. Due to broad mechanisms of toxicity, tolerance to biofuel molecules and feedstock inhibitors is a complex trait that involves a diversity of cellular adaptations and responses that probably contribute synergistically to the overall phenotype [50].

1.3.2 Phenotype improvement methods

The inherent biological complexity of chemical tolerance has generally precluded the use of rational engineering for phenotype improvement, thus most strategies are combinatorial in nature, following a paradigm of generating genotypic and phenotypic diversity in a population, then characterizing isolates with the desired properties (see Figure 1.11 for examples) [51]. A number of effective genome-scale approaches to improving stress tolerance phenotypes have been developed following this paradigm, but present methods are intrinsically limited to exploring small genotype search spaces. For example, knockout libraries (Figure 1.11A) have been screened for stress tolerant isolates, but this only identifies single loss-of-function mutations that contribute to improved tolerance [50]. Similarly, plasmid-based genomic library enrichment studies (Figure 1.11B) have been used to investigate genetic bases of stress tolerance, but these libraries are restricted to examining gene overexpression and cannot readily capture interactions between multiple distal genes due limitations on insert size [50]. Targeted mutagenesis of master transcriptional regulators (Figure 1.11C) has been used to generate libraries of mutants with global perturbations in gene expression, from which highly tolerant clones have been selected, but this approach inherently explores a very restricted genotype space [50, 52, 53].

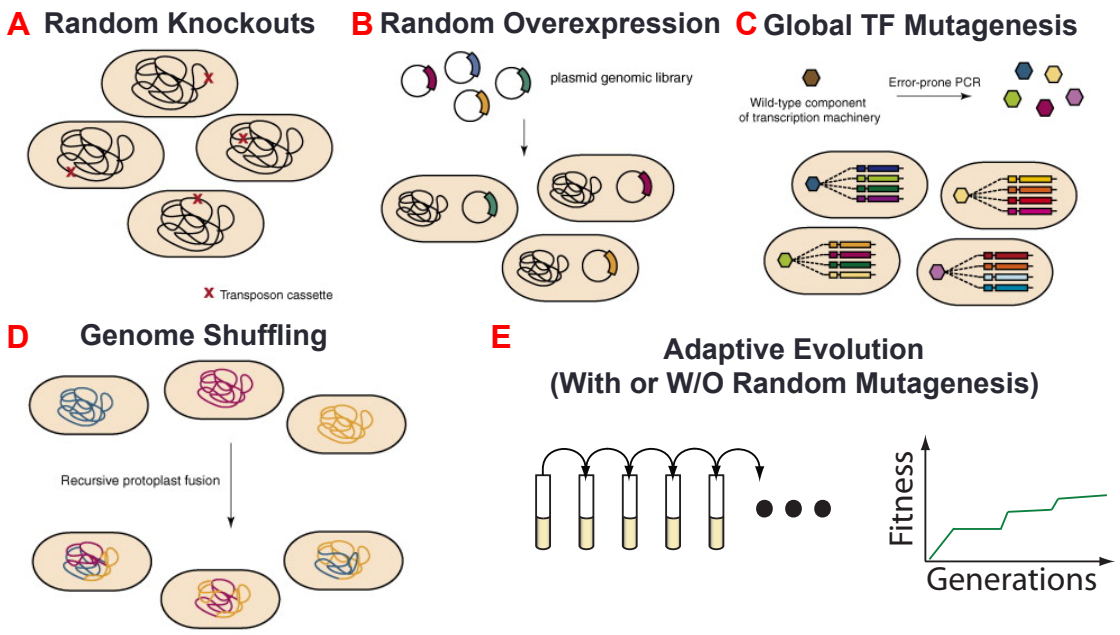

Figure 1.11: Overview of commonly used phenotype improvement methods. Adapted from [54].

1.3.3 Next-generation genome-scale phenotype improvement methods

There is a great need for phenotype improvement methods that expand accessible genotype search spaces and which can account for the multigenic nature of complex stress tolerance phenotypes. In recognition of this need, genome-scale methods that can capture interactions between distal genetic loci have recently been developed (Figure 1.12). An improved genomic library method, Coexisting/Coexpressing Genomic Libraries (CoGels), features plasmid-plasmid or plasmid-fosmid libraries that can coexist in a single cell, thus enabling detection of epistatic interactions between distal genetic loci (Figure 1.12B) [55]. Several combinations of genetic loci underlying the complex acid-tolerance phenotype of *E. coli* were identified by CoGels screeing, including the well know combination of *gadBC* and *adiC*, and the novel combination of *arcZ* and *recA* that remarkably increased acid tolerance by 9000-fold [55]. Another method, Trackable Multiplex Recombineering (TRMR), incorporates parallel DNA synthesis, homologous recombination, and molecular barcode technology to rapidly create and evaluate thousands of specific genetic modifications concurrently (Figure 1.12A) [56]. TRMR employs synthetic DNA cassettes containing a selectable marker, barcode sequences for tracking, a functional region for producing mutations or genetic modifications of interest, and targeting sequences for chromosomal integration via homologous recombination (Figure 1.12A) [56]. In a demonstration study, TRMR was applied to introduce approximately 4,000 upregulation and 4,000 downregulation gene mutations into a population of *E. coli* cells which were then subject to selection on corn stover hydroyslate, allowing for identification of mutations which improve tolerance. By using

excisable selection markers, TRMR could be executed recursively and thus may be capable of detecting interactions between distal mutations, though results from this approach have not yet been reported (Figure 1.12A) [56].

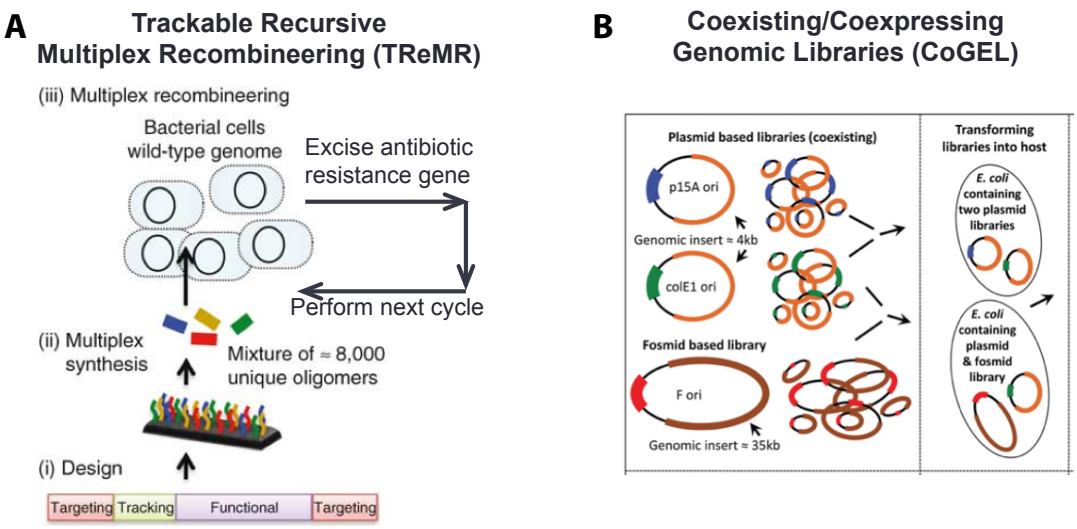

Figure 1.12: Examples of genome-scale phenotype improvement methods that are capable of capturing interactions between distal genetic loci. **(A)** Trackable Recursive Multiplex Recombineering (TRMR). Adapted from [56]. **(B)** Coexisting/Coexpressing Genomic Libraries (CoGels). Adapted from [55].

Recent advances in DNA sequencing technology and genome engineering have dramatically increased our ability to read and write genetic information, offering unprecedented opportunities for genome-scale research and engineering. An intriguing corollary to advances in DNA sequencing technology is the prospect of using whole genome resequencing to characterize genetic adaptations in evolved microbes. Experimental evolution of microbes has long been used as a tool for both fundamental studies of adaptation and for phenotype improvement in biotechnology applications [51, 57]. Resequencing the genomes of evolved microbes has proven to be extremely valuable for determining underlying mechanisms and genetic bases of adaptation, and has been applied to range selectable phenotypes [58]. In parallel with advances in DNA sequencing, new technologies have recently been developed for large-scale genome engineering, allowing for vastly expanded combinatorial mutation search spaces. Multiplex automated genome engineering (MAGE) is a key breakthrough in this area, allowing for rapid generation of combinatorial libraries of point mutations. MAGE entails subjecting a population of *E. coli* cells to repeated cycles of high efficiency homologous recombination with large libraries of mutagenic ssDNA oligonucleotides, generating up to $\approx 10^9$ genetic variants per day [59]. In their demonstration study, Wang *et al*. used MAGE to optimize a heterologous lycopene biosynthesis pathway in *E. coli*, generating $1.5\text{x}10^9$ different genetic variants in 35 MAGE cycles (about 4 days time)

from which variants with record-breaking lycopene yields were isolated, demonstrating the power of this technique [59]. The scale of MAGE has recently been extended through the development of conjugative assembly genome engineering (CAGE), enabling chromosomal regions modified in separate cells to be merged together in a single host [60]. MAGE has also been modified to permit construction of longer mutations (i.e. insertions $>$ 10 bp) via the co-selection MAGE (CoS-MAGE) method [61] and for selection of highly modified cells via co-operative oligonucleotide co-selection [62]. While MAGE and related methods afford genome-scale modification of existing genetic material, recent advances in DNA synthesis technologies have enabled de novo synthesis and assembly of multigene constructs, chromosomes, and even entire microbial genomes [63]. These new synthesis technologies will allow for bottom up design and construction of large-scale synthetic gene networks and perhaps even synthetic organisms.

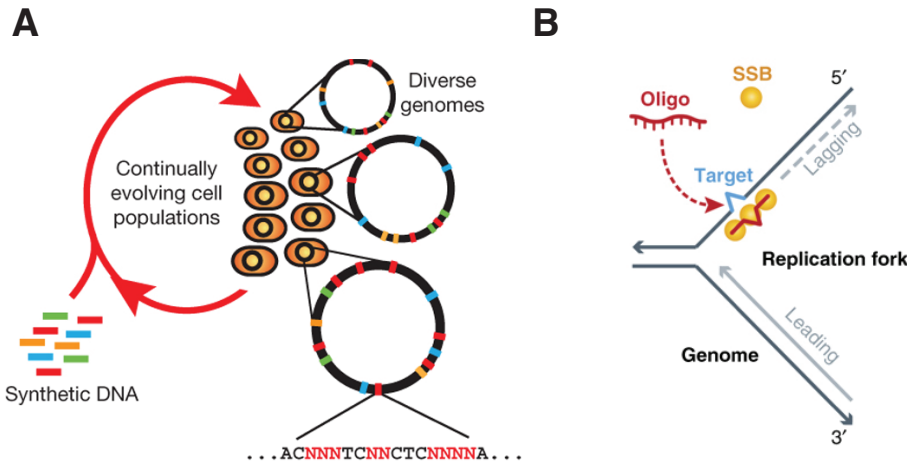

Figure 1.13: Multiplex Automated Genome Engineering (MAGE). **(A)** MAGE entails repeated cycles of high efficiency multiplexed oligonucleotide recombination, enabling iterative and rapid generation of combinatorial mutation libraries. Adapted from [59]. **(B)** Oligonucleotide recombination. Short oligonucleotides are targeted to the lagging strand of replicating DNA (with the aid single-strand binding proteins; SSB) and are integrated into daughter chromosomes. Under ideal conditions, allelic replacement efficiencies can be $>$25% (fraction recombinant cells out of total population) without selection. Adapted from [64].

Integrating approaches for both decoding and programming genomes offers much potential for elucidating and improving the genetic bases of complex stress tolerance phenotypes. In the second part of this dissertation, we will explore such an integrated approach for engineering isobutanol tolerant microbes.

1.4 Dissertation overview

The ultimate goal of this dissertation is to advance the development of cost-effective, efficient microbial systems for conversion of cellulosic feedstocks into high-quality biofuels. Tremendous research efforts are still needed to enable scalable and economically viable biofuel production, and engineering of microbes with desired properties (e.g. high efficiency and robustness) remains one of the biggest challenges [65], among others. We seek to address two major challenges in microbial biofuel production: achieving direct microbial conversion of lignocellulosic feedstocks to biofuel products, and improving microbial stress tolerance phenotypes to generate more robust biofuel production hosts. In both of these areas, the underlying biochemical complexity creates formidable obstacles for traditional engineering approaches. Ecology and evolutionary theory provide valuable frameworks for understanding complex biological systems. In this work, we weave design principles and tools derived from ecology and evolutionary theory into powerful methodologies for engineering novel microbial systems for biofuel production. Our approaches are general enough to be applied to a variety of organisms and biofuel products, but for proof-of-concept we specifically focus on production of and tolerance to isobutanol, widely considered to be one of the most promising next-generation biofuels.

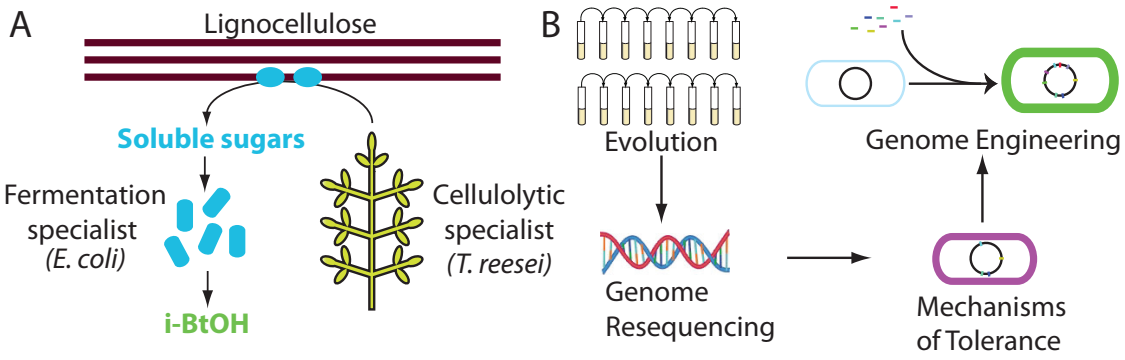

Figure 1.14: Dissertation overview. **(A)** Synthetic fungal-bacterial consortia for cellulosic isobutanol production. **(B)** Evolutionary-genomics approach for elucidating and improve microbial stress tolerance phenotypes.

Inspired by the ubiquity and power of natural microbial consortia, the first part of this dissertation explores a novel alternative direction for microbial biofuel production: the design and construction of robust and tunable microbial consortia consisting of multiple specialized species which work together to directly convert cellulose into biofuels (Figure 1.14A). The required biological functions are divided between two specialists: a cellulolytic specialist, which secretes cellulase enzymes to hydrolyze lignocellulosic biomass into soluble saccharides, and a fermentation specialist, which ferments soluble saccharides into biofuel products (Figure 1.14A). The modularity of the consortium platform offers several key advantages over the more conventional approach of using a single microbe, or "super-bug", which will be discussed in subsequent

chapters. As a proof-of-concept, we designed, modeled, and experimentally validated a synthetic microbial consortium for direct conversion of cellulose to isobutanol, using the fungus *Trichoderma reesei* RUTC30 as the cellulolytic specialist and isobutanol production strain *Escherichia coli* NV3 pSA55/69 as the fermentation specialist (Figure 1.14A) (Chapter 2). We also identified and characterized ecological mechanisms that permit stable and tunable consortium population compositions, enhancing the utility of this platform for real-world bioprocessing applications (Chapter 3).

In the second part of this dissertation, we leverage recent advances in DNA sequencing and genome engineering technologies to decode, improve, and refactor complex microbial stress tolerance phenotypes (Figure 1.14B). Our approach integrates experimental evolution of stress tolerance (Chapter 4), genome re-sequencing to identify acquired mutations (Chapter 4), genomic and functional dissection to reverse engineer mechanisms of tolerance (Chapter 4), and multiplex genome engineering to explore a large combinatorial genotype space for further improved variants (Chapter 5) (Figure 1.14B). We applied this methodology to study and improve isobutanol tolerance in *E. coli*. We evolved multiple *E. coli* lineages on isobutanol spiked media and then re-sequenced the genomes of highly tolerant isolates (Chapter 4). Through our genome resequencing efforts and subsequent followup studies, we identified a set of synergistic genetic loci correlated with isobutanol tolerance (Chapter 4). We also discovered a range of novel mechanisms of tolerance, most notably remodeling of the cell envelope to counteract deleterious effects of isobutanol and adaptations involving the *E. coli* stress response (Chapter 4). In the final stage of this project, we used multiplex genome engineering to perform targeted mutagenesis of candidate genetic loci, enabling rapid exploration of vast genotype spaces for improved variants (Chapter 5). Highly tolerant strains were isolated from mutant libraries and further characterized via detailed phenotype and genotype analysis, providing valuable insights into biochemical and genetic mechanisms of isobutanol tolerance, as well as generating highly tolerant *E. coli* strains that may be immediately useful for isobutanol production (Chapter 5).

Part I: Synthetic fungal-bacterial consortia for production of cellulosic biofuels and commodity chemicals

CHAPTER 2

Design and construction of synthetic *T. reesei* / *E. coli* consortia for direct conversion of cellulose to isobutanol

2.1 Summary

In nature microbes usually live in synergistic communities in which individual species with specialized roles cooperate to perform complex tasks. The powerful features of natural consortia have inspired great interest in engineering synthetic consortia for biotechnology applications. However, there are few reports of synthetic consortia being deployed in real-world applications. In this work, we bridge ecology and evolutionary theory with engineering principles to develop robust synthetic fungi-bacteria consortia for flexible biosynthesis of valuable products from lignocellulosic feedstocks. The required biological functions are divided between two specialists: the fungus *Trichoderma reesei*, which secretes cellulase enzymes to hydrolyze lignocellulosic biomass into soluble saccharides, and the bacterium *Escherichia coli*, which metabolizes soluble saccharides into desired products. We developed and experimentally validated a comprehensive mathematical model for *T. reesei* / *E. coli* (TrEc) consortia, providing insights on key determinants of performance. To illustrate bioprocessing applications, we demonstrate direct conversion of microcrystalline cellulose and ammonia fiber expansion (AFEX) pre-treated corn stover to isobutanol with the consortium, with titers up to 1.88 g/L and 62% of theoretical yield.

The majority of the work presented in this chapter has been submitted for publication to the Proceedings of the National Academy of Sciences and is currently in revision: J. Minty, M. Singer, S. Scholz, C.H. Bae, J. Ahn, C. Foster, J.C. Liao, and X. Lin. "Design and characterization of synthetic fungal-bacterial consortia for direct production of isobutanol from cellulosic biomass". *In revision (PNAS)*, 2013.

2.2 Introduction and background

In contrast to the CBP "superbug" paradigm, in almost all natural environments microbes live in synergistic communities in which individual species with specialized roles cooperate to survive and thrive together [25]. Natural microbial consortia hold many appealing properties in the context of bioprocessing, such as stability, functional robustness, and the ability to perform complex tasks [66, 30]. Inspired by the powerful features of natural consortia, there is rapidly growing interest in engineering synthetic consortia for biotechnology applications [66]. Examples include co-cultures of genetically modified *E. coli* for co-fermentation of hexose and pentose sugars [33, 35, 36] or direct conversion of lignocellulosic biomass to advanced biofuels [38], synthetic *S. cerevisiae* consortia for assembly of extracellular minicellulosomes and direct production of cellulosic ethanol [40], and a co-culture of *A. fermentans* and genetically engineered *S. cerevisiae* for converting lignocellulose to methyl halides [32], as reviewed in Chapter 1. These examples illustrate a broader trend of utilizing synthetic microbial consortia to compartmentalize pathways into different hosts for individual optimization and/or demonstrate that consortia may be engineered for division of labor to accomplish complex tasks proven difficult to achieve with monocultures.

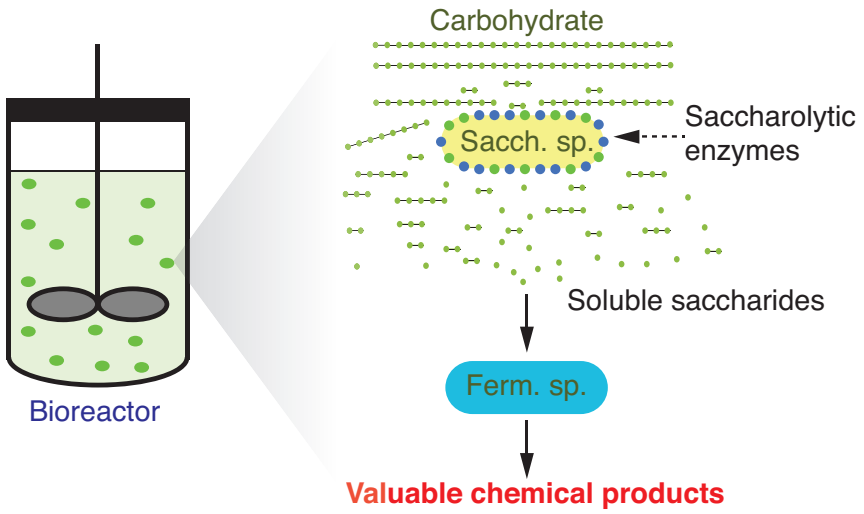

Figure 2.1: Proposed design of two-member saccharolytic/fermentation (S/F) consortia for consolidated bioprocessing of carbohydrates into fuels or other valuable chemical products. The saccharolytic specialist produces saccharifying enzymes that hydrolyze insoluble biomass polysaccharides into soluble mono and oligosaccharides, which are then metabolized by the fermentation specialist into desired products.

As an alternative to conventional approaches for engineering microbes for biofuel production, we will explore a promising new direction: engineering synthetic microbial consortia in which different species, each specialized for a specific task, work together to convert biomass feedstocks to fuels or other valuable products. We propose a general scheme in which the required biological

functions are divided between two microbial specialists: a saccharolytic specialist, which produces saccharifying enzymes to hydrolyze biomass carbohydrates into soluble saccharides, and a fermentation specialist, which metabolizes soluble saccharides into desired products (Figure 2.1). From an engineering standpoint, pursuing a microbial consortium approach of compartmentalizing each of these functionalities into separate organisms appears to be much more tractable than attempting to integrate all of them into a single host. The notion of using mixed microbial cultures to convert biomass carbohydrates into fuels or other products is not new, and has come in and out of vogue in bioprocessing research multiple times over the past few decades. Thus the general approach of using two-member saccharolytic/fermentation (S/F) mixed cultures (Figure 2.1) has been tried a number of times; various examples are given in Table 2.1. While these studies provide valuable proof-of-concept demonstrations, numerous barriers remain to utilizing synthetic consortia for CBP applications.

Product	**Saccharolytic sp.**	**Fermentation sp.**	**Substrate**	**Yield*** **(g/g)**	**Titer (g/L)**	**Notes and source**
Ethanol	*Kluyveromyces fragilis* LOCK 0027	*Zymomonas mobilis* 3881	Jerusalem artichoke (inulin)	0.48	99	[67]
	Fusarium oxysporum F3	*Saccharomyces cerevisiae* 2541	Sweet sorghum (glucose, sucrose, lignocellulose)	0.46	49	[68]
	Clostridium thermocellum M1570	*Thermoanaerobacterium saccharolyticum* ALK2	Avicel (cellulose)	0.41	38.1	Rich media; [19]
	Acremonium cellulolyticus C-1	*Saccharomyces cerevisiae* ATCC 4126	Solka Floc (cellulose)	0.18	46.3	Rich media; [69]
	Trichoderma reesei RUTC30	*Saccharomyces cerevisiae* Y1	Glucose and Solka Floc (cellulose)	0.27	40	Rich media with added glucose; [70]
Butanol	*Bacillus subtilis* WD 161	*Clostridium butylicum* TISTR 1032	Cassava starch	0.17	6.7	Rich media; [71]
	Clostridium thermocellum ATCC 27405	*Clostridium saccharoperbutylacetonicum* N1-4	Avicel (cellulose)	0.2	7.9	Rich media; [72]

Table 2.1: Examples of two-member saccharolytic/fermentation (S/F) consortia for producing biofuels from carbohydrates. *Yields given as g product / g estimated total carbohydrates consumed.

Most of the studies listed in Table 2.1 employ culture conditions and methods that are neither practical nor economical for industrial biofuel or chemical production. For instance, many of these studies use rich media containing costly supplements (e.g. yeast extract, peptone, amino acids, etc.), very high (>10% v/v) inoculation volumes, and in the case of cellulose, utilize model substrates instead of real biomass. Furthermore, many these prior works employ sequential inoculation schemes in which the saccharolytic specialist is cultured on the feedstock for a period of time before inoculating with the fermentation specialist; frequently the culture conditions are chosen to favor the saccharolytic organism at the beginning of the fermentation and then later modulated to favor the fermentation specialist. Such sequential inoculation strategies are

advantageous when the saccharolytic and fermentation specialists have different physiological and environmental preferences, and work well in bench-scale systems. However, sequential inoculation poses drawbacks for process scale-up. For batch fermentations, sequential inoculation requires preparation of inoculum cultures for each new batch, as opposed to the more cost-effective approach of recycling a fraction of the previous batch as inoculum for new batches (known as cell-recycling or repeated-batch fermentation). Additionally, sequential inoculation would be difficult to deploy in a continuous process, and would most likely require separate sequential bioreactors for the saccharolytic and fermentation specialists, respectively. This configuration would result in substantially increased costs compared to a single bioreactor CBP scheme (indeed, such a sequential bioreactor setup closely resembles an SSF or SSCF process). Finally, it is well known that mixed cultures created by arbitrarily combining different species are often unstable [45]; very few prior studies of S/F consortia have investigated the stability of these systems.

In addition to the above practical considerations, at a more fundamental level our limited understanding of the dynamics and interactions of microbial populations impedes the deployment of synthetic consortia for real-world applications. Evolutionary and ecology theory provides a framework to understand the population dynamics and stability of microbial consortia [47]. In this work, we bridge evolutionary and ecology theory with engineering principles to conduct an in-depth theoretical and experimental analysis of two-member S/F consortia for consolidated bioprocessing of lignocellulosic feedstocks. As a model system, we implemented a consortium for lignocellulose CBP consisting of the cellulolytic fungus *Trichoderma reesei* RUTC30 and *Escherichia coli* as a fermentation specialist. We developed a comprehensive mathematical model for *T. reesei-E. coli* (TrEc) consortia, semi-mechanistically capturing salient features and allowing us to elucidate key behaviors and ecological interactions. To illustrate the bioprocessing potential of this consortium, we demonstrate direct conversion of microcrystalline cellulose and pretreated corn stover to isobutanol by deploying an isobutanol producing *E. coli* strain as the fermentation specialist. Without costly nutrient supplementation, we achieved titers up to 1.86 g/L and yields up to 62% of the theoretical maximum, which represent the highest reported to date for conversion of cellulosic substrates to next-generation biofuels. While we offer isobutanol production as a proof-of-concept application, the modularity of our system allows it to be readily adapted to the large portfolio of existing metabolically engineered *E. coli* strains to produce a wide range of biofuels or commodity chemicals.

In the next three sections, we provide a brief background on microbial cellulose hydrolysis and cellulolytic microbes, and discuss the selection of our model *T. reesei-E. coli* (TrEc) system.

2.2.1 Microbial cellulose hydrolysis

Lignocellulose is highly recalcitrant to degradation, hindering biological utilization. Natural cellulolytic microbes produce sophisticated enzyme systems called cellulases and hemicellulases that effectively hydrolyze cellulose and hemicellulose, respectively [14]. Microbial cellulase systems are complex, and exhibit a great deal of variation between different organisms [14]. In all cases, cellulase systems consist of a set of enzymes that work cooperatively to hydrolyze cellulose and/or hemicellulose [14]. The most important enzyme components of cellulase systems are endoglucanases, exoglucanases, and β-glucosidases (Figure 2.2). Endoglucanases bind to insoluble cellulose (> 4 glucose units for crystalline cellulose) and hydrolyze internal $\beta(1 \rightarrow 4)$ bonds, liberating two shorter cellulose fragments [14] (Figure 2.2). Exoglucanases processively hydrolyze sugar units (usually glucose or cellobiose, a $\beta(1 \rightarrow 4)$ disaccharide) from the ends of cellulose polymers [14] (Figure 2.2). β-glucosidases hydrolyze soluble cellulose oligosaccharides to yield glucose [14] (Figure 2.2). Organisms usually produce many different variants of each enzyme type. There is a great deal of synergy and cooperativity between the individual components of cellulase systems, and oftentimes the overall hydrolysis rate depends on the ratio of each component in the system (i.e. maximum hydrolysis rate is achieved at an optimal ratio)[14].

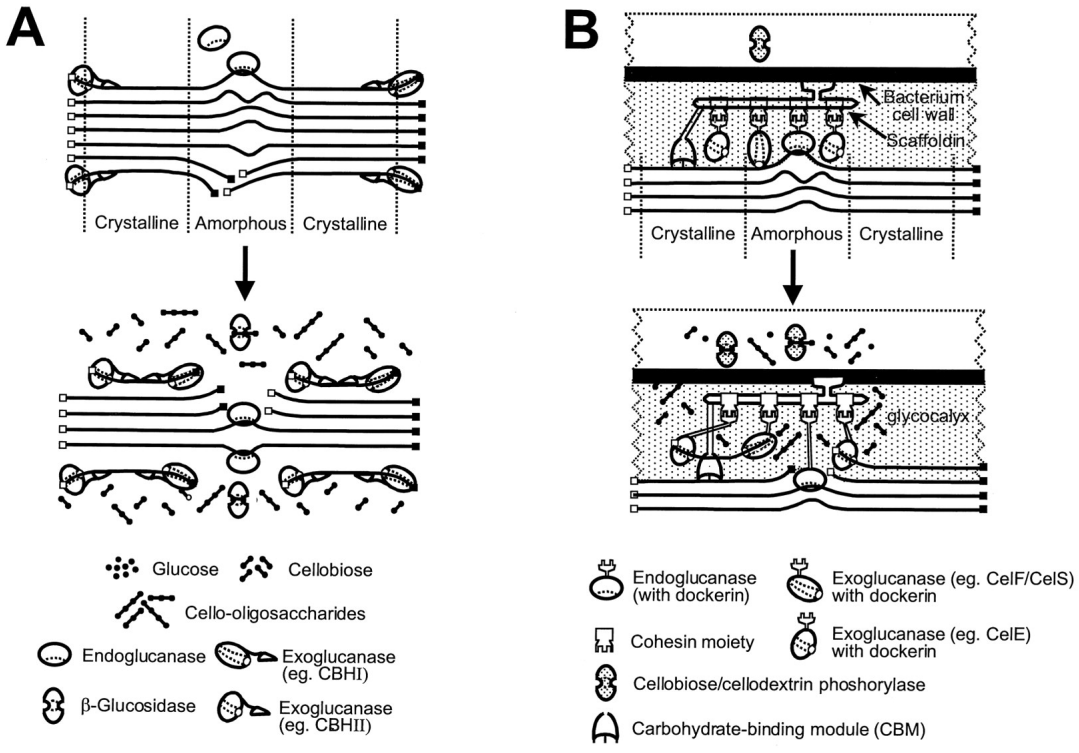

Figure 2.2: Microbial cellulase systems. Generalized depiction of **(A)** non-complexed and **(B)** complexed cellulase systems. The solid squares represent reducing ends, and the open squares represent non-reducing ends. Adapted from [14].

Microbial cellulase systems can be broadly classified as either non-complexed (Figure 2.2A) or complexed (Figure 2.2B) systems. In non-complexed cellulase systems, which are common among aerobic cellulolytic organisms, free cellulase enzymes are secreted from the cell into the external environment [14]. In complexed cellulase systems, cellulase enzymes are linked together via a scaffolding protein that remains attached to the cell surface, forming a complex known as a cellulosome [14]. In general, complexed cellulase systems have a higher specific activity (rate constant per unit mass protein) than non-complexed cellulase systems [14]. Microbes with complexed cellulase systems are thought to have selective access to cellulose hydrolysis products, due to the proximity of the cell to the substrate surface (Figure 2.2B) [14]. The mechanisms and kinetics of cellulase systems are not yet fully resolved, and still represent an active area of research. There is yet more uncertainty in the understanding of cellulose hydrolysis at the microbial level vs. the enzymatic level [15]. It is widely observed that cellulose hydrolysis rates are much higher in the presence of cellulolytic microbes compared to hydrolysis rates observed with isolated cellulase components, especially for organisms with complexed cellulase systems [14]. Furthermore, although complexed cellulase systems have a higher specific activity than non-complexed systems, organisms producing complexed cellulase systems are not generally superior cellulose utilizers. Complexed cellulase systems are typically produced by anaerobic organisms, and protein production is limited in such organisms due to the low energy conversion efficiency of anaerobic metabolism (i.e. ATP production via substrate level phosphorylation) [14]. In contrast, aerobic metabolism affords much higher energy yield (ATP production via oxidative phosphorylation), and consequently protein yield from substrate is much higher for aerobic metabolism [14, 9]. The higher production rates and yields of cellulase enzymes in aerobic organisms probably compensates for the lower specific activity of non-complexed cellulase systems. Indeed, at an ecological level, it has been repeatedly observed that cellulose hydrolysis proceeds much more rapidly in aerobic environments compared to anaerobic conditions [73].

2.2.2 Selection of the cellulolytic consortium member

Due to the inherent difficulties in expressing and secreting complex cellulase systems in non-native hosts (reviewed in Chapter 1), we propose selecting a native cellulase producing organism as the base strain for a cellulolytic consortium member. Selection of a base species for the cellulolytic consortium member requires careful consideration of organism physiology as well as engineering requirements. A primary criterion is efficient cellulase production and overall rapid cellulose degradation with minimal nutrition requirements. Other important criteria that need to be accounted for include physiological and ecological compatibility with the fermentation consortium member, and availability of genome sequences and molecular biology tools to facilitate further

engineering of the cellulolytic specialist (which may be necessary in later development stages of the consortium technology). Physiological compatibility is a key determinant in the overall consortium design - the consortium members need to grow well and perform their specialized tasks in the same environment, and therefore must be compatible in terms of preferences for pH, environmental redox state, temperature, nutrients, etc. The specialist organisms must also be compatible from an ecological standpoint (e.g. neutral or symbiotic interactions with minimal antagonism). Ideally, both consortium members should be tolerant to anticipated environmental stresses (including inhibitors present in the feedstock, and accumulation of toxic biofuel products, such as isobutanol). The ideal cellulolytic consortium member would meet all of the above criteria for productivity, compatibility, and stress tolerance requirements, however such an ideal organism may not exist in nature. For the cellulolytic specialist, we will therefore focus on selecting a relatively well-characterized organism (with available molecular biology tools), which meets our productivity/compatibility criteria to the greatest extent possible; if needed, such an organism could be engineered for further improvement of other properties (e.g. stress tolerance).

Energy metabolism, a key issue concerning the consortium culture environment, merits further discussion. Broadly speaking, microbial energy metabolism can be divided into anaerobic metabolism, aerobic metabolism, and microaerobic metabolism [9]. Anaerobic metabolism involves the production of ATP via substrate level phosphorylation, resulting in low ATP yield per unit substrate and the concomitant production of reduced terminal metabolites (anaerobic oxidative phosphorylation utilizing alternative electron acceptors, such as nitrate, is also possible but less relevant to the proposed microbial system) [9]. In contrast, aerobic metabolism is characterized by ATP generation via oxidative phosphorylation (utilizing oxygen as a terminal electron acceptor), resulting in high ATP and biomass yield from substrates and production of CO_2 and H_2O as metabolic byproducts [9]. Microaerobic metabolism varies between organisms and often exists as a superposition of aerobic and anaerobic metabolic states.

Many biofuels are produced via anaerobic metabolic pathways. For instance, ethanol is produced as the terminal metabolite of glucose fermentation via the Embden-Meyerhof-Parnas pathway of *S. cerevisiae* [9]. It is tempting to conclude that an anaerobic environment will be essential for the proposed microbial consortium to work efficiently. However, many promising next-generation biofuels and commodity chemicals (such as isoprenoids and fatty-acid derived hydrocarbons) are produced through biosynthetic non-fermentative pathways with substantial ATP and NADPH requirements, and thus it may actually be more efficient to produce these products aerobically. Furthermore, a strictly anaerobic consortium may have some fundamental disadvantages regarding cellulose hydrolysis and product yields. The low energy efficiency of

anaerobic metabolism limits protein production (and thus production of cellulase/cellulosomes) from substrates. In contrast, cellulase yield from substrate is much higher under aerobic conditions due to the high ATP yield from oxidative phosphorylation. The analysis of anaerobic vs. aerobic cellulose hydrolysis is obfuscated by the fact that complexed cellulase systems typical of anaerobic organisms usually have a much higher specific activity than the non-complexed cellulase systems of aerobic organisms [14]; this factor can potentially ameliorate the effects of decreased protein yield under anaerobic conditions. There are conflicting opinions in literature as to whether there are differences in aerobic or anaerobic microbial cellulose hydrolysis rates [14, 15, 73]; however, as discussed previously, it is widely observed in nature that cellulose degradation proceeds more rapidly in aerobic environments compared to anaerobic environments. In principle, the highest microbial cellulose hydrolysis rates could be achieved by maximizing both production of cellulases and maximizing the catalytic activity of cellulase systems. Such an approach could involve heterologous expression, secretion, and assembly of a highly active anaerobic cellulosome by an aerobic organism, or protein engineering of aerobic cellulase systems to maximize specific activity; this strategy may be feasible in the long term, but there has been only limited progress to date.

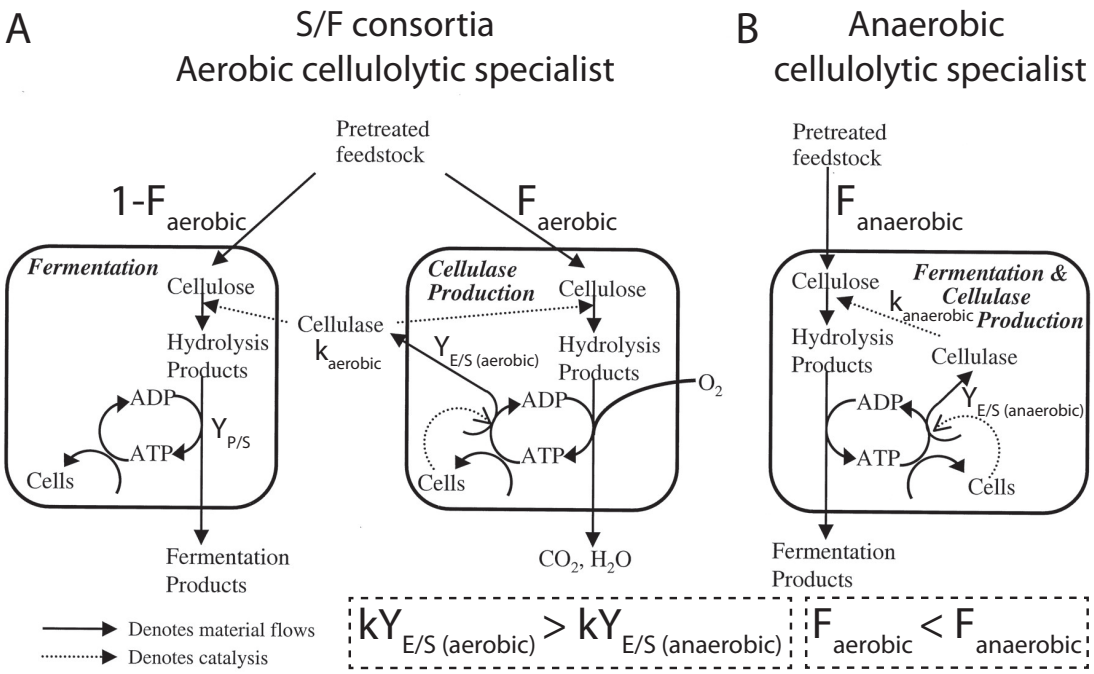

Figure 2.3: Summary of material flows and bioenergetics for CBP with **(A)** saccharolytic/fermentation (S/F) consortia with an aerobic cellulolytic specialist or **(B)** with anaerobic cellulolytic specialist. The term F denotes the fraction of feedstock consumed by the cellulolytic specialist, $Y_{E/S}$ is the cellulase-substrate yield coefficient (g-enzyme/g-substrate consumed), $Y_{P/S}$ is the product-substrate yield coefficient (g-product/g-substrate consumed), and k is the cellulase specific rate constant (IU/g-enzyme). Adapted from [14].

In addition to potentially lower cellulose hydrolysis rates, the lower cell/protein yields of anaerobic metabolism could also reduce biofuel/chemical product yields relative to an aerobic consortium. In S/F consortia, a certain fraction of the feedstock will ultimately be consumed by the saccharolytic specialist and used to produce cellulases with a yield of $Y_{E/S}$ (g-enzyme/g-substrate consumed), while the remainder will be consumed by the fermentation specialist and converted into desired biofuel or chemical products with a yield of $Y_{P/S}$ (g-product/g-substrate consumed) (Figure 2.3). To create an approximate basis for comparison, let us assume that a certain minimum mean cellulase activity level will be required for economically viable CBP (e.g. to achieve a target cellulose hydrolysis rate). The cellulase activity produced per unit mass of substrate consumed by the cellulolytic specialist is then $kY_{E/S}$, where $Y_{E/S}$ is as described previously and k is the cellulase specific rate constant (IU/g-enzyme). k is usually higher for complexed cellulase systems of anaerobic organisms compared to non-complexed cellulase systems of aerobic microbes; however $Y_{E/S}$ is much lower for anaerobic metabolism, and as a result $kY_{E/S}$ is typically lower for anaerobic cellulolytic microbes compared to aerobic microbes. For example, $kY_{E/S}$ for the well-known anaerobic cellulolytic microbe *C. thermocellum* is approximately 85% lower than that of *T. reesei* (values calculated from [15, 74]); thus to achieve the same cellulase activity, the fraction F of feedstock consumed by *C. thermocellum* would be approximately 6.7 times that of *T. reesei* (Figure 2.3). Product yields from the fermentation specialist are thus expected to be lower in an anaerobic S/F consortium relative to an aerobic system, due to the increased fraction of feedstock that needs to be consumed by the cellulolytic specialist. However, since the saccharolytic and fermentation specialists are competing for soluble hydrolysis products, the partition of substrate flow between the two organisms will ultimately depend on the substrate uptake kinetics for each microbe. Finally, we wish to note that in a CBP scheme using a single anaerobic cellulolytic "super-bug" (e.g. ethanol production with *C. phytofermentans* or *C. thermocellum*), substrate partition is not an issue since the desired biofuel or chemical products are produced by the cellulolytic organism itself [14].

While an aerobic consortium has potential advantages in terms of cellulose hydrolysis rates and partition of substrate flow between cellulolytic and fermentation specialists, aerobic bioprocessing may incur higher operating costs compared to an anaerobic process due to the requirement for aeration and agitation. These increased costs may be fully or partially offset by improved productivity and yields, although a detailed technoeconomic analysis needs to be done to fully evaluate these different bioprocessing options. Another major caveat is that the use of an aerobic organism for the cellulolytic consortium member conflicts with the requirement of fundamentally anaerobic metabolism for many biofuel and chemical products of interest. We wish to point out that there possible solutions to this dilemma, namely engineering biofuel

producing organisms to express an anaerobic metabolic phenotype under aerobic conditions. As an example, this could be implemented by removing key steps in the respiratory chain of a facultative anaerobe, thus disabling aerobic metabolism [75]. This approach was recently demonstrated in *E. coli* by knocking out the terminal cytochrome oxidases of the respiratory chain, and the resulting strains exhibited fermentative metabolism under aerobic conditions [75]. Overall, based on the considerations discussed in the preceding paragraphs, we think that utilizing an aerobic cellulolytic consortium member is a promising approach to maximizing cellulose hydrolysis and and achieving optimal productivity. However, as a contingency to the proposed aerobic bioprocessing platform, we will also consider anaerobic organisms as candidates.

For the fermentation specialist, we proposed using *Escherichia coli* as it is a well characterized model organism that has been metabolically engineered to produce a wide spectrum of biofuels and other valuable chemical products. *E. coli* grows well over a range reasonable of pH and temperature values, and can be feasibly adapted to pH/temperature conditions outside the wild-type optimal range; therefore use of this organism does not pose overly stringent environmental constraints on the selection of the cellulolytic consortium member. After our extensive considerations and a thorough literature review, the following organisms were identified as prime candidates for the cellulolytic consortium specialist: *Trichoderma reesei* RUT-C30 (aerobic), *Cytophaga hutchinsonii* (aerobic), and *Clostridium cellulolyticum* (anaerobic).

Each of the proposed organisms meets the criteria of being a productive and prototrophic cellulase producer, having a sequenced genome, available molecular biology tools, and anticipated environmental compatibility with *E. coli*. *T. reesei* is a mesophilic aerobic fungus originally isolated from decomposing organic matter in the South Pacific [14, 76]. *T. reesei* is notable for its aggressive growth characteristics and native ability to secrete large quantities of non-complexed cellulases, and is the most commonly used organism for industrial production of cellulases [14]. *T. reesei* is a model organism for cellulase production and has been subject to intense study for over 50 years [14]. The RUTC30 variant is a hypercellulolytic mutant capable of producing up to 0.25 g cellulase/g substrate consumed [77]. *C. hutchinsonii* is a gram-negative aerobic bacterium isolated from soil that reportedly rapidly degrades cellulose [78]. *C. hutchinonii* possesses interesting mechanisms of motility that aid in colonizing insoluble lignocellulosic substrates and may have a novel cellulase system [78]. *C. cellulolyticum* is a gram positive anaerobic mesophilic soil organism [21]. The *C. cellulolyticum* cellulosome has been intensely investigated and serves as a model system for mesophilic clostridial cellulosomes [21]. Clostridium cellulolyticum produces organic acids, ethanol, and hydrogen as fermentation products [21].

Out of the enumerated candidates, *T. reesei* RUTC30 appears to be the most promising candidate. This organism is well known for its cellulolytic abilities, and has the most extensively investigated cellulase system of all model organisms [76, 79]. *T. reesei* RUTC30 has increased resistance to alcohol toxicity compared to the WT strain, which will be of great benefit for the production of alcohol biofuels (e.g. isobutanol) [77]. The optimal temperature for *T. reesei* is between 25°C and 30°C, fully compatible with *E. coli*. *T. reesei* grows best at pH 4 to 5.5, which is lower than the optimal pH for *E. coli* (approximately 7); however, given that the pH preferences of these organisms are not too far apart, it is quite probable that co-cultures of *T. reesei* and *E. coli* could grow well at a consensus pH between the two optima (e.g. pH 6). Both *T. reesei* and *E. coli* are prototrophic, requiring only mineral salts, a nitrogen source, phosphorus source, and carbon source for growth; based on typical media recipes for these organisms, a low-cost fermentation medium suitable for both microbes could easily be devised. In terms of ecological compatibility, it is somewhat unorthodox to attempt co-culturing filamentous fungi and bacteria, given that many fungi are known to produce anti-bacterial metabolites. However, it has been demonstrated that *T. reesei* RUTC30 is not antagonistic towards gram negative bacteria [80].

Having discussed the selection of *T. reesei* RUTC30 as the cellulolytic consortium member, in the next section we provide a brief background on isobutanol production with *E. coli*, the selected fermentation specialist.

2.2.3 Isobutanol production with recombinant *E. coli* strains

For the fermentation specialist, we proposed using *Escherichia coli* since it is a well characterized model organism that has been metabolically engineered to produce a wide range of biofuels and other valuable chemical products. As a proof-of-concept demonstration, we will deploy the TrEc consortium with *E. coli* strains metabolically engineered to produce isobutanol, which is considered among the most promising next-generation biofuels. We will utilize isobutanol producing *E. coli* strains JCL260, NV3, and NV3r1 for the TrEc consortium, which are described briefly in this section. The James Liao lab at UCLA has recently engineered *E. coli* for isobutanol production via decarboxylation and reduction of 2-ketoisovalerate, an endogenous intermediate in valine biosynthesis [1, 3] (Figure 2.4). This was accomplished through heterologous expression of a branched chain 2-keto acid decarboxylase (*kivd* from *Lactococcus lactis*) and a broad substrate specificity alcohol dehydrogenase (ADH2 from *S. cerevisiae*) (Figure 2.4) [1] . To maximize isobutanol yield, genetic manipulations were carried out to increase metabolic flux towards 2-ketoisovalerate, including overexpression of highly specific acetolactate synthase *alsS* from

Bacillus subtilis (representing the first step in valine biosynthesis), overexpression of the valine biosynthesis genes *ilvCD*, and gene deletions (Δ*adhE*, Δ*ldhA*, Δ*frdAB*, Δ*fnr*, Δ*pta*, and Δ*pflB*) to eliminate flux to competing pathways (Figure 2.4) [1]. The resulting strain, JCL260 pSA55/69, is capable of converting glucose to isobutanol at high yield (86% of theoretical maximum) and titer (22 g/L).

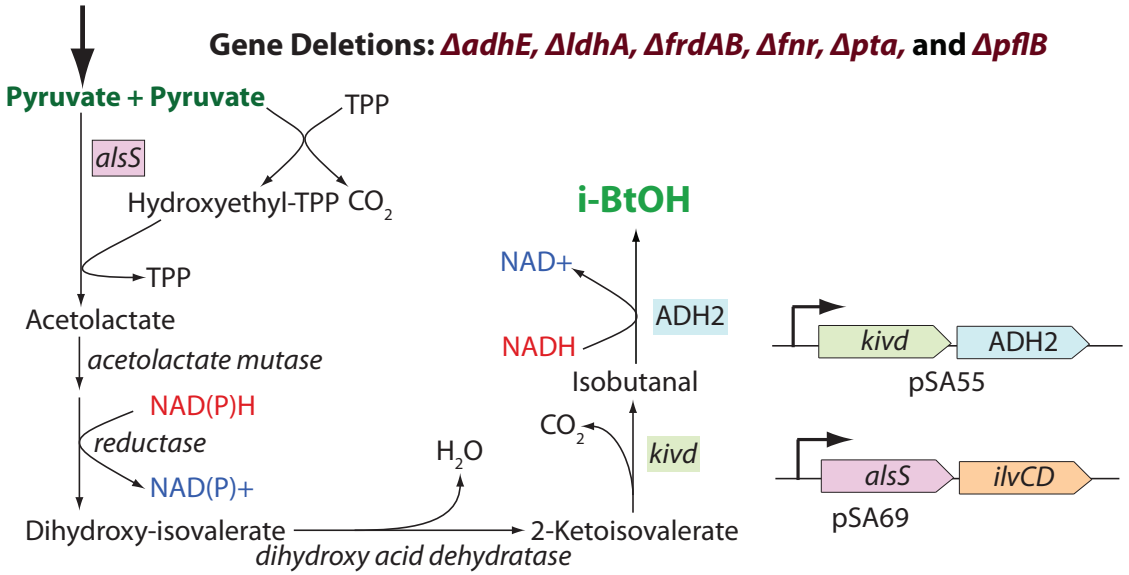

Figure 2.4: Engineering *E. coli* for isobutanol production. Metabolic pathway for isobutanol production and genetic manipulations used to create isobutanol producing strain JCL260 pSA55/69 [1].

In more recent work, the Liao lab used an evolutionary method as opposed to a rational design strategy to develop isobutanol producing strains of *E. coli* [3]. By applying repeated iterations random mutagenesis and selecting mutants with the ability to grow on glucose media with norvaline (a toxic valine analog), strains with enhanced isobutanol production (relative to the parental strain) were isolated [3]. The best-performing isolate, strain NV3, had a higher growth rate and reached a higher final cell concentration than other isobutanol producing strains, including JCL260 [3]. However, isobutanol titers for NV3 were much lower compared to JCL260 (13.6 g/L vs. 22g/L) [3]. Genome re-sequencing of NV3 was performed to investigate the genetic basis of isobutanol production in this evolved strain. Sequencing revealed a total 208 mutations, of which 62 were synonymous single nucleotide polymorphisms (SNPs) in open reading frames (ORFs). Mutations were distributed across many different genes and genetic loci, representing a wide array of cellular processes including metabolism and gene regulation [3]. A putative loss-of-function mutation was identified in master stress response regulator *rpoS* [3]. Repair of the *rpoS* mutation (strain NV3r1) increased isobutanol titer from 13.6 g/L to 21.2 g/L, comparable to rationally engineered strain JCL260 [3]. Since isobutanol is the product of

non-fermentative metabolism it can be produced in in the presence of oxygen, and for all three of the strains discussed here (JCL260, NV3, NV3r1), production is optimal under microaerobic conditions. Thus isobutanol production should be fully compatible (though not necessarily optimal) with the aerobic growth requirements of *T. reeesei* RUTC30, the selected cellulolytic member.

Having provided background and discussion on the selection of our model *T. reesei-E. coli* (TrEc) consortium, in the next section we present the results of theoretical and experimental studies of the TrEc consortium, beginning with an overview of our modeling framework.

2.3 Results

2.3.1 Model summary

To gain insights into the behavior and ecology of the *T. reesei* / *E. coli* (TrEc) consortium, we developed a comprehensive ordinary differential equation (ODE) modeling framework that captures salient features of the system. We derived rate expressions for microbial growth, uptake of soluble saccharides, production of cellulase enzymes (endoglucanase, exoglucanase, and β-glucosidase) by *T. reesei*, cellulose hydrolysis based on novel mechanistic models for each cellulase, isobutanol production by *E. coli*, and isobutanol toxicity (Figure 2.5A). An important subtlety is that soluble oligosaccharides are hydrolyzed to glucose via cell-wall localized β-glucosidases of *T. reesei* (Figure 2.5A) [81]. This leads to higher glucose concentration at the *T. reesei* cell surface compared to the bulk media, which we estimate using a mass transfer analysis (Figure 2.5A). The model was developed by writing differential mole/mass balances for each species of interest (for a batch reactor), including microbial biomass (*T. reesei* and *E. coli*), cellulase enzymes, insoluble cellulose polysaccharides, soluble oligo and monosaccharides, and isobutanol. Our model explicitly accounts for each possible cellulose saccharide S_{G_i}, where i is the number of glucose monomers. Thus the total number of ODEs will depend on the degree of polymerization (DP) distribution of S_{G_i}, which varies between different cellulosic substrates. For a given S_{G_i} distribution, a total of $5 + DP_{max}$ ODEs are required, where DP_{max} is the maximum DP of the cellulosic substrate. Parameter values were obtained from literature, with certain values experimentally measured in-house (See Tables B.1 to B.3).

We provide a brief summary of our modeling framework in this section. **See Appendix A for a full description of our model, including mass transfer analysis of the *T. reesei* cell surface, rate law expressions and derivations, and discussion of semi-mechanistic enzymatic cellulose**

hydrolysis models. For details regarding implementation of the model in MATLAB, see section 2.5.1.1. Note: equations numbers in this section correspond to those in Appendix A.

2.3.1.1 *T. reesei* model

***T. reesei* growth** *T. reesei* is a multicellular filamentous fungus that has different mycelial growth states. Vegetative growth and enzyme secretion are highly active at hyphal tips, while senescent mycelium is relatively dormant [82]. Assuming that growth at hyphal tips follows Monod kinetics and that *T. reesei* is capable of simultaneous utilization of multiple soluble sugars (i.e. glucose and soluble glucose oligosaccharides), *T. reesei* growth in the presence of isobutanol can be described with a segregated kinetic model:

$$\frac{dC_{Tr,v}}{dt} = \mu_{Tr} C_{Tr,v} - k_{v\to s} C_{Tr,v} \tag{A.40}$$

$$\frac{dC_{Tr,s}}{dt} = k_{v\to s} C_{Tr,v} - k_{Tr,d} C_{Tr,s} \tag{A.41}$$

In the first expression, $C_{Tr,v}$ is the vegetative mycelium concentration (g/L), μ_{Tr} is a generalized Monod function (1/h) depending on isobutanol concentration I (g/L) and concentration of soluble glucose saccharides S_{G_i} (g/L; i is the degree of polymerization), and $k_{v\to s}$ is the specific rate of conversion of vegetative mycelium to senescent mycelium (1/h). In the second expression, $C_{Tr,s}$ is the concentration of senescent mycelium (g/L) and $k_{Tr,d}$ is the specific death rate of the senescent mycelium (1/h). Our formulation of μ_{Tr} accounts for isobutanol toxicity and incorporates models for maintenance substrate uptake at low concentrations (derived in section A.2.1) and increased soluble saccharide concentrations at the *T. reesei* cell surface due to cell-wall localized β-glucosidase (derived in section A.2.2):

$$K_{Tr}^{I} = \begin{cases} \left(1 - \frac{I}{I_{Tr}^{*}}\right)^{n_{Tr}} & \text{if } I \le I_{Tr}^{*} \\ 0 & \text{if } I > I_{Tr}^{*} \end{cases} \tag{A.42}$$

$$\mu_{Tr} = K_{Tr}^{I} \sum_{i} p_{S_{G_i}} \left[\left(\mu_{max,Tr,S_{G_i}} + \frac{m_{Tr}}{Y_{S_{G_i}/C_{Tr}}} \right) \frac{S_{G_i} + \sum_{k=i+1}^{4} \theta_{k\to i} S_{G_k}}{K_{Tr,S_{G_i}} + S_{G_i} + \sum_{k=i+1}^{4} \theta_{k\to i} S_{G_k}} - \frac{m_{Tr}}{Y_{S_{G_i}/C_{Tr}}} \right] \tag{A.43}$$

$$\text{with } p_{S_{G_i}} = \frac{S_{G_i}}{\sum_{j} S_{G_j}}$$

Where K_{Tr}^{I} is an empirical inhibition function (dimensionless) [83], I is isobutanol concentration (g/L), I_{Tr}^{*} is the growth inhibiting concentration of isobutanol (g/L) for *T. reesei*, n_{Tr} is an empirically determined exponent, $\mu_{max,Tr,S_{G_i}}$ is maximum specific growth rate on substrate i (1/h), $p_{S_{G_i}}$ is the proportion of substrate i in the total substrate concentration ($S_{G_i}/\sum S_{G_k}$), S_{G_i} is substrate i concentration (g/L), $K_{Tr,S_{G_i}}$ is substrate i affinity (g/L), the coefficients $\theta_{k\to i}$ reflect increased soluble saccharide concentrations at the cell surface due cell-wall localized β-glucosidase (described in section A.2.2), m_{Tr} is the maintenance coefficient (g-substrate/g-biomass/h), and $Y_{S_{G_i}/C_{Tr}}$ is the substrate/biomass yield coefficient for substrate i (g-substrate/g-biomass). We assume that growth occurs via utilization of multiple substrates simultaneously, as opposed to diauxic substrate utilization. Available experimental data suggests that this is a reasonable assumption for *T. reesei*, especially the RUTC30 strain, which contains a loss-of-function mutation in catabolite repression gene *cre1* [84]. Our model assumes a total substrate maintenance requirement m_{Tr} rather than an individual maintenance term for each substrate i; this is reasonable for substrates with similar metabolism / energy yields (e.g. glucose and cellobiose), but could be revised for more diverse substrates.

***T. reesei* enzyme secretion** Assuming that enzyme secretion is stoichiometrically coupled to growth and that composition of secreted enzymes is constant, the following expression can be derived for cellulase production:

$$\frac{dE_T}{dt} = Y_{E_T/C_{Tr}}\mu_{Tr}C_{Tr,v} + k_{E_T}C_{Tr,s} \tag{A.44}$$

$$\frac{dE_i}{dt} = x_{E_i}\frac{dE_T}{dt} = x_{E_i}\left[Y_{E_T/C_{Tr}}\mu_{Tr}C_{Tr,v} + k_{E_T}C_{Tr,s}\right] \tag{A.45}$$

Where E_T is the total concentration of secreted enzymes (g/L), $Y_{E_T/C_{Tr}}$ is the enzyme/biomass yield coefficient (g-protein/g-biomass), k_{E_T} is the specific enzyme production rate of senescent mycelium (g-protein/g-biomass/h), E_i is concentration of enzyme i (g/L), x_{E_i} is the fraction of enzyme i in the total secretome (E_i/E_T), and the other terms are as described in previous sections. *T. reesei* produces a large suite of biomass degrading enzymes, but for the purpose of our cellulose hydrolysis model, we consider the most important enzymes [14]:

- cellobiohydrolase 1 ($CBH1$) and 2 ($CBH2$)
- endoglucanase 1 ($EG1$)
- β-glucosidase 1 (BGL)

***T. reesei* saccharide uptake** Assuming that saccharide uptake is stoichiometrically coupled to growth of vegetative mycelium and that both vegetative and senescent mycelia consume saccharides for maintenance, the following expression for saccharide uptake by *T. reesei* can be derived:

$$r_{S_{G_i}}^{Tr} = \left[Y_{S_{G_i}/C_{Tr}} K_{Tr}^{I} p_{S_{G_i}} \mu_{max,Tr,S_{G_i}} + m_{Tr}\right] \frac{S_{G_i}}{K_{Tr,S_{G_i}} + S_i} C_{Tr,v} + m_{Tr} p_{S_{G_i}} \frac{S_{G_i}}{K_{Tr,S_{G_i}} + S_{G_i}} C_{Tr,s} \tag{A.46}$$

Where $r_{S_{G_i}}^{Tr}$ is the total rate of saccharide i uptake by *T. reesei* (g/L/h), and all other terms are as described previously.

2.3.1.2 *E. coli* model

***E. coli* growth** We model *E. coli* growth with Monod kinetics [83], assuming that only glucose is utilized for growth (i.e. glucose oligosaccharides cannot be metabolized):

$$\frac{dC_{Ec}}{dt} = \left(\mu_{Ec} - k_{Ec,d}\right) C_{Ec} \tag{A.47}$$

Where C_{Ec} is *E. coli* concentration (g/L), μ_{Ec} is specific growth rate (1/h), and $k_{Ec,d}$ is the specific cell death rate (1/h). Our expression for μ_{Ec} accounts for inhibition from isobutanol and incorporates a model for maintenance substrate uptake at low concentrations (derived in section A.2.1):

$$K_{Ec,S_{G_1}}^{I} = \begin{cases} \left(1 - \frac{I}{I_{Ec,S_{G_1}}^{*}}\right)^{n_{Ec,S_{G_1}}} & \text{if } I \leq I_{Ec,S_{G_1}}^{*} \\ 0 & \text{if } I > I_{Ec,S_{G_1}}^{*} \end{cases} \tag{A.48}$$

$$\mu_{Ec} = K_{Ec,S_{G_1}}^{I} \left[\left(\mu_{max,Ec,S_{G_1}} + \frac{m_{Ec,S_{G_1}}}{Y_{S_{G_1}/C_{Ec}}}\right) \frac{S_{G_1}}{K_{Ec,S_{G_1}} + S_{G_1}} - \frac{m_{Ec,S_{G_1}}}{Y_{S_{G_1}/C_{Ec}}}\right] \tag{A.49}$$

Where $K_{Ec,S_{G_1}}^{I}$ is an empirical inhibition function (dimensionless) [83], I is isobutanol concentration (g/L), $I_{Ec,S_{G_1}}^{*}$ is the growth inhibiting concentration of isobutanol (g/L) for *E. coli*, $n_{Ec,S_{G_1}}$ is an empirically determined exponent, $\mu_{max,Ec,S_{G_1}}$ is maximum specific growth rate of *E. coli* on glucose (1/h), $K_{Ec,S_{G_1}}$ is glucose affinity (g/L), $Y_{S_{G_1}/C_{Ec}}$ is the glucose/biomass yield coefficient (g-substrate/g-biomass), and $m_{Ec,S_{G_1}}$ is the maintenance coefficient (g-substrate/g-biomass/h).

***E. coli* saccharide uptake** Substrate uptake is assumed to be stoichiometrically coupled to growth. Additionally, experimental data for *E. coli* demonstrates non-growth associated substrate

uptake (i.e. during stationary phase) for maintenance / isobutanol production [1]. We then model uptake of glucose as follows:

$$r_{S_{G_1}}^{Ec} = Y_{S_{G_1}/C_{Ec}} K_{Ec,S_{G_1}}^{I} \frac{\mu_{max,Ec,S_{G_1}} S_{G_1}}{K_{Ec,S_{G_1}} + S_{G_1}} C_{Ec} + m_{Ec,S_{G_1}} \frac{S_{G_1}}{K_{Ec,S_{G_1}} + S_{G_1}} C_{Ec} \tag{A.50}$$

Where $r_{S_{G_1}}^{Ec}$ is the rate of total glucose uptake by *E. coli* (g/L/h), and the other terms are as described previously.

***E. coli* isobutanol production** Unlike many metabolic products, isobutanol production is not stoichiometrically coupled to growth, since substantial isobutanol production is observed during stationary phase [1]. To account for this, all consumed substrates, both for growth and maintenance, will be assumed to be converted to isobutanol. For generality, we allow yield coefficients to vary between growth and non-growth associated substrate uptake:

$$\frac{dI}{dt} = Y_{I/S_{G_1}}^{growth} K_{Ec,S_{G_1}}^{I} \frac{\mu_{max,Ec,S_{G_1}} S_{G_1}}{K_{Ec,S_{G_1}} + S_{G_1}} C_{Ec} + Y_{I/S_{G_1}}^{maint} m_{Ec,S_{G_1}} \frac{S_{G_1}}{K_{Ec,S_{G_1}} + S_{G_1}} C_{Ec} \tag{A.51}$$

Where $Y_{I/S_{G_1}}^{growth}$ is the growth associated isobutanol/glucose yield coefficient (g-iBtOH/g-substrate) and $Y_{I/S_{G_1}}^{maint}$ is the non-growth (maintenance) isobutanol/glucose yield coefficient (g-iBtOH/g-substrate). In the case of *E. coli* K12, both yield coefficients would be 0 (i.e. no isobutanol production).

2.3.1.3 Enzymatic cellulose hydrolysis: general framework

There are numerous models reported in literature for microbial growth on cellulose [14]. However, few of these models accounts for the hydrolysis of cellulose to soluble saccharides. Competition between the *E. coli* and *T. reesei* for soluble saccharides is a crucial ecological interaction that needs to be accounted for to accurately predict population dynamics, behavior, and isobutanol production in the TrEc consortium. As a starting point for developing mechanistic models of cellulose hydrolysis, we utilize the general framework for enzymatic cellulose hydrolysis proposed by [14] and [85], which we describe in the following sections. Additionally, we also include generalized soluble saccharide mole balances that describe rates of production/consumption due to enzymatic hydrolysis and microbial uptake.

Endoglucanase Endoglucanases adsorb at random to cellulose molecules and cleave them to release two shorter chain polysaccharides [85]. This mechanism can be represented as [85]:

$$S_{G_i} + E_{EGm} \overset{K_{dis}^{EGm}}{\longleftrightarrow} S_{G_i} \cdot E_{EGm} \xrightarrow{k_{EGm}} E_{EGm} + S_{G_{i-j}} + S_{G_j} \tag{A.52}$$

Where E_{EGm} is endoglucanase m, K_{dis}^{EGm} is the dissociation constant for endoglucanase m (mM bonds), k_{EGm} is the rate constant of adsorbed E_{EGm} (mmol bonds/g-EGm·S_{G_i}/h), i and j are cellulose chain lengths, with $1 \leq j < i$, $S_{G_i} \cdot E_{EGm}$ is adsorbed E_{EGm}, and other terms are as previously described. The rate of hydrolysis of saccharide S_{G_i} by endoglucanase m is then [85]:

$$r_{S_{G_i}}^{EGm} = -k_{EGm} \left[S_{G_i} \cdot E_{EGm}\right] \tag{A.53}$$

Where $r_{S_{G_i}}^{EGm}$ is hydrolysis rate (mM bonds/h) and $[S_{G_i} \cdot E_{EGm}]$ is the mass concentration of E_{EGm} adsorbed to S_{G_i} (g/L). Cellulose saccharides S_{G_i} can be formed from endoglucanase hydrolysis of longer cellulose molecules S_{G_k}, with $k > i$. The rate of hydrolysis of S_{G_k} to S_{G_i} is equal to the overall rate of hydrolysis of S_{G_k} times the fraction of hydrolysis events that lead to a chain length S_{G_i}, $f_{G_k \to G_i}$. If all glycosidic bonds are cleaved at an equal rate, then $f_{G_k \to G_i} = 2/\left(k-1\right)$, leading to the following [85]:

$$r_{S_{G_k} \to S_{G_i}}^{EGm} = f_{G_k \to G_i} r_{S_{G_k}}^{EGm} = -\frac{2}{k-1} k_{EGm} \left[S_{G_k} \cdot E_{EGm}\right] \tag{A.54}$$

The overall rate of formation of S_{G_i} by endoglucanase is then the sum of the rate of hydrolysis of S_{G_i} and the rate of formation of S_{G_i} from S_{G_k} with $k > i$ [85]:

$$r_{S_{G_i}}^{EGm} = -k_{EGm} \left[S_{G_i} \cdot E_{EGm}\right] - \sum_{k>i} f_{G_k \to G_i} r_{S_{G_k}}^{EGm} \tag{A.55}$$

$$= -k_{EGm} \left[S_{G_i} \cdot E_{EGm}\right] + \sum_{k>i} \frac{2}{k-1} k_{EGm} \left[S_{G_k} \cdot E_{EGm}\right] \tag{A.56}$$

Where the upper limit of the summation is implicitly understood as DP_{max} (maximum polysaccharide length i for given type of cellulose) and other terms are as described previously. We developed expressions for $[S_{G_i} \cdot E_{EGm}]$ in terms of measurable variables by incorporating enzyme mass balances and substrate binding site balances with the endoglucanase adsorption equilibria. **See section A.6.1 for derivation and final rate law expression.**

Exoglucanase In contrast to endoglucanases, exoglucanases (often referred to as cellobiohydrolases) bind to the ends of cellulose chains and processively hydrolyze cellobiose units. Mechanis-

tically, this can be represented as [85]:

$$S_{G_i} + E_{CBHm} \xleftrightarrow{K_{dis}^{CBHm}} S_{G_i} \cdot E_{CBHm} \xrightarrow{k_{CBHm}} E_{CBHm} + S_{G_{i-2}} + S_{G_2} \tag{A.57}$$

Where E_{CBHm} represents cellobiohydrolase m, K_{dis}^{CBHm} is the dissociation constant for cellobiohydrolase m (mM bonds), k_{CBHm} is the rate constant of adsorbed E_{CBHm} (mmol bonds/g-CBHm·S_{G_i}/h), $S_{G_i} \cdot E_{CBHm}$ is adsorbed E_{EGm}, and the other terms are as described previously. The rate of hydrolysis of saccharide S_{G_i} by cellobiohydrolase is then [85]:

$$r_{S_{G_i} \to S_{G_2}}^{CBHm} = -k_{CBHm} \left[S_{G_i} \cdot E_{CBHm} \right] \tag{A.58}$$

Where terms are similar to those described for endoglucanase. S_{G_i} can also be formed from cellobiohydrolase hydrolysis of $i{+}2$ chain length cellulose molecules. The overall rate of formation of S_{G_i} by cellobiohydrolase is then the sum of the rate of hydrolysis of S_{G_i} and the rate of formation of S_{G_i} from $i+2$ saccharides [85]:

$$r_{S_{G_i}}^{CBHm} = -k_{CBHm} \left[S_{G_i} \cdot E_{CBHm} \right] + k_{CBHm} \left[S_{G_{i+2}} \cdot E_{CBHm} \right] \tag{A.59}$$

Where terms are as described previously. The overall rate of formation of S_{G_2} is the sum of cellobiohydrolase hydrolysis rates for all saccharides S_{G_i} for $i \geq 3$:

$$r_{S_{G_2}}^{CBHm} = k_{CBHm} \sum_{i \geq 3} \left[S_{G_i} \cdot E_{CBHm} \right] \tag{A.60}$$

Where the upper limit of the summation is implicitly understood as DP_{max} and other terms are as described previously. Analogous to our endoglucanase deriviation, we developed expressions for $[S_{G_i} \cdot E_{CBHm}]$ in terms of measurable variables by incorporating enzyme mass balances and substrate binding site balances with exoglucanase adsorption equilibria. **See section A.6.2 for derivation and final rate law expression.**

β-glucosidase β-glucosidase hydrolyzes cellobiose and other soluble cellulose oligosaccharides to glucose [86]. For soluble saccharides of DP $i = 2..4$ this can be mechanistically represented as:

$$S_{G_i} + E_{BGLm} \xleftrightarrow{K_{G_i,dis}^{BGLm}} S_{G_i} \cdot E_{BGLm} \xrightarrow{k_{BGLm,G_i}} E_{BGLm} + S_{G_{i-1}} + S_{G_1} \tag{A.61}$$

Where E_{BGLm} represents β-glucosidase m, $K_{G_i,dis}^{BGLm}$ is the dissocation constant for β-glucosidase m (mM substrate), k_{BGLm,G_i} is the rate constant of adsorbed E_{BGLm} (mmol/g-BGLm·S_{G_i}/h), $S_{G_i} \cdot E_{BGLm}$ is the β-glucosidase / substrate complex, and the other terms are as described previously.

The rate of hydrolysis of S_{G_i} or formation of S_{G_1} is then:

$$r_{S_{G_i}}^{BGLm} = -r_{S_{G_1}}^{BGLm} = -k_{BGLm,G_i}\left[S_{G_i} \cdot E_{BGLm}\right] \tag{A.62}$$

For cellobiose, the coefficient before $r_{S_{G_1}}^{BGLm}$ would be -2 instead -1 since two glucoses are produced per cellobiose. We adopt a multi-substrate Michaelis-Menten rate law, which has been previously shown to satisfactorily describe β-glucosidase kinetics [86]:

$$r_{S_{G_i}}^{BGLm} = -\frac{k_{BGLm,G_i} E_{BGLm} S_{G_i}}{K_{M,G_i}^{BGLm}\left(1 + \frac{S_{G_1}}{K_{G_1}^{BGLm}} \sum_{i=2}^{4} \frac{S_{G_i}}{K_{M,G_i}^{BGLm}}\right)} \tag{2.1}$$

Where K_{M,G_i}^{BGLm} is the Michaelis constant for S_{G_i} (mM), $K_{G_1}^{BGLm}$ is the glucose inhibition term (mM), and other terms are as described in previous sections. The total rate of glucose production via β-glucosidase is then:

$$r_{S_{G_1}}^{BGLm} = -2r_{S_{G_2}}^{BGLm} - \sum_{i=2}^{4} r_{S_{G_i}}^{BGLm} \tag{2.2}$$

2.3.1.4 Saccharide mass balances

In general, saccharide mass balances must account for both enzymatic cellulose hydrolysis and microbial saccharide uptake. However, insoluble cellulose molecules (chain length $i > 4$) are not utilized biologically. Thus for chain length $i > 4$ cellulose molecules, net rates of formation from endoglucanase and cellobiohydrolase need only be considered. Writing a mass balance for each cellulose molecule of chain length i with $i > 4$ yields [85]:

$$\frac{dS_{G_i}}{dt} = \sum_m r_{S_{G_i}}^{EGm} + \sum_n r_{S_{G_i}}^{CBHn} \text{ for } i > 4 \tag{A.63}$$

Where all terms are as described previously. For the case of the *T. reesei* cellulase system, consisting of endoglucanase 1 ($EG1$), cellobiohydrolase 1 ($CBH1$), and cellobiohydrolase 2 ($CBH2$), the mass balances reduce to:

$$\frac{dS_{G_i}}{dt} = r_{S_{G_i}}^{EG1} + r_{S_{G_i}}^{CBH1} + r_{S_{G_i}}^{CBH2} \text{ for } i > 4 \tag{A.64}$$

Where all terms are as described previously. For soluble saccharides, microbial saccharide uptake and β-glucosidase hydrolysis must also be considered. For a co-culture of *T. reesei* and *E. coli*, writing a mass balance on cellulose molecules of chain length $1 \leq i \leq 4$ yields:

$$\frac{dS_{G_i}}{dt} = r^{EG1}_{S_{G_i}} + r^{CBH1}_{S_{G_i}} + r^{CBH2}_{S_{G_i}} + r^{BGL}_{S_{G_i}} - \frac{1}{MW_{S_{G_i}}}\left(r^{Tr}_{S_{G_i}} + r^{Ec}_{S_{G_i}}\right) \text{ for } 1 \leq i \leq 4 \quad \text{(A.65)}$$

Where $MW_{S_{G_i}}$ is the molecular weight of S_{G_i} (g/mmol), and all other terms are as described previously. Most *E. coli* strains cannot metabolize cellulose oligosaccharides and are thus only able to use S_{G_1}; additionally, while it seems biologically plausible, there is little evidence to support significant uptake and metabolism of $i > 2$ glucose saccharides by *T. reesei*. We thus reduce the $i \leq 4$ saccharide balances to:

$$\frac{dS_{G_i}}{dt} = r^{EG1}_{S_{G_i}} + r^{CBH1}_{S_{G_i}} + r^{CBH2}_{S_{G_i}} + r^{BGL}_{S_{G_i}} \text{ for } 3 \leq i \leq 4 \quad \text{(A.66)}$$

$$\frac{dS_{G_2}}{dt} = r^{EG1}_{S_{G_2}} + r^{CBH1}_{S_{G_2}} + r^{CBH2}_{S_{G_2}} + r^{BGL}_{S_{G_2}} - \frac{1}{MW_{S_{G_2}}} r^{Tr}_{S_{G_2}} \quad \text{(A.67)}$$

$$\frac{dS_{G_1}}{dt} = r^{EG1}_{S_{G_1}} + r^{CBH1}_{S_{G_1}} + r^{CBH2}_{S_{G_1}} + r^{BGL}_{S_{G_1}} - \frac{1}{MW_{S_{G_1}}}\left(r^{Tr}_{S_{G_1}} + r^{Ec}_{S_{G_1}}\right) \quad \text{(A.68)}$$

Where all terms are as described previously.

2.3.2 Theoretical analysis of a synthetic *T. reesei* / *E. coli* consortium

Our model contains 50 parameters and includes variables for concentration of microbial biomass (vegetative and senescent *T. reesei* mycelium, and *E. coli*), enzymes, isobutanol, soluble oligosaccharides (one to four glucose monomers in size), and each possible cellulose polysaccharide from degree-of-polymerization (DP) $= 5$ up to the maximum DP of the substrate. We performed a global sensitivity analysis to dissect this functionally complex model and identify key parameters controlling consortium behavior (see section 2.5.1.2). The ODEs were numerically integrated with 1000 sets of parameter values and initial conditions (ICs) sampled from appropriate statistical distributions (Table B.1), with latin hypercube [87] selection. For each parameter or IC, partial rank correlation coefficients (PRCC) [87] were calculated with a set of output metrics. The most significant parameters and ICs ($p > 0.05$ and $|PRCC| > 0.1$) are shown in Figure 2.5B, with the same parameters/ICs labeled in Figure 2.5A; full results shown in Figure C.2.

Hierarchical clustering reveals three important sets of parameters/ICs (Figure 2.5B, colored dendrograms). Two of the parameter/IC clusters control partition of carbon flow between *T. reesei* and *E. coli*; parameters/ICs in cluster #1 (light blue dendrograms, Figure 2.5B) tend to be correlated with increased proportion of carbon channeled to *E. coli* (positive PRCC for $P_{C->Ec}$),

while cluster #3 (green dendrogram, Figure 2.5B) tends to be correlated with increased proportion to *T. reesei* (negative PRCC for $P_{C->Ec}$). Parameters/ICs controlling $P_{C->Ec}$ also influence population composition (X_{Ec}) and isobutanol production ($Y_{I/S}$ and Q_I), as would be intuitively expected. The parameters/ICs in cluster #2 (orange dendrograms, Figure 2.5B) primarily control the overall rates of enzymatic hydrolysis (and subsequently microbial growth), but have less influence on $P_{C->Ec}$. Initial population composition (quantified as *E. coli* population fraction, $X_{Ec}(t_0)$) and parameters associated with microbial growth and saccharide uptake have the strongest effect on $P_{C->Ec}$. Substrate hydrolysis rate, the limiting step for microbial growth and isobutanol production, is highly correlated with F_a (fraction of substrate accessible to enzymes), various initial conditions (such as initial cellulose and microbial biomass concentrations), and enzyme kinetic parameters. In addition to parameters/ICs, we also clustered the output metrics. The similarity between R_{cel} and R_{Tr} reflects the dependence of cellulose hydrolysis rate on cellulase production by *T. reesei*, while the other cluster contains output metrics trivially related to *E. coli*.

Isobutanol yield $Y_{I/S}$ and productivity Q_I correlate the most strongly with initial population composition $X_{Ec}(t_0)$ (Figure 2.5B; topmost row). To further investigate the effect of $X_{Ec}(t_0)$ on isobutanol production, we examined solutions over a range of $X_{Ec}(t_0)$ (Figure 2.5C). $Y_{I/S}$ and Q_I increase with $X_{Ec}(t_0)$ while R_{cel} decreases with $X_{Ec}(t_0)$. These trends illustrate an inherent tradeoff between R_{cel} and $Y_{I/S}$ which arises from the competitive nature of the TrEc consortium. Our model predicts that increasing $X_{Ec}(t_0)$ increases carbon flow to *E. coli* leading to increased $Y_{I/S}$, while growth and subsequently cellulase synthesis by *T. reesei* are reduced, leading to lower R_{cel} and increased fermentation times (see Figure C.1C for R_{cel} vs. $Y_{I/S}$ over sensitivity analysis sampling points). In addition to the tradeoff between R_{cel} and $Y_{I/S}$, all three output metrics show asymptotic trends with respect to $X_{Ec}(t_0)$, approaching saturation values when $X_{Ec}(t_0) \gtrsim 0.01$ (Figure 2.5C). We note that this behavior is consistent with the reported $X_{Ec}(t_0)$ PRCC values, since rank transformation linearizes all monotonic relationships (example shown in Figure C.1A and B). The predicted asymptotic effect is due in part to the kinetics of microbial saccharide uptake; the locally increased glucose concentration at the *T. reesei* cell surface affords higher uptake by *T. reesei* compared to *E. coli* under conditions of low glucose concentration, thus limiting substrate flow to *E. coli*. Co-culture dynamics in a sample numerical solution (Figure 2.5D) demonstrate this effect, with *E. coli* reaching saturation long before *T. reesei*.

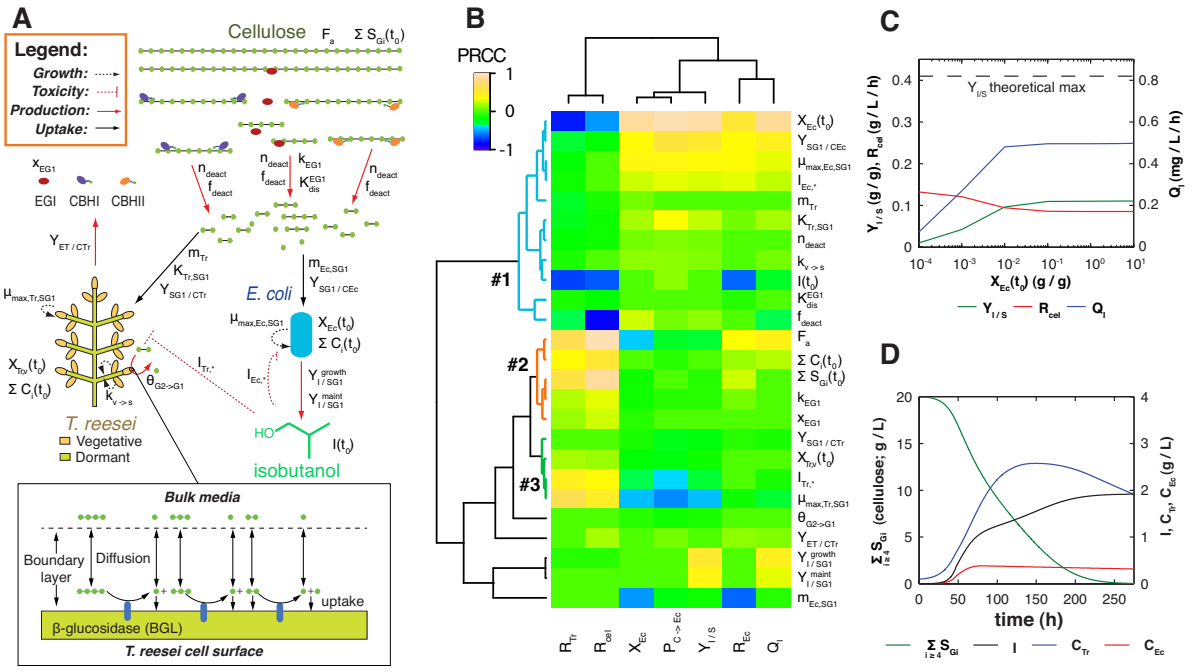

Figure 2.5: Design and theoretical analysis of the *T. reesei* / *E. coli* (TrEc) consortium. **(A)** Schematic of TrEc consortium. Key parameters, identified via sensitivity analysis (panel B), are labeled (see Appendix A and Table B.1 for details). *T. reesei* produces cellulases (endoglucanase I, EGI; cellobiohydrolase I, CBHI; cellobiohydrolase II, CBHII) which hydrolyze cellulose to soluble oligosaccharides. Oligosaccharides are further hydrolyzed to glucose via cell-wall localized β-glucosidase (BGL). Soluble saccharides serve as growth substrates for the microbes (cellobiose and glucose for *T. reesei*; glucose only for *E. coli*). *E. coli* ferments glucose into isobutanol, which inhibits microbial growth due to toxicity. **(B)** Global sensitivity analysis of TrEc consortium model. Partial rank correlation coefficients (PRCC) between model parameters and output metrics are shown with hierarchical clustering (Ward's method; Pearson correlation distance). Parameters are labeled in panel A. Output metrics: mean *T. reesei* growth rate (R_{Tr}; g/L/h), mean *E. coli* growth rate (R_{Ec}; g/L/h), mean cellulose hydrolysis rate (R_{cel}; g/L/h), *E. coli* population fraction at fermentation endpoint (X_{Ec}; g/g-total microbial biomass), fraction of substrate carbon consumed by *E. coli* ($P_{C->Ec}$; g/g-total), isobutanol yield ($Y_{I/S}$; g/g-cellulose), and isobutanol productivity (Q_I; g/g-cellulose/h). The most significant PRCCs ($p > 0.05$ and $|PRCC| > 0.1$) are shown here; for full results see Figure C.2. **(C)** Theoretical analysis of isobutanol production with the TrEc consortium. Numerical solutions were calculated over a range of $X_{Ec}(t_0)$ values, as described in section 2.5.1. Key fermentation metrics shown (R_{cel}, $Y_{I/S}$, and Q_I; as described for panel B). Initial conditions: 20 g/L microcrystalline cellulose and 0.1 g/L *T. reesei* biomass. **(D)** Sample numerical solution ($X_{EC}(t_0) = 0.01$) from the analysis in panel C, showing dynamics of total cellulose ($\sum_{i>4} S_{G_i}$; g/L), isobutanol (I; g/L), total *T. reesei* biomass ($C_{Tr} = C_{Tr,s} + C_{Tr,v}$; g/L) and *E. coli* biomass (C_{Ec}; g/L).

2.3.3 Comparison of experimental and modeling results for RUTC30 monoculture and K12/RUTC30 biculture

To study consortium ecology and validate our theoretical framework, we experimentally characterized a monoculture of *T. reesei* and bi-culture of *T. reesei* / *E. coli* in minimal media with model cellulose substrate. Since cellulose is insoluble and *T. reesei* is a multicellular organism, it is technically challenging to independently quantify cellulose, *T. reesei* biomass, and *E. coli* biomass in co-cultures. We ultimately developed a set of novel analytical methods to measure carbohydrate composition and microbial biomass, which are described in section 2.5.11. *E. coli* K12 was used in these co-culture studies since it is well characterized (i.e. model parameters readily available) and is a common chassis strain. To assess the model of *T. reesei* growth and enzymatic cellulose hydrolysis, we first performed a batch monoculture of *T. reesei* RUT30 on 20 g/L microcrystalline cellulose (MCC) and fit the model to experimental data via a simple regression (Figure 2.6A; regression described in section 2.5.1.4). Experimental *T. reesei* growth and cellulose degradation agree with model predictions, with only minor parameter adjustments required to achieve good fit (Figure 2.6A; section 2.5.1.4; parameter values listed in Table B.3). *T. reesei* growth rate is approximately exponential until 100 hours, and then decelerates (Figure 2.6A). Our model suggests that this growth pattern is due to changes in relative rates of cellulose hydrolysis and *T. reesei* saccharide uptake over the course of the culture. Initially, cellulose hydrolysis rate is predicted to exceed saccharide consumption by *T. reesei*, leading to accumulation of soluble saccharides and saturation of *T. reesei* growth rate. During later stages, enzymatic saccharide production is balanced by *T. reesei* uptake, with very low soluble saccharide concentrations (10^{-4} to 10^{-3} g/L) and non-exponential *T. reesei* growth; at this stage cellulose hydrolysis becomes the rate-limiting step.

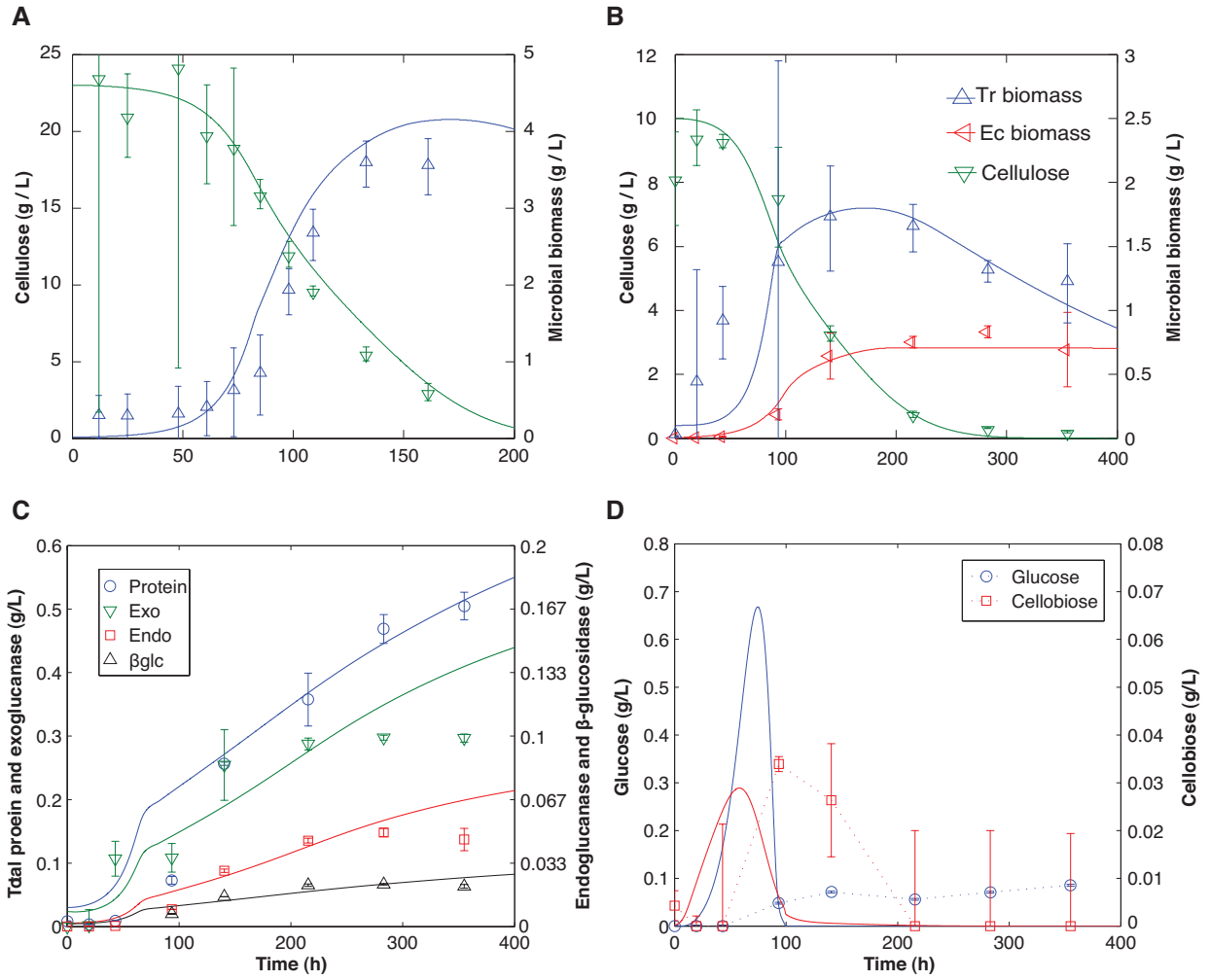

Figure 2.6: Comparison of experimental and modeling results for *T. reesei* RUTC30 monoculture and *T. reesei* RUTC30 / *E. coli* K12 biculture. Modeling results shown as smooth curves; experimental results shown as points. **(A)** *T. reesei* RUTC30 monoculture on 20 g/L microcystalline cellulose (MCC). **(B)** *T. reesei* RUTC30 / *E. coli* K12 bi-culture on 10 g/L MCC. **(C)** Total protein (Protein), exoglucanase (Exo), endoglucanase (Endo), and β-glucosidase (βglc) concentrations (g/L) for *T. reesei* RUTC30 / *E. coli* K12 bi-culture. **(D)** Glucose and cellobiose concentrations (g/L) for *T. reesei* RUTC30 / *E. coli* K12 bi-culture.

The full TrEc consortium model was validated with a batch biculture of *T. reesei* RUTC30 / *E. coli* K12 (hereafter referred to as "RUTC30/K12") on 10 g/L MCC (Figure 2.6B). Achieving good model fit required substantial adjustment of *E. coli* parameters from monoculture baseline values (section 2.5.1.4; parameter values listed in Table B.3), suggesting major physiological differences between biculture and monoculture growth for *E. coli*. Nonetheless, after regression, experimentally observed microbial growth and cellulose degradation agree with model predictions, affirming our modeling framework (Figure 2.6B). Our model predicts that co-culturing *E. coli* K12 with *T. reesei* RUTC30 will decrease *T. reesei* RUTC30 growth and cellulase production due to competition for soluble saccharides, resulting in lower R_{cel} and longer fermentation time relative to a *T. reesei* monoculture. This predicted ecological interaction appears to be recapitulated experimentally, with $R_{cel} = 0.13$ g/L/h and fermentation end-point time (defined as 90% cellulose conversion) of 161 h for a *T. reesei* RUTC30 monoculture, compared to $R_{cel} = 0.035$ g/L/h and fermentation end-point time of 216 h for a RUTC30/K12 biculture. In addition to microbial biomass and cellulose, we also measured total protein, β-glucosidase, endoglucanase, exoglucanase, cellobiose, and glucose concentrations in the RUTC30/K12 biculture (Figure 2.6C&D). Model predictions agree qualitatively well with experimentally measured enzyme concentrations during the early phases of the biculture, but diverge later (Figure 2.6C). In contrast to model predictions, glucose accumulates as microbial growth decelerates, reaching 0.085 g/L by the end of the culture (Figure 2.6D). Accumulation of soluble saccharides is commonly observed in batch cultures of cellulolytic microbes [14]; this phenomena may be due to cellulose hydrolysis rate exceeding saccharide uptake as microbial growth declines and ceases.

2.3.4 Isobutanol production with the TrEc consortium

2.3.4.1 Evaluation of different *E. coli* strains for isobutanol production

The studies described in the previous section establish the feasibility of co-culturing *T. reesei* and *E. coli* and provide a partial validation of our modeling framework. As a proof-of-concept application for the TrEc consortium, we demonstrate isobutanol production with bicultures of *T. reesei* RUTC30 and isobutanol producing *E. coli* strains. As described in section 2.2.3, the James Liao lab at UCLA has engineered various *E. coli* strains for isobutanol production, using both rational [1] and evolutionary [3] engineering strategies. For our initial investigations, we utilized strains JCL260 (rationally engineered; [1]), NV3 (evolved via norvaline selection; [3]), and NV3r1 (NV3 with WT *rpoS*; [3]) with plasmids pSA55 and pSA69 [1]. To assess isobutanol production under co-culture conditions, the aforementioned *E. coli* strains were bicultured with *T. reesei* RUTC30 in screw-cap flasks with minimal media and MCC as a cellulosic substrate (Figure 2.7A). Since our theoretical analysis suggests the initial population composition $X_{Ec}(t_0)$ is

a key determinant of isobutanol production (Figure 2.5B), we performed co-cultures over a range of inoculation ratios (Figure 2.7A). Co-culture isobutanol titers were relatively low and extremely variable compared to reported performance of *E. coli* monocultures on glucose, especially for strain JCL260. Strains NV3 and NV3r1 performed somewhat more consistently than JCL260, with NV3 reaching the highest mean isobutanol titers (Figure 2.7A). Isobutanol titer tended to increase with inoculation ratio, reaching a maximum at 0.1 or 1 Ec:Tr (gDW:gDW), depending on the strain; however, it is difficult to identify maxima due to high variance. To further investigate variance in isobutanol production, we performed a large number of replicate co-cultures (25 total) with JCL260 at 0.1 Ec:Tr (gDW:gDW) inoculation fraction (Figure 2.7A). Titers varied from 0 to 1700 mg/L, with most replicates falling in a range of 0 and 400 mg/L. The co-culture environment differs markedly from the monoculture conditions under which these isobutanol producing *E. coli* strains are typically utilized, and thus there are many possible reasons for the reduced titers and increased variance.

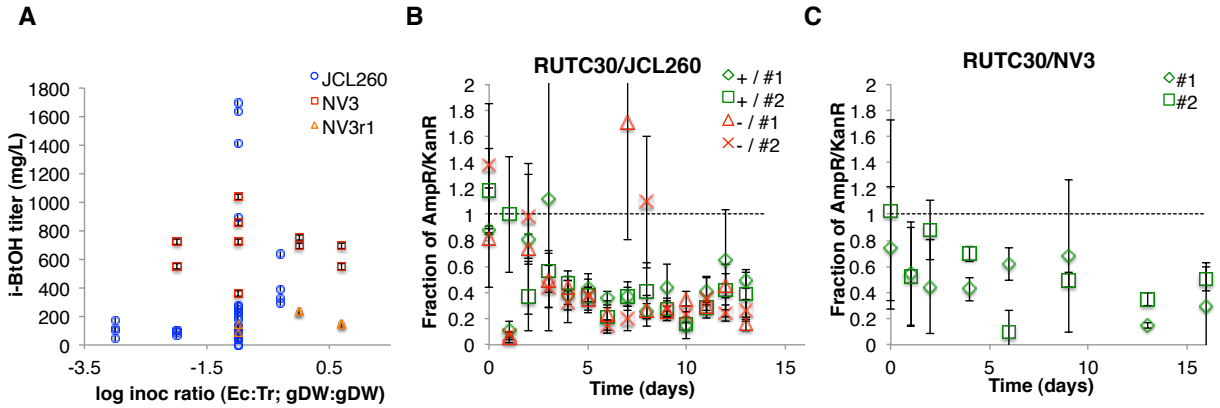

Figure 2.7: Performance of JCL260, NV3, and NV3r1 in biculture with RUTC30 on MCC. **(A)** Meta-summary of isobutanol titers at various inoculation ratios. Each point corresponds to a single culture. **(B)** Stability of plasmids pSA55 and pSA69 in *E. coli* JCL260 under co-culture conditions. Fraction of cells resistant to both ampicillin (AmpR; pSA55) and kanamycin (kanR; pSA69) was measured at each time point. Each series corresponds to single culture. Two of the cultures (green series) were periodically supplemented with ampicillin (pSA55) and kanamycin (pSA69) (as indicated by antibiotic bioassays) to maintain selective pressure while the other two cultures (red series) were unsupplemented controls. **(C)** Stability of plasmids pSA55 and pSA69 in *E. coli* NV3 under co-culture conditions. Each series corresponds to a single culture.

Plasmid instability represents a likely source of variation and reduced isobutanol titers. While we include appropriate concentrations of ampicillin (pSA55) and kanamycin (pSA69) in our culture media, antibiotic degradation is likely due to long culture times (i.e. >10 days). Furthermore, plasmid burden is probably higher under co-culture conditions due to low growth rates and isobutanol toxicity, potentially increasing selective pressure for plasmid loss. We investigated the dynamics of plasmid maintenance in JCL260 (Figure 2.7B) and NV3 (Figure 2.7C). For

both strains, plasmid loss follows an approximately asymptotic decay pattern, with residual plasmid maintenance levels between 20% and 50% (Figure 2.7B&C). We reasoned that antibiotic degradation may be the primary contributor to plasmid loss, and attempted to compensate for this by periodically supplementing ampicillin and kanamycin as indicated by a qualitative bioassay for antibiotic activity (Figure 2.7B). However, there was no difference in plasmid maintenance between supplemented and unsupplemented JCL260 cultures (Figure 2.7B) . Plasmid maintenance may be higher in NV3 compared to JCL260, however we cannot assert that these differences are statistically significant due to our limited data and high noise/error in our measurements. While plasmid maintenance is not substantially better in NV3 compared to JCL260, NV3 co-cultures produce higher average isobutanol titers than either JCL260 or NV3r1 co-cultures. Based on these results, NV3 was selected for use in all subsequent experiments. It is interesting to note that the only difference between NV3 and NV3r1 is the presence of a functional *rpoS* gene in NV3r1; thus *rpoS* loss-of-function may contribute to improved isobutanol production in co-culture, while conversely reducing isobutanol production in monoculture [3]. Since RpoS regulates a large number of genes related to stress tolerance and stationary phase, the link to isobutanol production in the co-culture environment is not immediately obvious.

In addition to plasmid stability, media composition is another factor that may affect co-culture isobutanol production. In our early exploratory co-culture studies, we discovered that *T. reesei* grows very poorly on *E. coli* minimal media, while conversely *T. reesei* minimal media (TMM) supports reasonable *E. coli* growth. Therefore we use *T. reesei* minimal media (pH 6.0, 10 to 20 g/L substrate) for TrEc co-culture experiments. In contrast, isobutanol producing *E. coli* strains are typically cultured on M9 media supplemented with trace nutrients and yeast extract (M9IP/YE media), with glucose as a substrate. To increase titers, feeding schemes are often used wherein cultures are periodically supplemented with concentrated feed solutions to replenish consumed glucose and nutrients. Thus media composition differs considerably in our co-culture studies compared to typical monoculture studies with isobutanol producing *E. coli* strains. TMM media has higher Mg, Ca, and SO_4 concentrations and lower Cl and PO_4 concentrations relative to M9IP/YE media. We typically perform co-culture experiments at pH 6.0 (midway between the optima of 5.0 for *T. reesei* and 7.0 for *E. coli*), while M9IP/YE media has pH 7.0. Furthermore, we eschew the use yeast extract or other costly nutrient supplements. While such supplements may dramatically improve growth and isobutanol production, it would not be economically feasible to use them in commercial production, and we think that our co-culture system should be tested under conditions representative of actual industrial fermentations. Finally, we typically use 10 to 20 g/L carbohydrate in our co-culture studies (for sake of minimizing total culture times), while the feeding schemes commonly used for isobutanol producing *E. coli* supply an equivalent of 100

g/L or more of glucose.

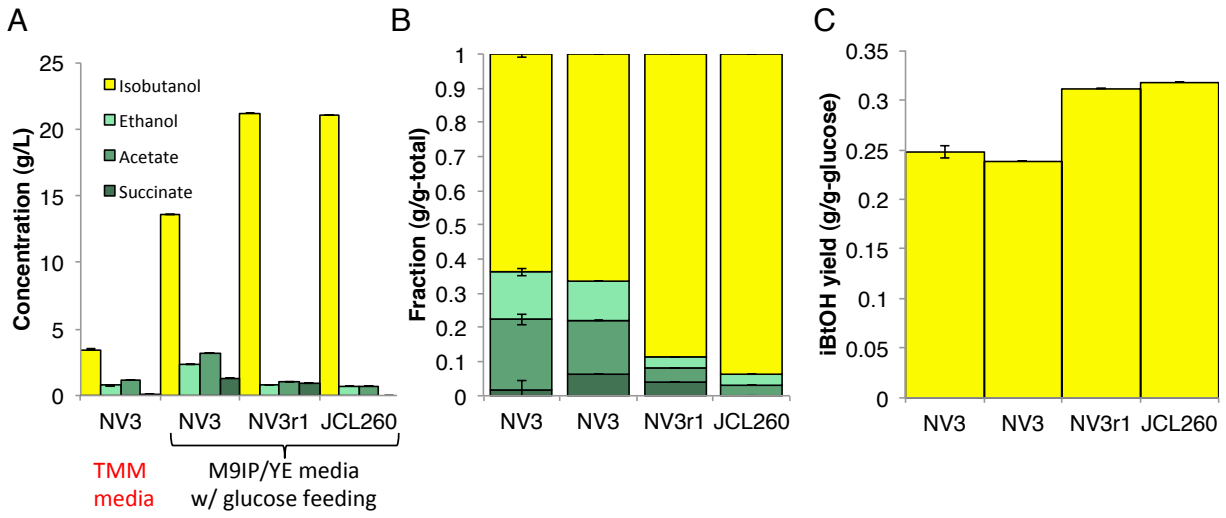

Figure 2.8: Performance of JCL260, NV3, and NV3r1 monocultures on different media. **(A)** Final concentrations of major fermentation products (isobutanol, ethanol, acetate, and succinate) for NV3 on *T. reesei* minimal media (TMM) with 20 g/L glucose and NV3, NV3r1, and JCL260 on M9 glucose media with trace nutrients and yeast extract (M9IP/YE); data for M9IP/YE cultures was obtained from [3]. **(B)** Relative proportions of major fermentation products for cultures from panel A. **(C)** Isobutanol yield for cultures from panel A.

To establish a reasonable reference for evaluating the performance of TrEc co-cultures, we characterized the production of isobutanol and other fermentation products by NV3 on TMM media with 20 g/L glucose as substrate (Figure 2.8A). The final isobutanol titer in TMM media was 3.5±0.1 g/L vs. 13.6 g/L reported for NV3 on M9IP/YE with glucose feeding [3]. Titers of other major fermentation products (ethanol, acetate, and succinate) are also lower in TMM media compared to M9IP/YE media with glucose feeding (Figure 2.8A). As would be expected, glucose consumption was reduced in TMM media, with approximately 14 g/L glucose consumed in 10 days compared to 57 g/L glucose in 4 days on M9IP/YE media. Interestingly, even though glucose consumption and product titers are reduced in TMM media, relative proportion of fermentation products (Figure 2.8B) and isobutanol yield (Figure 2.8C) are comparable between both media. These results suggest that NV3 has a similar metabolic phenotype in both TMM and M9IP/YE media, although substrate utilization is much lower for TMM media.

2.3.4.2 Isobutanol production with RUTC30/NV3 bicultures

To further investigate isobutanol production with the TrEc consortium, we conducted RUTC30/NV3 bicultures on both MCC and ammonia fiber expansion [88] pre-treated corn

stover (AFEX CS), representing model substrate and industrially relevant lignocellulosic feedstock, respectively. Bicultures were conducted over a series of inoculation ratios, ranging from 0.01 to 5 NV3:RUTC30 (g/g), with 20 g/L cellulosic substrate. All of the bicultures produced isobutanol; titers varied from 0.62 to 0.79 g/L for MCC bicultures, and from 1.46 to 1.86 g/L for AFEX CS bicultures (Figure 2.9A&C). Despite the wide range of inoculation ratios tested, titer ranges were relatively narrow for each cellulosic substrate, supporting asymptotic behavior predicted by the model (predicted titers shown in Figure C.1A).

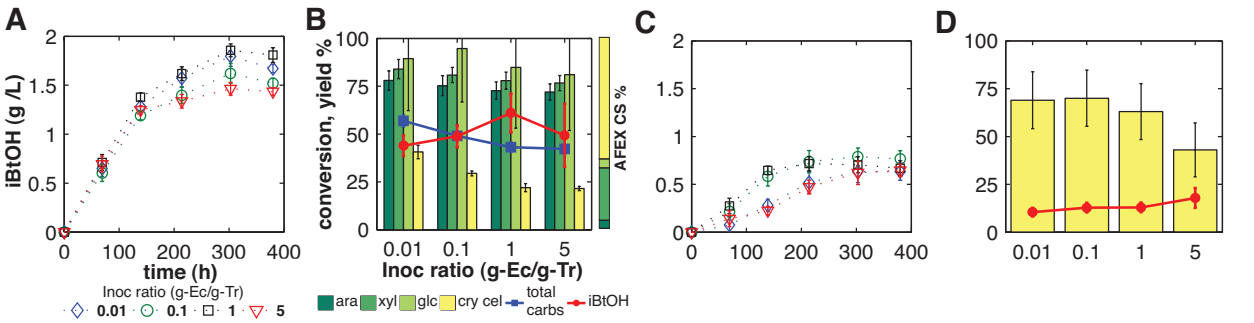

Figure 2.9: Isobutanol production with *T. reesei* RUTC30 and *E. coli* NV3 pSA55/69 bicultures (RUTC30/NV3) on microcrystalline cellulose (MCC) and ammonia fiber expansion [88] pre-treated corn stover (AFEX CS). Bicultures were conducted over a series of inoculation ratios, ranging from 0.01 to 5 NV3:RUTC30 (g/g), with 20 g/L cellulosic substrate. **(A)** Isobutanol production in RUTC30/NV3 bicultures on AFEX CS. **(B)** Conversion (% carbohydrate consumed) and isobutanol yield for major AFEX CS carbohydrates. Abbreviations: arabinose, ara; xylose, xyl; hemicellulose derived glucose, glc; crystalline cellulose, cry cel. For reference, relative proportions of AFEX CS major carbohydrates are shown in the stacked bar plot to right; see Table B.4 for precise composition.**(C)** Isobutanol production in RUTC30/NV3 bicultures on MCC. **(D)** Conversion (% carbohydrate consumed) and isobutanol yield for MCC bicultures.

In addition to assessing isobutanol production, we also characterized carbohydrate conversion (% carbohydrate consumed) and yield $Y_{I/S}$. AFEX CS carbohydrates include both hemicellulose (consisting primarily of mixed arabinose, xylose, and glucose polysaccharides) and crystalline cellulose; precise composition is given in Table B.4, with relative proportions of major carbohydrates shown in Figure 2.9B. RUTC30/NV3 bicultures achieved uniform and relatively high conversion of AFEX CS hemicellulose carbohydrates over all inoculation ratios tested (Figure 2.9B), while crystalline cellulose conversion was lower and tended to decline with increasing inoculation ratio (Figure 2.9B); since crystalline cellulose makes up a large proportion of AFEX CS carbohydrates (Figure 2.9B), total carbohydrate conversion trended with crystalline cellulose. $Y_{I/S}$ in AFEX CS bicultures tended to increase with inoculation ratio, due to declining conversion with relatively constant titer (Figure 2.9B). MCC bicultures followed similar trends to AFEX CS bicultures, with conversion declining and $Y_{I/S}$ increasing with NV3:RUTC30 inoculation ratio (Figure 2.9D). Carbohydrate conversion of RUTC30/NV3 bicultures was lower compared to that of the RUTC30/K12 biculture (Figure 2.9B&D vs. Figure 2.6B), suggesting that microbial substrate

uptake and isobutanol production stopped prematurely. Possible reasons for incomplete conversion include nutrient depletion or combined toxicity of *E. coli* fermentation products (Figure 2.10).

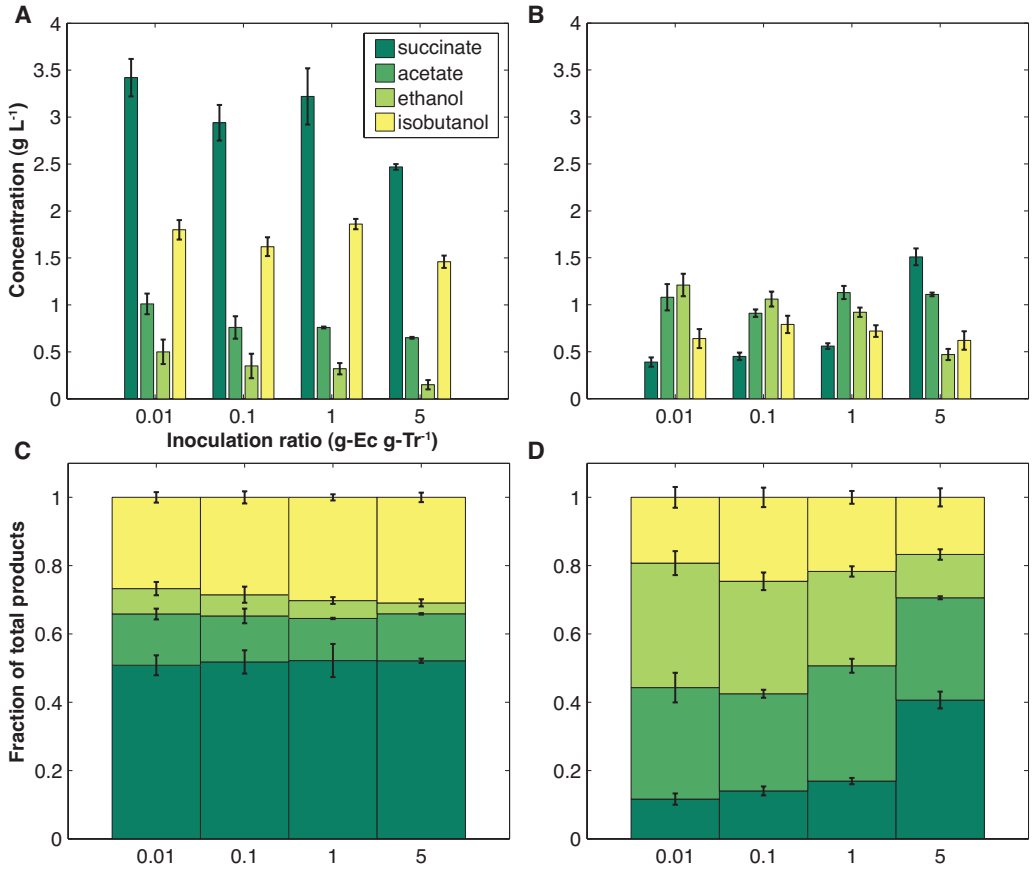

Figure 2.10: Endpoint concentrations and relative proportions of major fermentation products in *T. reesei* RUTC30 and *E. coli* NV3 pSA55/69 bicultures on AFEX pre-treated corn stover or MCC. Endpoint concentrations: **(A)** 20 g/L AFEX pre-treated corn stover bicultures and **(B)** 20 g/L MCC bicultures. Relative proportions: **(C)** 20 g/L AFEX pre-treated corn stover bicultures and **(D)** 20 g/L MCC bicultures.

In addition to isobutanol, we measured the end-point concentrations of other fermentation co-products (Figure 2.10). There are considerable differences between fermentation product distributions of RUTC30/NV3 bicultures (Figure 2.10) vs. NV3 monocultures on TMM glucose media (Figure 2.8B). Compared to monoculture, the relative proportion of isobutanol is substantially reduced while proportions of succinate, ethanol, and acetate are higher in NV3/RUTC30 bicultures, indicating an overall decline in isobutanol selectivity under co-culture conditions. In addition to differences between monoculture and co-culture, concentrations and proportions of fermentation products for RUTC30/NV3 bicultures varied between substrates. The concentrations and relative proportions of succinate were notably higher in AFEX CS bicultures compared to MCC bicultures, while the concentrations and relative proportions of ethanol and acetate were

higher in MCC bicultures (Figure 2.10). Interestingly, the distribution of fermentation products also varied with inoculation ratio. In MCC bicultures, the concentration and relative proportion of succinate increases with Ec:Tr inoculation ratio, while ethanol concentrations and relative proportions decrease (Figure 2.10B&D). For AFEX CS bicultures, the relative proportion of isobutanol increases with Ec:Tr inoculation ratio while ethanol concentrations and relative proportions decrease (Figure 2.10A&C); however the differences are slight and may not be significant.

2.4 Discussion and conclusion

2.4.1 Isobutanol production with the TrEc consortium: perspectives and future directions

To illustrate CBP applications of the TrEc consortium, we demonstrated isobutanol production from microcrystalline cellulose and AFEX pretreated corn stover with a biculture of *T. reesei* RUTC30 and *E. coli* NV3 pSA55/69. Overall, the highest isobutanol titer and yield (1.86 g/L and 62% theoretical, respectively) were produced on AFEX CS with an Ec:Tr ratio of 1 gDW:gDW (Figure 2.9A&B). These results illustrate the tremendous bioprocessing potential of the TrEc consortium (and S/F consortia in general), as the isobutanol titers and yields obtained are the highest reported to date for conversion of cellulosic substrates to next-generation biofuels and were achieved without detailed optimization of culture conditions or nutrient supplementation beyond minimal salts. Though these preliminary results are promising, isobutanol production by the consortium is lower than model predictions, and much additional work would be needed to achieve economically viable isobutanol titers, yields, and volumetric productivities with the TrEc consortium.

Our model predicts that RUTC30/NV3 bicultures with $X_{Ec}(t_0) \geq 0.1$ and 20 g/L MCC will reach an isobutanol titer of 2.2 g/L, corresponding to a yield of 0.11 g/g-MCC (27% theoretical; all MCC consumed) (Figure C.1A and Figure 2.5C&D). The experimentally observed values are lower than predicted; for RUTC30/NV3 bicultures with $0.01 \leq X_{Ec}(t_0) \leq 5$, isobutanol titers varied from 0.62 to 0.79 g/L isobutanol, with yields of 10$\pm$2% to 18$\pm$5% of theoretical and MCC conversions of 45$\pm$15% to 70$\pm$14% (Figure 2.9C&D). As discussed above, incomplete MCC utilization may be due to nutrient limitations or toxicity effects. The lower than expected isobutanol titers/yields are due in large part to reduced isobutanol selectivity of *E. coli* NV3 pSA55/69 under biculture conditions vs. monoculture (Figure 2.10B vs. Figure 2.8B). Based on a mass balance of primary fermentation products, we predict that if NV3 product distributions in

co-culture matched those in monoculture then isobutanol titers and yields would be approximately doubled, and thus closer values predicted by our model (Figure 2.11).

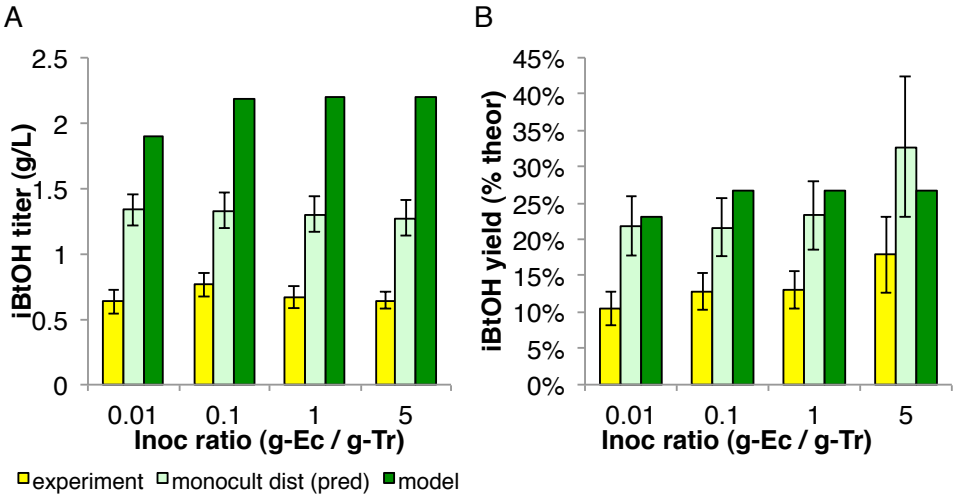

Figure 2.11: Analysis of isobutanol titers and yields in RUTC30/NV3 bicultures on 20 g/L MCC. Experimentally observed values (i.e. from Figure 2.9C&D) are denoted by "experiment", with model predictions as "model". Values denoted by "monocult dist (pred)" represent the expected titers/yields if NV3 product distributions in co-culture hypothetically matched those in monoculture, and were calculated from a mass balance on primary fermentation products (isobutanol, ethanol, acetate, and succinate). For each fermentation product j, the glucose G consumed to form product j was calculated based on the theoretical yield coefficient $Y_{j/G}$ (g-product j / g-glucose): $C_{G,j} = C_j / Y_{j/G}$, where $C_{G,j}$ and C_j are glucose and product concentrations, respectively. The total substrate consumed to form fermentation products was then calculated as $\sum C_{G,j}$, and isobutanol titer was predicted as $Y^{mono}_{iBtOH/G} \sum C_{G,j}$, where $Y^{mono}_{iBtOH/G}$ is the isobutanol yield (0.25 g/g-glucose) for NV3 in monocultures. Yield was calculated by dividing the predicted titer by the amount substrate consumed (Figure 2.9D). **(A)** Isobutanol titers. **(B)** Isobutanol yields (% theoretical, with theoretical yield of 0.41 g-isobutanol/g-glucose).

We have not yet elucidated the reason for different NV3 product distributions between co-culture and monoculture. The relative decrease in isobutanol may be due in part to plasmid loss, but there are numerous other possible explanations for the observed product distributions. We suspect that the NV3 metabolic phenotype in co-culture may differ greatly from the monoculture environment. For example, substrate concentrations and growth rates are much lower under co-culture conditions compared to monoculture. Previous studies have demonstrated that metabolite concentrations [89] and metabolic fluxes [90] in *E. coli* change with growth rate, thus *E. coli* metabolism may be quite different under co-culture conditions. Additionally, there may be unidentified interactions between *E. coli* and *T. reesei* that alter *E. coli* physiology and metabolism. Ultimately more investigation into *E. coli* metabolism under co-culture conditions is needed. We suggest that isobutanol selectivity of *E. coli* NV3 could be further improved via metabolic engineering to reduce flux down competing pathways (i.e. succinate, acetate, and ethanol).

Beyond increasing isobutanol selectivity, further development of the TrEc consortium for isobutanol production will likely require other significant improvements to *E. coli*, including investigating/improving isobutanol yields from pentose sugars (released from hemicellulose hydrolysis), stabilizing or chromosomally integrating plasmids, and improving tolerance to isobutanol and pretreatment inhibitors. In addition to *E. coli*, the performance of *T. reesei* will need to be improved as well. Increasing volumetric productivity (i.e. cellulose hydrolysis rate) and achieving complete substrate conversion will likely require increasing tolerance of *T. reesei* to isobutanol and pretreatment inhibitors, and possibly increasing the catalytic efficiency of cellulases. Once isobutanol titers are improved to $>$10 g/L, isobutanol toxicity is likely to become a key factor limiting further improvements in titer and productivity. Motivated by this issue, we investigate and improve isobutanol tolerance of *E. coli* in Part 2 of this dissertation. In conjunction with organism-level improvements, conditions for co-culture fermentations will also need to be optimized, including media composition, pH, temperature, and dissolved O_2 levels. Finally, to be useful for process design, our modeling framework would need to be expanded to include hydrolysis and microbial utilization of hemicellulose, and further refinements are needed to improve predictive accuracy for enzyme and soluble saccharide concentrations. Beyond the obvious improvements enumerated here, our theoretical analysis of TrEc consortium ecology suggests other strategies for controlling and enhancing performance, which will be briefly discussed in section 2.4.2.

2.4.2 Design and analysis of synthetic microbial consortia with ecology theory

Synthetic microbial consortia are valuable model systems for fundamental biological study and offer a plethora of potential biotechnology applications. However, our limited understanding of the dynamics and interactions of microbial populations hinders the deployment of synthetic consortia for real-world applications. In this work, we conducted an in-depth theoretical and experimental analysis of synthetic *Trichoderma reesei* RUTC30 and *Escherichia coli* (TrEc) consortia and demonstrated that the TrEc consortium could serve as a flexible system for consolidated bioprocessing of lignocellulosic biomass, using isobutanol production as a proof-of-concept application. While the general concept of using two-member saccharolytic/fermentation (S/F) consortia for consolidated bioprocessing of lignocellulosic feedstocks has been considered previously (Table 2.1), our theoretical and experimental studies elucidated important underlying ecological interactions and provide novel insights about the properties of the TrEc consortium. Aside from being of fundamental interest, such insights also suggest strategies to control consortium

performance via manipulation of ecological parameters. For instance, we demonstrate that carbon flow partition between *T. reesei* and *E. coli* determines the tradeoff between cellulose hydrolysis rate and product yield. Model sensitivity analysis identified key parameters controlling carbon flow partition, which could be adjusted via genetic or environmental manipulation to optimize rate and yield of the TrEc consortium. For example, substrate uptake kinetics of *E. coli* and *T. reesei* play a major role in determining carbon flow partition (Figure 2.5A&B). Product yield/titer can thus be increased by increasing the competitiveness of *E. coli* for glucose, for instance by increasing growth rate, substrate affinity, the biomass-substrate yield coefficient ($Y_{S_{G1}/C_{Ec}}$), and isobutanol tolerance, or conversely by decreasing the competitiveness of *T. reesei*. Volumetric productivity can be improved by simultaneously increasing the growth rates of both *T. reesei* and *E. coli*. While our investigation focused on the TrEc consortium as a model system, our insights and the theoretical framework that we developed could be readily extended to S/F consortia composed of different species and/or targeting different feedstocks (e.g. starch or inulin).

It is widely acknowledged that stability, robustness, and control of population composition remain key challenges in engineering synthetic consortia [66, 30, 24, 47]. In the next chapter, we identify and characterized ecological mechanisms that permit stable and tunable population compositions in the TrEc consortium, enhancing the utility of this platform for real-world bioprocessing applications.

2.5 Materials and methods

2.5.1 Model implementation and analysis

2.5.1.1 Implementation and numerical solutions of TrEc consortium model

The ODE modeling framework described in section 2.3.1 and Appendix A was implemented in MATLAB (MathWorks Inc) and solved numerically using the *ode15s* solver. Since we write mole balances for each possible saccharide S_{G_i}, the total number of ODEs depends on the degree of polymerization (DP) distribution of S_{G_i}. For a given S_{G_i} distribution, a total of $5 + DP_{max}$ ODEs are required, where DP_{max} is the maximum DP of a given cellulosic substrate. Parameters listed in Table B.1 and Table B.2 were used for initial modeling work; at later stages we experimentally measured certain parameters, and also performed a simple regression analysis to estimate parameters from experimental data (Table B.3). Initial conditions (C_{Ec}, $C_{Tr,v}$, $C_{Tr,s}$, I, and S_{G_i} at t_0) were chosen to be representative of typical experimental conditions. We approximate $S_{G_i}(t_0)$ with a log-normal distribution, which agrees qualitatively well with experimental data (e.g. see [91]):

$$S_{G_i}(t_0) = \overline{C} f\left(\log i; \overline{DP}, \sigma_{DP}^2\right) \tag{2.3}$$

$$= \frac{\overline{C}}{c_{v,DP}\overline{DP}\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log i - \overline{DP}}{c_{v,DP}\overline{DP}}\right)^2} \tag{2.4}$$

Where $f\left(\log i; \overline{DP}, \sigma_{DP}^2\right)$ is the probability density function of the normal distribution, $\overline{C} = \sum_{i>4} S_{G_i}$ is total cellulose (mM), i is DP (number of glucose monomers), $\overline{DP}$ is mean DP, and $c_{v,DP}$ is the coefficient of variation of $\overline{DP}$ (with $\sigma_{DP} = c_{v,DP}\overline{DP}$).

We empirically discovered that the functions for isobutanol toxicity ($K^I_{Ec,S_{G_1}}$ and K^I_{Tr}; equations A.48 and A.42, respectively) resulted in excessively long execution times for *ode15s* and other MATLAB ODE solvers. To make model analysis tractable, we tested the effect of different isobutanol toxicity functions on execution time. Exponential models were found to dramatically improve execution time:

$$K^I_{Ec,S_{G_1}} = e^{-\frac{I}{I_{Ec,*}}} \tag{2.5}$$

and

$$K^I_{Tr} = e^{-\frac{I}{I_{Tr,*}}} \tag{2.6}$$

Where $I_{Ec,*}$ and $I_{Tr,*}$ are isobutanol inhibition constants (g-isobutanol/L) for *E. coli* and *T. reesei*, respectively, and other terms are as described previously. The prior models for $K^I_{Ec,S_{G_1}}$ and K^I_{Tr} offer a better fit to experimental toxicity data, however the exponential models are still suitable for qualitative model analysis, as they reflect monotonically declining growth rate with isobutanol concentration.

2.5.1.2 Global sensitivity analysis of TrEc consortium model

We performed a global sensitivity analysis on the TrEc consortium model to map parameter space to consortium performance metrics, identify key parameters controlling consortium behavior, and quantify how parameter uncertainty affects model outputs. We applied the sensitivity analysis strategy proposed in [87], which we describe here briefly. The TrEc consortium model was numerically integrated with 1000 different sets of parameter values and initial conditions (ICs) sampled from appropriate statistical distributions using latin hypercube selection (LHS) [87]. We calculated partial rank correlation coefficients (PRCC) [87] between each parameter or IC and a set of output metrics, including mean *T. reesei* growth rate (R_{Tr}; g/L/h), *T. reesei* titer

($C_{Tr}(t_f)$; g/L), mean *E. coli* growth rate (R_{Ec}; g/L/h), *E. coli* titer ($C_{Ec}(t_f)$; g/L), mean cellulose hydrolysis rate (R_{Cel}; g/L/h), *E. coli* population fraction at fermentation endpoint ($X_{Ec}(t_f)$; g/g-total microbial biomass), fraction of substrate carbon consumed by *E. coli* ($P_{C\rightarrow Ec}$; g/g-total), isobutanol yield ($Y_{I/S}$; g/g-cellulose), isobutanol titer ($I(t_f)$; g/L), and isobutanol productivity (Q_I; g/g-cellulose/h).

LHS is a stratified sampling-without-replacement technique, where random parameter distributions are divided into N equally probable intervals, which are then sampled [87]. To explore the entire parameter range, each interval for each parameter is sampled exactly once without replacement [87]. A matrix is generated that consists of N rows corresponding to the sample size (i.e. total number of parameter sets), and of k columns corresponding to the total number of varied parameters [87]. Parameter and IC values were sampled from either a normal distribution or uniform distribution. For a normal distribution:

$$a_{i,j} = F^{-1}\left(p_{i,j}; \mu_j, (c_{v,j}\mu_j)\right) \tag{2.7}$$

Where $a_{i,j}$ is the value of parameter j in LHS sample i, $F^{-1}\left(p_{i,j}; \mu_j, (c_{v,j}\mu_j)\right)$ is the normal inverse cumulative distribution function, $p_{i,j}$ is the LHS probability for parameter j in sample i, μ_j is the mean value of parameter j, and $c_{v,j}$ is the coefficient of variation for parameter j (with $\sigma = c_{v,j}\mu_j$). For a uniform distribution:

$$a_{i,j} = a_{j,min} + p_{i,j}\left(a_{j,max} - a_{j,min}\right) \tag{2.8}$$

Where $a_{j,min}$ is the lower bound on parameter j, $a_{j,max}$ is the upper bound of parameter j, and other terms are as described above.

For parameters that can be arbitrarily varied (i.e. ICs) or that have a wide range of equally probable values, we sampled from a uniform distribution; all other parameters were sampled from a normal distribution. Table B.1 contains a list of all investigated parameters, sampling distribution (normal or uniform) for each parameter, and values of μ_j / $c_{v,j}$ or $a_{j,min}$ / $a_{j,max}$.

The TrEc consortium model was numerically integrated with each set of sampled parameter values, and the above output metrics (R_{Tr}, $C_{Tr}(t_f)$, R_{Ec}, $C_{Ec}(t_f)$, R_{Cel}, X_{Ec}, $P_{C->Ec}$, $Y_{I/S}$, $I(t_f)$, and Q_I) were calculated. LHS sampling and numerical integration were performed with MATLAB. We performed this analysis on a high-performance computing cluster (4 cores total; 2.67 GHz Intel Xeon X5650 processors; 4 GB RAM/core). Partial Rank Correlation Coefficients (PRCC) were calculated by rank transforming parameters and outputs, and then calculating partial

correlation coefficients between each parameter and model output [87]. PRCC represents a robust correlation metric for nonlinear but monotonic relationships between model inputs (parameters) and outputs, as long as minimal correlation exists between the inputs [87]. Parameter PRCCs are shown with hierarchical clustering (Wards method; Euclidean distance) in Figure C.2; statistically insignificant PRCCs ($p < 0.05$) are set to 0.

2.5.1.3 Theoretical analysis of isobutanol production with TrEc consortium model

To simulate and analyze isobutanol production by the TrEc consortium, we numerically solved the TrEc consortium model over a range $X_{Ec}(t_0)$ values using *E. coli* parameter values corresponding to isobutanol production strains JCL260 pSA55/69 [1] or NV3 pSA55/69 [3]. $Y_{S_{G_1}/C_{Ec}}$, $Y^{growth}_{I/S_{G_1}}$, $Y^{maint}_{I/S_{G_1}}$, $k_{Ec,d}$, $K_{Ec,S_{G_1}}$, and $m_{Ec,S_{G_1}}$ values for *E. coli* JCL260 pSA55/69 were estimated from data presented in [1], [2], and [3]; values for *E. coli* NV3 pSA55/69 were estimated from [3]. $\mu_{max,Ec,S_{G_1}}$ in glucose TMM media was experimentally determined for each strain (Table B.3). A complete set of parameter values for *E. coli* NV3 pSA55/69 and *E. coli* JCL260 pSA55/69 can be found in Table B.2.

2.5.1.4 Regression of model to experimental data

We performed a simple regression of the TrEc consortium model to experimental data obtained from a *T. reesei* RUTC30 monoculture on 20 g/L Avicel and a *T. reesei* RUTC30 / *E. coli* K12 biculture on 10 g/L Avicel. We attempted to fit experimentally measured insoluble cellulose concentration (gDW/L), *T. reesei* RUTC30 biomass concentration (gDW/L), and *E. coli* K12 biomass concentration (gDW/L) to model predictions. For each variable, a crude fit was evaluated as the residual sum of squares over all time points: $RSS = \sum_{t=t_0}^{t_f} \left(y_{exp}(t) - y_{model}(t)\right)^2$, where $y_{exp}(t)$ the is experimental value at time t and $y_{model}(t)$ is the predicted value. Starting with the parameter values listed in Table B.1, regression was performed by manually adjusting model parameters to minimize RSS for each variable. For the *T. reesei* RUTC30 monoculture, a reasonable fit between experimental data and modeling predictions was obtained by adjusting a single parameter, $Y_{S_{G_1}/C_{Tr}}$ (Table B.3). Obtaining reasonable fits for the RUTC30/K12 biculture required adjusting $\mu_{max,Tr,S_{G_1}}$, $K_{Tr,S_{G_1}}$, and $Y_{S_{G_1}}/C_{Tr}$ for *T. reesei* and $\mu_{max,Ec,S_{G_1}}$, $k_{Ec,d}$, $K_{Ec,S_{G_1}}$, $Y_{S_{G_1}}/C_{Ec}$, and $m_{Ec,S_{G_1}}$ for *E. coli* (Table B.3). In addition to cellulose and microbial biomass, we also quantified total protein, exoglucanase, endoglucanase, β-glucosidase, glucose, and cellobiose concentrations in the RUTC30/K12 biculture and attempted to fit our model to experimentally measured values; attaining reasonable fit required adjusting $Y_{E_T/C_{Tr}}$, k_{E_T}, k_{EG1}, k_{CBHI}, k_{CBH2}, k_{BGL,G_2}, k_{BGL,G_3}, and k_{BGL,G_4} (Table B.3).

2.5.2 Strains and media

Trichoderma reesei RUTC30 was used as the cellulolytic specialist for this study. *Escherichia coli* K12 MG1655 was used in *T. reesei* / *E. coli* biculture experiments for model validation and testing cooperator-cheater dynamics. Isobutanol production experiments were conducted with *Escherichia coli* JCL260 [1], NV3 [3], or NV3r1 [3] with plasmids pSA55 (ColE1 ori; Amp^R; $P_L lacO_1$::*kivd-ADH2*) and pSA69 (p15A ori; Kan^R; $P_L lacO_1$::*alsS-ilvCD*) [1]. *E. coli* strains were propagated in NG50 media (described in Chapter 4), supplemented with antibiotics for strains harboring pSA55/69 (100 μg/mL ampicillin and 30 μg/mL kanamycin). *T. reesei* monocultures, *E. coli* monocultures, and *T. reesei* / *E. coli* bicultures were grown in *Trichoderma* minimal media (TMM) formulated as follows (all concentrations in g/L unless otherwise noted): urea, 1; $(NH_4)_2SO_4$, 4; KH_2PO_4, 6.59; K_2HPO_4, 1.15; $FeSO_4$*$7H_2O$, 0.005; $MnSO_4$*H_2O, 0.0016; $ZnSO_4$*$7H_2O$, 0.0014; $CoCl_2$*$6H_2O$, 0.002; $MgSO_4$, 0.6; $CaCl_2$, 0.6; Tween-80, 0.0186% (v/v); carbon source (glucose, microcrystalline cellulose, or AFEX pre-treated corn stover) as indicated. For cultures with isobutanol-producing *E. coli* strains harboring plasmids pSA55/69, antibiotics were added as described above with 0.1 mM IPTG was added for induction. For flask culture experiments, TMM was buffered with 0.1 M maleate-NaOH at indicated pH [92]; K_2HPO_4 was eliminated and KH_2PO_4 reduced to 1.2 g/L. LB agar [93] was used for *E. coli* cell counting.

2.5.3 Preparation of inoculum cultures

E. coli was inoculated from cryostock into NG50 medium (with appropriate antibiotics) and incubated at 30°C with agitation until saturated. Cultures were then inoculated 1:100 (by volume) into TMM with 20 g/L glucose and incubated for a further 48 hours at 30°C with agitation. To remove residual glucose, cultures were washed by centrifuging at 12,000 rpm x 2 minutes, resuspending in TMM without carbon source, and then repeating centrifugation/resuspension.
T. reesei RUTC30 cryopreserved conidia were inoculated into TMM with 20 g/L glucose and incubated at 30°C with agitation for 48 hours. Cultures were then inoculated 1:50 (by volume) into TMM with indicated carbon source (microcrystalline cellulose or AFEX pre-treated corn stover) and incubated for a further 48 hours at 30°C with agitation.

2.5.4 Experimental measurement of μ_{max}

μ_{max} (maximum specific growth rate; 1/h) was measured for *E. coli* K12, *E. coli* NV3 pSA55/69, *E. coli* JCL260 pSA55/69, and *T. reesei* RUTC30, under conditions representative of co-culture experiments (glucose TMM media; growth temperature 30°C). μ_{max} was determined by fitting growth data to an exponential model ($\ln C = \ln C_0 + \mu_{max} t$; C is cell density at time t and

C_0 is initial density). *E. coli* cell density was determined by measuring optical density at 600 nm (OD_{600}); *T. reesei* cell density was determined via gravimetric analysis (described in section 2.5.11.1; gDW/L). Growth data was obtained by inoculating cultures in 20 g/L glucose TMM and measuring cell density over time intervals corresponding to the exponential growth phase. *E. coli* growth studies were conducted in microplates. Standard 96-well microplates were filled with 200 μL medium per well and seeded with 2 μL of prepared inoculum culture (described in preceding section) per well. OD_{600} was measured every 10 minutes for 48 to 96 hours using a Spectramax M5 or Versamax plate reader (Molecular Devices, LLC), with 30 °C incubation temperature and agitation between reads. For *T. reesei*, we conducted growth studies using 1 L flask cultures. Flasks were filled with 200 mL medium, seeded with 2 mL prepared inoculum culture, and then incubated at 30 °C with agitation. Triplicate 10 mL samples were taken for gravimetric dry mass analysis 12, 28, 21, 24, and 27 hours after inoculation. Experimentally measured μ_{max} values are reported in Table B.3.

2.5.5 Characterization of isobutanol toxicity

To characterize isobutanol toxicity in *E. coli* and *T. reesei*, μ_{max} was measured in 20 g/L glucose TMM media spiked with isobutanol at various concentrations, using procedure described in preceding section. To prevent evaporation of isobutanol from cultures, microplates (*E. coli*) were sealed with an impermeable and optically clear adhesive film; flasks (*T. reesei*) were tightly wrapped with parafilm (sterilized with 70% v/v ethanol). μ_{max} values for *E. coli* and *T. reesei* at various isobutanol concentrations are reported in Table B.3.

2.5.6 Experimental measurement of K_S, $Y_{S/C}$, and m

2.5.6.1 Theory

Modeling and experimental results demonstrate that both growth rates and soluble saccharide (i.e. glucose and cellodextrin) concentrations are low in *T. reesei* / *E. coli* bicultures on cellulosic substrates, in contrast to saturated growth rates and high glucose concentrations in batch monoculture experiments. Since $K_{S_{G_1}}$ (Monod glucose affinity; g/L) is considered an extant kinetic parameter [94], it should be measured under conditions similar to the co-culture environment. We thus chose to measure $K_{S_{G_1}}$, $Y_{S_{G_1}/C}$ (glucose/biomass yield coefficient; g-glucose/g-biomass), and $m_{S_{G_1}}$ (maintenance coefficient; g-glucose/g-biomass/h) with chemostat experiments; growth rate and substrate concentrations can be precisely controlled via dilution rate, allowing us to simulate biculture conditions. Chemostat cultures can reach a steady state in which cell densities and substrate concentrations are constant [83]:

$$\frac{dC}{dt} = 0 \tag{2.9}$$

$$\frac{dS_{G_1}}{dt} = 0 \tag{2.10}$$

$$\mu = D \tag{2.11}$$

Where C is cell density, μ is specific growth rate (1/h) and D is dilution rate (media flow rate / culture volume), and other terms are similar to those described above. For single substrate (i.e. glucose) growth, we model μ as

$$\mu = \left(\left(\mu,max, S_{G_1} + \frac{m_{S_{G_1}}}{Y_{S_{G_1}/C}} \right) \frac{S_{G_1}}{K_{S_{G_1}} + S_{G_1}} - \frac{m_{S_{G_1}}}{Y_{S_{G_1}/C}} \right) \tag{2.12}$$

Where all terms are similar to those described in section 2.3.1. We can then apply the above steady-state criteria to derive relations for expressions for $K_{S_{G_1}}$, $Y_{S_{G_1}/C}$, and $m_{S_{G_1}}$ in chemostat cultures:

$$K_{S_{G_1}} = \frac{\left(\mu_{max,S_{G_1}} - D\right) S_{G_1}}{D} \tag{2.13}$$

and

$$Y^{app}_{S_{G_1}/C} = \frac{S_{G_1,0} - S_{G_1}}{C} \tag{2.14}$$

$$= Y_{S_{G_1}/C} + \frac{m_{S_{G_1}}}{D} \tag{2.15}$$

Where $Y^{app}_{S_{G_1}/C}$ is the apparent yield coefficient (g-glucose/g-cells), $S_{G_1,0}$ is glucose concentration in the chemostat feed (g/L) and other terms are as described previously.

2.5.6.2 Chemostat experiments

Chemostat studies were conducted in a BioFlo 3000 bioreactor (New Brunswick Scientific), using 1.5 L, 2 L, or 2.5 L culture volumes. Glucose TMM media was used in all studies, with 4 g/L glucose for *T. reesei* cultures and 0.3 g/L glucose for *E. coli* cultures; concentrations were chosen to guarantee that glucose would be the sole limiting nutrient (estimated from data collected during μ_{max} characterization experiments). Bioreactor was maintained at pH 6, temperature 30°C, agitation 200 rpm, and air flow at 2 vvm (volume air per volume culture per minute) for all studies. Chemostat studies for *E. coli* were performed at dilution rates of $D = 0.04$, 0.063, 0.069,

0.116, 0.174, 0.25, and 0.35 1/h; for *T. reesei*, experiments were performed at $D = 0.019$ and 0.04 1/h. Chemostat studies were initiated by adding prepared inoculum to bioreactor at 1:100 (*E. coli*) or 1:50 (*T. reesei*) volume ratio, and pre-culturing in batch mode for 48 hours. After 48 hours, chemostat mode was started by setting the media feed pump to produce desired dilution rate D; a level probe was used to control the harvest pump rate and thus maintain constant culture volume. Sigma antifoam 204 was added as needed, with 0.02% (v/v) added per supplementation. Samples were taken in triplicate one to two times per turnover period (D^{-1}; h). *E. coli* cultures were analyzed for OD_{600}, viable cell concentration (cells/L), total dry biomass (gravimetric analysis; gDW/L), and glucose (g/L), as per procedures described in section 2.5.11; *T. reesei* cultures were analyzed for total dry biomass (gravimetric analysis; gDW/L) and glucose (g/L). Glucose analysis samples were prepared by dispensing culture directly from the bioreactor into syringes filled with steel beads chilled at -20°C and then filtering through 0.22 μm membranes to sterilize; this procedure quenches microbial metabolism, ensuring accurate measurement of low residual glucose concentrations [95]. Steady state was assumed when biomass and glucose concentrations stabilized to constant values (within error limits) for two to three consecutive samples. After taking steady-state samples, feed pump was adjusted to next desired dilution rate and sampling was continued. To avoid adaptive evolution, chemostat experiments were run for < 15 turnovers. $K_{S_{G_1}}$ was calculated from D, glucose concentration (S_{G_1}), and μ_{max}. To determine $Y_{S_{G_1}/C}$ and $m_{S_{G_1}}$, we collected data at various values of D and performed a linear regression of $1/D$ against $Y^{app}_{S_{G_1}/C}$, taking $Y_{S_{G_1}/C}$ as the y-intercept and $m_{S_{G_1}}$ as the slope. Values of $K_{S_{G_1}}$, $Y_{S_{G_1}/C}$, and $m_{S_{G_1}}$ determined from chemostat studies are reported in Table B.3.

2.5.7 Bioreactor batch cultures

Batch cultures were conducted in a BioFlo 3000 bioreactor (New Brunswick Scientific) with 3 L TMM media buffered with phosphate. Bioreactor was maintained at pH 6, temperature 30°C, agitation 200 rpm, and air flow 6.0 L/min (2 vvm). Cultures were sampled periodically and analyzed as described in section 2.5.11 for total dry mass, carbohydrate composition, *E. coli* cell counts (if applicable), total protein, β-glucosidase activity, exoglucanase activity, and endoglucanase activity.

2.5.8 Isobutanol production with *E. coli* monocultures

Duplicate cultures of *E. coli* JCL260 NV3 pSA55/69 were inoculated in 10 mL TMM media (pH 6) with 20 g/L glucose as growth substrate. All cultures were supplemented with 100 μg/mL ampicillin, 30 μg/mL kanamycin, and 0.1 mM IPTG. Cultures were incubated in 50 mL Falcon tubes at 30°C with agitation. 0.1 mM IPTG was added every five days to compensate for degradation. Samples were taken every two days for analysis of fermentation products, described in

section 2.5.11.

2.5.9 Isobutanol production with RUTC30/NV3 bicultures

Duplicate *T. reesei* RUTC30 / *E. coli* NV3 pSA55/69 bicultures were inoculated in 50 mL TMM media (pH 6) with 20 g/L microcrystalline cellulose or 20 g/L AFEX pre-treated corn stover as growth substrate. All cultures were supplemented with 100 μg/mL ampicillin, 30 μg/mL kanamycin, and 0.1 mM IPTG. Cultures were incubated in 250 mL screw-cap flasks at 30°C with agitation. 0.1 mM IPTG was added every five days to compensate for degradation. Samples were taken every three days for analysis of fermentation products, described in section 2.5.11. Substrate conversion and isobutanol yield were estimated via carbohydrate analysis of endpoint samples, described in section 2.5.11.

2.5.10 Characterization of pSA55/69 stability

Four replicate *T. reesei* RUTC30 / *E. coli* JCL260 pSA55/69 co-cultures were inoculated on 10 g/L microcrystalline cellulose TMM media buffered with 0.1 M maleate-NaOH (pH 6), at a ratio of 0.1 JCL260:RUTC30 (g/g), and cultured as described in above sections. Two cultures served as controls, while the other two were assayed for antibiotic degradation and supplemented with ampicillin (to 100 μg/mL) or kanamycin (to 30 μg/mL) as needed; these cultures were also supplemented with IPTG (to 0.1 mM) every 5 days to compensate for degradation. Samples were taken every 24 hours. Isobutanol concentration was determined as described in section 2.5.11, and plasmid maintenance was assayed by performing serial dilution plating and cell counting on non-selective LB media and LB media supplemented with 100 μg/mL ampicillin and 30 μg/mL kanamycin; fraction of cells retaining pSA55/69 was taken as fraction of amp^R / kan^R cells out of total. We devised a bioassay to assess antibiotic degradation. Co-culture supernatants were sterile filtered and supplemented with 5x LB / 50 g/L glucose (final concentration of 1x LB / 5 g/L glucose) to ensure that nutrients were not growth limiting, then incoculated with *E. coli* DH5α, *E. coli* DH5α pSA55 (amp^R), *E. coli* DH5α pSA56 (kan^R), or *E. coli* JCL260 pSA55 pSA69 (amp^R / kan^R) and incubated overnight at 37 °C. Growth of non-resistant strains was taken to indicate antibiotic degradation, and co-cultures were supplemented with depleted antibiotics as indicated.

2.5.11 Analytical techniques

2.5.11.1 Gravimetric analysis: flask culture samples

8 to 10 mL culture samples were washed by centrifuging at 12,000 rpm x 15 minutes, resuspending in 45 mL dH2O, and repeating centrifugation. Cell pellets were transferred to pre-weighed

aluminum boats and dried for 24 hours at 70°C; total dry mass was taken as the difference between boat with cell pellet and empty boat weight.

2.5.11.2 Gravimetric analysis: bioreactor samples

Culture samples (40 mL for batch and *E. coli* chemostat cultures; 20 mL for *T. reesei* chemostat cultures) were centrifuged at 12,000 rpm x 15 minutes and resuspended in 45 mL 20% (w/v) citric acid. Samples were then vortexed for 1 minute and incubated at room temperature for 10 minutes to dissolve phosphate and carbonate precipitates [19]. Samples were washed by centrifuging at 12,000 rpm x 15 minutes, resuspending in 45 mL dH2O, and repeating centrifugation. Cell pellets were lyophilized at 0.02 millibar / -45°C in pre-weighed tubes; total dry mass was taken as the difference between tube with cell pellet and empty tube weight.

2.5.11.3 Total secreted protein assay

Total protein concentration in culture supernatant was measured using the Bradford assay as described in [96], with bovine serum albumin (BSA) as a calibration standard. Undiluted, 1:10, $1{:}10^2$, and $1{:}10^3$ diluted samples (in 0.15 M NaCl) were analyzed; BSA calibration standards ranged from 0 to 10 μg/mL in assay mix.

2.5.11.4 β-glucosidase assay

β-glucosidase activity was assayed using p-nitrophenyl-β-D-glucopyranoside (pNPG) as a substrate, using a previously described procedure [97] with modifications. Undiluted, 1:10, and $1{:}10^2$ diluted culture samples in 50 mM citrate buffer (pH 4.8) were pre-incubated at 50°C for 5 minutes. Reactions were initiated by adding pNPG to 2.5 mM and incubating for 10 minutes at 50°C . Reactions were quenched and color developed by adding NaOH-glycine buffer (pH 10.8) to 0.2 M. Concentration of released p-nitrophenol was determined by measuring absorbance at 405 nm; p-nitrophenol calibration standards ranged from 0 to 0.25 mM in assay mix. One international unit (IU) of enzyme activity was defined as the amount of enzyme required to produce 1 μmol of p-nitrophenol per minute.

2.5.11.5 Endoglucanase assay

Endoglucanase activity was assayed using carboxymethylcellulose (CMC) as a substrate, using a previously described procedure [98] with modifications. Undiluted, 1:10, $1{:}10^2$, and $1{:}10^3$ diluted culture samples in 50 mM citrate buffer (pH 4.8) were pre-incubated at 50°C for 5 minutes. Reactions were initiated by adding CMC to 1% (w/v) and incubating for 30 minutes at 50°C.

Reactions were quenched and analyzed for reducing sugars by adding one equivalent volume of 3,5-dinitrosalicylic acid (DNS) reagent mix [98] and incubating at 95°C for 5 minutes. Concentration of 3-amino,5-nitrosalicylic acid (produced stoichiometrically from reducing sugars) was determined by measuring absorbance at 540 nm; glucose was used as a calibration standard with 0 to 1.67 mM in assay mix. One international unit (IU) of enzyme activity was defined as the amount of enzyme required to produce 1 μmol glucose-equivalent per minute.

2.5.11.6 Exoglucanase assay

Exoglucanase activity was assayed using microcrystalline cellulose as a substrate [99], with the procedure otherwise identical to the endoglucanase assay.

2.5.11.7 Carbohydrate analysis

Carbohydrate analysis of co-culture samples was performed by the Great Lakes Bioenergy Research Center (GLBRC); procedures are described in detail in [100]. Briefly, co-culture samples were washed and lyophilized as described in preceding sections for gravimetric analysis. Lyophilized co-culture material was finely ground and treated with trifluoroacetic acid (TFA) to hydrolyze the hemicellulose / matrix polysaccharide fraction [100]. The resulting soluble saccharides were derivatized to alditol acetates and analyzed via GC/MS [100]. The insoluble fraction remaining after TFA hydrolysis (consisting of crystalline cellulose, lignin, and other recalcitrant components) was analyzed for crystalline cellulose using the Updegraff procedure [100]. Our analysis methods cannot distinguish between carbohydrates originating in lignocellulose and those from microbial biomass (i.e. *T. reesei* and *E. coli*). To account for microbial carbohydrate contributions, we analyzed carbohydrate composition of pure microbial biomass samples (*T. reesei* RUTC30 grown on 20 g/L Avicel TMM media; *E. coli* K12 grown on 20 g/L glucose TMM media). Composition data for detectable carbohydrates in microbial biomass and AFEX pre-treated corn stover (AFEX CS) are shown in Table B.4. After determining microbial biomass concentrations (described below), we subtracted the microbial contribution to each measured carbohydrate:

$$S_{i,actual} = S_{i,meas} - x_{i,Ec}C_{Ec} - x_{i,Tr}C_{Tr} \tag{2.16}$$

Where $S_{i,meas}$ is the measured concentration of carbohydrate i (arabinose, xylose, mannose, galactose, hemicellulose dervied glucan, and crystalline cellulose; g/L), $x_{i,Ec}$ is the fraction of i in *E. coli* biomass (g-*i*/g-Ec), $x_{i,Tr}$ is the fraction of i in *T. reesei* biomass (g-*i*/g-Tr), $S_{i,actual}$ is the corrected carbohydrate concentration (g/L), and other terms are as described previously. Uncultivated media samples (Avicel or AFEX CS TMM media) were analyzed to determine initial concentration of major carbohydrates, $S_i|_{t_0}$ (g/L); $S_{i,actual}$ and $S_i|_{t_0}$ were used in calculation of conversions,

yields, etc.

2.5.11.8 Estimation of *T. reesei* biomass

Mannose was selected as a biomarker for *T. reesei* biomass; mannose makes up a substantial fraction of total *T. reesei* biomass while being a relatively minor component of *E. coli* and AFEX CS (Table B.4). To develop a correlation between *T. reesei* biomass and mannose, *T. reesei* RUTC30 was cultured on 20 g/L Avicel TMM media and samples were taken periodically for carbohydrate analysis. *T. reesei* fraction of total dry mass (X_{Tr}; g/g-total) was found to be linear with mannose fraction of total mass (X_{mann}; g/g-total), with $X_{Tr} = 15.02X_{mann} + 0.1173$ ($R^2 = 0.88$; $\sigma_{est} = 0.07$); results shown in Figure 2.12. *T. reesei* biomass concentration (g/L) is calculated as $C_{Tr} = X_{Tr}C_{Tot}$, where terms are as described above.

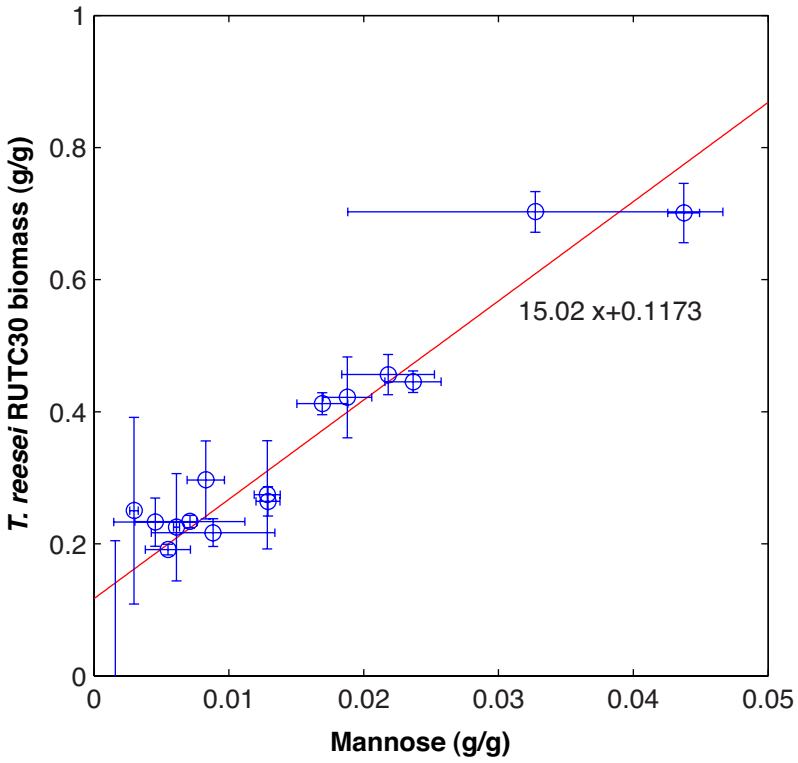

Figure 2.12: Correlation between *T. reesei* RUTC30 biomass and mannose. *T. reesei* RUTC30 was cultured on 20 g/L Avicel TMM media and samples were taken periodically for carbohydrate analysis. *T. reesei* mass fraction (i.e. non-cellulose fraction; g/g-total) shown as a function of mannose mass fraction (g/g-total); $R^2 = 0.88$ and $\sigma_{est} = 0.07$ for linear regression.

2.5.11.9 Estimation of *E. coli* biomass

E. coli biomass was estimated by cell counting (direct measurement of viable cell concentration) or via subtractive mass balance (indirect measurement of total cell concentration). For cell counting, 10 fold serial dilutions of culture samples were prepared and 100 μL aliquots plated onto LB agar; to determine pSA55/69 cell counts, LB agar was supplemented with 100 μg/mL ampicillin and 30 μg/mL kanamycin. Plates were incubated at 37°C overnight and counted. An estimated cell mass of $1.5x10^{-13}$ g/cell (determined from chemostat experimental data) was used to convert cells/L to gDW/L. To estimate *E. coli* biomass via mass balance, we write a total mass balance on insoluble co-culture components and solve for *E. coli* biomass concentration:

$$C_{Ec} = \frac{C_{Tot} - \sum S_{i,meas} - (1 - \sum x_{i,Tr})\, C_{Tr} - C_{res}}{1 - \sum x_{i,Ec}} \tag{2.17}$$

Where C_{Tot} is the total dry mass concentration (gDW/L), C_{res} is the residual insoluble fraction (i.e. undegraded lignin, etc; g/L), and the other terms are as defined previously.

2.5.11.10 Quantification of soluble saccharides and fermentation products

Isobutanol and soluble saccharide concentrations (> 50 mg/L) were quantified using high performance liquid chromatography (HPLC). Culture samples were incubated at 99°C for 10 minutes to thermally denature enzymes, and then filtered through a 0.22 μm membrane. Samples were then analyzed on an Agilent 1100 HPLC equipped with a Rezex ROA ion-exchange column, using 5 μL injection volume, 60°C column temperature, 0.005N H_2SO_4 mobile phase at 0.5 mL/min, and a refractive index detector (RID) for analyte quantification. Soluble saccharide concentrations in the range from 5 to 50 mg/L were quantified using hexokinase-based assays. Glucose was quantified with a D-Glucose HK kit (Megazyme Internatonal, cat #K-GLUHK) following manufacturer's protocol, but with reagent and sample volumes scaled down by a factor of 10. Oligosaccharide concentrations were measured via enzymatic hydrolysis to glucose and subsequent glucose quantification. Oligosaccharides were hydrolyzed by combining samples with an equivalent volume of 2 IU/mL *A. niger* β-glucosidase (Megazyme International) in 0.1 M acetate buffer (pH 4) and incubating at 49°C for 20 minutes. Samples were neutralized with 1 M NaOH and then assayed for glucose as described above. For chemostat studies, where glucose concentrations ranged < 5 mg/L, glucose was quantified using a fluorometric glucose assay kit (BioVision, cat #K606) following manufacturers protocol; depleted chemostat media was used as a blank and for preparation of calibration standards.

2.5.11.11 Identification of unknown fermentation products

HPLC analysis of *T. reesei* RUTC30 / *E. coli* NV3 pSA55/69 bicultures revealed numerous fermentation products. The four most abundant products have retention times (RT) of 15.4, 19.1, 26.3, and 39 minutes, collectively representing $> 75\%$ of the total peak area. We identified and quantified these products by running calibration standards for candidate compounds; candidates were selected on basis of expected fermentation products and a manufacturer-supplied RT index for the Aminex HPX-87H column (equivalent to Rezex ROA column; US/EG Bulletin 1847, Bio-rad) . The RT 15.4, 19.1, 26.3, and 39 min peaks were subsequently identified as succinate, acetate, ethanol, and isobutanol, respectively.

CHAPTER 3

Exploiting cooperator-cheater dynamics to develop novel mechanisms for stabilizing and tuning *T. reesei* / *E. coli* consortia

3.1 Summary

Synthetic microbial consortia with tunable and stable populations are highly desirable for biotechnology applications, but it has proven challenging to engineer consortia with these properties. In the previous chapter, we developed a synthetic *T. reesei* / *E. coli* (TrEc) consortium for consolidated bioprocessing of lignocellulosic biomass. As a continuation of this work, we identify and characterize ecological mechanisms that permit stable and tunable population compositions in the TrEc consortium, enhancing the utility of this platform for real-world bioprocessing applications. We show that cooperator-cheater dynamics within the TrEc consortium lead to stable coexistence between *T. reesei* and *E. coli*, and provide mechanisms for tuning population composition. More broadly, we suggest that cooperator-cheater dynamics could be applied as a general tool for designing and implementing stable and tunable synthetic consortia.

The majority of the work presented in this chapter has been submitted for publication to the Proceedings of the National Academy of Sciences and is currently in revision: J. Minty, M. Singer, S. Scholz, C.H. Bae, J. Ahn, C. Foster, J.C. Liao, and X. Lin. "Design and characterization of synthetic fungal-bacterial consortia for direct production of isobutanol from cellulosic biomass". *In revision (PNAS)*, 2013.

3.2 Introduction and background

Microbial consortia have tremendous potential for biotechnology applications, particularly in the area of bioprocessing. Microbial consortia with stable population compositions and function

would be highly advantageous for industrial bioprocessing (as well as other applications) and could enable lower-cost process configurations. For instance, stable microbial consortia could be used in repeated batch fermentations (in which a fraction of the previous batch is used as inoculum for new batches) or in continuous processes; operating costs for these configurations are generally lower compared to standard batch fermentation. In contrast, microbial consortia with unstable population compositions could at best only be used in standard batch processes, and each batch would require preparation of individual inoculum cultures for each organism utilized. In addition to stability, tunable population composition is also a highly desirable feature, as it would allow consortia performance to be optimized and adjusted for different products, feedstocks, and/or processing conditions. For example, in TrEc consortia the carbon flow partition between *T. reesei* and *E. coli* determines the tradeoff between cellulose hydrolysis rate and product yield (as discussed in section 2.3.2), and so population composition can potentially be tuned to achieve optimal rate and yield.

Synthetic biology frequently employs a reductionist approach of "bottom-up" engineering of biological devices from well characterized genetic parts. This powerful strategy has been successfully extended to engineering microbial consortia. Still, it is widely acknowledged that stability, robustness, and control of population composition remain key challenges in engineering synthetic consortia [24, 30, 47, 66]. Addressing the aforementioned challenges requires additional consideration of the stability of ecological interactions within engineered consortia. Evolutionary and ecology theory provides a framework to understand the stability of social interactions in microbial consortia, and indeed some of these theories have been empirically tested with synthetic microbial consortia [47]. Despite this, many efforts towards engineering synthetic consortia do not consider long-term stability. In this chapter, we seek to investigate the TrEc consortium from an ecology and evolutionary perspective in order to identify and design mechanisms that enable stable and tunable population compositions. In the next section, we provide a brief background on the current state-of-the-art in population regulation in synthetic consortia.

3.2.1 Population control and coordination in synthetic consortia

3.2.1.1 Synthetic cell-cell signaling

In nature, many microbial species communicate via specific signaling compounds, often for the purpose of sensing population levels and synchronizing behavior. Although these communication systems may serve different functions, they are generally referred to as quorum sensing (QS) systems [101]. A classic example is the use of N-acyl-homoserine lactone (AHL) QS signals

by the squid symbiont *Vibrio fischeri* to coordinate expression of bioluminescence genes after colonization of squid light organs [101]. *V. fischeri* cells produce and secrete AHLs, and once concentrations reach a threshold level, transcription of bioluminescence *lux* operon is induced [101]. Thus *V. fischeri* uses AHLs to coordinate population-dependent expression of the *lux* operon, which is only induced after a threshold population (and thus AHL level) is reached [101]. The *V. fischeri* QS system ensures that bioluminescence is only induced at the high cell densities that occur inside the squid symbiont and prevents planktonic low-density populations from wasting cellular resources on bioluminescence [101]. Numerous other examples of QS systems in nature abound, including both intra-species as well as inter-species signaling [101].

Several natural signaling systems have been co-opted for use in engineered genetic circuits, thus enabling synthetic intercellular communication. Some noteworthy examples were given in Chapter 1, including a synthetic predator-prey system (AHL signaling; Figure 1.9A) [41], multicellular Boolean logic gates (*S. cerevisiae* α factor signaling; Figure 1.9C) [43], and an oscillatory circuit synchronized across an entire cell population (AHL and H_2O_2 signaling) [44]. Synthetic intercellular communication has even been extended to inter-species and inter-kingdom signaling; for example, Weber *et al*. employed acetaldehyde-based signaling between and among mammalian, bacteria, yeast, and plant cells to engineer several demonstration synthetic ecosystems [102]. Synthetic cell-cell signaling represents a valuable tool for engineering synthetic microbial consortia. To date synthetic intercellular communication has been used to construct canonical ecological and logic systems for proof-of-concept and fundamental study, but there are few reports of using synthetic signaling in consortia for biotechnology applications, despite much interesting potential. One key challenge is that synthetic genetic circuits are subject to mutational inactivation, which often occurs in relatively few generations of growth depending on the size and host burden of the circuit [46]. As a result, consortia with programmed intercellular communication are frequently highly unstable and oftentimes cannot be cultivated outside of microfluidic devices, since larger cell populations increase the probability of competitive loss-of-function mutants [66, 103]. These stability issues need to be addressed before consortia with synthetic intercellular communication can be developed for bioprocessing applications.

3.2.1.2 Synthetic ecologies

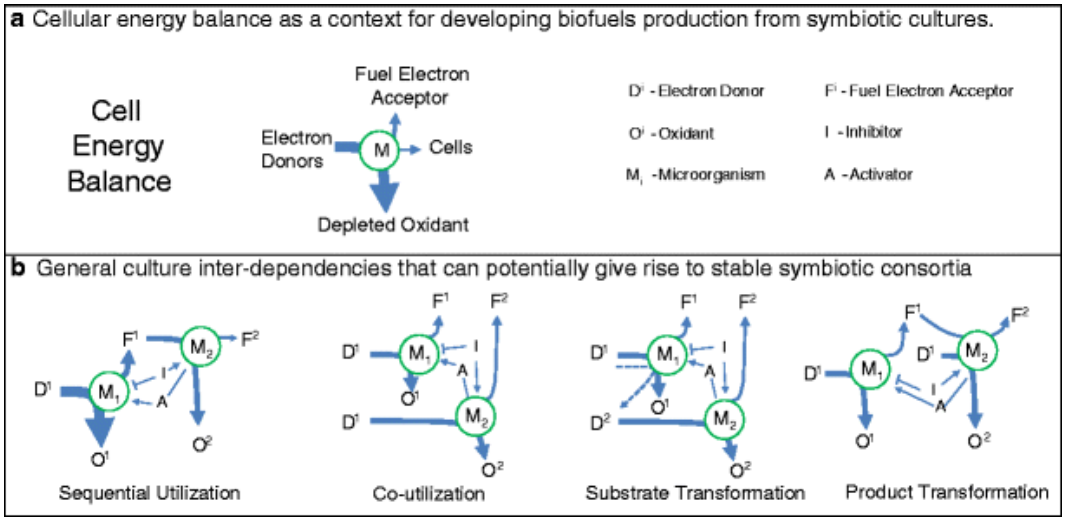

Figure 3.1: Symbiotic interaction schemes for designing biofuel producing consortia. **(A)** Energy balance for single biofuel producing microbe. **(B)** Interdependencies that could give rise to stable consortia (dual cultures shown for simplicity). Inhibitors and activators are generalized representations of various types of molecules such as quorum sensing (QS) signals and exchanged metabolites. Adapted from [30].

As a complement to synthetic genetic circuits and signaling, it is possible to engineer synthetic microbial consortia wherein populations interact via ecologically stable motifs. By incorporating designs with ecological and evolutionary stability, such systems may stably persist over long time scales. Furthermore, in principle it should be possible to tune population composition by modulating ecological interaction parameters (e.g. genetically or through environmental manipulation). One broad means of achieving ecological stability is to engineer mutualistic interactions between consortia members, wherein the different consortia members are interdependent on one another. There are a number of symbiotic interaction topologies that could be used to design synthetic consortia. Some examples relevant to consolidated bioprocessing applications include sequential substrate utilization, co-utilization of substrate, substrate transformation, and product transformation (Figure 3.1) [30]. We note that while the examples to be discussed in this section involve two-species consortia, these general interaction patterns could be extended to more members.

In sequential substrate utilization, one species metabolizes substrate to waste products that serve as substrate for a second species; in product transformation, the second species co-utilizes both the primary substrate and waste products (Figure 3.1B) [30]. Sequential utilization and product transformation interactions can be mutualistic if the waste products produced by the first species are toxic, since the first species provides substrate for the second while the second species aids the community by removing toxins. The product transformation approach was demonstrated

by Bayer *et al*. in engineering a consortium for production of methyl halides from cellulose; the cellulolytic microbe *Actinotalea fermentans* ferments cellulose to toxic waste products ethanol and acetate, which are in turn converted to methyl halides by an engineered *S. cerevisiae* strain that also presumably co-utilizes cellulose hydrolysis products (Figure 3.2A) [32]. Bernstein *et al*. employed the sequential utilization scheme in engineering a two-member synthetic *E. coli* consortium for serial degradation of glucose (Figure 3.2B) [104]. A primary producer metabolizes glucose to acetate and other waste products, while a glucose-negative *E. coli* strain consumes the products of the primary producer (Figure 3.2B) [104]. The consortium has higher biomass productivity compared to monocultures of either member strain, illustrating the mutualistic nature of the system [104]. While the product transformation and sequential substrate utilization motifs employed in these two examples are expected to be stable, neither of these studies examined population dynamics over long time scales.

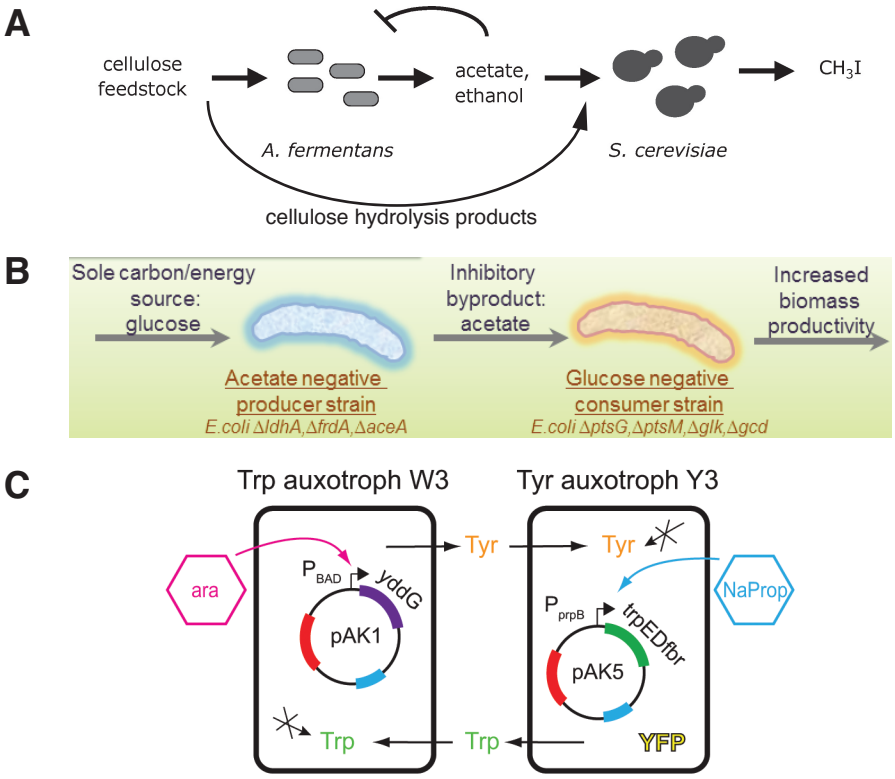

Figure 3.2: Synthetic consortia based on mutualistic interactions. **(A)** Product transformation system for conversion of cellulose to methyl halides. *Actinotalea fermentans* ferments cellulose to ethanol and acetate, which are in turn converted to methyl halides by an engineered *S. cerevisiae* (co-utilization of cellulose hydrolysis products is also inferred). Consumption of ethanol/acetate ameliorates toxicity effects on *Actinotalea fermentans*. Adapted from [32]. **(B)** Two-member synthetic *E. coli* consortium for sequential degradation of glucose. The primary producer metabolizes glucose to acetate and other waste products, while a glucose-negative *E. coli* strain consumes the products of the primary producer, alleviating inhibition and thus enhancing biomass productivity. Adapted from [105]. **(C)** Tunable symbiosis between two *E. coli* strains auxotrophic for Trp (strain W3) and Tyr (strain Y3). Equilibrium population composition depends on the growth requirements and export rates for each amino acid; by controlling amino acid export rates (via transcriptional regulation of plasmid-based exporters or pathways), population composition can be tuned. Adapted from [106].

In addition to the interaction topologies shown in Figure 3.1, cross-feeding, wherein species are obligately interdependent on one another for nutrients or growth substrates, represents another important class of ecologically stable interactions; substrate transformation (Figure 3.1B) is a specific instance of this general topology. An example of cross-feeding was previously discussed in Chapter 1, where Shou *et al*. engineered a synthetic cross-feeding symbiosis between two *S. cerevisiae* strains (Figure 1.9B) [42]. In another noteworthy example, Kerner *et al*. engineered a tunable symbiosis between two *E. coli* strains (Figure 3.2C) [106]. Each strain is auxotrophic for a different amino acid, and when co-cultured the strains cross-feed each other (Figure 3.2C) [106]. In minimal media the strains are mutually dependent on one another for survival, and the equilibrium population composition depends on the growth requirements and export rates for each amino acid [106]. By modulating amino acid export rates, Kerner *et al*. were able to tune the

co-culture composition [106]. This study demonstrates that ecological parameters can be adjusted to tune the population composition of a synthetic consortium. Furthermore, the cross-feeding scheme used in this work was shown to be stable over time-scales on the order of hundreds of generations (N. Lin, personal communication).

The synthetic consortia discussed above are all based on cooperative interactions between different species or strains. However, other stable interaction topologies are possible, and we have only begun to explore the full range of possible ecological designs for synthetic consortia. Furthermore, mechanisms for population coordination have not yet been applied to biofuel producing consortia, despite the apparent importance of this design parameter. In the next section, we examine the TrEc consortium from an evolutionary game theory perspective and discuss possible ecological mechanisms that may permit stable and tunable population compositions.

3.2.2 TrEc consortia: an evolutionary game theory perspective

Evolutionary and ecology theory provide a framework for understanding and predicting the outcomes of social interactions in microbial consortia. We can apply evolutionary game theory models from this framework to identify or design mechanisms that enable stable and tunable population compositions in the TrEc consortium. In the context of game theory, *T. reesei* acts as a cooperator and *E. coli* as an obligate cheater within the TrEc consortium. Cellulase secretion by *T. reesei* is a cooperative behavior since cellulase production is metabolically expensive and *T. reesei* does not have exclusive access to cellulose hydrolysis products. *E. coli* behaves as a cheater by utilizing cellulose hydrolysis products without bearing the burden of cellulase production. Game theory suggests that cooperators and cheaters can stably coexist provided that cooperators receive high enough net benefits in their interactions with cheaters, a defining feature of the so-called "Snowdrift game" [107]. Cooperator-cheater coexistance was recently demonstrated in *S. cerevisiae* populations growing on sucrose (Figure 3.3) [107]. For *S. cerevisiae* to utilize sucrose, this disaccharide must first be hydrolyzed by invertase [107]. Sucrose hydrolysis occurs in the periplasmic space and 99% of the hydrolysis products diffuse before they can be imported by the cell, thus making invertase production in *S. cerevisiae* a cooperative behavior (Figure 3.3A) [107]. Experimental and theoretical studies with cooperating (invertase secreting) and cheating (invertase negative) *S. cerevisiae* demonstrated that cooperator-cheater coexistance is possible if i) cooperators have privileged access to hydrolysis products and ii) fitness benefits are a concave function of substrate concentration [107] (Figure 3.3B&C). Furthermore, it was demonstrated that equilibrium population composition depended on the relative the costs and

benefits of cooperation (Figure 3.3C) [107].

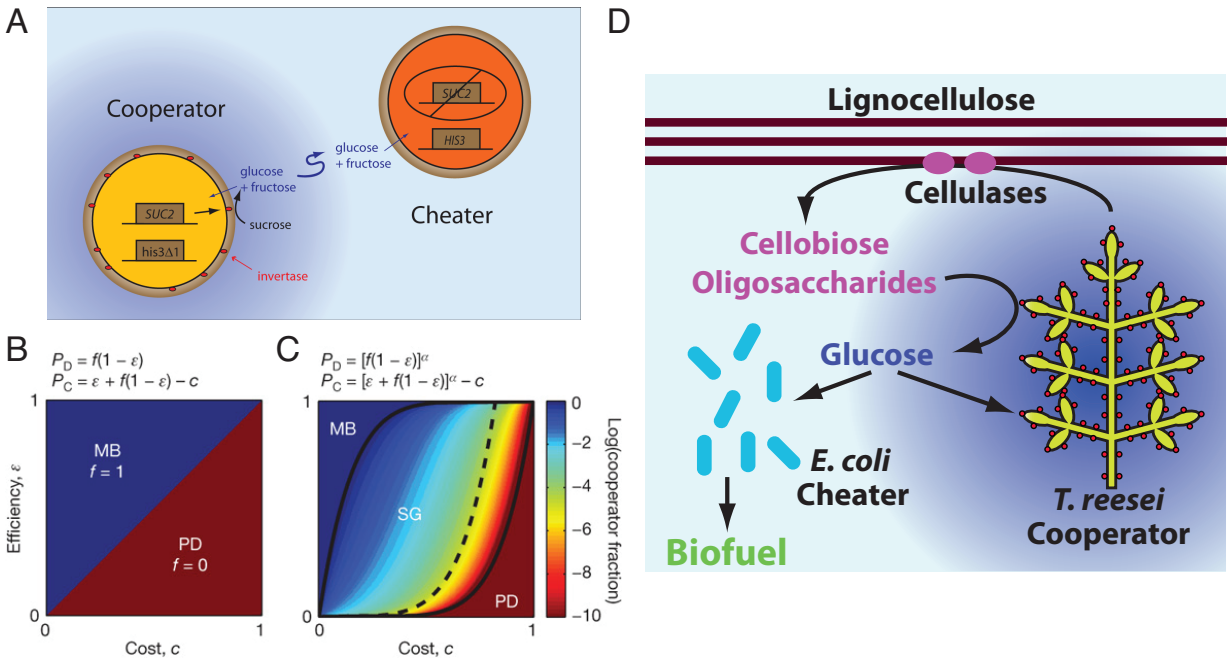

Figure 3.3: Cooperater-cheater coexistence. Panels A, B, and C adapted from [107]. **(A)** Sucrose metabolism in *S. cerevisiae*. Cooperator strains produce periplasmic invertase, which hydrolyzes sucrose to glucose and fructose. 99% of the hydrolysis products diffuse before they can be imported by the cell, however periplasmic localization of invertase leads to increased glucose/fructose concentrations near cooperator cells thus affording them privileged access. Cheater cells are invertase negative, and thus consume glucose/fructose without bearing the burden of invertase production. **(B)** Defection (cheating) and cooperation payouts, P_D and P_C respectively, and fraction of cooperators f at equilibrium for a linear model in which cooperation has cost c and leads to total benefits of unity that are captured with an efficiency ϵ. This model leads to a mutually beneficial (MB) game (i.e. fixation of cooperators, $f = 1$) at $\epsilon > c$, while $\epsilon < c$ leads to a prisoner's dilemma (PD) game (i.e. fixation of defectors). **(C)** A modified model with concave benefits ($\alpha < 1$; $\alpha = 0.15$ in figure) yields a central region of parameter space that is a Snowdrift (SG) game (i.e. stable coexistence of cooperators and cheaters, $0 < f < 1$). **(D)** Characteristics of the TrEc consortium that may give rise to cooperator-cheater dynamics.

The TrEc consortium will likely satisfy the requirements for cooperator-cheater coexistence discussed above (Figure 3.3D). *T. reesei* produces cell wall associated β-glucosidases [81], which probably afford privileged access to glucose in a manner similar to periplasmic invertase in *S. cerevisiae* sucrose metabolism (Figure 3.3D). Additionally, *T. reesei* forms biofilms on cellulose substrate surfaces under certain culture conditions, which could provide an additional means for privileged access to hydrolysis products [108]. For both *T. reesei* and *E. coli*, growth rates on soluble saccharides are approximately concave functions of concentration (e.g. Monod kinetics), and thus the requirement for concave fitness benefits is automatically satisfied. Since equilibrium population composition has been shown to depend on relative cooperation/cheating benefits [107], it may be possible to modulate these parameters to tune composition. The TrEc consortium has numerous complexities not present in the *S. cerevisiae* system, such as complex substrate

hydrolysis kinetics, isobutanol production and toxicity, extremely low growth rates and soluble saccharide concentrations, and potential unidentified interactions between *T. reesei* and *E. coli*. In this chapter, we investigate cooperator-cheater dynamics in the TrEc consortium and explore tuning population composition via modulation of relative cooperation/cheating benefits.

3.3 Results

3.3.1 Simplified TrEc consortium model for stability analysis

We developed an abridged model of the TrEc consortium to facilitate stability analysis and investigation of cooperator-cheater dynamics within the TrEc consortium. We model organism growth and substrate uptake with Monod kinetics; for simplicity we neglect maintenance substrate uptake. To reduce model complexity, we simplified cellulase kinetics to a single Michaelis-Menten rate law in terms of total cellulose concentration and total cellulase concentration; we assume cellulose is hydrolyzed to cellobiose, which is then hydrolyzed to glucose via β-glucosidase; all other polysaccharide intermediates are lumped into the cellulose term. A description of the model is given in the following sections.

3.3.1.1 *T. reesei* growth and substrate uptake

We model *T. reesei* growth and substrate uptake using Monod kinetics. For simplicity we neglect different mycelium types and assume that glucose is the sole growth substrate. An important subtlety is that cellobiose is hydrolyzed to glucose via cell-wall localized β-glucosidase of *T. reesei* [81], as depicted in Figure 3.3. This leads to locally increased concentration of glucose at the cell surface relative to the bulk media, thus affording privileged access to *T. reesei*. **We performed a mass-transfer analysis to estimate the concentration of glucose at the cell surface; see section A.2.2 for derivation.**

$$\frac{dC_{Tr}}{dt} = \frac{\mu_{max,Tr}\left(S_{G_1} + \theta_{G_2 \to G_1} S_{G_2}\right)}{K_{S,Tr} + S_{G_1} + \theta_{G_2 \to G_1} S_{G_2}} C_{Tr} \tag{3.1}$$

$$r^{Tr}_{S_{G_1}} = Y_{S_{G_1}/C_{Tr}} \mu_{Tr} C_{Tr} \tag{3.2}$$

$$= Y_{S_{G_1}/C_{Tr}} \frac{\mu_{max,Tr}\left(S_{G_1} + \theta_{G_2 \to G_1} S_{G_2}\right)}{K_{S,Tr} + S_{G_1} + \theta_{G_2 \to G_1} S_{G_2}} C_{Tr} \tag{3.3}$$

Where all terms are analogous to those defined in section 2.3.1.

3.3.1.2 *E. coli* growth and substrate uptake

We model *E. coli* growth and substrate uptake using Monod kinetics:

$$\frac{dC_{Ec}}{dt} = \frac{\mu_{max,Ec} S_{G_1}}{K_{S,Ec} + S_{G_1}} C_{Ec} \tag{3.4}$$

$$r^{Ec}_{S_{G_1}} = Y_{S_{G_1}/C_{Ec}} \mu_{Ec} C_{Ec} \tag{3.5}$$

$$= Y_{S_{G_1}/C_{Ec}} \frac{\mu_{max,Ec} S_{G_1}}{K_{S,Ec} + S_{G_1}} C_{Ec} \tag{3.6}$$

Where all terms are analogous to those defined in section 2.3.1.

3.3.1.3 Enzyme production by *T. reesei*

We assume that enzyme secretion by *T. reesei* is directly coupled to growth:

$$\frac{dE_{cel}}{dt} = Y_{E_{cel}/C_{Tr}} \mu_{Tr} C_{Tr} \tag{3.7}$$

$$= Y_{E_{cel}/C_{Tr}} \frac{\mu_{max,Tr} \left(S_{G_1} + \theta_{G_2 \to G_1} S_{G_2}\right)}{K_{S,Tr} + S_{G_1} + \theta_{G_2 \to G_1} S_{G_2}} C_{Tr} \tag{3.8}$$

and

$$\frac{dE_{BGL}}{dt} = Y_{E_{BGL}/C_{Tr}} \mu_{Tr} C_{Tr} \tag{3.9}$$

$$= Y_{E_{BGL}/C_{Tr}} \frac{\mu_{max,Tr} \left(S_{G_1} + \theta_{G_2 \to G_1} S_{G_2}\right)}{K_{S,Tr} + S_{G_1} + \theta_{G_2 \to G_1} S_{G_2}} C_{Tr} \tag{3.10}$$

Where E_{cel} is total cellulase (i.e. endoglucanase and exoglucanase) concentration (g/L), E_{BGL} is total β-glucosidase concentration (g/L), $Y_{E_{cel}/C_{Tr}}$ is the cellulase / biomass yield coefficient (g-cellulase/g-biomass), $Y_{E_{BGL}/C_{Tr}}$ is the β-glucosidase / biomass yield coefficient (g-β-glucosidase/g-biomass) and other terms are as defined in section 2.3.1. Due to the direct coupling between *T. reesei* growth and enzyme production, we can write enzyme concentration as:

$$E_{cel} = Y_{E_{cel}/C_{Tr}} C_{Tr} \tag{3.11}$$

$$E_{BGL} = Y_{E_{BGL}/C_{Tr}} C_{Tr} \tag{3.12}$$

3.3.1.4 Cellulose hydrolysis

We describe cellulase kinetics using a Michaelis-Menten rate law. For simplicity we neglect different types of cellulases, declining reactivity with conversion, and product inhibition, and lump all polysaccharides into a single cellulose concentration variable:

$$r_{S_{G_2}}^{cel} = -r_C^{cel} = \frac{k_{cel} E_{cel} S_C}{K_{M,cel} + S_C} \tag{3.13}$$

Where $r_{S_{G_2}}^{cel}$ is cellobiose production rate (g/L/h), r_C^{cel} is cellulose hydrolysis rate (g/L/h), k_{cel} is the rate constant (g/g-cellulase/h), S_C is total cellulose concentration (all polysaccharides with $DP > 2$; g/L), and $K_{M,cel}$ is cellulase affinity (g/L).

3.3.1.5 Cellobiose hydrolysis

We describe β-glucosidase kinetics using a Michaelis-Menten rate law:

$$r_{S_{G_1}}^{BGL} = -r_{S_{G_2}}^{BGL} = \frac{k_{BGL} E_{BGL} S_{G_2}}{K_{M,BGL} + S_{G_2}} \tag{3.14}$$

Where $r_{S_{G_1}}^{BGL}$ is glucose production rate (g/L/h), k_{BGL} is the rate constant (g/g-β-glucosidase/h), $K_{M,BGL}$ is β-glucosidase affinity (g/L), and other terms are as described in section 2.3.1.

3.3.1.6 Cellobiose mass balance

We write a mass balance on cellobiose accounting for production via cellulose hydrolysis and consumption via β-glucosidase hydrolysis:

$$\frac{dS_{G_2}}{dt} = r_{S_{G_2}}^{cel} - r_{S_{G_2}}^{BGL} \tag{3.15}$$

$$= \frac{k_{cel} E_{cel} S_C}{K_{M,cel} + S_C} - \frac{k_{BGL} E_{BGL} S_{G_2}}{K_{M,BGL} + S_{G_2}} \tag{3.16}$$

Where all terms are as described in section 2.3.1.

3.3.1.7 Glucose mass balance

We write a mass balance on glucose accounting for production via cellobiose hydrolysis and microbial consumption:

$$\frac{dS_{G_1}}{dt} = r^{BGL}_{S_{G_1}} - r^{Ec}_{S_{G_1}} - r^{Tr}_{S_{G_1}} \tag{3.17}$$

$$= \frac{k_{BGL}E_{BGL}S_{G_2}}{K_{M,BGL} + S_{G_2}} - Y_{S_{G_1}/C_{Ec}}\frac{\mu_{max,Ec}S_{G_1}}{K_{S,Ec} + S_{G_1}}C_{Ec} \tag{3.18}$$

$$- Y_{S_{G_1}/C_{Tr}}\frac{\mu_{max,Tr}\left(S_{G_1} + \theta_{G_2\rightarrow G_1}S_{G_2}\right)}{K_{S,Tr} + S_{G_1} + \theta_{G_2\rightarrow G_1}S_{G_2}}C_{Tr}$$

Where all terms are as described in section 2.3.1.

3.3.2 Steady state analysis of simplified TrEc model

3.3.2.1 Criteria for steady state population composition

The system of ODEs for the simplified TrEc consortium model does not have a non-trivial steady-state. However, we can consider the case of a pseudo-steady state in which microbial population composition is fixed:

$$\frac{d}{dt}\left(\frac{C_{Ec}}{C_{Tr}}\right) = 0 \tag{3.19}$$

Simplifying and rearranging:

$$\frac{1}{C_{Ec}}\frac{dC_{Ec}}{dt} - \frac{1}{C_{Tr}}\frac{dC_{Tr}}{dt} = 0 \tag{3.20}$$

...

$$\frac{\mu_{max,Ec}S_{G_1}}{K_{S,Ec} + S_{G_1}} = \frac{\mu_{max,Tr}\left(S_{G_1} + \theta_{G_2\rightarrow G_1}S_{G_2}\right)}{K_{S,Tr} + S_{G_1} + \theta_{G_2\rightarrow G_1}S_{G_2}} \tag{3.21}$$

This pseudo-steady state can be satisfied if S_{G_2} and S_{G_1} are constant:

$$\frac{dS_{G_2}}{dt} = r^{cel}_{S_{G_2}} - r^{BGL}_{S_{G_2}} = 0 \tag{3.22}$$

$$= \frac{k_{cel}E_{cel}S_C}{K_{M,cel} + S_C} - \frac{k_{BGL}E_{BGL}S_{G_2}}{K_{M,BGL} + S_{G_2}} = 0 \tag{3.23}$$

and

$$\frac{dS_{G_1}}{dt} = r_{S_{G_1}}^{BGL} - r_{S_{G_1}}^{Ec} - r_{S_{G_1}}^{Tr} = 0 \tag{3.24}$$

$$= \frac{k_{BGL} E_{BGL} S_{G_2}}{K_{M,BGL} + S_{G_2}} - Y_{S_{G_1}/C_{Ec}} \frac{\mu_{max,Ec} S_{G_1}}{K_{S,Ec} + S_{G_1}} C_{Ec} - Y_{S_{G_1}/C_{Tr}} \frac{\mu_{max,Tr} \left(S_{G_1} + \theta_{G_2 \to G_1} S_{G_2}\right)}{K_{S,Tr} + S_{G_1} + \theta_{G_2 \to G_1} S_{G_2}} C_{Tr} = 0 \tag{3.25}$$

The above expressions cannot be satisfied since C_{Ec} , C_{Tr}, E_{cel}, E_{BGL}, and S_C are all potentially changing with time. We can make a simplifying assumption that for all times $S_C \gg K_{M,cel}$, making it possible for the system to reach a pseudo-steady state. Physically, this could be interpreted as having an unlimited cellulose supply, or alternately that the TrEc consortium is serial passaged before cellulose becomes limiting. Assuming $S_C \gg K_{M,cel}$, substituting $E_{cel} = Y_{E_{cel}/C_{Tr}} C_{Tr}$, $E_{BGL} = Y_{E_{BGL}/C_{Tr}} C_{Tr}$, $C_{Ec} = X\left(C_{Ec} + C_{Tr}\right)$, and $C_{Tr} = (1 - X)\left(C_{Ec} + C_{Tr}\right)$ (where X is *E. coli* population fraction; g-*E. coli*/g-total), we can further simplify:

$$\frac{dS_{G_2}}{dt} = k_{cel} Y_{E_{cel}/C_{Tr}} (1 - X) - \frac{k_{BGL} Y_{E_{BGL}/C_{Tr}} (1 - X) S_{G_2}}{K_{M,BGL} + S_{G_2}} = 0 \tag{3.26}$$

and

$$\frac{dS_{G_1}}{dt} = \frac{k_{BGL} Y_{E_{BGL}/C_{Tr}} (1 - X) S_{G_2}}{K_{M,BGL} + S_{G_2}} - Y_{S_{G_1}/C_{Ec}} \frac{\mu_{max,Ec} S_{G_1}}{K_{S,Ec} + S_{G_1}} X - Y_{S_{G_1}/C_{Tr}} \frac{\mu_{max,Tr} \left(S_{G_1} + \theta_{G_2 \to G_1} S_{G_2}\right)}{K_{S,Tr} + S_{G_1} + \theta_{G_2 \to G_1} S_{G_2}} (1 - X) = 0 \tag{3.27}$$

Note that $\frac{d}{dt}\left(\frac{C_{Ec}}{C_{Tr}}\right) = 0$ implies that X is constant. The three criteria for steady-state population composition are then:

Criterion 1: Constant population fraction

$$\frac{\mu_{max,Ec} S_{G_1}}{K_{S,Ec} + S_{G_1}} = \frac{\mu_{max,Tr} \left(S_{G_1} + \theta S_{G_2}\right)}{K_{S,Tr} + S_{G_1} + \theta S_{G_2}} \tag{3.28}$$

Criterion 2: Constant cellobiose concentration

$$k_{cel} Y_{E_{cel}/C_{Tr}} - \frac{k_{BGL} Y_{E_{BGL}/C_{Tr}} S_{G_2}}{K_{M,BGL} + S_{G_2}} = 0 \tag{3.29}$$

Criterion 3: Constant glucose concentration

$$\frac{k_{BGL}Y_{E_{BGL}/C_{Tr}}\left(1-X\right)S_{G_2}}{K_{M,BGL}+S_{G_2}}-Y_{S_{G_1}/C_{Ec}}\frac{\mu_{max,Ec}S_{G_1}}{K_{S,Ec}+S_{G_1}}X$$
$$-Y_{S_{G_1}/C_{Tr}}\frac{\mu_{max,Tr}\left(S_{G_1}+\theta_{G_2\rightarrow G_1}S_{G_2}\right)}{K_{S,Tr}+S_{G_1}+\theta_{G_2\rightarrow G_1}S_{G_2}}\left(1-X\right)=0 \tag{3.30}$$

The steady-state criteria constitute a system of three equations with three unknown variables (X, S_{G_2}, S_{G_1}) that can be solved with analytical or numerical methods.

3.3.2.2 Non-dimensionalization and analytical solution

To facilitate analysis, we can simplify the above steady-state criteria by non-dimensionalizing the variables. We start with criterion 2, the simplest of the above expressions. Applying $K_{S,Tr}$ as the characteristic cellobiose concentration scale:

$$k_{cel}Y_{E_{cel}/C_{Tr}}-\frac{k_{BGL}Y_{E_{BGL}/C_{Tr}}S_2^*}{K_{BGL}^*+S_2^*}=0 \tag{3.31}$$

Where $S_2^*=S_{G_2}/K_{S,Tr}$ and $K_{BGL}^*=K_{M,BGL}/K_{S,Tr}$. This expression can be readily solved for S_2^*:

$$S_2^*=\frac{K_{BGL}^*}{\alpha-1} \tag{3.32}$$

$$\alpha=Y_{BGL/cel}\frac{k_{BGL}}{k_{cel}} \tag{3.33}$$

$$Y_{BGL/cel}=Y_{E_{BGL}/C_{Tr}}/Y_{E_{cel}/C_{Tr}} \tag{3.34}$$

As would be intuitively expected, the steady-state cellobiose concentration is thus purely a function of cellulase and β-glucosidase kinetics and the relative proportion of these two enzymes. The dimensionless term α describes the relative ratio of β-glucosidase activity to cellulase activity.

Applying $K_{S,Tr}$ as the characteristic glucose and cellobiose concentration scale and $\mu_{max,Tr}$ as the characteristic time scale to criterion 1 yields:

$$\mu_{Ec}^*=\frac{\mu^*S_1}{K_S^*+S_1^*}=\mu_{Tr}^*=\frac{S_1^*+\theta_{G_2\rightarrow G_1}S_2^*}{1+S_1^*+\theta_{G_2\rightarrow G_1}S_2^*} \tag{3.35}$$

Where $S_1^*=S_{G_1}/K_{S,Tr}$, $K_S^*=K_{S,Ec}/K_{S,Tr}$, $\mu^*=\mu_{max,Ec}/\mu_{max,Tr}$, and other terms are as described previously. Solving for S_1^* yields:

$$S_1^* = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} \tag{3.36}$$

Where $a = \mu^* - 1$, $b = \mu^* \left(1 + \theta_{G_2 \to G_1} S_2^*\right) - \theta_{G_2 \to G_1} S_2^* - K_S^*$, $c = -\theta_{G_2 \to G_1} S_2^* K_S^*$, and other terms are as described previously. The physically meaningful root(s) of the above equation satisfy $0 \leq S_1^* < \infty$. With expressions for S_1^* and S_2^*, we can solve criterion 3 for X. We first apply the above dimensional scalings to criterion 3 and divide by $Y_{S_{G_1}/C_{Tr}}$ to yield:

$$\frac{k_{BGL} Y_{E_{BGL}/S_{G_1}} S_2^*}{\mu_{max,Tr} \left(K_{BGL}^* + S_2^*\right)} (1 - X) - Y_{Tr/Ec} \frac{\mu^* S_1^*}{K_S^* + S_1^*} X - \frac{S_1^* + \theta_{G_2 \to G_1} S_2^*}{1 + S_1^* + \theta_{G_2 \to G_1} S_2^*} (1 - X) = 0 \tag{3.37}$$

Where $Y_{E_{BGL}/S_{G_1}} = Y_{E_{BGL}/C_{Tr}} / Y_{S_{G_1}/C_{Tr}}$ and $Y_{Tr/Ec} = Y_{S_{G_1}/C_{Ec}} / Y_{S_{G_1}/C_{Tr}}$. Solving for X and simplifying yields:

$$X = \frac{\mu_{Tr}^* - \beta r_{BGL}^*}{\mu_{Tr}^* - Y_{Tr/Ec} \mu_{Ec}^* - \beta r_{BGL}^*} \tag{3.38}$$

Where:

$$\mu_{Tr}^* = \frac{S_1^* + \theta_{G_2 \to G_1} S_2^*}{1 + S_1^* + \theta_{G_2 \to G_1} S_2^*} \tag{3.39}$$

$$\mu_{Ec}^* = \frac{\mu^* S_1^*}{K_S^* + S_1^*} \tag{3.40}$$

$$r_{BGL}^* = \frac{S_2^*}{K_{BGL}^* + S_2^*} \tag{3.41}$$

$$\beta = \frac{k_{BGL} Y_{BGL/S_{G_1}}}{\mu_{max,Tr}} \tag{3.42}$$

The μ_{Tr}^* term represents dimensionless *T. reesei* growth rate, μ_{Ec}^* is dimensionless *E. coli* growth rate, and βr_{BGL}^* is a dimensionless ratio of cellobiose hydrolysis rate to glucose uptake rate by *T. reesei*. $X > 0$ requires that $\mu^* > Y_{Tr/Ec} \mu^* + \beta r_{BGL}^*$.

Steady state carbon flow partition can be calculated as:

$$P_{C \to Ec} = \frac{Y_{Tr/Ec} \mu_{Ec}^* X}{Y_{Tr/Ec} \mu_{Ec}^* X + \mu_{Tr}^* (1 - X)} \tag{3.43}$$

Where $P_{C \to Ec}$ is the ratio of glucose uptake by *E. coli* to total microbial glucose uptake.

3.3.3 Theoretical analysis of cooperator-cheater dynamics in TrEc consortia

We used the modeling framework developed in the previous section to explore cooperator-cheater dynamics in TrEc consortia. We analyzed RUTC30/K12 bicultures as a reference cases, as well as a modified model for NV3/RUTC30 bicultures that includes isobutanol production and toxicity effects.

3.3.3.1 RUTC30/K12 bicultures

For the dimensionless pseudo-steady state TrEc consortium model, equilibrium population composition (in terms of *E. coli* fraction X_{Ec}) is a function of 7 different parameter values. Parameter definitions and descriptions are summarized in Table 3.1, while baseline parameter values (corresponding to *T. reesei* RUTC30 and *E. coli* K12) are given in Table D.1. It may be possible to modulate some or all of these parameters to tune the population composition of the TrEc consortium. To explore the cooperator-cheater tuning concept, we used the dimensionless TrEc consortium model to examine population dynamics/equilibria and carbon flow partition over a design space of plausible parameter values.

Parameter	Definition	Description
μ^*	$\frac{\mu_{max,Ec}}{\mu_{max,Tr}}$	Ec:Tr growth rate ratio (cheater benefits)
K_S^*	$\frac{K_{S,Ec}}{K_{S,Tr}}$	Ec:Tr glucose affinity ratio (inverse cheater benefits)
$\theta_{G_2 \to G_1}$	$\frac{\Delta S_{G_1}}{S_{G_2}}$	Increase in glucose concentration relative to the bulk medium at the *T. reesei* cell surface due to cellobiose hydrolysis (cooperator privileged access)
$Y_{Tr/Ec}$	$\frac{Y_{S_{G_1}/C_{Ec}}}{Y_{S_{G_1}/C_{Tr}}}$	*T. reesei* biomass-substrate yield coefficient to *E. coli* yield coefficient ratio (relative resource utilization efficiency of cooperator)
β	$\frac{k_{BGL} Y_{BGL/S_{G_1}}}{\mu_{max,Tr}}$	Glucose production rate to *T. reesei* uptake rate ratio (relative resource production/consumption by cooperator)
α	$Y_{BGL/cel} \frac{k_{BGL}}{k_{cel}}$	Cellobiose hydrolysis rate to cellulose hydrolysis rate ratio (relative consumption/production of intermediate substrate)
K_{BGL}^*	$\frac{K_{M,BGL}}{K_{S,Tr}}$	Dimensionless β-glucosidase cellobiose affinity

Table 3.1: Summary of dimensionless parameters in TrEc consortium model. All terms are as described in section 3.3.1. See Table D.1 for baseline values for *T. reesei* RUTC30 and *E. coli* K12.

Our analysis suggests that a wide range of equilibrium population compositions can be achieved by tuning μ^*, K_S^*, $\theta_{G_2 \to G_1}$, and/or β (Figure 3.4); for sake of simplicity, we held $Y_{Tr/Ec}$, α, and K_{BGL}^* constant, though in principle these parameters could be tuned as well. Carbon flow partition (fraction of substrate carbon consumed by *E. coli*; $P_{C \to Ec}$) is proportional to

population composition, underscoring that population tuning could be used to directly control biosynthesis of desired products (Figure D.1). Equilibrium X_{Ec} increases with β, reflecting that growth substrate available to the cheater increases as relative substrate production (compared to substrate consumption) by the cooperator increases (Figure 3.4). Furthermore, for a given parameter set there is likely a minimum β value below which stable coexistence is not possible and *E. coli* becomes extinct (Figure 3.4). Parameters associated with relative cooperation/cheating benefits affect population equilibria in an intuitive manner; increasing cooperator privileged access $\theta_{G_2 \to G_1}$ tends to decrease equilibrium X_{Ec}, while equilibrium X_{Ec} tends to increase with increasing cheating benefits (increasing μ^* or decreasing K_S^*) (Figure 3.4). Equilibrium X_{Ec} may be asymptotic with μ^*, and furthermore μ^* has little effect on population equilibria at extreme values of K_S^* (i.e. when K_S^* is the dominant term) (Figure 3.4).

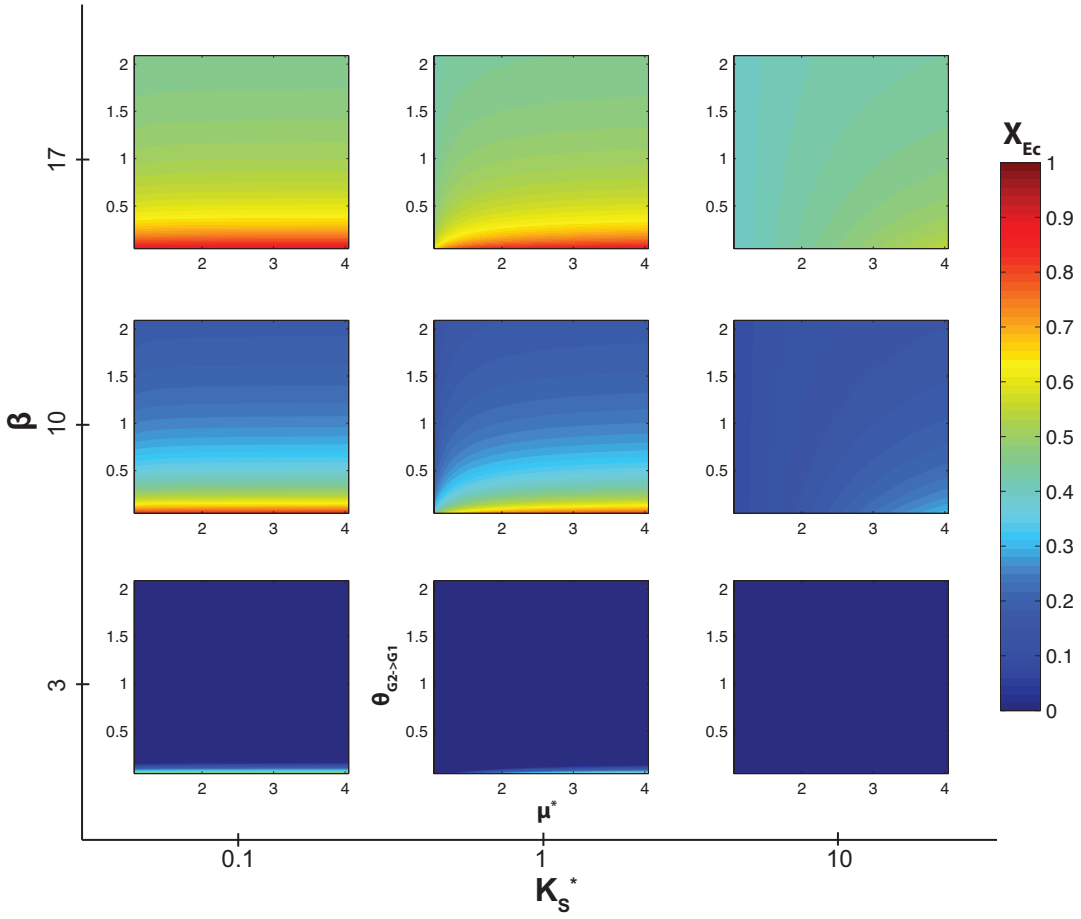

Figure 3.4: Steady state analysis of simplified RUTC30/K12 TrEc consortium model. X_{Ec} (*E. coli* population fraction; g/g-total) as a function of β, K_S^*, μ^*, and $\theta_{G_2 \to G_1}$ over biologically reasonable ranges for each parameter. Parameter values are as follows: $\alpha = 8.7$, $K_{BGL}^* = 23.25$, $Y_{Tr/Ec} = 1.5$; $\beta = 3, 10, 17$, $K_S^* = 0.1, 1, 10$, $\mu^* = 1.05...4.05$, and $\theta_{G_2 \to G_1} = 0.05...2.05$. For explanation of parameters see section 3.3.1 and Table 3.1.

In addition to equilibria, we also used our modeling framework to explore population stability and dynamics. In principle, stability can be determined from the eigenvalues of an ODE system linearized around fixed points. However, due to the complexity of our model these calculations are intractable. Instead, we conducted a numerical analysis of the ODE system presented in section 3.3.1, with a simplifying assumption that for all times $S_C \gg K_{M,cel}$ (see section 3.3.2.1 for discussion). In dynamic simulations over the μ^* and $\theta_{G_2 \to G_1}$ space (holding other parameters constant), equilibria were reached within 12 to 60 doubling times regardless of initial conditions, suggesting that equilibrium states are both stable and readily accessible (sample trajectories shown in Figure 3.5). Population trajectories tended to undergo damped oscillations as they converged to equilibrium (possibly due to feedback delay between cellulose hydrolysis, soluble saccharide concentrations, and microbial growth), with equilibration time increasing as $\mu^* \to 1$ (Figure 3.5).

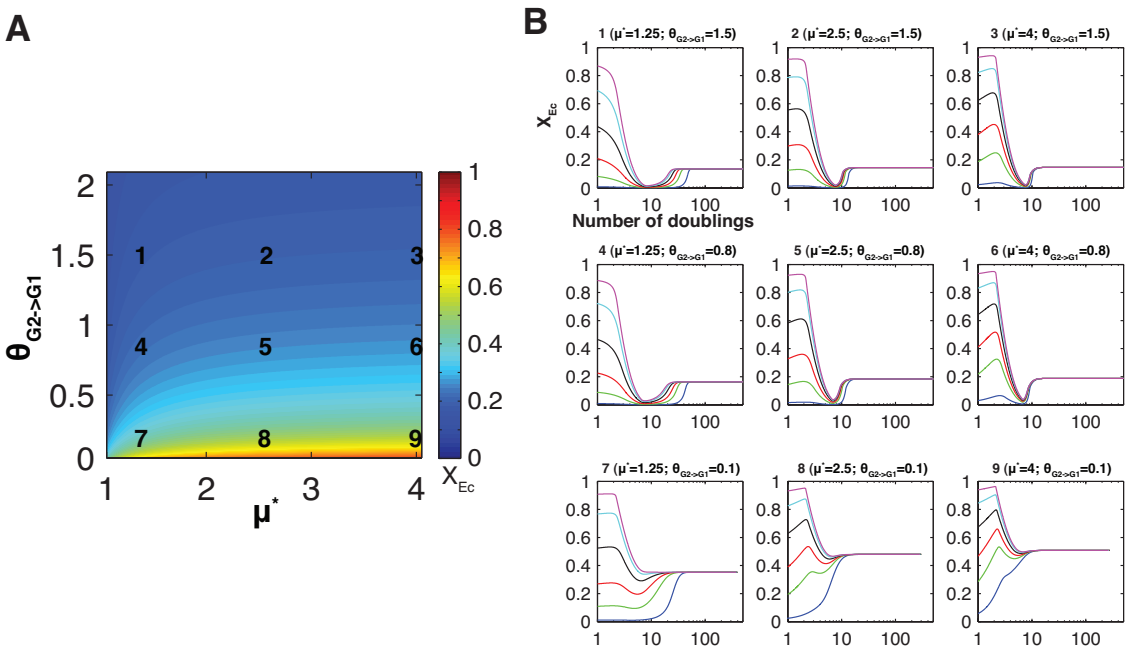

Figure 3.5: Population dynamics in simplified RUTC30/K12 TrEc consortium model. **(A)** Steady state population composition (*E. coli* population fraction X_{Ec}; g/g-total) as a function of μ^* and $\theta_{G_2 \to G_1}$, with $\alpha = 8.7$, $K^*_{BGL} = 23.25$, $Y_{Tr/Ec} = 1.5$, and $\beta = 10$. **(B)** Numerical ODE solutions corresponding to μ^* and $\theta_{G_2 \to G_1}$ values at each labeled point in A. For each μ^*, $\theta_{G_2 \to G_1}$ point, ODEs were solved with various initial X_{Ec} values. For explanation of parameters see section 3.3.1, Table 3.1, and Table D.1.

3.3.3.2 RUTC30/NV3 bicultures: isobutanol production/toxicity effects

The above theoretical analysis does not account for isobutanol production/toxicity effects. Since these effects could impact population stability/dynamics, we extended our modeling framework to account for isobutanol production and toxicity. We will assume growth-coupled production

of isobutanol by *E. coli*. For sake of simplicity, the reduced version of model does not include maintenance substrate uptake or isobutanol production. Isobutanol production rate is then given by:

$$\frac{dI}{dt} = Y_{I/S_{G_1}} r^{Ec}_{S_{G_1}} \tag{3.44}$$

Where I is isobutanol concentration (g/L), $Y_{I/S_{G_1}}$ is the isobutanol yield coefficient (g-isobutanol / g-glucose), and $r^{Ec}_{S_{G_1}}$ is the *E. coli* glucose uptake rate (g/L/h). For *E. coli* NV3 pSA55/69, $Y_{I/S_{G_1}} = 0.25$ g-isobutanol/g-glucose. We model isobutanol toxicity with empirical growth inhibition functions similar to Equations A.42 & A.48:

$$K^I_{Sp} = \begin{cases} \left(1 - \frac{I}{I^*_{Sp}}\right)^{n_{Sp}} & \text{if } I \leq I^*_{Sp} \\ 0 & \text{if } I > I^*_{Sp} \end{cases} \tag{3.45}$$

Where Sp is species (*E. coli* or *T. reesei*), I is isobutanol concentration (g/L), I^*_{Sp} is the growth inhibiting isobutanol concentration (g/L) for Sp, and n_{Sp} is an empirical exponent for species Sp. Overall growth rates are given by:

$$\frac{dC_{Sp}}{dt} = K^I_{Sp} \mu_{Sp} C_{Sp} \tag{3.46}$$

Where C_{Sp} is concentration of species Sp and μ_{Sp} is the growth rate (i.e. Monod function, see Equations 3.1 & 3.4 in section 3.3.1). I^*_{Sp} and n_{Sp} values for *E. coli* NV3 pSA55/69 and *T. reesei* RUTC30 are given in Table 3.2.

Parameter	**NV3 pSA55/69**	**RUTC30**
I^*_{Sp}	10 g/L	6 g/L
n_{Sp}	1.5	1.5

Table 3.2: Isobutanol toxicity parameters used in modified cooperator-cheater model. Estimated from data in Table B.3 and Chapter 4.

Other parameter values for NV3 pSA55/69 are given in Table B.2; for remaining model parameters, we used the same values as in the previous section (Table D.1). Adding isobutanol production (Equation 3.44) to our modeling framework makes it impossible for batch cultures to reach non-trivial pseudo-steady states, due to accumulation of isobutanol in the system. Thus we can cannot apply the steady state analysis described in section 3.3.2.1. Instead, we conducted a numerical analysis of the ODE model wherein we simulated iterative serial cultures. This scheme is analogous to repeated batch fermentation, where a fraction of the previous batch is used as inoculum for new batches (Figure 3.6A, inset). For our theoretical analysis, this was simulated by

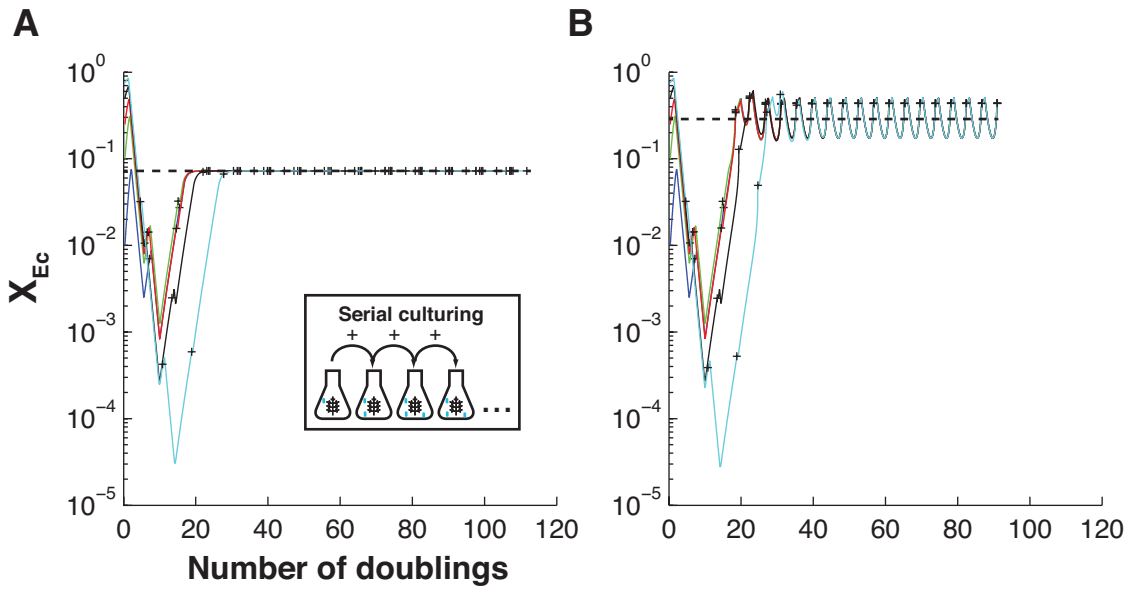

Figure 3.6: Population dynamics in simplified RUTC30/NV3 TrEc consortium model. **(A)** Numerical ODE solutions for NV3/RUTC30 cooperator-cheater model over a range of $X_{Ec}(t_0)$ (initial *E. coli* population fraction; g/g-total) values with $Y_{I/S_{G_1}} = 0$ (i.e. no isobutanol production/toxicity). All solutions converge to $X_{Ec} \approx 0.07$. Inset depicts serial culturing scheme used in numerical simulations, with "+" indicating passaging points. **(B)** Numerical ODE solutions for NV3/RUTC30 cooperator-cheater model over a range of $X_{Ec}(t_0)$ values, with $Y_{I/S_{G_1}} = 0.25$ g-isobutanol/g-glucose and isobutanol toxicity parameters given in Table 3.2. All solutions converge to $X_{Ec} \approx 0.12$

numerically integrating the ODE system to a fixed endpoint, dividing end-point variable values by a dilution factor (typically 50 fold), and using these values as initial conditions (ICs) for the next iteration (Figure 3.6A, inset).

To examine the effects of isobutanol production/toxicity, we compare numerical solutions for NV3/RUTC30 bicultures without isobutanol production/toxicity (i.e. $Y_{I/S_{G_1}} = 0$, but with remaining parameters otherwise the same) and with isobutanol production (Figure 3.6A&B, respectively). Models without and with production/toxicity effects both converge to stable equilibria (Figure 3.6). However, dynamics and equilibrium X_{Ec} values are distinctly different, with isobutanol production/toxicity effects leading to steady oscillations and approximately doubling the equilibrium X_{Ec} (Figure 3.6). As isobutanol accumulates in a culture, relative growth rates (i.e. μ^*) change due to tolerance differences between *E. coli* and *T. reesei* (see Table 3.2) leading to varying population fraction and preventing the system from reaching a steady state. However, imposing serial passaging at fixed endpoints creates a "seasonality" effect wherein isobutanol concentrations are periodically diluted, thus reseting isobutanol inhibition in the system. At steady state, oscillatory population dynamics are such that X_{Ec} declines at the beginning of the culture, reaches a minimum, and then begins to increase due to isobutanol accumulation and increased μ^*; the end-point X_{Ec} then becomes the initial X_{Ec} for the next iteration (Figure 3.6B).

Thus seasonality of isobutanol concentration is the driving force responsible for the observed oscillations. Interestingly, despite initial divergences, all of the population composition trajectories become synchronized, likely due to the fixed passaging points (Figure 3.6B). In addition to inducing oscillations, isobutanol production/toxicity also increases equilibrium X_{Ec} (Figure 3.6). The relative increase in X_{Ec} is simply due to the higher isobutanol tolerance of *E. coli* compared to *T. reesei* (Table 3.2), which substantially improves the mean relative fitness of *E. coli* in the co-culture.

3.3.4 Experimental validation of cooperator-cheater dynamics in TrEc consortia

We experimentally investigated population dynamics and equilibria in TrEc consortia by conducting serial passaging studies with *T. reesei* / *E. coli* bicultures. To simplify our measurements of microbial composition, we used 10 g/L cellobiose (a soluble cellulose disaccharide) as growth substrate instead of cellulose. We note that our mass transfer model for oligosaccharide hydrolysis at the *T. reesei* cell wall (section A.2.2) was derived with an assumption of low oligosaccharide concentrations (typical of growth on cellulosic substrates). For the simplified model presented in this section, glucose concentration at the *T. reesei* cell surface is modeled as $S_{G_1} + \theta_{G_2 \rightarrow G_1} S_{G_2}$; in the limit of saturating S_{G_2} concentrations (i.e. $S_{G_2} \gg K_{M,BGL}$), our derivation reduces to $\theta_{G_2 \rightarrow G_1} S_{G_2} \rightarrow \delta$, where δ is a constant. Population equilibria will thus differ with cellobiose as the primary growth substrate compared to cellulose; however the substitution of cellobiose for cellulose retains requisite features for cooperator-cheater coexistence (concave fitness benefits and cooperator privileged access) and should yield dynamics qualitatively similar to cellulose. To explore the cooperator-cheater tuning concept, we conducted bicultures over several different μ^* (Ec:Tr μ_{max} ratio) values. This parameter was chosen as it is relatively straightforward to manipulate experimentally via modulation of environmental conditions or by selection of strains with different intrinsic growth rates. *T. reesei* and *E. coli* have different optimal pH ranges (*T. reesei* optimum: pH 5 [109]; *E. coli* optimum: pH 7 [110]), allowing us to modulate μ^* by varying culture pH. Experimentally measured μ^* for RUTC30/K12 as a function of pH can be found in Table B.5. We can also adjust μ^* by selecting strains with different intrinsic growth rates; for instance in pH 6.0 TMM media, *E. coli* K12 has $\mu_{max,S_{G_1}} = 0.41 \pm 0.02$ while for NV3 pSA55/69 has $\mu_{max,S_{G_1}} = 0.14 \pm 0.02$ (Table B.3).

Experimental results for *T. reesei* / *E. coli* bicultures at different μ^* values are shown in Figure 3.7. As predicted, *T. reesei* / *E. coli* bicultures inoculated at different initial X_{Ec} all

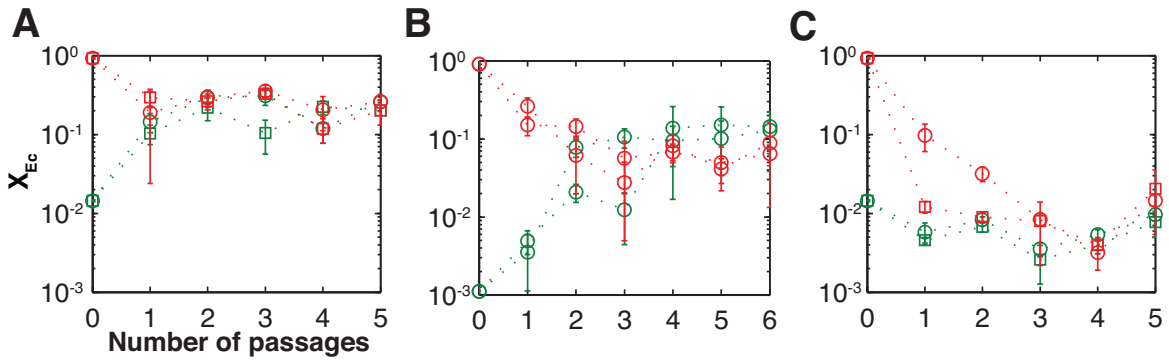

Figure 3.7: Experimental investigation of cooperator-cheater coexistence in the TrEc consortium. Replicate *T. reesei* / *E. coli* cultures were inoculated at different population compositions, corresponding to high (0.9) or low (0.01 or 0.001) initial X_{Ec} (red or green points/lines, respectively) and serial passaged on 10 g/L cellobiose media. **(A)** Experimentally observed population trajectories for RUTC30/K12 bicultures at pH=6.0 ($\mu^* \approx 4$). Equilibrium $X_{Ec} = 0.24 \pm 0.04$ **(B)** Experimentally observed population trajectories for RUTC30/NV3 bicultures ($\mu^* \approx 1.5$). Equilibrium $X_{Ec} = 0.1 \pm 0.2$. **(C)** Experimentally observed population trajectories for RUTC30/K12 bicultures at pH=5.3 ($\mu^* \approx 1$). Equilibrium $X_{Ec} = 0.007 \pm 0.003$.

converge to stable equilibria (Figure 3.7). Furthermore, equilibrium X_{Ec} increases with cheating benefits: bicultures at $\mu^* \approx 4$ (K12/RUTC30 pH 6, Figure 3.7A) converge to $X_{Ec} = 0.24 \pm 0.04$, bicultures at $\mu^* \approx 1.5$ (NV3/RUTC30 pH 6, Figure 3.7B) converge to $X_{Ec} = 0.1 \pm 0.2$, and bicultures with $\mu^* \approx 1$ (K12/RUTC30 pH 5.3, Figure 3.7C) converge to $X_{Ec} = 0.007 \pm 0.003$. Consistent with our theoretical analysis (Figure 3.5B), equilibration time appears to increase as $\mu^* \to 1$ (Figure 3.7). Population convergence is rapid for $\mu^* \approx 4$ cultures (Figure 3.7A), occurring within a single passage (approximately 6 population doublings). Convergence is comparatively slower for $\mu^* \approx 1.5$ and $\mu^* \approx 1$ bicultures, requiring 3 to 4 passages (approximately 18 to 24 population doublings) (Figure 3.7B&C). Results with NV3/RUTC30 bicultures (Figure 3.7B) demonstrate that cooperator-cheater dynamics are preserved when isobutanol production/toxicity effects are present. Our model predicts oscillating population compositions for NV3/RUTC30 bicultures (Figure 3.6B). However, in our experiments X_{Ec} was measured only once per passage, and thus our measurements do not have sufficient resolution to detect the predicted oscillations.

3.4 Discussion and conclusion

Overall, our experimental results support theoretical predictions of cooperator-cheater dynamics within TrEc consortia, and illustrate the feasibility of tuning population composition via modulation of ecological parameters. Beyond TrEc consortia, cooperator-cheater dynamics could be applied as a general tool for designing and programming stable and tunable synthetic consortia. The most obvious application would be to extend cooperator-cheater tuning to other S/F

consortia (e.g. Table 2.1). By selecting saccharolytic specialists with cell-wall localized hydrolytic enzymes, cooperator-cheater tuning could be utilized in S/F consortia for a variety of products and feedstocks (lignocellulose, starch, inulin, etc.). Since many saccharolytic microbes have cell-wall localized hydrolytic enzymes (e.g. the ubiquitous cellulosomes of anaerobic microbes), there are numerous possible microbe combinations for creating S/F consortia with cooperator-cheater dynamics. Furthermore, the use of saccharolytic specialists with different intrinsic ecological parameter values (i.e. Table 3.1) represents a possible means of tuning/optimization. In addition to S/F consortia, cooperator-cheater dynamics could be applied in other ways to engineer consortia for bioprocessing. For example, cooperator-cheater dynamics could be applied to tune bicultures of invertase positive/negative strains of *S. cerevisiae* metabolically engineered to produce different biofuel/chemical products, such that products can be made at a precisely defined, pre-programmed ratio. In contrast to previous studies of cooperator-cheater coexistence [107], our theoretical analysis suggests that oscillatory population dynamics may occur when product (e.g. biofuel/chemical) toxicity effects are present. From an applied standpoint, these oscillatory dynamics may be undesirable in industrial bioprocessing. However, our analysis suggests that consortia can maintain a "repeatable" population structure in the sense that oscillations reach a steady period/amplitude synchronized with serial passaging points; thus it may still be feasible to deploy consortia with these dynamics in repeated batch fermentations or other bioprocesses. For future work, we suggest conducting more detailed studies of population dynamics in isobutanol-producing TrEc consortia to further explore theoretical predictions.

In addition to population dynamics, selection of parameters for tuning is another important factor to examine for potential bioprocessing applications. In our experimental studies, we used μ^* to tune population composition since this parameter is straightforward to manipulate. However, other parameters could potentially be adjusted to tune population composition. For instance, α, β and $\theta_{G_2 \to G_1}$ could be modulated by genetically engineering *T. reesei* to alter expression of β-glucosidase genes (and/or other cellulases in the case of α). As another example, K_S^* could be adjusted by changing organism substrate affinity, for instance by genetically engineering *E. coli* or *T. reesei* to express alternative substrate transporters with different K_M values. Ultimately, desired performance characteristics and accessible parameter ranges will determine the optimal parameter to use for tuning. For example, there are parameter ranges over which μ^* has little effect on population equilibria (e.g. $K_S^* < 1$; Figure 3.4), thus it may be beneficial to select a different parameter for tuning in these cases (e.g. β or K_S^*). As an interesting corollary, we propose that parameter ranges over which population equilibria are insensitive to μ^* may actually be desirable in bioprocessing applications, as population equilibria would potentially be buffered against variations in μ^* (for instance, due to fluctuations in environmental conditions like temperature,

pH, etc.). Finally, as a caveat we note that our parameter set is not dimensionally independent; parameter subsets with overlap include α, β and $\theta_{G_2 \to G_1}$, which all depend on k_{BGL} as described above, K_S^* and K_{BGL}^*, which both depend of $K_{S,Tr}$, and β and μ^*, which both depend on $\mu_{max,Tr}$. Thus it may be difficult to adjust certain parameter combinations independently. We suggest exploring the effects of parameters beyond μ^* in future work with TrEc consortia or other S/F consortia employing cooperator-cheater dynamics.

Beyond population dynamics and selection of parameters for tuning, it is also important to consider the long-term evolutionary stability of cooperator-cheater systems. While ecological motifs such as metabolic cross-feeding [42, 106] and cooperator-cheater dynamics may exhibit stable short-term behavior, these systems may not necessarily be stable over long evolutionary timescales, especially in homogenous environments. A more detailed discussion about future directions for improving the stability of TrEc consortia through engineered spatial structure and/or mutualisms will be given in Chapter 6.

3.5 Materials and methods

3.5.1 Model analysis

For dynamic analysis, the ODEs described in section 3.3.1 were implemented in MATLAB and solved numerically using $ode15s$. All other details regarding model analysis (equations, parameter values, etc.) can be found in sections 3.3.1, 3.3.2.1, and 3.3.3.

3.5.2 Cooperator-cheater coexistence experiments

Duplicate *T. reesei* / *E. coli* bicultures were inoculated at high $X_{Ec}(t_0)$ (= 0.9 g/g-total) or low $X_{Ec}(t_0)$ (= 0.01 or 0.001 g/g-total) in 10 g/L cellobiose TMM media (described in section 2.5.2) buffered with maleate-NaOH at indicated pH, with approximately 0.05 gDW/L total initial microbial biomass. Cultures were incubated at 30°C with agitation for 48 hours, then inoculated 1:50 (by volume) into fresh media; serial culturing was repeated for five to ten passages. At each passage, cultures were analyzed for total dry mass and *E. coli* cell counts, described in section 2.5.11. 10 mL cultures in 50 mL Falcon tubes were used for RUTC30/K12 co-cultures, while 50 mL cultures in 250 mL screw-cap flasks were used for RUTC3/NV3 bicultures. RUTC30/NV3 bicultures were supplemented with 100 μg/mL ampicillin, 30 μg/mL kanamycin, and 0.1 mM IPTG for maintenance/induction of plasmids pSA55/69 in *E. coli* NV3.

Part II: Genome evolution and engineering for elucidating the genetic architecture of isobutanol tolerance in *Escherichia coli*

CHAPTER 4

Evolution combined with genomic study elucidates genetic bases of isobutanol tolerance in *Escherichia coli*

4.1 Summary

In the second part of this dissertation, we seek to develop novel evolutionary-genomics based methods for elucidating and improving complex stress tolerance phenotypes. Understanding the genetic architecture underlying complex phenotypes is of great fundamental interest and also has important ramifications in bioprocessing, since product toxicity frequently limits titers and volumetric productivity. Microbial stress tolerance is a complex multigenic trait that is intractable to traditional genetic study and rational engineering efforts. In this work, we experimentally evolve isobutanol tolerant *E. coli* strains and then use whole genome resequencing and followup functional dissection studies to reverse engineer genetic bases and mechanisms of tolerance. Consistent with the complex, multigenic nature of solvent tolerance, we observe adaptations in a diversity of cellular processes. Many adaptations appear to involve epistasis between different mutations, implying a rugged fitness landscape for isobutanol tolerance. We observe a trend of evolution targeting post-transcriptional regulation and high centrality nodes of biochemical networks. Collectively, the adaptations we observe suggest mechanisms of isobutanol tolerance based on remodeling the cell envelope and surprisingly, stress response attenuation.

The majority of the work presented in this chapter was published in *Microbial Cell Factories* as follows: J. Minty, A. Lesnefsky, F. Lin, Y. Chen, T. Zaroff, A. Veloso, B. Xie, C. McConnell, R. Ward, D. Schwartz, J.M. Rouillard, Y. Gao, E. Gulari, and X. Lin. Evolution combined with genomic study elucidates genetic bases of isobutanol tolerance in *Escherichia coli*. *Microbial Cell Factories*, 10:18, 2011. doi:10.1186/1475-2859-10-18

4.2 Introduction and background

4.2.1 Experimental evolution with microbes

Experimental evolution has emerged as a powerful tool for studying adaptation in microbes. Experimental evolution provides a valuable means for investigating fundamental dynamics and bases of adaptation, as well as testing theories about the evolutionary process [57]. As a pioneering example, Lenski and colleagues began a long term evolution experiment (LTEE) in 1988 consisting of 12 replicate *E. coli* lineages evolving in glucose minimal media [111]. The Lenski LTEE continues as of this writing (April 2013) and has surpassed 58,000 generations. The LTEE has been used to study and test numerous important theories in evolutionary biology, and provides experimental evidence for key questions such as how rates of evolution vary over time, the extent to which evolutionary trajectories are repeatable, and the relationship between evolution at the phenotypic and genotypic levels [111]. In the decades since the founding of the LTEE, there has been tremendous growth in the use of experimental evolution with microbes for fundamental and applied studies [111]. Recent advances in DNA sequencing technology have led to dramatically increased throughput, enabling rapid and relatively inexpensive resequencing of microbial genomes [112]. An intriguing corollary to the development of next-generation sequencing methods is the prospect of using whole genome resequencing to characterize the genetic bases of adaptation in evolved microbes. This approach was employed in several foundational studies to investigate adaptations in experimentally evolved *E. coli* populations, including short-term evolution on glycerol [113, 114] or lactate [115] minimal media, and genome evolution over 40,000 generations of the LTEE experiment [116] (examples of mutations found in these studies shown in Figure 4.1A).

These studies have provided numerous valuable and intriguing insights into genotypic and phenotypic adaptation. In the glycerol adaptation study, sequenced isolates from five separate populations all acquired mutations in glycerol kinase *glpK* early in evolution [113]. Phenotypic and functional analysis suggest that *glpK* mutations provide substantial fitness benefits by increasing GlpK activity [113]. These results indicate that glycerol kinase is a rate limiting step in glycerol metabolism for WT *E. coli*, and thus increasing GlpK activity constitutes a key mechanism of adaptation. In addition to elucidating mechanisms of adaptation, these studies also provide information on the modes and dynamics of evolution. For example, genomic study of *E. coli* isolates evolved on glycerol [113, 114] or lactate [115] minimal media revealed that epistatic interactions between mutations were common under these conditions. In particular, several mutations had neutral or deleterious effects individually, but confer fitness benefits when combined with other mutations [58]. Additionally, diversity of genotypic outcomes may depend on

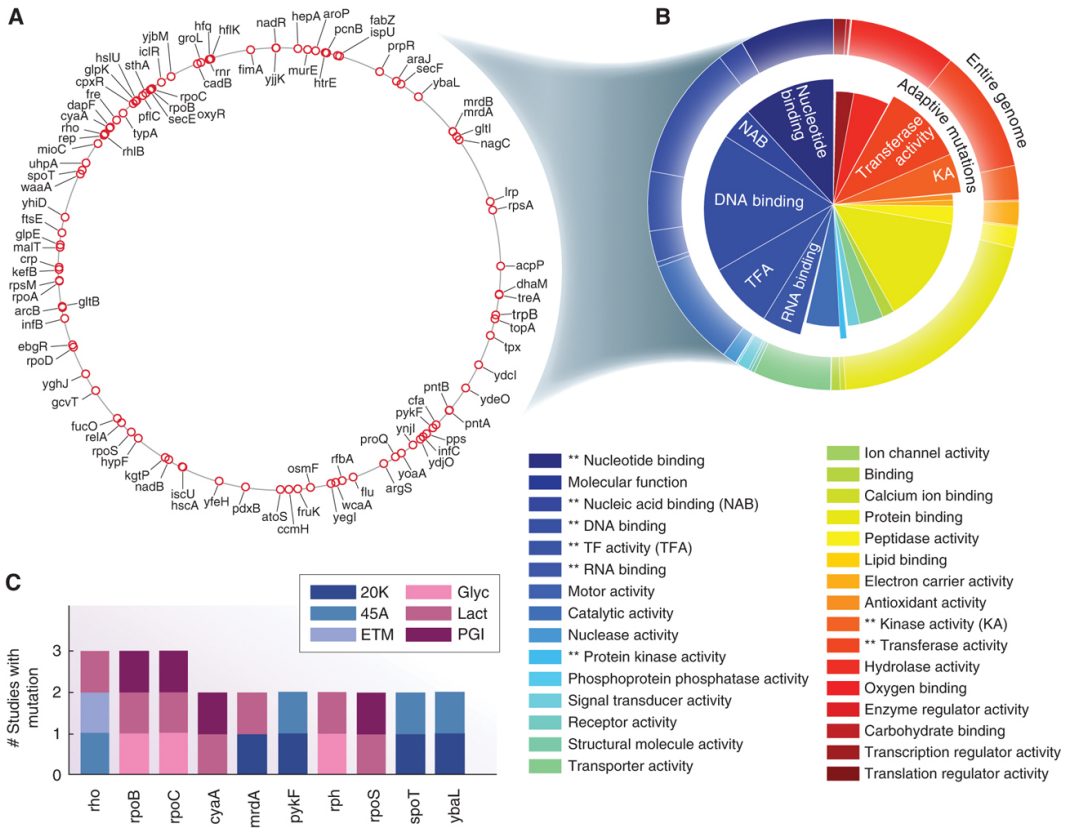

Figure 4.1: Mutations identified in selected *E. coli* experimental evolution studies. Adapted from [58]. **(A)** Chromosome map of point mutations found in open reading frames in multiple experimental evolution studies ([113], [116], [115], [118], [119], and [120]). **(B)** Enrichment analysis (based on gene ontology categories) for loci shown in panel A. Protruding wedges represent statistically enriched categories (see [58] for further details). **(C)** Genes mutated in multiple studies. 20K= Lenski Long Term Evolution Experiment (LTEE) at 20,000 generations [116], 45A=evolution at high temperature [119], ETM=evolution of ethanol tolerance [121], Glyc=evolution on glycerol minimal medium [113], Lact=growth on lactate minimal media [115], and PGI=evolution on glucose minimal media following the deletion of *pgi* [118].

selective pressures and environment; for example, genotypic adaptations in replicate lineages were more diverse in lactate evolved populations compared to glycerol evolved populations, hinting at different fitness landscapes for these two phenotypes [115]. The LTEE genome resequencing study examined evolution over time in a single lineage, providing information on long-term evolutionary dynamics [116]. Surprisingly, discordances in evolutionary rates were observed, wherein genome evolution proceeded at a nearly linear rate while fitness improvement decelerated over time [116]. While this rate discordance suggests a preponderance of neutral mutations, there is evidence that most of the mutations acquired up 20,000 generations are in fact beneficial [116]. The mechanisms that underlie the observed rate discordance are probably complex, may include effects from clonal interference [116] and/or negative epistasis between beneficial mutations [117].

A particularly striking finding observed in many experimental evolution studies is that

mutations in genes with global regulatory functions (including global transcriptional and post-transcriptional regulators, and genes controlling DNA supercoiling) often provide large fitness benefits and constitute an important mode of adaptation (Figure 4.1B) [57, 122, 113, 115]. Mutations in global regulators can result in large-scale expression changes and have profound fitness effects [58]. Most randomly occurring mutations in global regulators are probably deleterious, but the centrality of these regulators creates potential for mutations with high fitness benefits. Global regulators with high mutation frequencies in experimental evolution studies include RNA polymerase subunits *rpoB*/*rpoC*, transcriptional terminator *rho*, σ^S master stress response regulator *rpoS*, and stringent response regulator *spoT* (Figure 4.1C). Studies with reconstructed single mutants have demonstrated that global regulatory mutations frequently make the greatest single contribution to fitness out of acquired mutations [113, 115]. Global regulatory mutations generally exhibit pleiotropic effects; while the net sum of fitness effects arising from expression changes across many genes may be beneficial, some of the specific expression changes elicited by global regulatory mutations may be maladaptive [123]. This creates the potential for subsequent mutations that reduce maladaptive side effects of global regulatory mutations, and thus epistatic compensatory mutations are likely to occur in populations that fix global regulatory mutations [123].

In addition to providing valuable contributions to the fundamental understanding of evolutionary biology, studies of mechanisms, modes, and dynamics of genotypic adaptation are of great utility in applied fields such as biological engineering or medicine. For example, the observation of ubiquitous adaptations in global regulators may have inspired the highly successful strain improvement strategy of mutagenizing global transcription factors to generate libraries with a diversity of global gene expression patterns, from which variants with substantially improved phenotypes can be isolated (as discussed in Chapter 1). Beyond the fundamental studies described above, experimental evolution and subsequent genome resequencing could be used to explore the genetic basis of more complex microbial phenotypes of medical or industrial relevance, such as resistance to antibiotics or product tolerance in biofuel fermentations. In this work, we comprehensively investigate adaptations to exogenous isobutanol stress in *E. coli* through experimental evolution followed by genome resequencing and functional dissection studies. An overview of this work in the broader context of this dissertation is given in the following section.

4.2.2 Overview

Toxicity of alcohol biofuels limits final titer and volumetric productivity in fermentations, motivating efforts to engineer microbes with improved tolerance [48]. Due to broad mechanisms of toxicity, alcohol tolerance is a complex trait that involves a diversity of cellular adaptations and responses that probably contribute synergistically to the overall phenotype [50]. The inherent biological complexity of alcohol and other chemical tolerance phenotypes precludes the use of rational engineering for tolerance improvement, thus most strategies are combinatorial in nature, following a paradigm of generating genotypic and phenotypic diversity in a population, then characterizing isolates with desired properties (as discussed in Chapter 1). However, most currently used methods explore relatively small subsets of the possible genotype space and often cannot capture interactions between distal genetic loci (Chapter 1). Advances in DNA sequencing, DNA synthesis, and genome-scale genetic manipulation techniques have greatly expanded our ability to read and write genetic information. Integrating methods for identifying genes/mutations conferring desired phenotypes with genome-scale combinatorial engineering represents a promising strategy for improving microbial chemical tolerance phenotypes. In the second part of this dissertation, we develop a novel integrated approach that entails experimental evolution of stress tolerance followed by genome resequencing to identify acquired mutations, genomic and functional dissection to reverse engineer mechanisms of tolerance, and targeted genome engineering for further phenotype improvement (Figure 4.2). The proposed methodology transcends many limitations in previous approaches by dramatically expanding accessible genotype space and allowing for the identification of synergistic interactions between genetic loci.

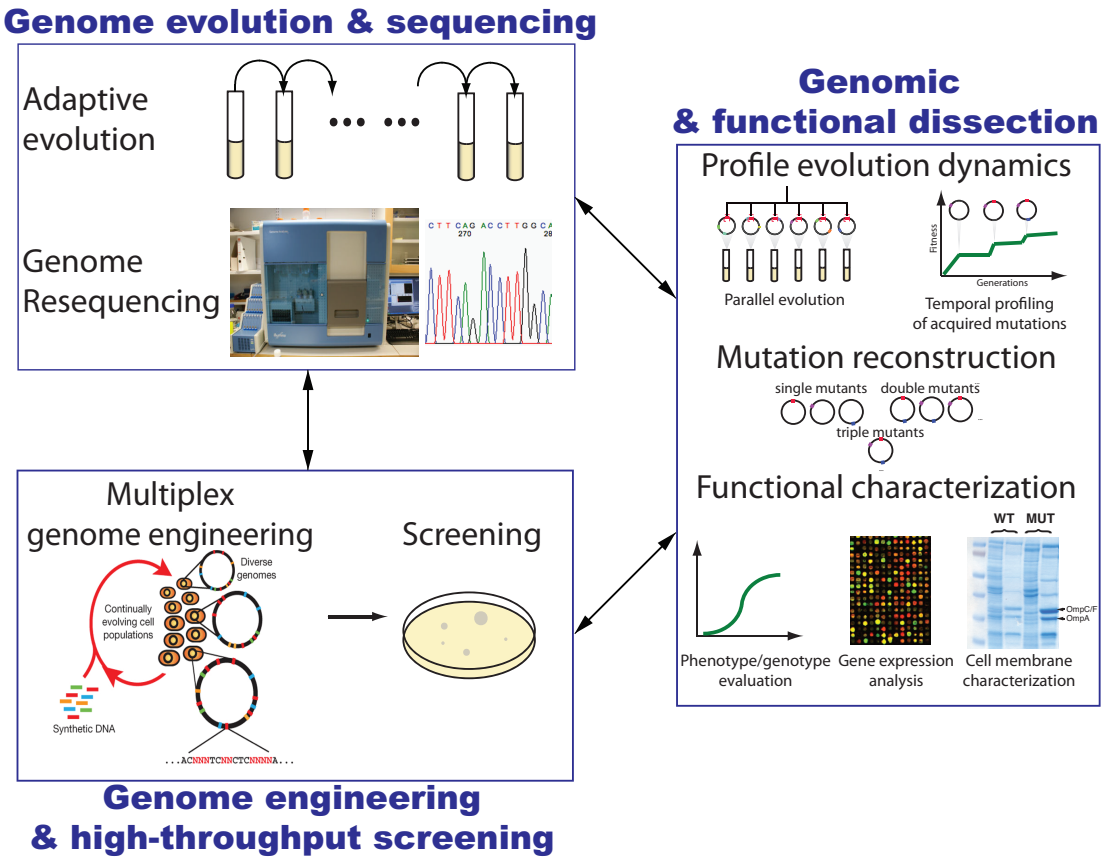

Figure 4.2: Overview of evolutionary-genomics approach for tolerance phenotype elucidation and engineering.

In Part I of this dissertation, we demonstrated direct conversion of lignocellulosic feedstocks to isobutanol using the TrEc consortium (Chapter 2). While isobutanol is widely regarded as one of the most promising next-generation biofuels, it is highly toxic to microbes. Motivated by this, we apply our proposed evolutionary-genomics methodology to study and improve isobutanol tolerance in *E. coli*. Improving the isobutanol tolerance of *T. reesei* will also be crucial to our efforts to engineer a robust and efficient consortium for cellulosic isobutanol production. However, it will be difficult to fully apply the approach outlined above to *T. reesei*, as there are fewer available molecular genetics tools for this organism. Instead we have focused on improving isobutanol tolerance in *T. reesei* through classical strain improvement techniques (evolution, breeding, and genome shuffling), which will not be discussed in this dissertation.

In this chapter, we comprehensively investigate adaptations to exogenous isobutanol stress in *E. coli* through experimental evolution followed by genome resequencing and and followup genetic, phenotypic, and biochemical studies. Using this approach, we have identified key loci involved in isobutanol tolerance. Consistent with the complex, multigenic nature of isobutanol

tolerance, we find diverse and often surprising genetic adaptations to isobutanol stress that were not obvious from other approaches to investigating tolerance. The divergent growth phenotypes of the end populations and studies with single and multiple mutation reconstructions suggest a rugged fitness landscape with many epistatic interactions. This work was conducted from June 2008 to March 2011. In this time period, numerous other studies employing experimental evolution of stress tolerance followed by genome resequencing were initiated in parallel with our work. Some examples closely related to our studies include Atsumi *et al.* 2010 (isobutanol adaptation in *E. coli*) [124], Reyes *et al.* 2012 (n-butanol adaptation in *E. coli*) [125], and Dragosits *et al.* 2013 (n-butanol adaptation in *E. coli*) [126]. The emergence of many such studies in parallel is a testament to the utility and power of this approach, and provides a large set of additional experimental data with which to compare and interpret our results.

4.3 Results

4.3.1 Experimental evolution and phenotypic characterization of end populations

E. coli EcNR1, a derivative of *E. coli* K12 MG1655 harboring a λ Red prophage integrated at the bio locus, was evolved by serial passaging of six independent populations for approximately 500 generations on isobutanol spiked NG50 media (M9 minimal medium supplemented with 50 g/L carbon source and 0.25 mg/L biotin). An initial isobutanol concentration of 0.75% (w/v) (corresponding to approximately 75% growth inhibition) was used for all populations, providing strong selective pressure. Isobutanol concentration was gradually increased during evolution to maintain approximately constant selective pressure.

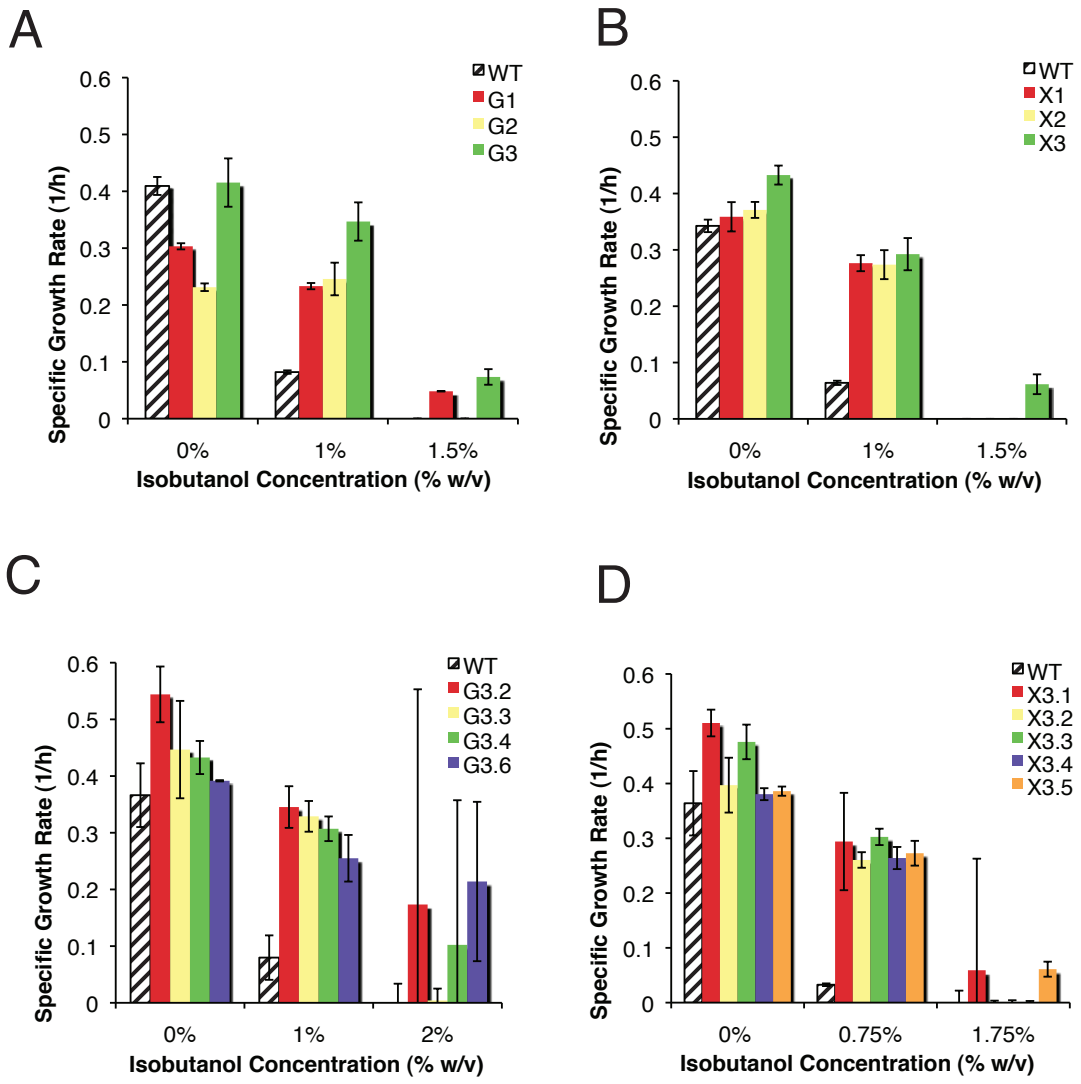

Figure 4.3: Isobutanol tolerance of evolution end populations was evaluated by measuring growth rate at various isobutanol concentrations, with the parent *E. coli* EcNR1 included as a reference. Populations were phenotyped with the same carbon source they were evolved in. **(A)** populations evolved on glucose as sole carbon sources (three lineages, G1, G2, G3), **(B)** populations evolved on xylose as sole carbon sources (three lineages, X1, X2, X3), **(C)** selected clonal isolates obtained from G3 population, **(D)** selected clonal isolates obtained from X3 population.

Populations were evolved on two different carbon sources, with three populations evolved with 50 g/L glucose as the sole carbon source (designated glucose #1, glucose #2, and glucose #3, abbreviated G1, G2, G3) and another three populations evolved with 50 g/L xylose as the sole carbon source (designated xylose #1, xylose #2, and xylose #3, abbreviated X1, X2, X3). Glucose and xylose are important constituents of lignocellulosic feedstocks and are metabolized by different pathways; thus we explore adaptations in different metabolic contexts relevant to biofuel production [50]. Cultures from each evolving population were periodically archived by cryopreservation, and phenotyped by measuring maximum specific growth rate (μ_{max}; 1/h) at

various isobutanol concentrations using a microplate spectrophotometer.

All of the evolved populations show significantly improved fitness at high isobutanol concentrations relative to the parent *E. coli* EcNR1 strain (WT) (Figure 4.3A and B). Interestingly, the populations show divergent growth phenotypes. Clonal isolates from two highly tolerant populations, G3 (glucose #3 population) and X3 (xylose #3 population), were further phenotyped, revealing significant heterogeneity within these populations (Figure 4.3C and D). Three clones from G3 were capable of growth at 2% isobutanol in glucose media and two clones from X3 grew at 1.75% isobutanol in xylose media, representing 60% and 40% improvements in tolerance respectively, compared to WT (Figure 4.3C and D). Two representative clones with high fitness, G3.2 and X3.5, were chosen for further characterization.

Evolution often produces adaptations that show tradeoffs in relative fitness across different environments [57]. To investigate specificity of adaptation, the fitness (relative to WT) of clones G3.2 and X3.5 at 0% and 1% (w/v) isobutanol was assessed on minimal glucose, minimal xylose, and rich LB media (Figure 4.4A and B). At 0% (w/v) isobutanol both G3.2 and X3.5 show improved fitness on xylose minimal medium and decreased fitness on LB medium, relative to WT (Figure 4.4A). At 1% (w/v) isobutanol, G3.2 and X3.5 show markedly improved relative fitness on both glucose and xylose minimal media, and decreased fitness in LB medium (Figure 4.4B). These results suggest that the two isolates characterized have accumulated adaptations to isobutanol stress specific to minimal media, and these adaptations appear to exhibit antagonistic pleiotropy in rich medium. This minimal-rich medium antagonisitic pleiotropy we observed underscores the importance of carefully selecting evolution conditions. On the other hand, although G3.2 and X3.5 were evolved on glucose and xylose media respectively, neither of these strains appears to have developed carbon-source specific adaptations in 0% and 1% isobutanol environments. We further assayed fitness in glucose and xylose minimal media at higher isobutanol concentrations (Figure 4.4C). At 1.5% (w/v) isobutanol, we observed relative fitness trends suggesting greater specificity of adaptation for G3.2 and X3.5 to their respective carbon sources, but we could not substantiate that these differences were statistically significant due to the error bars in our measurements (Figure 4.4C). Interestingly, at 0% isobutanol X3.5 appears to have higher relative fitness than G3.2 in all media types tested (Figure 4.4A). ATP yield from xylose metabolism is lower compared to glucose metabolism, and we speculate that low ATP yield increases selective pressure for more energy efficient use of carbon sources [127]. This may explain how adaptations to a low ATP yield substrate such as xylose could also be beneficial to growth on other carbon sources.

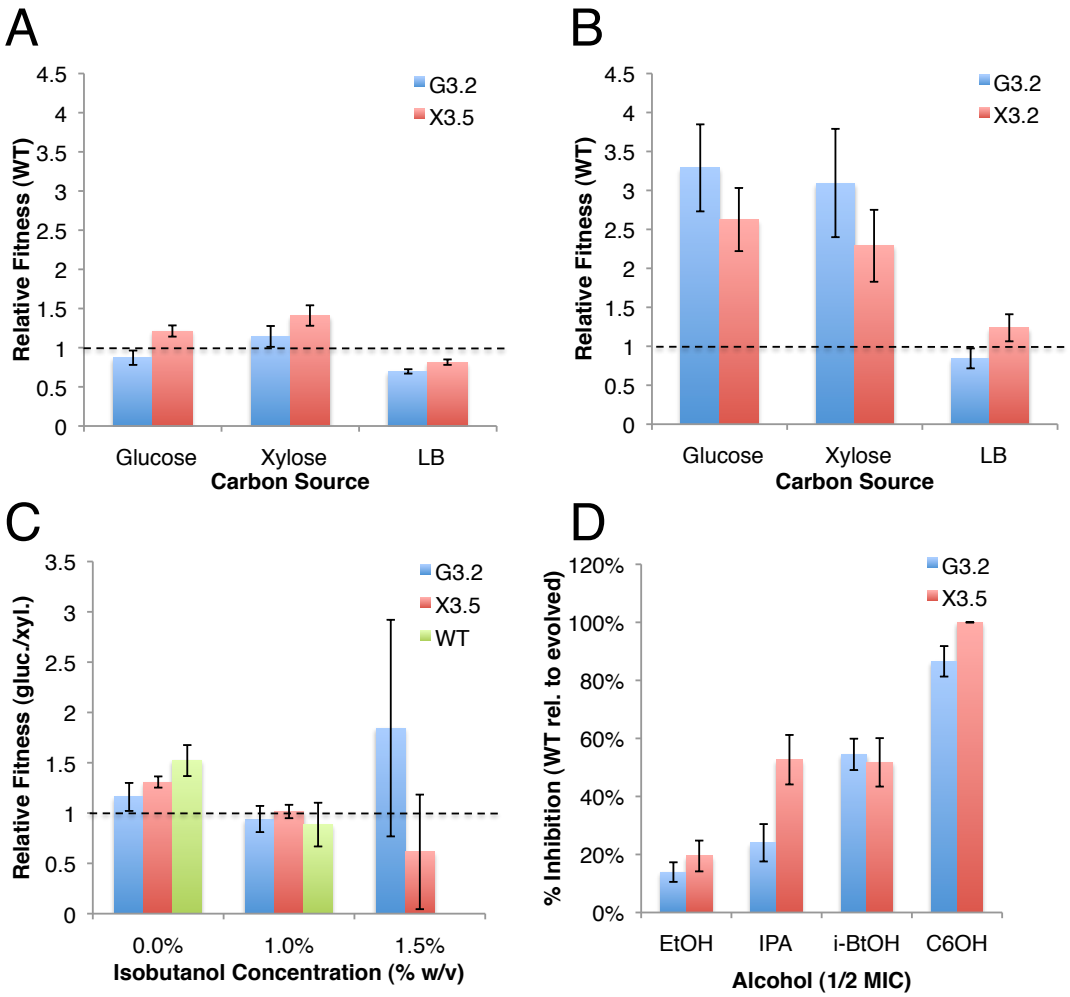

Figure 4.4: Specificity of adaptation with different carbon sources and alcohols. Specificity of adaptation of clones G3.2 (isolated from G3 evolution end population) and X3.5 (isolated from X3 evolution end population) was examined by assessing isobutanol tolerance of each clone and the parent *E. coli* EcNR1 on glucose minimal medium, xylose minimal medium, and LB medium. Tolerance to ethanol, isopropanol, and n-hexanol was also determined for these strains and the parent *E. coli* EcNR1. **(A)** Fitness (relative to parent *E. coli* EcNR1) of G3.2 and X3.5 on glucose, xylose, and LB medium with 0% (w/v) isobutanol. **(B)** Fitness (relative to parent *E. coli* EcNR1) of G3.2 and X3.5 on glucose, xylose, and LB medium with 1% (w/v) isobutanol, **(C)** Fitness of G3.2 and X3.5 (in glucose vs. xylose) at 0%, 1%, and 1.5% (w/v) isobutanol, **(D)** Percent relative inhibition of *E. coli* EcNR1 compared to G3.2 and X3.5 (defined as $1 - \mu_{WT}/\mu_{MUT}$, where μ_{WT} is the maximum specific growth rate of *E. coli* EcNR1, and μ_{MUT} is the maximum specific growth rate of G3.2 or X3.5) on ethanol (3.5% v/v), isopropanol (2.5% v/v), isobutanol (0.5% w/v), and hexanol (0.25% v/v). The alcohol concentrations correspond to approximately $1/2$ the minimum growth inhibiting concentration (MIC) for the parent *E. coli* EcNR1 strain.

In addition to investigating specificity of adaptation to different carbon sources, we also examined the tolerance of G3.2 and X3.5 to various alcohols with potential for microbial biofuel production, including ethanol, isopropanol, and n-hexanol (Figure 4.4D). While all alcohols share the same general mechanisms of toxicity via chaotropic effects and interactions with membrane lipid bilayers, specific biophysical effects are known to vary with alcohol chain length [128].

Molecular dynamics simulations and experiments with model lipid bilayers have demonstrated that long chain alcohols ($\geq$C8) tend to condense and stiffen lipid bilayers, while short chain alcohols ($\leq$C2) have opposite effects [128]; lipid bilayer interactions with intermediate length and branched alcohols (such as isobutanol) have not been well characterized. We examined the percent relative inhibition of WT compared to G3.2 and X3.5 (defined as $1 - \mu_{WT}/\mu_{MUT}$, where μ_{WT} is the maximum specific growth rate of *E. coli* EcNR1, and μ_{MUT} is the maximum specific growth rate of G3.2 or X3.5) at 3.5% (v/v) ethanol, 2.5% (v/v) isopropanol, 0.5% (w/v) isobutanol, and 0.25% (v/v) n-hexanol; concentrations were chosen to correspond to approximately 50% of the minimum growth inhibiting concentration (MIC) on glucose minimal medium at 30°C. For all alcohols assayed, G3.2 and X3.5 displayed higher tolerance than WT; interestingly, the relative inhibition of WT increased with increasing chain length (hexanol $>$ isobutanol $\geq$ isopropanol $\geq$ ethanol), indicating the adaptations to isobutanol stress may be selective to the effects of longer chain alcohols.

4.3.2 Genome resequencing of isobutanol tolerant clones

To identify the genetic bases of adaptation to isobutanol stress, we resequenced the genomes of highly tolerant clones from our evolved populations with the Illumina Solexa platform, using 36 base pair single-end or paired-end read configurations. 612 to 756 million base pairs (MB) of raw sequence was generated for each sequenced genome with single-end read configuration, with approximately four times as much sequence generated with paired-end reads. Coverage averaged approximately 125x and 500x for the 4.65 MB *E. coli* EcNR1 genome, using single-end and paired-end reads, respectively. Reads were mapped to the *E. coli* EcNR1 reference genome sequence using Novoalign v2.04.02 and MAQ v0.7; single nucleotide polymorphisms (SNPs) and short insertion/deletions (indels) were called from the consensus sequence [129, 130]. Larger indels were detected by examining coverage distribution. Unmapped reads were collected and de novo assembled using Velvet v0.7.51 to detect breakpoints near sites of structural variation (SV) [131].

We resequenced the genomes of G3.2, G3.6, and X3.5, three highly isobutanol tolerant clones from the evolution end populations (discovered mutations summarized in Figure 4.5 and Table E.1; full mutation lists available in Additional file 1 and the reference genome sequence in Additional file 2). It was discovered that the G3 lineage had acquired a 19 bp deletion in *mutL*, a component of the methyl-directed mismatch repair system (MMR). MMR loss-of-function mutations lead to an approximately 100-fold increase in mutation rate, giving rise to the so-called mutator phenotypes [57]. Subsequently G3.2 and G3.6 were highly mutated, having 48 and 64

mutations respectively, with 20 mutations in common between these two clones (Figure 4.5B and Table E.1). To narrow down candidate mechanisms of genetic adaptation, we resequenced the genome of a non-mutator clonal isolate from generation 266 of the G3 lineage (G3.266.7) and identified 8 mutations in this clone (Figure 4.5B and Table E.1). For X3.5, 11 mutations were revealed (Figure 4.5A and Table E.1).

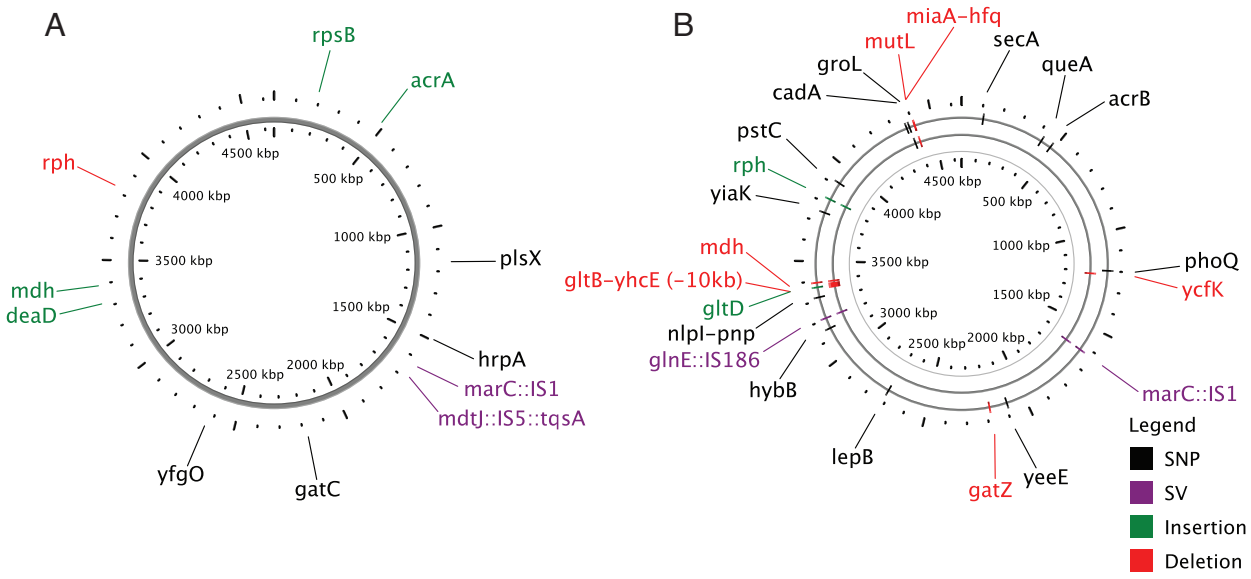

Figure 4.5: Chromosome maps of mutations discovered through whole genome resequencing of isobutanol tolerant clones. Genome resequencing of clones G3.2 and G3.6 (from G3 end population), X3.5 (from X3 end population), and G3.266.7 (from G3 lineage, generation 266) was done using the Illumina Solexa platform. SNPs (single nucleotide polymorphisms) and short indels (nucleotide insertions or deletions) were detected by mapping reads were onto the reference *E. coli* EcNR1 genome sequence. Structural Variation (SV) breakpoints were detected by de novo assembly of unmapped reads for single-end sequence data or with BreakDancer v0.0.1 for paired-end sequence data. **(A)** Chromosome map of mutations detected in X3.5; **(B)** Chromosome map of overlapping mutations in G3.2/G3.6 and G3.266.7. Mutations in G3.2/G3.6 are shown on the outer ring, while mutations in G3.266.7 are shown on the inner ring. All mutations in G3.266.7 except for the *gltB-yhcE* 10kb deletion and *ycfK* 1688 bp deletion are on the same line of descent as G3.2 and G3.6 clones.

A total of 131 mutations were discovered across clones X3.5, G3.2, G3.6, and G3.266.7 (full list available in Additional file 1). 96 mutations were SNPs, 25 mutations were short indels, and 10 mutations were SVs. Most mutations occurred in the coding region of genes. The detected SVs consisted of transposon insertions (*marC::IS1* in all sequenced isolates, *glnE::IS186* in the G3 clones, and *mdtJ::IS5::tqsA* in X3.5), an approximately 10kb deletion between *gltB* and *yhcE* in G3.266.7, and a 1688 bp deletion in the *ycfK* gene of the e14 prophage in G3.266.7. Mutations were found in diverse genetic loci representing many cellular processes. BiNGO (Biological Network Gene Ontology tool) was used to assess any overrepresented Gene Ontology (GO) terms in the full mutation set, but the only statistically significant finding was an enrichment of membrane proteins (corrected p-value = $7.23\text{x}10^{-3}$), with a borderline significant enrichment of

RNA helicases (corrected p-value = $7.22\text{x}10^{-2}$) [132].

4.3.3 Parallel evolution

Comparison of the genotypes of X3.5, G3.2, G3.6, and G3.266.7 reveals a number of parallel genotypic adaptations. In particular, mutations in *rph*, *acrAB*, *marC*, *mdh*, and the *gatYZABCD* operon were found in all of these clones (Table E.1). *E. coli* K12 MG1655 (the parent strain of *E. coli* EcNR1) has a 1 bp deletion in the *rph-pyrE* operon, resulting in reduced levels of orotate phosphoribosyltransferase (the product of *pyrE*) and subsequently suboptimal pyrimidine biosynthesis levels [115]. Thus restorative mutations are commonly observed in *rph-pyrE* during experimental evolution studies with *E. coli* K12 MG1655, and are general adaptations to growth on minimal media (e.g. see Figure 4.1). The AcrAB proteins are components of the AcrAB-TolC multidrug efflux pump, a membrane transporter which translocates a wide range of substrates out of the cytoplasmic membrane and periplasmic space; efflux via the AcrAB-TolC complex has been previously identified as an important mechanism of tolerance to organic solvents such as toluene, immediately suggesting a possible role for acrAB-tolC in isobutanol tolerance [50]. Possible links to isobutanol tolerance are not as obvious for *marC* (a predicted membrane protein of unknown function), *mdh* (NADH dependent malate dehydrogenase), and the *gatYZABCD* operon, which encodes proteins involved in galactitol transport and catabolism.

To investigate possible parallel genotypic adaptations in our other evolved lineages, the *acrAB* operon, *tolC*, and *mdh* were sequenced in 8 clonal isolates from each of the evolved endpoint populations (Table 4.1). The *marC* locus was also sequenced in each endpoint population; examination of PCR products revealed indel mutations (discernable by product size) at near 100% allele frequency, allowing for whole population samples to be sequenced (Table 4.1). We also sequenced the post-transcriptional regulator *hfq* in our endpoint populations since an *hfq* mutation was found in G3, and modulation of *hfq* has been observed as a common mechanism of adaptation in other experimental evolution studies. *rph* and *gatYZABCD* were not investigated further since *rph-pyrE* adaptations have been characterized in previous works, while the relatively large size of the *gatYZABCD* operon was prohibitive for Sanger sequencing.

acrAB mutations were discovered in X1, X2, X3, G1, and G3 populations (Table 4.1). Each population fixed only a single mutation in *acrA* or *acrB*, and allele frequency was near 100% (8/8 clones) except for G1, which had an allele frequency of approximately 25% (2/8 clones) and X3, with an approximate 50% allele frequency (4/8 clones). We did not detect *acrAB* mutations in

Population	Gene					
	acrA	***acrB***	***tolC***	***mdh***	***hfq***	***marC***
X1	-	481310 A→C (V773G)	-	-	4407590 T→G (I24M)	1625925 IS1 insertion (Disruption)
X2	484383 T→G (N154T)	-	-	3390659 +4:GATT (Frameshift)	-	1626084 IS5 insertion (Disruption)
X3	483735 +1:A (Frameshift)	-	-	3390936 +5:AACCT (Frameshift)	-	1626081 IS1 insertion (Disruption)
G1	484669 G→T (R59S)	-	-	-	-	1625925 IS1 insertion (Disruption)
G2	-	-	-	-	-	1626100 -6:CCACCA (V13,V14 Deletion)
G3	-	480665 G→A (P988L)	-	3390726 -1:C (Frameshift)	4407505 -7:AGGAAAA (RBS deletion)	1625925 IS1 insertion (Disruption)

Table 4.1: Parallel genotypic adaptation in isobutanol tolerant *E. coli* EcNR1 endpoint populations was investigated by direct sequencing of loci of interest in sets of clonal isolates from each population. Each entry is formatted as follows: mutation position first line, nucleotide changes second line, and protein effect third line. Mutation positions are given as absolute genomic coordinates in the *E. coli* EcNR1 reference sequence (Additional file 2). SNPs are indicated by base transition/transversion. Small insertions are indicated by a "+", with the size (number of bp) of the insertion and sequence of inserted bases. Small deletions are designated by "-" with a format similar to that for small insertions. Ribosome binding site is abbreviated as RBS.

the G2 population, which intriguingly had the lowest fitness out of the six endpoint populations. *tolC* mutations were not detected in any of the populations. The fixation of *acrAB* mutations in five out of six independent populations suggests strong selective pressure and parallel evolution at this locus. Mutations affected amino residues at a variety of positions in the protein structure (Figure 4.6A). The *acrAB* mutations acquired in the isobutanol tolerant lineages bear noteworthy similarities to mutations reported to affect substrate specificity in *acrA* and *mexB*, a *Pseudomonas aerogenosa* structural homolog to *acrB* (Figure 4.6B) [133]. Mutations N154T and R59S of AcrA are spatially proximal to D111N and V244M AcrA mutations reported to affect substrate specificity of AcrA-MexB (Figure 4.6). Mutation V773 of AcrB is in the vicinity of the TolC docking region of MexB / AcrB, where mutation A802V of MexB is known to affect substrate specificity (Figure 4.6). Mutation P988L of AcrB is located in a turn between transmembrane α-helices; several MexB mutations associated with changes in substrate specificity (T329I, T557I, and T489I) also occur in turns between transmembrane α-helices (Figure 4.6).

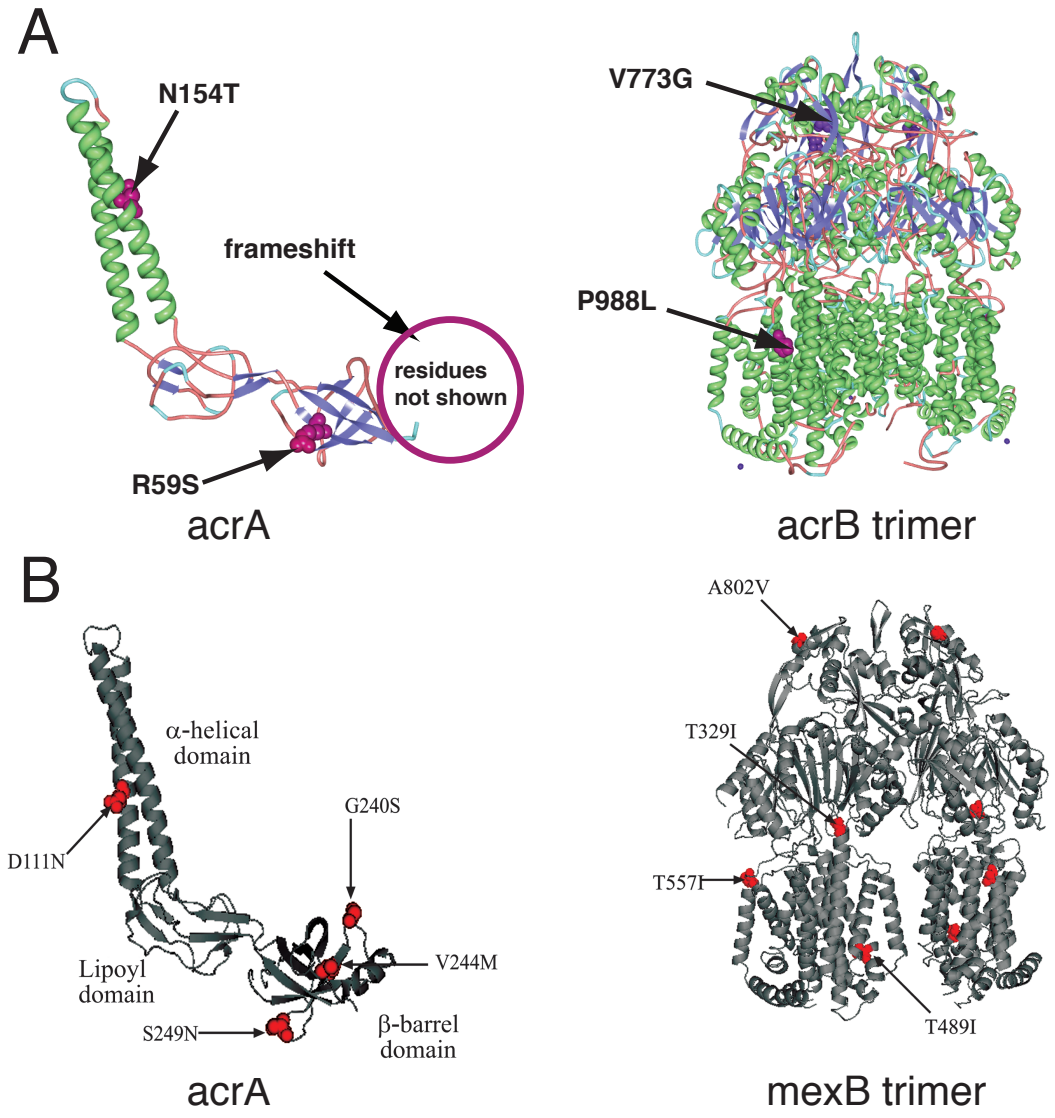

Figure 4.6: AcrAB mutations detected in end populations. The *acrAB* locus was sequenced in clonal isolates from each end population (G1, G2, G3, X1, X2, X3). **(A)** Discovered mutations mapped onto AcrA and AcrB protein structures, **(B)** Mutations in AcrA and MexB (an AcrB homolog in *Pseudomonas aerogenosa*) associated with changes in substrate specificity or AcrA-MexB assembly.

marC mutations were detected in all endpoint populations, providing strong evidence of parallel adaptation at this locus (Table 4.1). All detected *marC* mutations were transposon (IS1 or IS5) insertions, with the exception of an in-frame six bp deletion in G2 (Table 4.1). Transpositions occurred at positions 1625925 and 1626081/1626084, suggesting that these sites are insertion hotspots. Transpositions into *marC* likely cause loss-of-function from disruption, and could also affect expression of the divergently transcribed *marRAB* operon. Functional effects of the *marC* six bp deletion in G2 are not immediately obvious; this mutation results in deletion of two residues (V13 and V14) from a transmembrane helix.

Mutations in *mdh* were also common in the evolved populations, with mutations detected in X2, X3, and G3 at approximately 100% allele frequency (8/8 clones) (Table 4.1). All *mdh* mutations were insertions or deletions resulting in frameshifts. Since substantial numbers of amino acid residues are affected in each case, these mutations are likely to cause loss-of-function of *mdh*. *hfq* mutations were less common in the endpoint populations, with mutations detected in X1 and G3 only. The X1 Hfq mutation I24M is located in the 3-proximal purine nucleotide selectivity pocket (R-site) [134]. The R-site is involved in binding polyA RNA, but possible functional effects of the I24M mutation are not immediately obvious [134]. In G3, the ribosome binding site of *hfq* is partially deleted, potentially leading to lower intracellular Hfq protein levels through reduced translation initiation rate of *hfq* mRNA.

4.3.4 Genotypic evolutionary dynamics

We investigated the dynamics of genotypic adaptation in the G3 and X3 lineages by phenotyping and genotyping population samples from intermediate generations (Figure 4.7, Additional file 3). Phenotyping was done by assessing growth rate at various isobutanol concentrations, while intermediate generation genotyping was conducted by screening whole-population samples for mutations identified in sequenced clones, using Sanger sequencing of PCR amplified loci of interest or allele specific PCR. Due to the large number of mutations in the G3 end population clones, we screened only for those mutations identified in G3.266.7 and the *acrB* and *gatZ* loci (Figure 4.7A). All mutations detected in X3.5 were screened in the intermediate generations (Figure 4.7B).

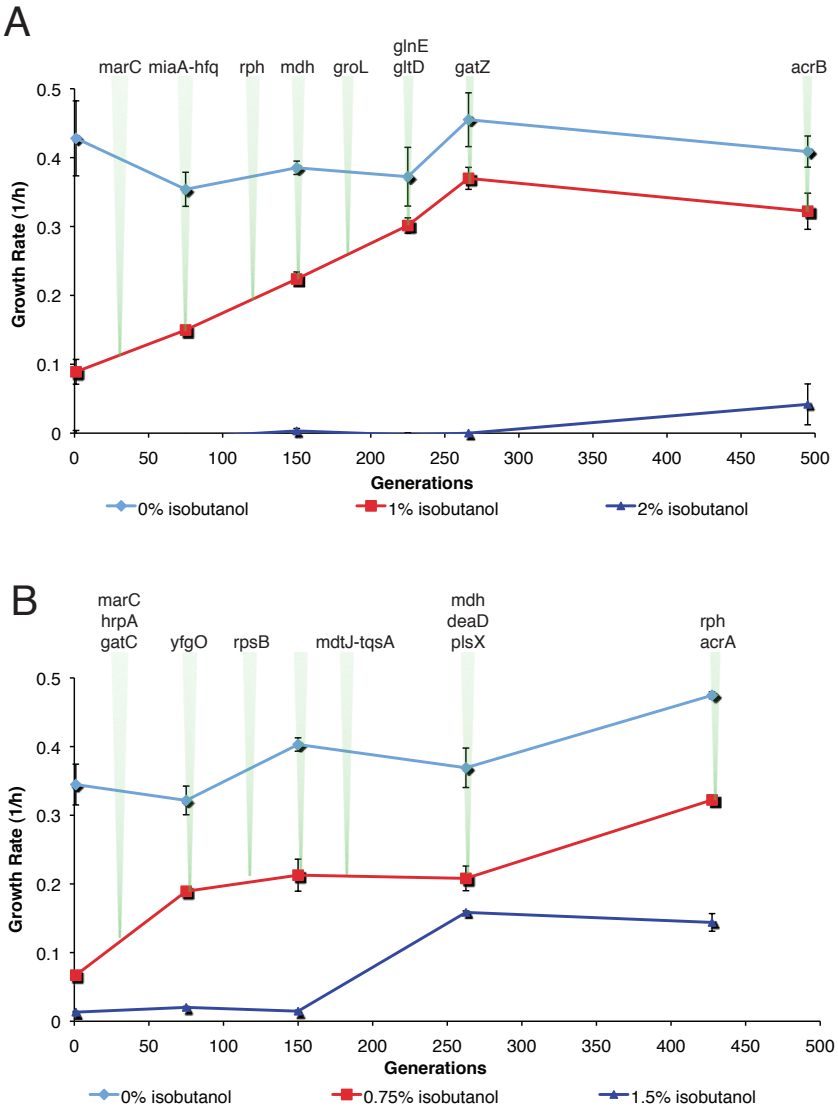

Figure 4.7: Fitness trajectories and temporal order of mutations. Intermediate generations of the G3 and X3 lineages were phenotyped and genotyped. Phenotyping was done by measuring growth rate at various isobutanol concentrations, while genotyping was done via allele specific PCR or direct Sanger sequencing of PCR products from whole population samples. Green arrows denote intermediate generations that were genotyped. Mutations are listed above the first generation in which they were detected. **(A)** Mutations in G3.266.7 and *gatZ* and *acrB* mutations (detected in G3.26/G3.6) were traced through various evolutionary time points. **(B)** All mutations detected in X3.5 were traced through various evolutionary time points.

The phenotype/genotype trajectories reveal that genotypic adaptations in each lineage had pleiotropic effects across different isobutanol concentrations. In both the X3 and G3 lineages, the first mutations acquired (*marC*/*miaA-hfq* in G3 and *marC*/*gatC*/*hrpA*/*yfgO* in X3) appear to drastically increase growth rates at intermediate isobutanol concentrations (1% and 0.75% w/v for G3 and X3, respectively), while having neutral or negative effects at 0% isobutanol (Figure 4.7).

The initial *marC/miaA-hfq* mutations fixed in the G3 lineage appear to have a slightly negative effect on growth rate at 0% isobutanol (Figure 4.7A). Subsequent mutations in the G3 lineage (*rph*, *mdh*, *groL*, *glnE*, *gltD*, and *gatZ*) appear to monotonically increase the growth rate at 1% (w/v) isobutanol while gradually restoring growth rate at 0% isobutanol (Figure 4.7A). In the G3 lineage, the 0% and 1% (w/v) isobutanol growth rate trajectories appear to plateau after about 260 generations, while growth rate at 2% (w/v) isobutanol increases to the endpoint population (Figure 4.7A). In contrast, the growth rate trajectories at 0% and 0.75% (w/v) isobutanol in X3 increase to the end of the evolution, while the growth rate at 1.5% (w/v) isobutanol is relatively constant after generation 266 (Figure 4.7B). Interestingly, in X3 there was a period during the evolution between generations 150 and 266 where growth rate changes at 0% and 0.75% (w/v) isobutanol were flat, while there was a rapid increase in the growth rate at 1.5% (w/v) isobutanol. The growth rate increase in 1.5% (w/v) isobutanol is correlated with an *mdtJ::IS5::tqsA* mutation appearing at generation 180 and a *mdh/deaD/plsX* mutation cluster appearing in generation 266.

4.3.5 DNA microarray study of gene expression changes in G3.2

To gain insights into potential regulatory adaptations to isobutanol stress, we performed a gene expression study with G3.2, a highly isobutanol tolerant sequenced clone. We examined gene expression in G3.2 and the parent *E. coli* EcNR1 (WT) in 0% and 0.5% (w/v) isobutanol glucose minimal medium. For each strain/culture condition (G3.2/0% isobutanol, G3.2/0.5% isobutanol, WT/0% isobutanol, WT/0.5% isobutanol), three biological replicates were employed. Cultures were inoculated in media containing respective amounts of isobutanol, grown to mid log phase and harvested for transcriptome measurement. RNA samples were labelled and hybridized to a custom *E. coli* microarray as described in the Materials and methods section. A total of 4280 genes were included on the microarrays. After a pre-processing procedure that included background adjustment and normalization, 4235 genes with acceptable signals were subject to further analysis. Two filters were first employed to select genes with notable changes across the conditions, which resulted in a list of 2026 genes. Two-sample students t-test was then conducted to determine statistically significant differences in gene expression. The full set of microarray results is included in Additional file 4. As illustrated in Figure 4.8A, 326 and 381 genes were differentially regulated by isobutanol stress in WT and G3.2 respectively. Differential transcriptional response between WT and G3.2 to isobutanol stress was observed for 223 genes, with the most significantly perturbed genes (ranked by p-value) shown in Figure 4.8B (see Additional file 4 for full results). Real time quantitative reverse transcription polymerase chain reaction (qRT-PCR) was used to validate two genes with large expression changes (*gadA* and *fimI*) and two genes with subtle expression

changes (*fabA* and *rfaJ*) (Figure E.1). Target expression levels were determined by fitting a MAK2 model to qRT-PCR data, and expression was normalized to housekeeping gene *rpoD*, which was found to be invariant across all strains/conditions in our microarray data set and has been used in other studies to normalize gene expression data in gram negative bacteria [135, 136]. Expression levels measured by qRT-PCR correlated well with microarray data (Figure E.1).

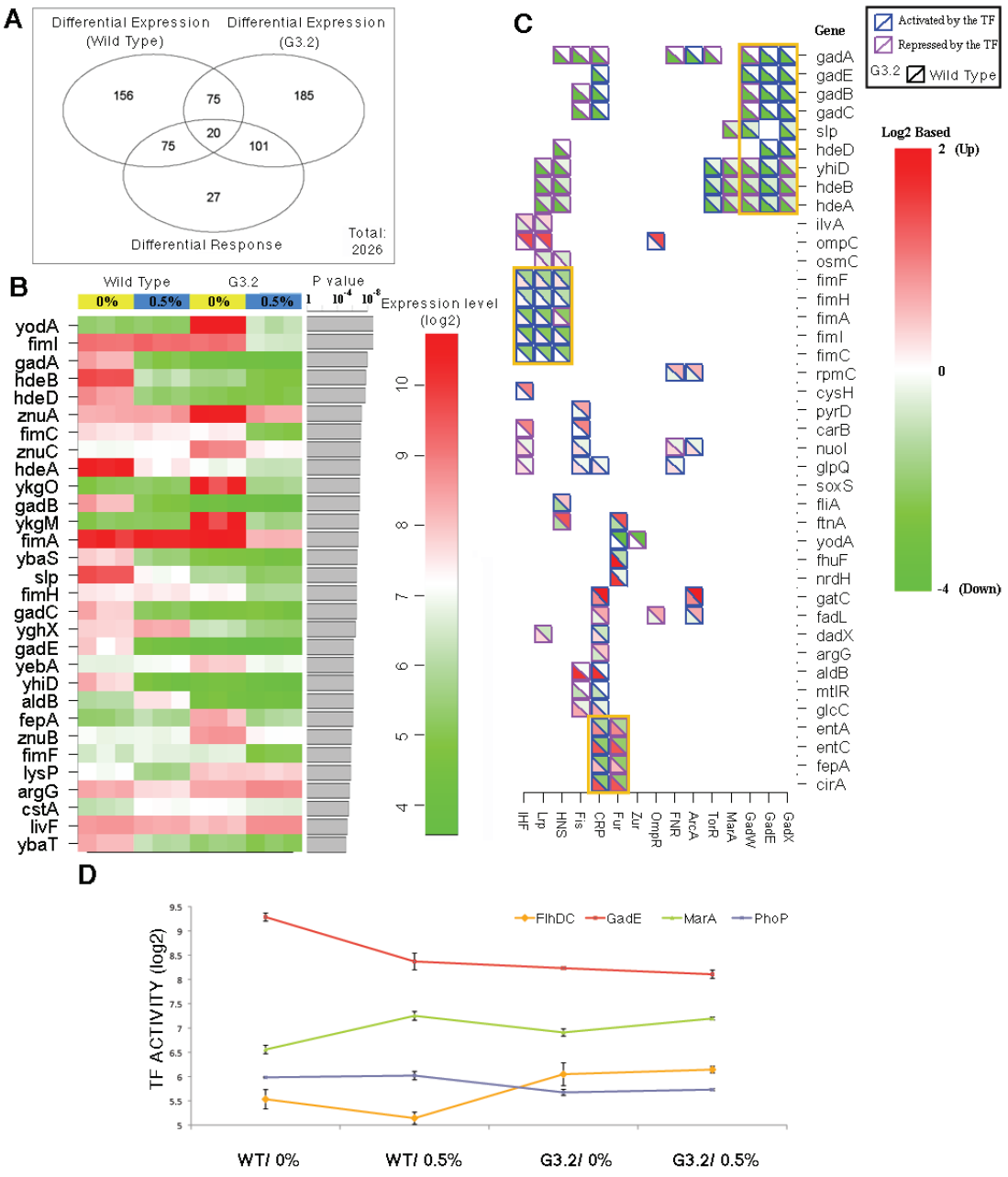

Figure 4.8: Microarray study of gene expression changes in G3.2. DNA microarray study was conducted to study changes in gene expression between isobutanol tolerant clone G3.2 and the parent *E. coli* EcNR1. G3.2 and *E. coli* EcNR1 (WT) were grown to mid-exponential phase in both 0% and 0.5% (w/v) isobutanol spiked minimal media. **(A)** Summary of genes expression changes. **(B)** Top 30 genes with the most significant differences in transcriptional response between G3.2 and WT. **(C)** Top 40 genes with the most significant differences in transcriptional response between G3.2 and WT, and controlling transcription factors. **(D)** Activities of transcription factors predicted by Network Component Analysis (NCA).

BiNGO was used to assess any overrepresented Gene Ontology (GO) in the full set of genes with differential transcriptional response, using p = 0.05 as a cutoff for significance.

Overrepresented ontologies included transition metal ion transport, amine transport, amino acid metabolic processes, glutamine family amino acid metabolic processes, chemical homeostasis, and various cell envelope related components and processes (including flagella and fimbriae, polysaccharide biosynthesis, and lipid metabolism); a full list of overrepresented gene ontologies and related genes is available in Additional file 4. We further investigated changes in regulation by examining transcription factors known to control genes differentially regulated between WT and G3.2. Acid fitness island genes (*gadA*, *gadE*, *gadB*, *gadC*, *slp*, *hdeD*, *yhiD*, *hdeB*, and *hdeA*), regulated by GadE, GadX, and GadW, are strongly repressed at both 0% and 0.5% (w/v) isobutanol in G3.2 (Figure 4.8B and C, Additional file 4) . Fimbrial biogenesis genes (*fimF*, *fimH*, *fimA*, *fimI*, *fimC*), regulated by IHF, Lrp, and HNS, are strongly repressed in G3.2 by isobutanol; genes associated with iron acquisition (*entA*, *entC*, *fepA*, and *cirA*), regulated by Fur and CRP, are found to be repressed by isobutanol in G3.2 as well (Figure 4.8B and C, Additional file 4).

To dissect the apparently complex regulatory changes evolved in G3.2, we applied Network Component Analysis (NCA) to the microarray data to identify transcription factors with significant activity changes in G3.2 compared to WT (Figure 4.8D, Additional file 4). Based on previous study of isobutanol response network in *E. coli* [137] and preliminary examination of our microarray data, we selected 16 transcription factors (TFs) that are potentially involved in isobutanol tolerance (ArcA, PdhR, Fnr, Fur, FlhDC, OmpR, CRP, GadE, MarA, Nac, LexA, PurR, Fis, IHF, PhoB and PhoP) for this analysis. Due to limited data (i.e. four strain/isobutanol conditions), we used a subset of four TFs in each NCA analysis and repeated the analysis for different combinations of TFs. Only TFs with consistent and significant predicted activity changes across different combinations of TFs and different replicates were retained for further analysis. GadE, PhoP, FlhDC, and MarA were subsequently found to be the most significantly perturbed TFs in G3.2 compared to WT (Figure 4.8D).

NCA reveals constitutively reduced activity in G3.2 of GadE, a regulator of the acid fitness island genes, and PhoP, a regulator of genes involved in Mg^{2+} homeostasis, resistance to antimicrobial peptides, acid resistance (including acid fitness island genes), and LPS modification (Figure 4.8D). FlhDC, a master regulator of flagellum biosynthesis, has increased activity in G3.2 and is not repressed by isobutanol, as in WT (Figure 4.8D). MarA, which regulates genes associated with response to oxidative stress, organic solvents, and heavy metals, shows increased activity at 0% isobutanol in G3.2 relative to WT, and reduced upregulation in response to isobutanol. In a previous study of the isobutanol response network in *E. coli* [137], it was concluded that activities of ArcA, PhoB, and Fur were significantly increased by isobutanol stress due to isobutanol induced quinone/quinol malfunction. We performed NCA for various combinations of

ArcA, PhoB, and Fur with FlhDC, GadE, MarA, and PhoP to determine whether these results are recapitulated in our study. We found that for most tested TF combinations, ArcA, PhoB, and Fur activities are increased by isobutanol in WT EcNR1, consistent with previous results (Additional file 4 and [137]). Responses of ArcA, PhoB, and Fur in G3.2 differ from WT, suggesting that these transcriptional responses to isobutanol stress may have changed during evolution (Additional file 4). Especially notable is the differential response of Fur to isobutanol, with upregulation in WT versus downregulation in G3.2 observed for many tested TF combinations (Additional file 4). Many of the top differentially expressed genes identified in our microarray study are regulated by IHF, HNS, Fis, and CRP (which were incidentally identified as being significantly perturbed by isobutanol in [137]), suggesting that these TFs may also be involved in the differential transcriptional response between WT and G3.2 (Figure 4.8C).

Integrated examination of genotype and microarray expression data yields insights into the genetic basis of gene expression and transcription factor activity patterns in G3.2. One of the first mutations fixed in the G3 lineage is *miaA-hfq* 4407505 -7:AGGAAAA, a partial ribosome binding site deletion that is likely to reduce *hfq* mRNA translation. Hfq is a global regulator that functions by mediating binding between a variety of sRNAs and their target mRNAs, which can alter target protein levels via effects on translation initiation or mRNA degradation [138]. Hfq is required for translation of rpoS (σ38) mRNA, the master transcriptional regulator for general stress response; thus G3.2 is expected to have lower RpoS activity [138]. Previous work indicates that in minimal medium, *flhDC* is strongly repressed by RpoS, while *gadE* is strongly upregulated by RpoS; the activity changes observed for these transcription factors are consistent with reduced RpoS activity in G3.2 [139]. Many other gene expression changes in G3.2 are also consistent with reduced RpoS activity (see Additional file 4). In addition to *rpoS*, *hfq* regulates numerous other genes involved in a variety of cellular processes, however since *hfq* regulation is post-transcriptional, many of these effects cannot be captured in a DNA microarray study [138]. Besides possible changes in post-transcriptional regulation, microarray data indicates that *rpoS* is differentially regulated at the transcriptional level in G3.2 compared to WT. *rpoS* is upregulated in WT by isobutanol stress, consistent with previous gene expression studies [137] (Additional file 4). In contrast, in G3.2 *rpoS* expression appears to be slightly repressed by isobutanol; furthermore the basal expression level of *rpoS* in G3.2 is lower compared to WT, providing additional evidence of reduced RpoS activity in G3.2.

NCA analysis revealed constitutively reduced activity of the PhoP and GadE transcriptional regulators. PhoP is part of a Mg^{2+} responsive two-component signal transduction system, with sensor kinase PhoQ phosphorlyating (and thus activating) PhoP in response to low Mg^{2+}

levels [140, 141]. Interestingly, G3.2 has a *phoQ* 1197581 A→G mutation, causing L209P in transmembrane region 2 in the PhoQ protein, which may lead to reduced activity of the PhoPQ system. Transcriptional changes caused by PhoPQ perturbation are potentially adaptive, since PhoP is involved in stress response and regulates genes related to Mg^{2+} homeostasis, resistance to antimicrobial peptides, acid resistance, and LPS modification [140, 141]. In a previous NCA study [137], GadE activity was found to be strongly repressed by isobutanol. This finding was recapitulated in our NCA results for WT, while in G3.2 GadE is constitutively repressed (Figure 4.8D). The evolution of constitutive GadE repression in G3.2 hints that the GadE regulon (comprised of the major acid resistance genes) may be maladaptive to isobutanol stress. There is substantial overlap between the PhoP, GadE, Hfq, and RpoS regulons, pointing towards possible co-evolution between these different regulators.

4.3.6 Investigating phenotypic and functional effects of mutations

Previous investigations have identified the cell envelope as a primary target of solvent toxicity. G3.2 contains mutations in numerous genes and regulators associated with the cell envelope, including *secA* and *lepB* (components of the Sec apparatus, which translocates periplasmic and membrane targeted proteins from the cytosol), *hfq* (involved in sRNA mediated regulation of many membrane proteins), *fepE* and *yjgQ* (involved in LPS biosynthesis), and *phoPQ* (regulator of various LPS modification genes). Additionally, the DNA microarray study revealed that many genes related to cell envelope components and processes were differentially expressed in G3.2. We investigated possible cell envelope adaptations by profiling cellular fatty acid composition and cell envelope proteins in the parent *E. coli* EcNR1 strain and G3.2 during growth at 0.5% isobutanol (Figure 4.9A and B). Cellular fatty acid composition was determined using gas chromatography-flame ionization detector (GC-FID) quantification, and was found to differ considerably between G3.2 and WT EcNR1 (Figure 4.9A). The cyclopropane fatty acid fraction is significantly reduced in G3.2, probably as a result of downregulation of *cfa* (cyclopropane fatty acyl phospholipid synthase) in this strain (Figure 4.9A; see Additional file 4 for *cfa* expression data from the DNA microarray study). Additionally, the overall unsaturated:saturated fatty acid ratio is increased in G3.2 (Figure 4.9A), due mainly to an increase in the proportion of C16:1 and C18:1 fatty acids relative to C16:0 (data not shown).

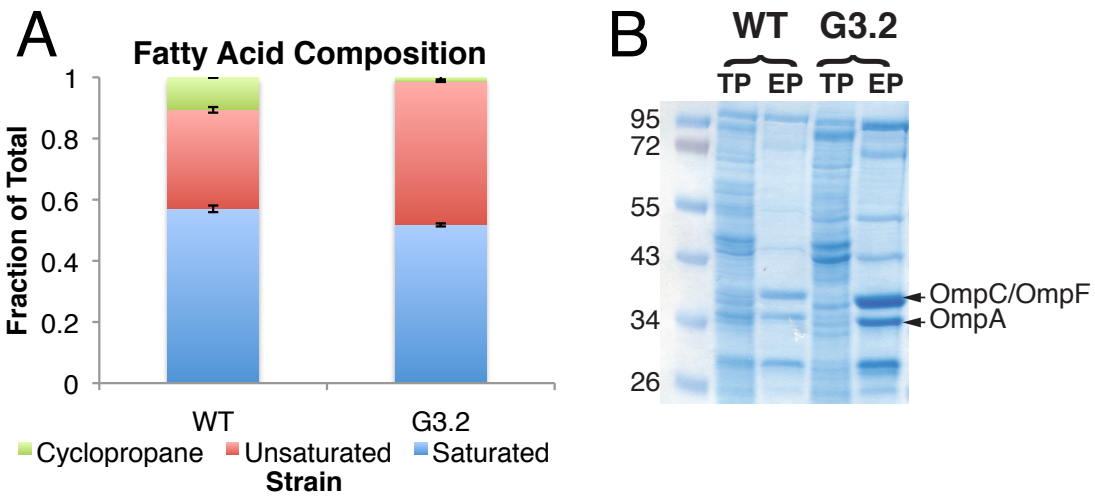

Figure 4.9: Cell envelope composition of G3.2 compared to WT EcNR1. Possible cell envelope adaptations in G3.2 were investigated by profiling cellular fatty acid composition and cell envelope proteins in G3.2 and the parent *E. coli* EcNR1 during growth at 0.5% (w/v) isobutanol. **(A)** Fatty acid composition of G3.2 and WT EcNR1. Relative proportions of cyclopropane, unsaturated, and saturated fatty acids were determined by GC-FID analysis. **(B)** SDS-PAGE profile of cell envelope proteins in G3.2 and WT EcNR1. Cell envelopes were isolated from $5x10^9$ cells and analyzed by SDS-PAGE. For comparison, 20 μg total cellular proteins (TP) from each strain (WT and G3.2) were analyzed alongside the isolated cell envelopes (EP). As a reference, bands corresponding to outer membrane proteins OmpA and OmpC/OmpF are indicated with arrows. Experiment was repeated to verify results (not shown).

To determine cell envelope protein profiles, cell envelopes were isolated from $5x10^9$ cells by sonication and differential centrifugation and then analyzed with sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) (Figure 4.9B). SDS-PAGE analysis reveals an overall increase in cell envelope proteins (on a per cell basis) in G3.2 compared to WT. (Figure 4.9B). To examine changes in relative protein abundance between G3.2 and WT, protein bands were quantified by densitometry analysis (using ImageJ software) and normalized to the sum of intensities of the major protein bands. The 72 kDa, 55 kDa, and OmpC/OmpF bands were found to be notably upregulated in G3.2 relative to WT, with relative increases of 1.2, 2.2, and 1.3 fold, respectively (Figure 4.9B). Upregulation of OmpC/OmpF is consistent with DNA microarray results, which show upregulation of *ompF* in G3.2 (Additional file 4).

In addition to characterizing possible cell envelope adaptations in G3.2, we conducted detailed investigations of phenotypic and functional effects of key mutations identified in isobutanol tolerant clones. Selected mutations were reconstructed in *E. coli* EcHW24 (EcNR1 Δ*mutS*) singly and in various combinations using ssDNA mediated recombination [59]. We focused on characterizing parallel genotypic adaptations, including *marC*, *acrAB*, *mdh*, and *rph* mutations identified in G3.2 and X3.5, as well as the first five mutations to appear in the G3 lineage, *marC*, *miaA-hfq*, *rph*, *mdh*, and *groL*, which are associated with monotonically increasing isobutanol

tolerance (Figure 4.6A). *marC*, *acrAB*, *mdh*, and *rph* single mutants were constructed to study the phenotypic and functional effects of these mutations in isolation, while *marC*, *miaA-hfq*, *rph*, *mdh*, and *groL* mutations were constructed singly and in various combinations to study fitness benefits and investigate possible epistatic interactions among these mutations. Phenotypic effects were investigated by measuring growth of mutants in isobutanol spiked minimal medium, and functional assays were performed for *acrAB* and *mdh*. The parent *E. coli* EcNR1 and *E. coli* EcNR1 single gene knockouts (Δ*acrA::kan*, Δ*acrB::kan*, Δ*mdh::kan*) were employed as controls in phenotype and functional assays.

marC mutations were detected in every evolution endpoint population. All detected *marC* mutations were transposon (IS1 or IS5) insertions, with the exception of an in-frame six bp deletion in G2. *marC* transposon insertions could not be produced with ssDNA mutagenesis, so we instead approximated the effect of transposon insertions by knocking out *marC*, reasoning that this could mimic effects of gene disruption caused by transposon insertion; additionally, deletion of *marC* was found to improve isobutanol tolerance in an independent study (James C. Liao, UCLA personal communications). Consistent with our expectations, Δ*marC::kan* was found to significantly improve maximum specific growth rates and final densities in 0.5% (w/v) isobutanol minimal medium relative to the parent *E. coli* EcNR1 (Table 4.2). Growth rate improvement of Δ*marC::kan* was higher in xylose medium (39±2% above WT growth rate) compared to glucose medium (20±5% above WT growth rate) at 0.5% (w/v) isobutanol (Table 4.2). In contrast, Δ*marC::kan* improved final cell densities more in glucose medium compared to xylose medium (40±10% vs. 7.6±0.5% improvement over WT; Table 4.2). Δ*marC::kan* had a slight negative effect on maximum specific growth rate and final cell densities at 0% (w/v) isobutanol in both xylose and glucose media (Table 4.2).

acrAB mutations were identified in five out of six independent evolved populations, suggesting that mutations at this locus are likely to have positive adaptive effects. Consistent with this expectation, *acrA* 483735 +1:A (identified in X3.5) and acrB 480665 G→A (identified in the G3 lineage) dramatically increased maximum specific growth rates and final cell densities in 0.5% (w/v) isobutanol minimal medium relative to the parent *E. coli* EcNR1, while having more subtle effects on growth in 0% isobutanol (Table 4.2). Δ*acrA::kan* and Δ*acrB::kan* produced fitness benefits of similar or greater magnitude, implying that loss-of-function of *acrAB* is associated with improved isobutanol tolerance (Table 4.2). This result is surprising given that the AcrAB-TolC efflux pump is an important mechanism of tolerance to other organic solvents and antibiotics. AcrAB-TolC efflux pump activity was measured via ethidium bromide (EtBr) accumulation in reconstructed single mutants and clonal isolates harbouring *acrAB* mutations

Locus	Gene/mutation	Clone	0% i-BtOH		0.5% i-BtOH		Functional Effect
			$\Delta\mu_{max}$	ΔOD$_{max}$	$\Delta\mu_{max}$	ΔOD$_{max}$	
marC	Δ*marC::kan (xylose media)*	X1,X2,X3 (IS1/IS5 insertion)	-5.0 ±0.5%	-2.4±0.1%	39±2%	7.6±0.5%	-
	Δ*marC::kan (glucose media)*	G1,G3 (IS1 insertion)	-3.2 ±0.2%	-7.8±0.3%	20±5%	40±10%	-
acrAB	*acrA* 483735 +1:A	X3.5	14±1%	3.3±0.1%	49±9%	72.4±0.8%	Reduced EtBR efflux; 188±7% increase in intracellular EtBr
	Δ*acrA::kan*	N/A (control)	6.5±0.6%	7.6±0.5%	32±5%	103±1%	Reduced EtBR efflux; 210±10% increase in intracellular EtBr
	acrB 480665 G→A	G3.2	11.2±0.9%	2.1±0.1%	22±3%	31±1%	Reduced EtBR efflux; 21±8% increase in intracellular EtBr
	Δ*acrB::kan*	N/A (control)	5.7±0.6%	-3.6±0.1%	8.2±1.1%	64±2%	Reduced EtBR efflux; 340±20% increase in intracellular EtBr
mdh	*mdh* 3390936 +5:AACCT	X3.5	-3.7±0.3%	-1±0.1%	0.4±0.1%	13±2%	Loss of function; no detectable Mdh activity
	mdh 3390726 -1:C	G3.2	2.3±0.1%	4.4±0.1%	-8.1±3%	-1.2±0.1%	Loss of function; no detectable Mdh activity
	Δ*mdh::kan*	N/A (control)	-1.6±0.2%	4±0.1%	10.8±1.6%	10±1%	Loss of function; no detectable Mdh activity
rph	*rph* 3823220 +4:GTCG	G3.2	39±2%	11.9±0.1%	49±16%	-12.9±0.1%	-

Table 4.2: Phenotypic and functional effects of selected *marC*, *acrAB*, *mdh*, and *rph* mutations. Selected point mutations in *acrAB*, *mdh*, and *rph* identified in isobutanol tolerant clones were reconstructed in the parent *E. coli* EcNR1 strain using ssDNA mediated mutagenesis. *marC* transposon insertions could not be produced with ssDNA mutagenesis; instead, we approximated the effect of transposon insertions by knocking out *marC*. Reconstructed mutants were phenotyped by measuring specific growth rate in minimal media with 0% and 0.5% (w/v) isobutanol; percent change in growth rate and maximum OD_{600} relative to the parent *E. coli* EcNR1 are reported. Functional effects of *acrAB* mutations were assessed with an *in vivo* ethidium bromide accumulation assay and *mdh* mutations were assessed by directly measuring malate dehydrogenase enzyme activity in cell lysates.

from evolution end populations [142]. Since AcrAB-TolC is the primary efflux pump for EtBr, mutations altering AcrAB-TolC activity or substrate specificity would be expected to affect the accumulation of intracellular EtBr [142]. Increased EtBr accumulation (consistent with reduced AcrAB-TolC activity) was observed in all examined end population clonal isolates harbouring *acrAB* mutations (full data set in Figure E.2). The *acrA* 483735 +1:A single mutant had an EtBr accumulation profile similar to Δ*acrA::kan* and X3.5. In contrast, the EtBr accumulation profile in G3.2 was similar to Δ*acrB::kan*, but *acrB* 480665 G→A (identified in the G3.2) showed only modest changes in EtBR accumulation relative to the parent strain, implying that G3.2 may have additional mutations affecting efflux pump activity.

The AcrAB-TolC multidrug efflux pump has been well characterized in its role for antibiotic and solvent tolerance, but a recent study suggests that AcrAB-TolC may also function as an exporter for a hitherto unidentified quorum sensing signal (QSS) [143]. There is strong evidence that the QSS exported by AcrAB-TolC is associated with upregulation of *rpoS* transcription; Δ*acrAB* mutants have reduced *rpoS* expression and altered temporal patterns of expression [143]. Our gene expression study of G3.2 provides evidence of reduced RpoS activity in this strain. Interestingly, two evolved populations, X1 and G3, were found to have mutations in *hfq*, which is required for translation of *rpoS* mRNA, suggesting that RpoS modulation might be a common adaptive effect of these different mutations. We assayed RpoS activity via iodine staining in the parent *E. coli* EcNR1 strain, each evolution endpoint population (G1, G2, G3, X1, X2, X3), a Δ*acrA::kan* mutant, and a constructed single mutant containing the *miaA-hfq* mutation found in the G3 lineage (*miaA-hfq* 4407505 -7:AGGAAAA). RpoS positively regulates glycogen biosynthesis, which can be measured by staining cells with iodine cells with higher glycogen levels stain darker [144]. While this assay is an indirect measure of RpoS activity and is subject to many confounding factors (such as other regulation of glycogen biosynthesis), it is commonly used in literature and has been demonstrated to be well correlated with RpoS activity [144].

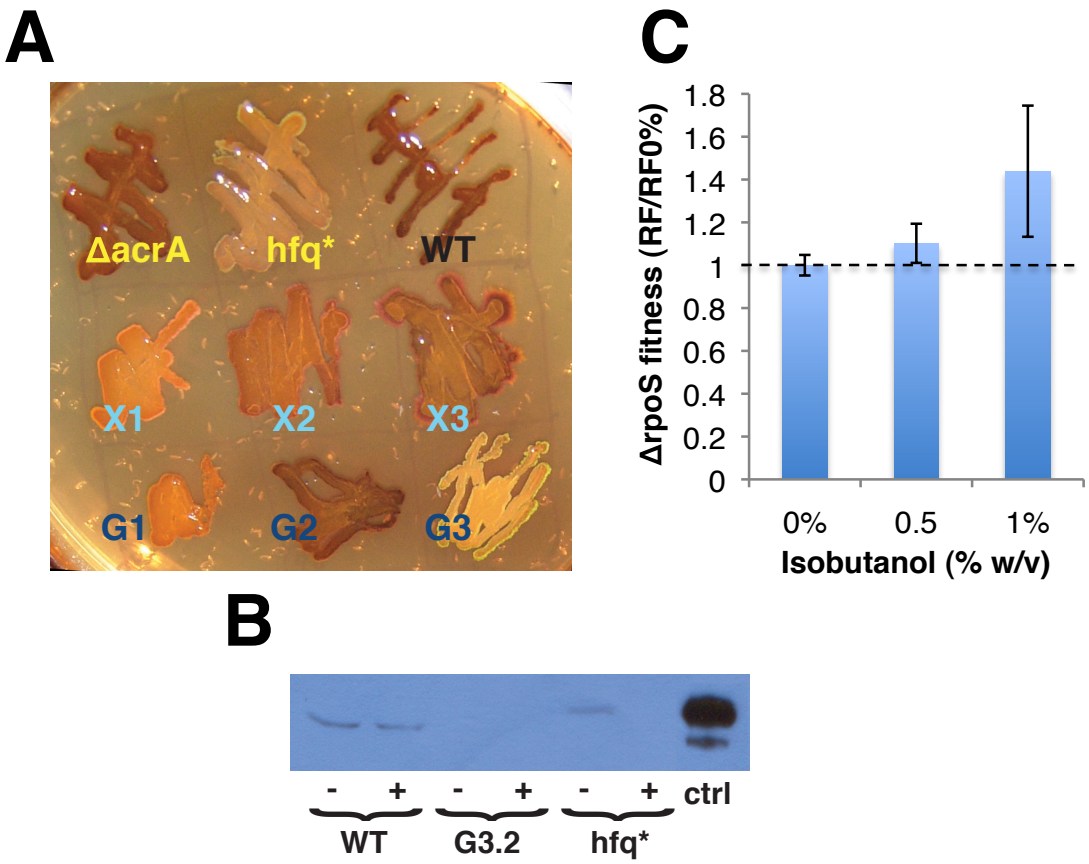

Figure 4.10: Survey of RpoS activity in evolved populations and phenotype study of a ΔrpoS mutant. RpoS activity was assayed in evolved populations and selected single mutants using an I_2 staining assay and Western blot analysis. **(A)** I_2 staining assay. Overnight cultures were streaked on glucose minimal medium agar spiked with 0.35% (w/v) isobutanol, incubated at 30°C for 48 hours, and then stained with USP tincture of iodine. Samples are as follows (from top left, to bottom right): *E. coli* EcNR1 Δ*acrA::kan* (ΔacrA), *E. coli* EcHW24 *miaA-hfq* 4407505 -7:AGGAAAA (hfq*), *E. coli* EcNR1 (WT), Xylose #1 end population (X1), Xylose #2 end population (X2), Xylose #3 end population (X3), Glucose #1 end population (G1), Glucose #2 end population (G2), Glucose #3 end population (G3). **(B)** Western blot analysis of RpoS in total cellular protein extracted from cultures of *E. coli* EcNR1 (WT), G3.2, and *E. coli* EcHW24 *miaA-hfq* 4407505 -7:AGGAAAA (hfq*) grown to early exponential phase either with (+) or without (-) 0.5% (w/v) isobutanol in NG50 medium. Experiment was repeated several times to verify results; representative Western blot shown. Purified *E. coli* RpoS (NeoClone) was used as a positive control (ctrl). **(C)** Phenotype study of a ΔrpoS mutant. *E. coli* BW25113 Δ*rpoS::kan* (obtained from the Keio collection [145]; strain # JW5437-1) and parent strain *E. coli* BW25113 were grown in 0%, 0.5%, and 1% (w/v) isobutanol glucose media. To facilitate comparison, we report normalized relative fitness (RF/RF0 %), defined as relative fitness divided by relative fitness at 0% (w/v) isobutanol; relative fitness (RF) was calculated as $\mu_{\Delta rpoS}/\mu_{WT}$ where $\mu_{\Delta rpoS}$ is the maximum specific growth rate (1/h) of *E. coli* BW25113 Δ*rpoS::kan* and μ_{WT} is the maximum specific growth rate (1/h) of *E. coli* BW25113.

Iodine staining results are show in Figure 4.10A. Single mutant *miaA-hfq* 4407505 -7:AGGAAAA (hfq* in Figure 4.10A) stains lighter than the parent *E. coli* EcNR1 strain (WT in Figure 4.10A), consistent with the expected reduction of Hfq and RpoS activity in this mutant. Both of the end populations harbouring *hfq* mutations, X1 and G3, stain much lighter than WT suggesting reduced Hfq activity and subsequently RpoS levels in both end populations (Figure 4.10A). G1 and X2 also show significantly lighter staining than WT, suggesting reduced

RpoS activity in these strains as well (Figure 4.10A). Staining in X3 is only slightly lighter than WT, while G2 and Δ*acrA::kan* (unexpectedly) stain very similarly to WT. Curiously, the association between Δ*acrAB* and reduced RpoS reported in the literature was not evidenced in our iodine staining assay. We suspect that this discrepancy may be due to differences in assay techniques. Previous studies of RpoS activity of Δ*acrAB* mutants were done with liquid cultures, with RpoS activity assayed by real-time PCR or Western blotting, while our assay was done on solid medium using iodine staining to measure intracellular glycogen levels, which are directly controlled by RpoS [143]. Concentrations of the QSS exported by AcrAB-TolC are likely to vary dramatically between liquid and solid cultures due to cell density differences, and could thus confound assay results.

As a follow up to the I_2 staining assay, we directly checked RpoS expression by Western blot analysis of RpoS in the parent strain (WT), G3.2, and single mutant *miaA-hfq* 4407505 -7:AGGAAAA (hfq*) grown with and without 0.5% (w/v) isobutanol (Figure 4.10B). RpoS Western blot analysis was repeated several times to verify results; Figure 4.10B shows a representative Western blot. RpoS expression is evident in the parent EcNR1 strain (WT) at both 0% and 0.5% (w/v) isobutanol, while RpoS expression was not detected in G3.2 under either condition (Figure 4.10B). Interestingly, in hfq* RpoS is detectable at 0% isobutanol, but not at 0.5% (w/v) isobutanol. These results directly demonstrate reduced RpoS expression at 0.5% (w/v) isobutanol in G3.2 and the *miaA-hfq* single mutant relative to the parent EcNR1 strain, consistent with I_2 staining results (Figure 4.10A). We attempted Western blot analysis of RpoS in other strains (including evolution endpoint populations G1, G2, G3, X1, X2, and X3; sequenced isobutanol tolerant clone X3.5; and Δ*acrA::kan* single mutant); however, due to inconsistent outcomes between experiments, we are unable to draw conclusions about RpoS expression levels in these other strains (results not shown). To ascertain whether reduced RpoS activity is indeed adaptive to isobutanol stress, we examined the isobutanol tolerance phenotype of a Δ*rpoS::kan* mutant (Figure 4.10C). Δ*rpoS::kan* caused a growth defect at 0% and 0.5% (w/v) isobutanol relative to the WT strain, while at 1% (w/v) isobutanol the relative fitness of Δ*rpoS::kan* is slightly higher than WT; to facilitate comparison, we report normalized relative fitness (calculated by dividing relative fitness by relative fitness at 0% w/v isobutanol). These results suggest that attenuated RpoS activity may indeed be adaptive to isobutanol stress, since normalized relative fitness is increased at 0.5% and 1% (w/v) isobutanol (Figure 4.10C). However, complete loss-of-function of *rpoS* appears to incur significant costs that overshadow adaptive effects at isobutanol concentrations below 1% (w/v) (Figure 4.10C).

mdh mutations appear in three out of six evolution end populations, suggesting that these

mutations may be adaptive. However, in 0% and 0.5% (w/v) isobutanol spiked minimal medium, we found relatively minor differences in growth between the parent *E. coli* EcNR1 strain, *mdh* 3390726 -1:C (found in G3.2) single mutant, *mdh* 3390936 +5:AACCT (found in X3.5) single mutant, and Δ*mdh::kan* (Table 4.2). Thus *mdh* mutations do not appear to improve isobutanol tolerance in isolation, hinting that fitness benefits may come via epistatic interactions with other mutations. All of the *mdh* mutations identified in evolution end populations were indels causing frameshifts, suggesting that these mutations lead to loss-of-function. To assess functional effects of *mdh* mutations, NADH dependent malate dehydrogenase activity was measured in crude cell lysates of G3.2, *mdh* 3390726 -1:C single mutant, X3.5, *mdh* 3390936 +5:AACCT single mutant, Δ*mdh::kan*, and the parent *E. coli* EcNR1. NADH dependent malate dehydrogenase activity was not detectable in G3.2, *mdh* 3390726 -1:C single mutant, X3.5, *mdh* 3390936 +5:AACCT single mutant, or Δ*mdh::kan*, while assay of the parent *E. coli* EcNR1 yielded enzyme activity of 3.8±0.2 U/mg-wet-cells, consistent with our expectation that 3390726 -1:C and 3390936 +5:AACCT lead to loss-of-function of *mdh*.

Restorative mutations are commonly observed in *rph-pyrE* during experimental evolution studies with *E. coli* K12 MG1655 [115], and indeed all sequenced clonal isolates from our evolution end populations had *rph* mutations. We investigated the adaptive benefits of the *rph* 3823220 +4:GTCG mutation acquired in the G3 lineage. This mutation was found to substantially improve maximum specific growth rate in both 0% and 0.5% (w/v) isobutanol spiked glucose minimal medium, consistent with the notion that *rph* mutations are a general adaptation to growth on minimal media (Table 4.2).

Genotypic adaptation to isobutanol stress is complex and involves diverse genetic loci, as revealed in our genome resequencing results. The apparent multigenic nature of isobutanol tolerance suggests that epistasis, interactions between different genes, is probably an important factor in many of the evolved genetic adaptations. To study fitness benefits and investigate possible epistasis, the first five mutations fixed in the G3 lineage (Figure 4.7A), *marC*, *miaA-hfq*, *rph*, *mdh*, and *groL*, were reconstructed singly and in various combinations in *E. coli* EcHW24, using multiplex recursive ssDNA mediated mutagenesis [59]. As explained above, the *marC::IS1* mutation could not be created using ssDNA recombination, so instead we knocked out *marC* (*marC::kan*) to approximate gene disruption effects caused by IS1 insertion. The resulting mutant set was phenotyped by measuring the maximum specific growth rate in 0%, 0.5%, and 1% (w/v) isobutanol glucose minimal media; results are presented as relative fitness, defined as mutant maximum specific growth rate divided by maximum specific growth rate of the parent *E. coli*

EcHW24. Epistasis is assumed to follow a simple multiplicative fitness model $w = \varepsilon \prod_i^n w_i$, where $w =$ relative fitness of a particular mutation combination, $\varepsilon =$ total epistatic interaction parameter, and $w_i =$relative fitness of single mutants; log epistasis is calculated as $\log \varepsilon = \log w - \sum_i^n \log w_i$ (Figure 4.11) [146].

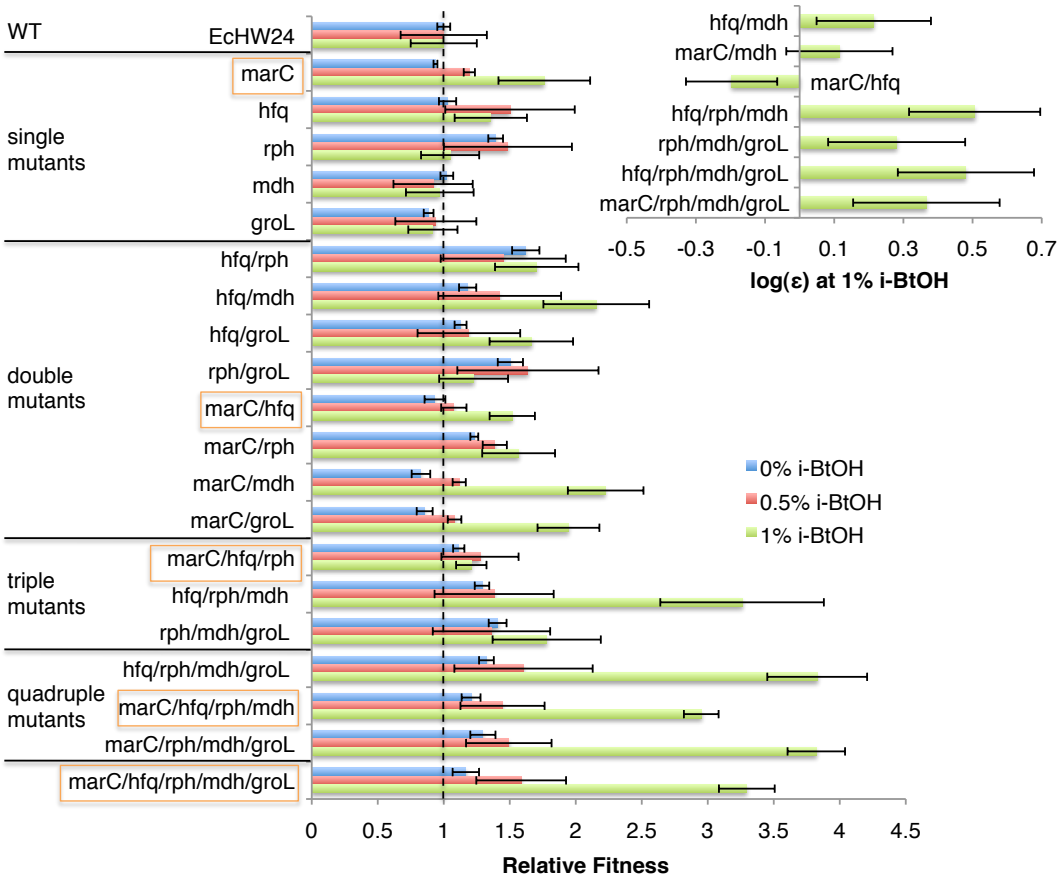

Figure 4.11: Fitness effects of first five mutations fixed in G3 lineage. The first five mutations fixed in the G3 lineage (*marC*, *miaA-hfq*, *rph*, *mdh*, and *groL*) were constructed singly and in various combinations in *E. coli* EcHW24 (EcNR1 Δ*mutS*) using ssDNA mediated recombination; the *marC::IS1* mutation was approximated by knocking out *marC*. The constructed mutants and *E. coli* EcHW24 were phenotyped by measuring the maximum specific growth rate in 0%, 0.5%, and 1% (w/v) isobutanol spiked minimal glucose media; relative fitness was calculated as mutant maximum specific growth rate (1/h) divided by *E. coli* EcHW24 maximum specific growth rate (1/h). Mutation combinations corresponding to the order of appearance in G3 are highlighted. Epistasis is assumed to follow a simple multiplicative fitness model $w = \varepsilon \prod_i^n w_i$, where $w =$ relative fitness of a particular mutation combination, $\varepsilon =$ total epistatic interaction parameter, and $w_i =$relative fitness of single mutants; log epistasis is calculated as $\log \varepsilon = \log w - \sum_i^n \log w_i$

As would be expected, there is a general trend of improved relative fitness with increasing numbers of mutations (Figure 4.11; *miaA-hfq* abbreviated *hfq* and *marC::kan* abbreviated

marC). Relative fitness improvements at 1% (w/v) isobutanol are the most dramatic, with the *miaA-hfq*/*rph*/*mdh*/*groL* and *marC*/*rph*/*mdh*/*groL* quadruple mutants having a 3.8 fold increase in growth rate compared to *E. coli* EcHW24; fitness changes at 0% and 0.5% (w/v) isobutanol are much smaller, and appear to plateau with introduction of an *rph* mutation (Figure 4.11). Individually, the *marC*, *miaA-hfq*, *rph*, *mdh*, and *groL* mutations have relatively modest effects. *mdh* and *groL* single mutants have fitness essentially identical to the parent *E. coli* EcHW24 at all tested isobutanol concentrations (Figure 4.11). The *rph* single mutant has improved relative fitness at 0% and 0.5% isobutanol, while the *marC* and *miaA-hfq* single mutants have improved relative fitness at 0.5% and 1% isobutanol (Figure 4.11). Notable improvements in relative fitness in 1% (w/v) isobutanol were observed for some double mutants, in particular, *miaA-hfq*/*rph*, *miaA-hfq*/*mdh*, *marC*/*mdh*, and *miaA-hfq*/*groL* (Figure 4.11). We suspect that there may be positive epistasis between *miaA-hfq* and each of *mdh*, *rph*, and *groL*; however due to limited growth rate measurement precision, we can assert statistically significant epistasis only for *miaA-hfq* and *mdh* (Figure 4.11). Likewise, positive epistasis between *marC* and each of *mdh* and *groL* seems plausible (Figure 4.11), but cannot be ascertained due to limited measurement precision. Interestingly, *marC* and *miaA-hfq* demonstrate significant negative epistasis at 1% isobutanol (Figure 4.11).

Many higher order mutation combinations show substantial relative fitness improvements at 1% (w/v) isobutanol (Figure 4.11). The quadruple mutants *miaA-hfq*/*rph*/*mdh*/*groL* and *marC*/*rph*/*mdh*/*groL* have the greatest relative fitness, with a 3.8 fold improvement in growth rate, followed by the *marC*/*miaA-hfq*/*rph*/*mdh*/*groL* pentuple mutant and *hfq*/*rph*/*mdh* triple mutant, each having a 3.3 fold improvement in growth rate (Figure 4.11). Results for the reconstructed *marC*/*hfq*/*rph*/*mdh* quadruple mutant and full pentuple mutant are approximately consistent with the evolution trajectory (Figure 4.7A). Epistasis analysis reveals significant positive epistatic interactions for *miaA-hfq*/*rph*/*mdh*, *rph*/*mdh*/*groL*, *miaA-hfq*/*rph*/*mdh*/*groL*, and *marC*/*rph*/*mdh*/*groL* (Figure 4.11). Comparison of fitness effects of *rph*/*mdh*/*groL* vs. *miaA-hfq*/*rph*/*mdh*/*groL* and *marC*/*rph*/*mdh*/*groL* provides compelling evidence that an *miaA-hfq* or *marC* genetic background exhibits positive epistasis with *rph*/*mdh*/*groL* (Figure 4.11).

4.4 Discussion

Due to broad mechanisms of toxicity, microbial solvent tolerance is a complex phenotype, involving adaptations in diverse cellular processes [50]. This inherent complexity suggests that genotypic adaptation to solvent stress will involve a rugged fitness landscape with many epistatic

interactions between genes [147]. Fitness landscape topology and epistasis have important ramifications for efforts towards engineering complex phenotypes. Many of the approaches previously employed to investigate genetic bases of adaptation to solvent stress are inherently limited to exploring restricted regions of the fitness landscape and often fail to capture interactions between distal genes; thus these approaches may fail to uncover many important adaptations [50]. In our study, we used experimental evolution of multiple lineages of *E. coli* under isobutanol stress followed by genome resequencing and phenotypic characterization, allowing us to investigate the full bases of adaptation in our evolved lineages. Our results reveal many novel patterns of genotypic adaptation and suggest several important tolerance mechanisms, informing future efforts towards engineering more robust strains of *E. coli* for isobutanol production and also providing general insights into the evolution of complex stress tolerance phenotypes.

4.4.1 Genotypic patterns of adaptation: epistasis, global effect mutations, and parallel evolution

Consistent with the complex nature of solvent tolerance, our genome resequencing results reveal genetic adaptations in a diversity of cellular processes (Table 4.2). The apparent multigenic nature of isobutanol tolerance suggests that epistatic interactions and coevolution between different genetic loci are probably important factors in many of the evolved adaptations. Examining the genotypes of X3.5 and G3.2 reveals several possible examples of epistasis and coevolution between genes that encode interacting proteins or that participate in functionally related cellular processes. G3.2 has mutations in *secA* and *lepB*, two components of the Sec protein translocation apparatus that exports periplasmic and membrane proteins from the cytosol (Table E.1, Figure 4.5) [148]. The SecA S233P mutation is in the preprotein binding domain of SecA, which recognizes and binds nascent cytosolic peptides targeted for export, while LepB is a signal peptidase that cleaves the N-terminal leader peptide proteins after secretion; these mutations may collectively alter the peptide specificity of the Sec complex [148, 149]. Both clones sequenced from the G3 lineage (G3.2 and G3.266.7) have a probable loss-of-function mutation (indel leading to frameshift) in *gltD*, a subunit of glutamate synthase, and *glnE*, a regulator of glutamine synthase activity; these mutations suggest a rewiring of nitrogen metabolism towards increased glutamine synthesis [150]. The X3 lineage acquired mutations in *rpsB*, the 30S ribosomal subunit S2, and *deaD* (*csdA*), an RNA helicase involved in ribosome biogenesis that is known to be a multicopy suppressor of temperature-sensitive *rpsB* mutants [151]. The *rpsB* mutation, which occurs before *deaD* in the X3 lineage, is associated with improved growth at 0% and 0.75% (w/v) isobutanol, while the appearance of a *mdh*/*deaD*/*plsX* mutation cluster is associated with improved growth

at 1.5% (w/v) isobutanol (Figure 4.7B). Possible functional effects of the *rpsB*/*deaD* mutations are not clear. However, recent studies have identified alterations in rRNA processing that contribute to acid tolerance in *Clostridium acetobutylicum* and mutations in ribosomal machinery in *Pseudomonas putida* that contribute to chemical tolerance, setting a precedent for ribosomal mechanisms of complex stress tolerance [152, 153].

Epistasis can also occur though more cryptic mechanisms. The evolution trajectory (Figure 4.7A) showed that fitness at 1% (w/v) isobutanol increases monotonically as *marC*, *miaA*-*hfq*, *rph*, *mdh*, and *groL* were acquired in the G3 lineage, yet *rph*, *mdh*, and *groL* single mutants did not have significant fitness effects at 1% (w/v) isobutanol (Figure 4.11). Our mutation reconstruction analysis demonstrates that significant positive epistasis is correlated with an *miaA*-*hfq* or *marC* genetic background (Figure 4.11). Curiously, we detected significant negative epistasis between *miaA*-*hfq* and *marC*. Negative epistasis often occurs between genes with overlapping functions [154], suggesting that *miaA*-*hfq* and *marC* could have a shared mechanism for improving isobutanol tolerance; this interpretation is further supported by the fact that *miaA*-*hfq* and *marC* each show positive epistasis with subsequent mutations in the G3 lineage (i.e. *rph*/*mdh*/*groL*). Since *marC* is a poorly characterized gene of unknown function, possible mechanistic links between *miaA*-*hfq* and *marC* or between *marC* and *rph*/*mdh*/*groL* are not apparent. One possibility is that *marC* mutations (such as deletion or transposon insertions) could affect expression of the divergently transcribed *marRAB* locus, which is involved in regulation of genes associated with response to oxidative stress, organic solvents, and heavy metals; some of the genes in the *marRAB* regulon are coregulated by *hfq*, including *acrAB* [155]. Indeed, our gene expression study of G3.2 reveals slightly reduced levels of *marA* and *marB* transcripts, and NCA identified transcription factor MarA as having significantly perturbed activity (Figure 4.8D and Additional file 4). An independent study of isobutanol tolerance in *E. coli* reported that fitness benefits of Δ*marRAB* and Δ*marC* were comparable, but deletion of the full *marCRAB* locus yielded the greatest improvement in isobutanol tolerance (James C. Liao, UCLA personal communications). More investigation and characterization of *marC* is needed to elucidate mechanisms underlying the observed negative epistasis between *miaA*-*hfq* and *marC* and positive epistasis between *marC* and *rph*/*mdh*/*groL*.

The functional basis of epistasis between *rph*/*mdh*/*groL* and *miaA*-*hfq* is also not immediately obvious, as *hfq*, *rph*, *mdh*, and *groL* participate in seemingly disparate cellular processes: *hfq* is a global regulator that mediates binding between sRNAs and their target mRNAs, *rph* has RNase PH activity, *mdh* is TCA cycle enzyme malate dehydrogenase, and *groL* is part of the groEL chaperone [150]. *miaA*-*hfq* 4407505 -7:AGGAAAA is the second mutation acquired in

the G3 lineage, and is associated with significantly improved fitness at 1% isobutanol both in the evolutionary trajectory (Figure 4.7A) and as a reconstructed single mutant (Figure 4.11). The mutation *miaA-hfq* 4407505 -7:AGGAAAA is a partial ribosome binding site deletion that is likely to reduce translation initiation rate of *hfq* mRNA (evaluated with the Ribosome Binding Site Calculator, Beta version) and thus reduce Hfq protein levels (evidenced by iodine staining assay, Figure 4.10A) [156]. Since Hfq is a global regulator, a reduction in activity is likely to perturb expression of many proteins. The net effect of these perturbations is clearly beneficial at 1% (w/v) isobutanol, but some of the specific expression perturbations caused by reduced Hfq are likely to be maladaptive. This suggests that some of the subsequent mutations in G3 may be compensatory to the perturbations caused by reduced Hfq activity, and that the initial fixation of *hfq* in G3 may have been a crucial determinant of the evolutionary trajectory in this lineage. Interestingly, many of the mutations subsequently acquired in the G3 lineage are in genes known to be regulated (either directly or indirectly) by Hfq, including *acrB*, *phoPQ*, *gltD*, *mdh*, *groL*, and the *gat* operon [138, 150]. For example, reduced Hfq activity is associated with increased *mdh*, *gltD*, and *phoPQ* expression; mutations in these genes could thus be compensatory by reducing expression or protein activity [138, 150]. Indeed, the *mdh* mutation in G3 was verified to be a loss-of-function mutation, and our gene expression study indicates reduced PhoPQ activity in G3.2, probably due to the *phoQ* mutation. Reduced Hfq activity is also associated with reduced GroL levels and the acquired *groL* mutation could likewise be compensatory, perhaps through increasing GroL activity [138].

The role of the *miaA-hfq* mutation in the evolution of isobutanol tolerance in G3 suggests that global regulatory network perturbation is an important genetic mechanism of adaptation. Indeed, an accumulating body of research points towards regulatory network perturbation as a general and important mechanism of adaptive evolution under a variety of contexts and selective pressures, as discussed in section 4.2.1 and [57, 157, 122]. Investigations of transcription factor network evolution suggest higher evolvability and rates of divergence for central transcription factors compared to peripheral regulators [158]. Furthermore, the concept of regulatory network perturbation as a mode of adaptation has been utilized for engineering complex stress tolerance phenotypes; targeted mutagenesis of *rpoD* and *rpoA* in *E. coli* was used to generate mutant libraries with perturbed global gene expression patterns, from which variants with dramatically improved ethanol tolerance (*rpoD* mutagenesis) and n-butanol tolerance (*rpoA* mutagenesis) were iteratively isolated [53, 52, 54].

Many of the adaptive global effect perturbations identified in previous studies involved transcriptional regulatory changes, often through mutations in transcription factors or genes

controlling DNA supercoiling. Indeed, our gene expression study of G3.2 revealed a number of transcriptomic adaptations, probably due in part to changes in RpoS and PhoP activity. However, our results also suggest that global changes in post-transcriptional regulation might constitute important modes of adaptation in our evolved lineages. Gene Ontology analysis of mutations accumulated in G3.2 and X3.5 indicates a significant overrepresentation of genes with RNA helicase activity, including *secA* (G3.2), *rhlB* (G3.2; see Additional file 1), *hrpA* (X3.5), and *deaD* (X3.5). RNA helicases can participate in various modes of post-transcriptional regulation, including mRNA processing, translation, or degradation [159]. Regulatory effects related to RNA helicase activity of *secA* (G3.2), *rhlB* (G3.2), *hrpA* (X3.5), and *deaD* (X3.5) are not well characterized, and thus it is difficult to speculate about specific mechanisms of adaptation related to these mutations; however acquisition of *hrpA* and a *mdh*/*deaD*/*plsX* mutation cluster in the X3 lineage is correlated with significant improvements in isobutanol tolerance (Figure 4.7B). These results suggest that RNA helicases may be interesting targets for targeted mutagenesis to improve isobutanol tolerance and possibly other complex stress tolerance phenotypes.

In addition to mutations in RNA helicase genes, we discovered other possible post-transcriptional regulatory adaptations in our evolved lineages. X3.5 acquired a mutation in *rpsB*, ribosomal subunit S2, which could potentially affect translation; however other effects are possible, as noted in our epistasis discussion. The X1 and G3 lineages acquired mutations in post-transcriptional regulator *hfq*, which probably result in reduced activity, as evidence by iodine staining assay (Figure 4.10A). Hfq is a global regulator that functions by mediating binding between a variety of sRNAs and their target mRNAs, which can alter target protein levels via effects on translation initiation or mRNA degradation [138]. Interestingly, many stress response regulons incorporate sRNA mediated regulation, including the RpoS regulated general stress response (*rpoS* translation is mediated through *rprA* and *dsrA* sRNAs), oxidative and antibiotic stress (*oxyS*, *gcvA*, and *micF* sRNAs), osmotic shock (*omrA* and *omrB* sRNAs), cell envelope stress (*micA* and *rybB* sRNAs), and iron limitation (*ryhB* sRNA) [160]. Stress response tuning via *hfq* mutations, possibly dominated by modulation of RpoS, may thus constitute a mechanism of adaptation to isobutanol stress, but due to the global regulatory role of Hfq, many other adaptive effects are possible. Beneficial *hfq* mutations have recently been discovered in *E. coli* lineages evolved on lactate minimal medium, under glucose limitation, and under phosphate limitation, underscoring that *hfq* evolution represents a flexible and general mechanism of adaptation, and *hfq* may be an interesting mutagenesis target for engineering improved stress tolerance phenotypes [115, 161, 162].

The role of centrality in the evolution of biochemical networks has been investigated in

a number of studies. Investigations of protein-protein interaction and metabolic networks suggest that central proteins tend to evolve more slowly than peripheral proteins, in contrast to findings for transcriptional regulatory networks [158, 163, 164]. Intriguingly, we found numerous potentially adaptive mutations in genes known to have high centrality in protein-protein interaction and metabolic networks. Examples of adaptation at central protein-protein interaction nodes include adaptive mutations in *groL*, a chaperone involved in folding of many different proteins, and mutations in *secA*/*lepB*, components of the Sec complex which exports numerous proteins out of the cytosol. We also discovered potentially adaptive mutations in high centrality metabolic network nodes, such as *mdh* and *gltD*/*glnE*. Our results suggest that more investigation is warranted into the role of centrality in the evolution of biochemical networks.

Parallel evolution occurs when independent lineages evolve similar traits, and is considered strong evidence of selective pressure [57]. We have identified several instances of parallel genotypic adaptation in our evolved lineages. In particular, mutations in *rph*, *gatYZABCD* operon, *mdh*, *acrAB*, and *marC* were found in all of the resequenced evolution endpoint clones (Table 4.1). *marC* mutations (consisting mostly of transposon insertions) were discovered in all six evolution endpoint populations, while deletion of a genomic region containing *marC* (*hipA*-*flxA*) was reported in an independent study of evolution of isobutanol tolerance (Table 4.1) [124]. Loss-of-function of *marC* appears to be broadly adaptive to isobutanol stress under various growth conditions, including glucose and xylose minimal media (Table 4.2) and yeast extract supplemented glucose media [124]. In addition to *marC*, *acrAB* mutations were also prevalent in evolution endpoint populations (occurring in five out of six populations; Table 4.1), and adaptive *acrAB* mutations have also been reported in other investigations of isobutanol tolerance [165, 124]. In both our work and independent studies, it was found that loss-of-function of *acrAB* was correlated with significantly improved isobutanol tolerance. Like *marC*, effects of *acrAB* mutations appear to be broadly adaptive since isobutanol tolerance is improved for a variety of growth conditions, including xylose and glucose minimal media (this study), yeast extract supplemented glucose media [124], and rich LB media [165].

In contrast to *marC* and *acrAB* mutations, in isolation *mdh* mutations did not improve isobutanol tolerance. Yet the appearance of *mdh* indel mutations in three out of six independent lineages strongly suggests adaptive effects, perhaps through epistatic interactions. In the case of the G3 lineage, we find significant positive epistasis between *hfq* and *mdh* (Figure 4.11). However, *mdh* mutations also appear in strains without *hfq* mutations or evidence of altered Hfq activity, suggesting that *mdh* mutations may be epistatic with other genetic backgrounds. Phenotype analysis of reconstructed mutants hints that there may be functional overlap between *hfq* and

marC, and we observe possible epistasis in a constructed *marC*/*mdh* double mutant, although the measured epistasis is not statistically significant (Figure 4.11); given these results and the prevalence of *marC* mutations in the evolution endpoint populations, epistasis between *marC* and *mdh* seems plausible but more investigation is needed.

We did not investigate *rph* and *gatYZABCD* parallel evolution as thoroughly as the mutations discussed above. As described previously, restorative mutations in *rph* are general adaptations of *E. coli* K12 MG1655 to growth in minimal medium rather than specific adaptations to isobutanol stress. However, it should be noted that *rph* mutations have been reported to be epistatic with a variety of genetic backgrounds, and our mutation reconstruction analysis points towards such a possibility (Figure 4.11). The *gatYZABCD* operon is strongly upregulated in response to isobutanol stress (Figure 4.8 and Additional file 4). *gatYZABCD* genes are involved in galactitol transport and catabolism, and do not have an obvious role in isobutanol tolerance. An independent study found that deletion of *gatY* was correlated with improved isobutanol tolerance [124]. We suspect that *gatYZABCD* overexpression in response to isobutanol is spurious and provides no stress tolerance benefit, perhaps leading to selective pressure for mutations that reduce expression levels or lead to loss-of-function. We did not investigate *gatYZABCD* in our other evolved lineages due to the relatively large size of this locus.

During the course of our isobutanol tolerance evolution study, we became aware of a similar project concurrently underway in another laboratory [124]. It is informative to compare the findings reported in [124] with our own results. In this parallel study, *E. coli* JCL260 (an isobutanol production strain) was evolved on yeast extract - glucose media supplemented with isobutanol and then sequenced [124]. A total of 27 mutations were identified in SA481 (the sequenced evolved isolate), consisting of 25 transposon (IS10) insertions, one SNP, and one large genomic deletion [124]. Mutation repair analysis and subsequent gene deletion studies revealed five key genetic loci involved in isobutanol tolerance: *tnaA*, *gatY*, *acrA*, *yhbJ*, and the *hipA-flxA* genomic region [124]. It was demonstrated that deletion of these genetic loci and of *marCRAB* (contained within the *hipA-flxA* genomic region) conferred isobutanol tolerance [124]. In our study, we discovered parallel mutations in *acrAB*, *marC*, and the *gatYZABCD* operon, and demonstrated that deletion of *acrA*, *acrB*, and *marC* conferred isobutanol tolerance, consistent with the independently reported results [124]. Furthermore, in [124] single mutations reportedly had minor impacts on isobutanol tolerance (with the exception of Δ*acrA*), but mutation combinations (Δ*acrA*/Δ*gatY*, Δ*acrA*/Δ*tnaA*, Δ*tnaA*/Δ*gatY*/Δ*acrA*, Δ*tnaA*/Δ*gatY*/Δ*acrA*/Δ*marCRAB*, and Δ*tnaA*/Δ*gatY*/Δ*acrA*/Δ*marCRAB*/Δ*yhbJ*) showed synergistic effects on isobutanol tolerance; these findings are consistent with our suggestion that epistasis is an important factor in the

evolution of isobutanol tolerance and possibly other complex stress tolerance phenotypes.

Beyond the five key mutations discussed in [124], there are other notable parallels and relationships between the mutations in SA481 and those identified in our study. We identified parallel loss-of-function mutations in malate dehydrogenase *mdh* (catalyzing NAD dependent oxidation of malate to oxaloacetate) in three out of six evolution endpoint populations (Table 4.2), and our results indicate that *mdh* mutations provide fitness benefits via epistasis with *hfq* and possibly *marC* mutations (Figure 4.11). SA481 contains an IS10 insertion in malate dehydrogenase *maeA* (catalyzing NAD dependent decarboxylation of malate to pyruvate) [124]; the preponderance of malate dehydrogenase mutations in isobutanol tolerant mutants suggests that rewiring of metabolic pathways around the malate node may be important in the evolution of isobutanol tolerance. Mutations affecting the *mdtJI-tqsA* locus were identified both in our study (*mdtJ::IS5::tqsA* in X3.5) and in SA481 (*tqsA::IS10*) [124]. TqsA is a transporter for quorum sensing signal autoinducer 2 (AI-2). Interestingly, the *tnaA* mutation SA481 may also be related to quorum sensing, since TnaA (L-cysteine desulfhydrase/tryptophanase) catalyzes catabolism of tryptophan to indole, a putative quorum sensing molecule [150]. Additionally, the adaptive effects of *acrAB* mutations are suspected to be related to the quorum sensing role of the AcrAB-TolC system. In aggregate, these observations suggest that adaptations involving quorum sensing systems may be important in isobutanol tolerance, and merit future investigation. As a final parallel, we demonstrated that RpoS is downregulated in G3.2 and identified mutations in RpoS regulators (including *hfq* and *acrAB*). Interestingly, in SA481 the RpoS regulator *rssB* (which regulates proteolytic degradation of RpoS) is mutated, as well as *acrA*, providing additional evidence that RpoS modulation may be adaptive to isobutanol stress.

While there are many overlaps between our results and those reported in [124], there are also significant differences between the genetic loci identified in these two studies. *tnaA* (L-cysteine desulfhydrase/tryptophanase) and *yhbJ* (which senses glucosamine-6-phosphate and regulates glucosamine-6-phosphate synthase *glmS*) were reported to contribute to isobutanol tolerance in [124], but were not identified in our study. Likewise, we discovered epistatic mutations in *miaA-hfq*, *mdh*, *rph*, and *groL* that confer high isobutanol tolerance in glucose minimal media, yet none of these genes were identified in [124]. Since evolved adaptations often show tradeoffs in relative fitness across different environments, differences in conditions between these two studies could account for different evolutionary trajectories. In particular, it should be noted that in our study we used a different parent strain (*E. coli* EcNR1 vs. JCL260 in [124]) and different media (glucose or xylose minimal media vs. yeast extract supplemented media in [124]); we also note that adaptations in our evolved strains exhibit antagonistic pleiotropy in rich media (Figure 4.4A

and B).

In addition to the above work by Atsumi *et al*., two independent n-butanol tolerance evolution studies were reported after the completion of our work, providing additional experimental data with which to compare and interpret our results. Reyes *et al*. evolved *E. coli* K12 BW25113 in chemostats with isobutanol spiked M9 glucose media and performed genome resequencing on multiple tolerant isolates, leading to the identification of mutations in *relA*, *rho*, *nus*, *feoA*, and SNPs in Fur binding sites in the promoter regions of the *fes-ybdZ-entF-fepE* and *fepA-entD* operons [125]. In the other study, Dragosits *et al*. evolved *E. coli* K12 MG1655 populations in five different environments (acidic, osmotic, oxidative, and n-butanol stress, and a control medium-only environment) with M9 glucose media [126]. Genome resequencing of isolates from the n-butanol adapted population revealed mutations in *acrA*, *fepA*, *marC*, and *rpoB* [126]. The identification of *acrA* and *marC* mutations across three different studies (our work, Atsumi *et al*. [124], and Dragosits *et al*. [126]) substantiates the important adaptive effects conferred by these mutations, and interestingly suggests that the underlying biochemical mechanisms provide cross-tolerance between n-butanol and isobutanol toxicity.

The most noteworthy outcome of these studies is the occurrence of numerous transcriptional changes and mutations involving iron-ion transport and metabolism (i.e. *feoA* and Fur binding site mutations in Reyes *et al*. [125]; *fepA* in Dragosits *et al*. [126]). Previous studies of ethanol tolerance in *E. coli* have found that iron-ion transport and metabolism tend to be upregulated in ethanol adapted strains compared to WT [121, 166]. *fepA* was upregulated and shown to be a key adaptation in both n-butanol and osmotic stress tolerant strains in Dragosits *et al*. [126]. In Reyes *et al*., differential transcriptional regulation of iron related genes was observed in every characterized isolate, with NCA suggesting significantly changed Fur activity in six out of eight isolates [125]. However, the direction of transcriptional changes varied between isolates, with some showing increased expression of iron ion transport and metabolism genes, and others showing decreased expression [125]; this variation suggests that different modes of adaptation (potentially involving epistasis) with iron acquisition genes are possible. Despite this, overexpression of iron-related genes *feoA* and *entC* was demonstrated to improve n-butanol tolerance in the parental strain [125], suggesting a causal link between these transcriptional changes and alcohol tolerance. In our studies, genes associated with iron acquisition (*entA*, *entC*, *fepA*, and *cirA*; regulated by Fur and CRP) were found to be repressed by isobutanol in G3.2 vs. upregulated by isobutanol in WT; however despite these regulatory differences, expression at 0.5% isobutanol tended to be similar between WT and G3.2 (Additional file 4). Collectively, these results and prior literature [121, 166] suggest that mutations in iron related genes may play a complex role in multiple types

of stress tolerance. The biochemical tolerance mechanisms conferred by increased expression of iron related genes have not been fully elucidated. Reyes *et al*. suggest that upregulation of iron acquisition genes increases intracellular Fe^{3+} leading to activation of the BasS/BasR system, which in turn induces LPS modification genes [125]. Consistent with this, expression of the BasS/BasR regulated *arn* operon has been shown to improve both n-butanol [125] and ethanol [167] tolerance of *E. coli*.

4.4.2 Remodeling the cell envelope: possible mechanism of adaptation to isobutanol stress

An accumulated body of evidence indicates that the cell membrane is a primary target of alcohol toxicity. Alcohols have been demonstrated to intercalate the membrane lipid bilayer, leading to detrimental changes in the physicochemical properties of membrane [50]. Examining the genotypic and phenotypic adaptations of our evolved lineages in the context of this known mechanism of toxicity reveals a trend of evolution targeting various features of the cell envelope through a diversity of processes. We observe adaptations that may lead to alterations in cell envelope protein composition, downregulation of fimbriae biogenesis and upregulation of flagellar biogenesis, and alterations in peptidoglycan, membrane lipid composition, and lipopolysaccharide (LPS) composition (Figure 4.12A). Collectively, these adaptations suggest that evolution may be remodeling the cell envelope to counteract the detrimental effects of isobutanol on the cell membrane.

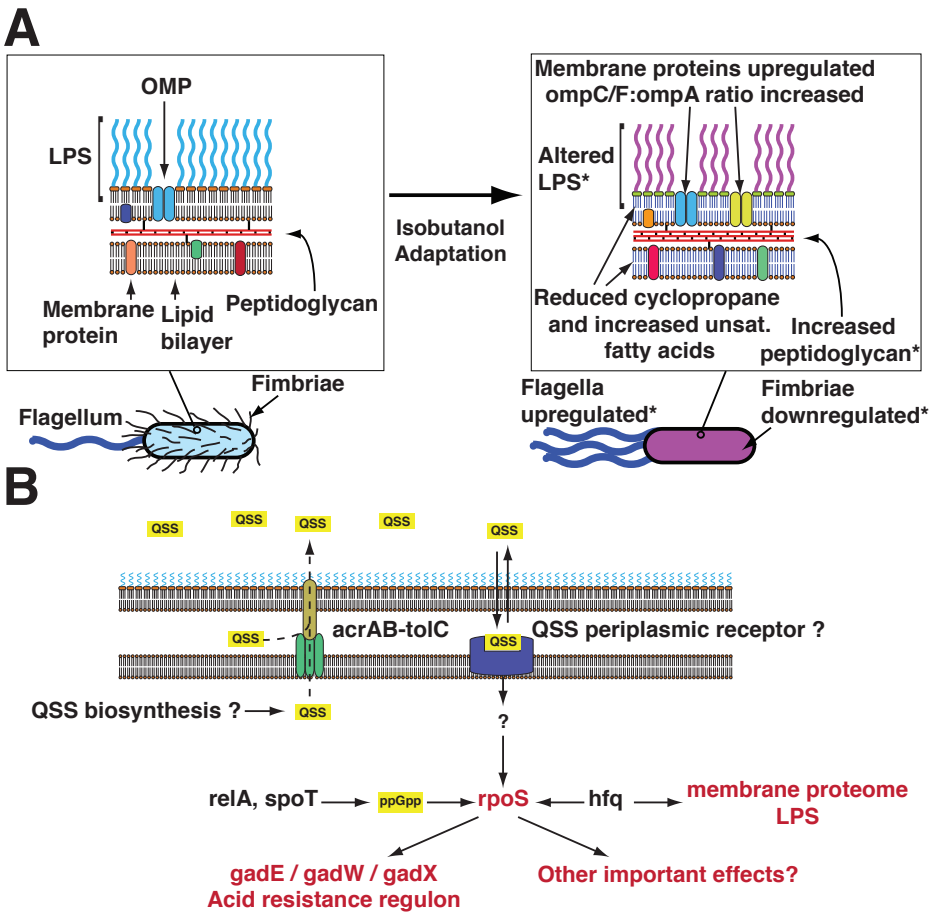

Figure 4.12: Possible mechanisms of evolved isobutanol tolerance. Examining genotypic adaptations to isobutanol stress in the context of known modes of solvent toxicity suggests that remodeling the cell envelope and stress response attenuation might be important mechanisms of adaptation in our evolved lineages. **(A)** Postulated mechanisms of adaptation involving the cell envelope. Asterisked (*) items were inferred from genoptyic data without direct experimental validation. We observe adaptations that may lead to alterations in cell envelope protein composition, downregulation of fimbriae biogenesis and upregulation of flagellar biogenesis, and alterations in peptidoglycan, membrane lipid composition, and lipopolysaccharide (LPS) composition. Collectively, these adaptations suggest that evolution may be remodeling the cell envelope to counteract the detrimental effects of isobutanol on the cell membrane. **(B)** Our results suggest that attenuation of RpoS activity may be a convergent adaptive effect associated with *hfq*, *acrAB-tolC*, *relA*, and *spoT* mutations. Hfq is required for translation of *rpoS* mRNA, RelA and SpoT synthesize signalling molecule ppGpp which upregulates RpoS, and one study suggests that AcrAB-TolC exports an unidentified quorum sensing signal (QSS) that upregulates RpoS, possibly via a periplasmic receptor [138, 143, 168]. All depicted regulatory interactions involve upregulation. Signalling molecules (ppGpp and QSS) are boxed in yellow. Effects/gene targets that might be ultimately linked to isobutanol tolerance are red. AcrAB-TolC quorum sensing model adapted from [143], using similar notation.

Examination of the genotypes of G3.2, G3.6, and X3.5 yields evidence of significant selective pressure on membrane proteins; Gene Ontology analysis indicates a significant overrepresentation of membrane proteins among mutated genes (corrected p-value = $7.23\text{x}10^{-3}$) [132]. We have also identified various mutations that could cause global changes in cell envelope proteome composition. Such changes are potentially adaptive to isobutanol stress, perhaps countering

effects of isobutanol on membrane integrity and mechanical properties. Two lineages, X1 and G3, acquired *hfq* mutations that appear to reduce activity, and in G3, the *miaA-hfq* mutation was shown to improve tolerance at 1% (w/v) isobutanol (Table 4.2 and Figure 4.11). Many important membrane proteins, including *ompA*, *ompC*, *ompW*, *ompF*, *ompT*, *lamB*, *cirA*, *fecA*, and *fepA* feature Hfq mediated sRNA regulation; indeed, almost half of the Hfq binding sRNAs with known targets regulate the expression of outer membrane proteins [160, 169]. In most cases sRNAs are involved in downregulation, with sRNAs binding to target mRNA (mediated by Hfq) leading to translation inhibition and/or increase mRNA degradation [160]. Thus mutants with reduced Hfq activity would be expected to have altered membrane protein composition and a general increase in outer membrane proteins, which has been verified in other studies and found to cause pleiotropic phenotype effects [138, 160, 161, 170]. Consistent with this expectation, 1-D SDS-PAGE characterization of cell envelope protein composition of G3.2 revealed an overall increase in membrane proteins in G3.2 compared to WT (Figure 4.9B). Additionally, several protein bands (corresponding to 72 kDa, 55 kDa, and the OmpF/OmpC bands) appear to be upregulated in G3.2 (Figure 7B). *ompF* is negatively regulated by *micF* sRNA, which both inhibits translation of *ompF* and reduces *ompF* mRNA levels. *ompF* mRNA was found to be upregulated in the DNA microarray study (Additional file 4) and SDS-PAGE analysis (Figure 4.9B) supports increased levels of OmpF protein, consistent with the notion that reduced Hfq activity would increase both *ompF* mRNA and OmpF protein levels by abolishing *micF* regulation. OmpC may also be upregulated in G3.2 (Figure 4.9B; due to the similar molecular weights of OmpF and OmpC, we were not able to resolve these proteins separately with SDS-PAGE). Interestingly, a recent study reported that a mutant deficient in *rybB* sRNA, which downregulates OmpC and OmpW, had improved SDS tolerance providing additional precedent for stress tolerance mechanisms involving modulation of sRNA mediated outer membrane protein regulation [171].

In G3.2 and G3.6, we identified mutations in *secA*/*lepB*, two components of the Sec protein translocation apparatus. The Sec apparatus is responsible for translocating many periplasmic and membrane targeted proteins from the cytosol; examples of known Sec substrates include MalE, LamB, OmpA, OmpF, OppA, PhoE, MBP, DegP, FhuA, FkpA, OmpT, OmpX, TolB, TolC, YbgF, YcgK, YgiW and YncE [172]. As discussed previously, the *secA*/*lepB* mutations may alter the peptide binding specificity of the Sec apparatus, and thus membrane and periplasmic protein composition could be altered by changes in Sec mediated translocation, perhaps by increasing export of some proteins while reducing export of others. However, we cannot predict how the S233P mutation in the preprotein binding domain of SecA will affect peptide specificity, and thus more investigation into this hypothesis is needed. Interestingly the signal recognition particle (SRP), which is involved in cotranslational targeting of peptides to the Sec translocase, has been

implicated in mechanisms of acid tolerance in *Clostridium acetobutylicum* and *Steptococcus mutans* [50, 152, 173].

In addition to adaptations involving general modulation of membrane and periplasmic protein composition, we observe adaptations that appear to specifically target fimbriae and flagellar biogenesis (Figure 4.12A). G3.2 shows strong transcriptional upregulation of flagellar biosynthesis genes, and NCA identified increased activity of related transcription factor FlhDC (Figure 4.8D). Differential transcriptional regulation of flagellar biosynthesis genes is commonly observed in *E. coli* strains evolved for chemical stress tolerance, however regulation patterns are not consistent across different types of stress. For instance, gene expression studies of an acid adapted strain showed upregulation of flagellar biosynthesis genes [126], while in contrast downregulation of flagellar biosynthesis was observed in an ethanol adapted strain [174]. Furthermore, regulation of flagellar genes can vary even among strains evolved under the same conditions; for example, in a transcriptional analysis of eight different n-butanol tolerant strains, flagellar biosynthesis genes were upregulated in three strains and downregulated in five [125]. Whether upregulation of flagellar biosynthesis in G3.2 is indeed adaptive or simply a spurious consequence of reduced RpoS activity in this strain needs to be investigated. In contrast to flagellar genes, G3.2 shows strong transcriptional downregulation of the *fim* operon (*fimAICDFGH*), responsible for fimbriae biogenesis (Figure 4.8C). A study of ethanol tolerance with gene overexpression and transposon mutagenesis libraries demonstrated that loss-of-function of *fim* genes was correlated with improved ethanol tolerance, while *fim* overexpression was correlated with negative fitness effects [121]. These results strongly suggest that *fimAICDFGH* downregulation observed in G3.2 is indeed adaptive. Interestingly, the X3 lineage acquires a mutation in *hrpA*, a gene known to be involved in processing the *daa* fimbriae operon in pathogenic *E. coli* strains [175]. The *marC*/*gatC*/*hrpA* mutations occur early in the X3 lineage and are correlated with a significant fitness increase (Figure 4.7B). However, *E. coli* EcNR1 does not have the *daa* operon, so the role of *hrpA* in fimbrial biogenesis in this strain is not clear, and additional studies indicate that *hrpA* may have other important in vivo functions [176].

Modification of cell envelope peptidoglycan, lipoprotein, and LPS content is a commonly observed adaptation of bacteria to solvent stress [50]. Recent studies have revealed that LPS modification is a ubiquitous mechanism of alcohol tolerance in *E. coli* and possibly other gram-negative bacteria. Lipid characterization revealed altered Lipid A composition (presence of 3-hydroxytridecanoic acid) in an n-butanol tolerant *E. coli* mutant [177]. Increased expression of the BasS/BasR regulated *arn* operon has been shown to improve n-butanol [125] and ethanol [167] tolerance of *E. coli*. The *arn* operon consists of enzymes which add positively charged

4-amino-4-deoxy-L-arabinose to the lipid A moiety of LPS [125], perhaps reducing the extent to which alcohols partition into the membrane via steric hindrance or decreased surface hydrophobicity. In addition to the *arn* operon, expression of *lpcA* (D-sedoheptulose 7-phosphate isomerase; first committed step in the biosynthesis of a core LPS component) has been shown to improve ethanol tolerance in *E. coli* [167], suggesting that a general increase in LPS may be beneficial for tolerance. In addition to LPS, solvent tolerant *E. coli* strains have also been reported to have increased lipoprotein content, possibly strengthening the cell membrane against the fluidizing effects of solvent [178]. Finally, modification of peptidoglycan content and structure has also been correlated with alcohol tolerance in *E. coli*. Enhanced peptidoglycan biosynthesis has been implicated in tolerance to ethanol and isobutanol, possibly by changing the rigidity or structural strength of the cell [121, 124]. A study of ethanol tolerance with gene overexpression and transposon mutagenesis libraries demonstrated that loss-of-function of peptidoglycan biosynthesis *mur* genes was correlated with reduced fitness, while overexpression of *mur* genes was correlated with improved tolerance [124]. An independent study of isobutanol tolerance using experimental evolution and genome resequencing demonstrated that upregulation of *glmS*, which is responsible for the synthesis of peptidogylcan and LPS precursor glucosamine-6-phosphate (GlcN-6-P), resulted in improved isobutanol tolerance [124].

Our evolved lineages show several possible mechanisms of adaptation involving peptidoglycan, lipoprotein, and LPS biosynthesis. G3.2 was found to contain mutations in *fepE* and *yjgQ*, which are involved in LPS biosynthesis (Additional file 1). The G3 lineage also acquired probable loss-of-function mutations in *glnE* and *gltD*, correlated with improved fitness, that might alter nitrogen metabolism towards increased biosnythesis of glutamine, a precursor for GlcN-6-P (Figure 4.7A) [150]. Interestingly, our gene expression study of G3.2 also shows significant downregulation of glutamine and glutamate catabolic genes (*ybaS* and *gadAB*, respectively), further suggesting increased glutamine production in G3.2 (Additional file 4) [150]. These changes suggest that G3.2 may be increasing glutamine flux for peptidoglycan and LPS biosynthesis, but glutamine is a central metabolite so numerous other adaptive effects would also be possible [150]. Another adaptation potentially related to regulation of LPS biosynthesis and modification was observed in phosphoethanolamine transferase *eptB*. This gene is negatively regulated by sRNA *mgrR*, and thus *eptB* might be upregulated in X1 and G3, which harbour *hfq* mutations associated with reduced activity; additionally, *eptB* transcriptional upregulation was observed in the G3.2 gene expression study (Additional file 4) [141]. *eptB* modifies the LPS by adding a phosphoethanolamine (pEtN) moiety to the terminal 3-deoxy-d-manno-octulosonic acid (KDO) of LPS [141]. Possible adaptive effects of this modification to isobutanol stress are not clear, but interestingly *eptB* upregulation is associated with increased resistance to polymyxin B, a

detergent-like antibiotic that targets the cell membrane [141].

Our gene expression study of G3.2 reveals differential transcriptional regulation for multiple genes involved in peptidoglycan, lipoprotein, and LPS biosynthesis. Many genes associated with LPS biosynthesis are differentially expressed in G3.2; Gene Ontology analysis revealed significant overrepresentation of cellular polysaccharide biosynthetic process genes (corrected p-value = $3\text{x}10^{-2}$). NCA identified significantly reduced activity of PhoP and GadE transcription factors, and incidentally many of the differentially regulated peptidoglycan, lipoprotein, and LPS genes are part of these regulons (Figure 4.8 and Additional file 4). We observed downregulation of *slp* (starvation lipoprotein, acid resistance regulon), *pagP* (palmitoyl transferase for lipid A, PhoP regulon), *ybjG* (undecaprenyl pyrophosphate phosphatase, PhoP regulon), and *slyB* (outer membrane lipoprotein, PhoP regulon) (see Additional file 4); downregulation of these genes may be adaptive to isobutanol tolerance, but the collective effect of these perturbations is unclear. Other potentially important lipoprotein and LPS genes differentially regulated in G3.2 included *nlpD* (putative outer membrane lipoprotein, downregulated in G3.2) and numerous members of the *rfa* gene cluster, which comprise the pathway for LPS core-oligosacchride assembly (*rfaL*, *rfaQ*, *rfaG*, *rfaS*, *rfaB*, *rfaI*, *rfaJ*, *rfaY*, and *rfaZ*, all upregulated) (Additional file 4). Interestingly, despite the established role of the *arn* operon in n-butanol [125] and ethanol [167] tolerance, transcriptional downregulation of *arn* genes was observed in the G3.2 gene expression study (Additional file 4). This discrepancy may arise due differences in membrane interactions between straight-chain alcohols (ethanol and n-butanol) and branched alcohols (isobutanol), or alternately G3.2 may have evolved alternative LPS adaptations that confer tolerance through different biochemical effects.

Bacteria are known to adapt to solvent stress by altering membrane lipid composition, including cis-to-trans isomerisation of fatty acids, changing the proportions of saturated/unsaturated fatty acids, altering composition of phospolipid head groups, and altering fatty acid acyl chain length [50]. In our evolved lineages, we observe several possible adaptations to isobutanol stress involving membrane lipid composition. The X3 lineage acquires a mutation in *plsX* that may be adaptive to isobutanol stress (Figure 4.7B). The function of *plsX* has not been fully elucidated, but it is suspected to play a role in fatty acid metabolism, possibly regulating the intracellular concentration of acyl-[acyl carrier protein] (acyl-ACP)[179]. Many genes associated with lipid metabolism are differentially expressed in G3.2. Gene Ontology analysis revealed significant overrepresentation of lipid catabolic process genes (corrected p-value = $4.6\text{x}10^{-2}$); in particular many lipid catabolism genes were downregulated in G3.2, including *tesB* (acyl-CoA thioesterase II), *hdhA* (7-alpha-hydroxysteroid dehydrogenase), *gabT* (4-aminobutyrate aminotransferase), *pgpB* (phosphatidylglycerophosphatase B), *fadJ* (fused enoyl-CoA hydratase and epimerase and

isomerase), *fadE* (acyl coenzyme A dehydrogenase), and *fadB* (3-hydroxyacyl-CoA dehydrogenase). Other notable expression changes in lipid metabolism genes include downregulation of *cfa* (cyclopropane fatty acyl phospholipid synthase), *ybhO* (cardiolipin synthase 2), and *aidB* (predicted acyl-CoA dehydrogenase), and upregulation of *fabA* (3-hydroxydecanoyl-ACP dehydrase) (Additional file 4). We profiled the fatty acid composition of G3.2, revealing a significant decrease in cyclopropane fatty acids and increased unsaturated fatty acids (mainly to due increased proportion of C18:1 and C16:1) relative to WT (Figure 4.9A). The decrease in cyclopropane fatty acids is attributable to downregulation of *cfa*, while the increase in unsaturated fatty acids is probably a collective effect of multiple gene expression changes (with *fabA* upregulation possibly playing an important role) [180]. Previous studies have demonstrated that *E. coli* responds to short chain (C2-C4) n-alkanol exposure by increasing the unsaturated:saturated fatty acids ratio, suggesting that the increase in unsaturated fatty acids in G3.2 is indeed adaptive to isobutanol stress [50]. In one study, a mutation in *fabF* (β-ketoacyl-ACP synthases II) was shown to specifically increase C18:1n-7 content and confer improved n-butanol tolerance in *E. coli* [181], however in a different study decreased proportions of unsaturated fatty acids were observed in n-butanol adapted strains [125]. Interestingly, recent *in vitro* biophysical studies of model yeast membranes (ternary lipid bilayers consisting of saturated lipids, unsaturated lipids, and ergosterol) in the presence of ethanol have demonstrated that increasing the proportion of unsaturated lipids protects the membrane by preventing the formation of interdigitated lipid phases [182]. The decrease in cyclopropane fatty acids observed in G3.2 is somewhat counterintuitive, since several studies have implicated cyclopropane fatty acids in n-butanol tolerance [50, 183, 125]. As discussed above for LPS modifications, differences in lipid composition between n-butanol and isobutanol tolerant strains may reflect differences in membrane effects between straight-chain vs. branched alcohols, and further investigation is needed.

Collectively, we have observed many genotypic and gene expression changes that suggest evolution may be remodeling the cell envelope to counteract the detrimental effects of isobutanol on the cell membrane. However, further investigation is needed to profile cell envelope changes in isobutanol tolerant strains and to ascertain that the observed changes are indeed adaptive to isobutanol stress. Analysis of cell envelope protein and fatty acid composition (Figure 4.9) will need to be extended to other isobutanol tolerant lineages to investigate parallel adaptations; additionally, the cell envelope proteome could be resolved in greater detail by performing two-dimensional electrophoresis. Peptidoglycan and LPS composition will need to be profiled across different isobutanol tolerant lineages to validate inferences drawn from the gene expression and genotype data and to investigate parallel adaptations. In addition to profiling cell envelope composition in evolved isolates, mutations in genes related to cell envelope composition (e.g. *hfq*,

secA/*lepB*, *glnE*/*gltD*, *phoQ*, *plsX*, etc.) could be reconstructed singly and in combinations with other mutations from their respective lineages, and the resulting mutant library could be profiled. These investigations will help to elucidate causal links between genotype, isobutanol tolerance, and cell envelope composition.

4.4.3 Stress response attenuation: surprising adaptations

We observe evidence of RpoS downregulation in many of our evolved isobutanol tolerant lineages, implying that reduction of RpoS activity may be adaptive to isobutanol stress (Figure 4.10A and B). This is a surprising result, given that RpoS is a master regulator of the general stress response and many prior studies show that reduction of RpoS activity increases sensitivity to a variety of environmental stresses [144]. In the G3 evolution endpoint population, reduced activity of Hfq (which is required for translation of *rpoS* mRNA) is probably the primary mechanism of RpoS downregulation (Figure 4.10A and B). The X1 population also contains an *hfq* mutation and shows evidence of reduced RpoS activity as well (Figure 4.10A). *acrAB-tolC* mutations associated with reduction or loss-of-function were common in our evolution endpoint populations, being discovered in five out of six populations total (Table 4.1); additionally, adaptive *acrAB* mutations were reported in an independent genomic investigation of isobutanol tolerance [124]. A recent study provides evidence that AcrAB-TolC exports an unidentified quorum sensing signal (QSS) that is associated with transcriptional upregulation of *rpoS*, possibly via a periplasmic receptor (Figure 4.12B) [143]. This suggests that adaptive effects associated with reduced AcrAB-TolC activity (Table 4.2) may be linked to downregulation of RpoS. Interestingly, a transposon mutagenesis study demonstrated that loss of function of *relA* or *spoT* is correlated with enhanced tolerance to n-butanol and isobutanol [165], and Reyes *et al.* identified a *relA* mutation in an n-butanol adapted strain [125]. *relA* and *spoT* both synthesize guanosine tetraphosphate (ppGpp), an alarmone involved in stringent response [168]. ppGpp upregulates *rpoS*, and many genes of the RpoS regulon require ppGpp for transcription; thus the correlation of reduced *relA* or *spoT* activity with improved n-butanol and isobutanol tolerance may also be related to downregulation of RpoS activity [168].

Our results suggest that attenuation of RpoS activity may be a convergent adaptive effect associated with *hfq*, *acrAB-tolC*, *relA*, and *spoT* mutations (summarized in Figure 4.12B). To determine whether RpoS attenuation is indeed adaptive, we examined the isobutanol tolerance phenotype of a Δ*rpoS::kan* single mutant (Figure 4.10C). Δ*rpoS::kan* was found to cause a growth defect at 0% and 0.5% (w/v) isobutanol relative to the WT strain, while at 1% (w/v)

isobutanol the relative fitness of Δ*rpoS::kan* is slightly higher than WT; to facilitate comparison, we report normalized relative fitness (calculated by dividing relative fitness by relative fitness at 0% w/v isobutanol). These results suggest that RpoS attenuation may indeed be adaptive to isobutanol stress. However, complete loss-of-function of *rpoS* appears to incur significant costs that overshadow adaptive effects at isobutanol concentrations below 1% (w/v) (Figure 4.10C), indicating that RpoS activity may need to be finely tuned to achieve optimal isobutanol tolerance. The growth defect of Δ*rpoS::kan* at 0% (w/v) isobutanol was unexpected, since a previous study examining *E. coli* K12 MG1655 and *E. coli* K12 MG1655 Δ*rpoS* reported nearly identical growth phenotypes for the parent and mutant strain in glucose minimal medium [139]; reasons for the discrepancy between our results and those reported in [139] might include differences in media formulation and genetic background of the host strains (*E. coli* K12 BW25113 in our study vs. *E. coli* K12 MG1655 in [139]).

Since RpoS is a global regulator affecting the expression of many genes, it is difficult to speculate on the specific adaptive effects of reduced RpoS activity. It is likely that only a subset of the gene expression changes elicited by reduced RpoS activity are adaptive, and certain gene expression changes may in fact be maladaptive, as suggested by the fitness costs apparent in the Δ*rpoS::kan* single mutant. In minimal medium, RpoS has been demonstrated to be a dominant activator of acid resistance genes (GadE/GadX/GadW regulon) and repressor of flagellar genes (FlhDC regulon) [139]. In G3.2, we observe gene expression changes in these regulons consistent with reduced RpoS activity, and NCA identified GadE and FlhDC among the most significantly perturbed transcription factors (Figure 4.8 and Additional file 4). A recent genomic study of ethanol tolerance demonstrated that overexpression of GadE/GadX/GadW regulon and *cadAB* acid resistance genes decreased ethanol tolerance, while transposon mutagenesis leading to loss-of-function of these genes was associated with increased fitness [121]. This strongly suggests that downregulation of the GadE/GadX/GadW regulon observed in G3.2 is in fact beneficial for isobutanol stress, and indeed GadE/GadX/GadW downregulation might be an important adaptive effect provided by reduced RpoS activity (Figure 4.12B). Given the dominance of RpoS in regulating GadE/GadX/GadW, we would expect GadE/GadX/GadW downregulation to be recapitulated in other populations with reduced RpoS activity [139]. Interestingly, G3.2 also has a *cadA* 4363790 A→G mutation, further suggesting that acid resistance genes are under selective pressure (Table E.1). It is not clear why attenuation of acid stress response would be adaptive to isobutanol stress. One possibility is that reduced expression glutamine/glutamate catabolic genes (i.e. *ybaS* and *gadAB*; key components of the acid stress response) may increase intracellular pools of glutamine for biosynthesis of LPS and peptidoglycan, as discussed previously. More investigation into the relationship between the GadE/GadX/GadW regulon and alcohol stress will

be needed.

4.4.4 Caveats and limitations

Given that evolved adaptations can be highly specific to a particular environmental context or genetic background, microbial tolerance to biofuel products should be evolved under conditions that closely approximate those used for biofuel production. We want to point out that we used aerobic cultivation conditions in our isobutanol tolerance evolution study, while isobutanol production in engineered *E. coli* strains is optimal under microaerobic conditions [1]. Additionally, we evolved a WT *E. coli* strain for tolerance to exogenous isobutanol, whereas isobutanol will be formed endogenously during production with engineered *E. coli* strains. Nonetheless, we feel our results still provide a valuable advancement in understanding mechanisms of isobutanol tolerance, and also provide interesting insights into the evolution of complex stress tolerance phenotypes. We have identified several mutations that appear to provide broad fitness benefits in a variety of conditions and genetic contexts (such as *marC* and *acrAB*, which were also identified in [124]), and it seems plausible that some of the adaptations we identified will be beneficial under microaerobic conditions and/or for endogenous isobutanol production as well.

4.4.5 Conclusions

In this work, we used experimental evolution of *E. coli* followed by genome resequencing and a gene expression study to elucidate genetic mechanisms of adaptation to isobutanol stress. Comparison between strains evolved in glucose and xylose minimal media revealed little carbon source specificity of adaptation, but we find that adaptations exhibit significant antagonistic pleiotropy between rich and minimal media. By examining genotypic adaptation in multiple independent lineages, we find evidence of parallel evolution in *marC*, *hfq*, *mdh*, *acrAB*, *gatYZABCD*, and *rph*. Many isobutanol tolerant lineages show reduced RpoS activity, perhaps related to mutations in *hfq* or *acrAB*. Consistent with the complex, multigenic nature of solvent tolerance, we observe adaptations in a diversity of cellular processes. Many of the adaptations appear to involve epistasis between different mutations, implying a rugged fitness landscape for isobutanol tolerance. We observe a common trend of evolution targeting post-transcriptional regulation and high centrality nodes of biochemical networks, and suggest that post-transcriptional regulators, such as *hfq*, RNA helicases, and sRNAs may be interesting mutagenesis targets for engineering complex stress tolerance phenotypes. Collectively, the genotypic adaptations we observe suggest mechanisms of adaptation to isobutanol stress based on remodeling the cell envelope and surprisingly, stress

response attenuation. The tolerant strains generated by this work could serve as useful chassis for engineering isobutanol production strains. Since we conducted our evolution in a minimal media environment, our strains may actually perform better in industrial fermentation conditions (or in the TrEc consortium) compared to current isobutanol production strains that depend on costly nutrient supplements to achieve reasonable growth and performance [1, 3].

Combining methods for identifying genes/mutations conferring desired phenotypes with genome-scale combinatorial engineering represents a potentially powerful approach for improving microbial tolerance phenotypes. In the next chapter, we use the mutations and mechanisms of tolerance identified in this work to select genetic loci for targeted mutagenesis using Multiplex Automated Genome Engineering (MAGE), allowing us to refactor isobutanol tolerance and rapidly explore large genotype spaces for improved variants.

4.5 Addendum: X2.5 genome resequencing

After this study was completed, we received a generous opportunity from MYcroarray LLC for additional genome resequencing work. At the time of this work (2011), MYcroarray was in the process of developing a sequence enrichment by depletion service for next-generation sequencing applications. In sequence enrichment by depletion, unwanted sequences are removed from solution by hybridization to complementary baits. MYcroarray has developed an approach wherein the sequencing library is denatured and hybridized to a collection of biotinylated RNA baits complementary to both strands of the sequences to be depleted. Sequences hybridized to the baits are then removed from solution using streptavidin coated magnetic beads, leaving enriched sample in the solution[1]. We collaborated with MYcroarray by providing them with an additional isobutanol tolerant isolate for beta testing their enrichment by depletion service.

To select a clonal isolate for whole genome resequencing, we first performed an *in vivo* mutator assay on each endpoint population (G1, G2, G3; X1, X2, X3) using frequency of spontaneous rifampin resistance as a metric for mutation rate, with WT *E. coli* EcNR1 and MutS- *E. coli* EcHW24 as references. We found that all glucose lineages had developed mutator phenotypes; thus we avoided selecting isolates from these endpoint populations as they are likely to have high mutation loads. We conducted additional phenotyping and genotyping assays on clonal isolates from X1, X2, and X3. Using our previously developed allele specific PCR reactions, we discovered that all isolates from the X3 endpoint population were isogenic to X3.5 at each

[1]Referenced from http://www.mycroarray.com/myreads/myreads.html

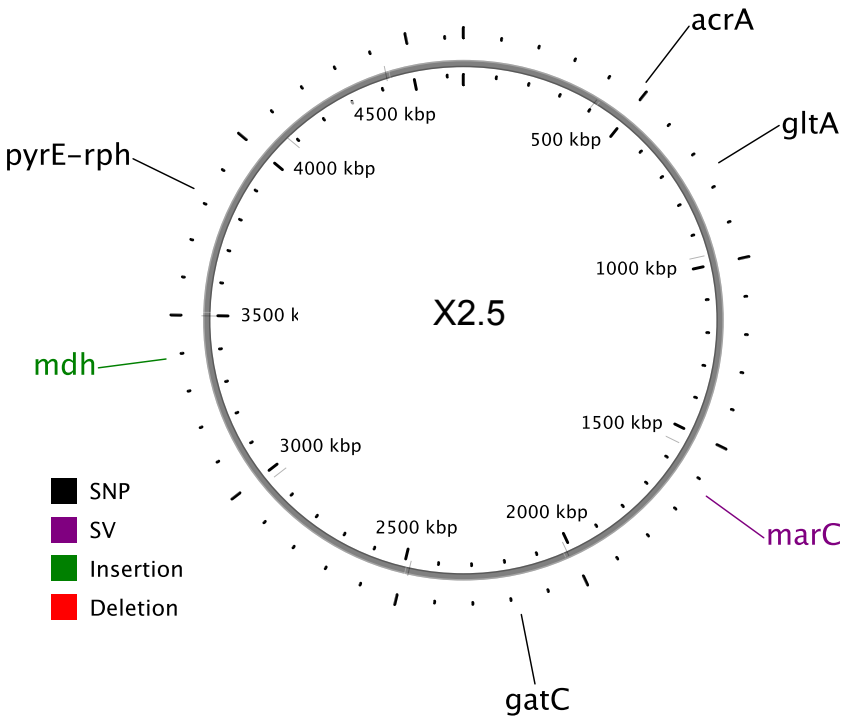

Figure 4.13: Chromosome map (*E. coli* EcNR1 reference sequence) of mutations discovered through whole genome resequencing of clone X2.5.

genotyped locus[2]; thus X3 is likely dominated by the X3.5 genotype. We then focused on populations X1 and X2 as possible sources of isobutanol tolerant isolates with unique genotypes. We performed a simplified phenotype analysis by examining maximum OD_{600} in 0.75% (w/v) isobutanol NX50 media for clonal isolates from X1 and X2. Clone X2.5 grew out to a noticeably higher OD_{600} than the other isolates and was selected for genome resequencing.

MYcroarray performed three sequencing runs with X2.5 genomic DNA using the Roche 454 GS Junior platform, yielding approximately 35 Mb sequence per run. Data was analyzed using the 454 DataAnalysis package. After the first run, depletion was performed to remove genomic regions with high coverage (>10x) and thus enrich for regions with poor coverage. This procedure was repeated again after the second run. We were able to map >99.9% of the *E. coli* EcNR1 genome, however only 40% of genome had >10x coverage. We performed Sanger sequencing to validate putative mutations; confirmed mutations are shown in Figure 4.13 and listed in Table E.1. The mutations in *acrA*, *marC*, and *mdh* were identified in our previous studies of parallel evolution between lineages, while *pyrE-rph*, *gatC*, and *gltA* mutations were not previously known in X2.5. Mutations in *rph* and *gatC* were identified in the X3 and G3 lineages (as discussed in the preceding sections) and thus their acquisition in X2.5 provides further evidence of parallel evolution at these

[2]It is possible, however, that these isolates contain additional mutations not present in X3.5.

loci. Out of the mutations identified in X2.5, *gltA* 753345 T→C (E116G in the GltA protein) is the most unique. *gltA* encodes citrate synthase, an essential gene. Thus unlike many of the other mutations identified in our study, it is unlikely to be a loss-of-function mutation. The functional effects of the E116G mutation in the GltA protein are unknown. As discussed previously, mutations in *mdh* (malate dehydrogenase) were common in our evolution end-populations, and an *mdh* insertion mutation was identified in X2.5. Interestingly, malate dehydrogenase and citrate synthase operate sequentially in the TCA cycle, suggesting that rewiring TCA cycle metabolism may be an important mechanism of adaption.

We conducted no further phenotypic, genotypic, or functional characterization of X2.5 beyond the work presented here.

4.6 Materials and methods

4.6.1 Base strains, media, and growth conditions

E. coli EcNR1 was used as the parent strain in our evolution studies. *E. coli* EcNR1 is a derivative of *E. coli* K12 MG1655 containing a modified λ prophage integrated at the *bioA*/*bioB* locus [59]. *E. coli* EcHW24 was used as a host strain for producing chromosomal mutations with ssDNA mediated recombination. *E. coli* EcHW24 is a MutS- derivative of *E. coli* EcNR1 containing 2864887 T→G and 2864892 G→T SNPs that produce premature stop codons in *mutS*. NG50 medium, consisting of M9 salts at 1x concentration, 50 g/L glucose, and 0.25 mg/L biotin was used in the adaptive evolution of *E. coli* EcNR1 populations with glucose as a sole carbon source. NX50 medium, formulated similarly to NG50 with glucose replaced by 50 g/L xylose, was used in the adaptive evolution of *E. coli* EcNR1 populations on xylose. NG50 and NX50 agar media were prepared by supplementing NG50 and NX50 media with 15 g/L agar. LB Lennox broth (10 g/L tryptone, 5 g/L yeast extract, and 5 g/L NaCl) and LB agar (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, and 15 g/L agar) were used for propagating strains during genetic manipulations and genomic DNA extraction. Clonal isolates from evolved populations were obtained by isolation streaking culture samples on LB agar plates or NG50/NX50 agar plates and propagating selected colonies. All *E. coli* strains used in this study were grown at 30°C with 150 to 200 rpm shaking.

4.6.2 Adaptive evolution

Isobutanol tolerant lines of *E. coli* EcNR1 were evolved by serial passaging of three independent populations on isobutanol spiked NG50 medium (glucose as sole carbon source) and three independent populations on isobutanol spiked NX50 medium (xylose as sole carbon source). Cultures were grown in tightly capped 15 mL Falcon tubes containing 3 mL of respective medium. The initial isobutanol concentration was 0.75% (w/v) for all populations, and was gradually increased during the evolution to maintain an approximately constant selective pressure. Cultures were passaged when populations reached mid log phase, and the fresh cultures were inoculated to yield an initial optical density at 600 nm (OD_{600}) = 0.002. Each lineage was periodically checked for contamination by isolation streaking culture samples on LB agar. Samples from each population were cryopreserved every 5 to 10 passages by centrifuging 2 x 1.4 mL samples from each culture at 14,000 rpm x 1 minute, washing the cell pellets with fresh medium, centrifuging again, and resuspending each cell pellet in 150 μL fresh medium + 150 μL cryopreservation solution (65% v/v glycerol and 0.1 M $MgSO_4$). Cell suspensions were transferred to 96-well microplates, sealed with adhesive film, and stored at -80°C. The evolution proceeded for 180 days, corresponding to approximately 500 generations for the glucose lineages and 430 generations for the xylose lineages, assuming approximately 7.5 generations per passage.

4.6.3 Phenotype evaluation

Isobutanol tolerance was quantified by measuring the maximum specific growth rate (μ_{max}, 1/h) and saturating OD_{600} at various isobutanol concentrations, using a microplate spectrophotometer. Inoculum was prepared by centrifuging 1 mL of overnight culture at 12,000 rpm x 2 min and resuspending cell pellets in a volume of fresh medium such that OD_{600} = 2. Standard 96-well microplates were filled with 200 μL medium per well (spiked with isobutanol or alcohols as appropriate) and seeded with 2 μL of prepared inoculum per well. Microplates were covered with adhesive film to prevent isobutanol evaporation and microplate lids were affixed with tape. OD_{600} was measured every 10 minutes for 48 hours using Molecular Devices Spectramax M5 or Molecular Devices Versamax plate readers, with 30°C incubation temperature and agitation between reads. μ_{max} was calculated via linear regression of $\ln(OD_{600})$ vs. time (h) after subtracting blank values; regression was done over the time intervals corresponding to log growth phase.

4.6.4 Genome resequencing

Genomic DNA (10 to 20 μg) was isolated from clonal isolates chosen for genome resequencing (G3.2, G3.6, G3.266.7, and X3.5) using a DNEasy spin column kit (Qiagen, Germantown, MD, USA) according to the manufacturers protocol. Genomic DNA libraries were prepared using the Illumina genomic DNA library generation kit following the manufacturer's protocol (Illumina Inc., San Diego, CA, USA). Briefly, 5 to 10 μg genomic DNA was fragmented using a Covaris Acoustic System to ˜200 bp DNA fragments. The ends of the fragmented DNA were then repaired with T4 DNA polymerase, Klenow DNA polymerase, and T4 PNK to convert the overhangs into phosphorylated blunt ends. Klenow fragment (3 to 5 exo minus) was used to add an A base to the 3 end of the blunt phosphorylated DNA fragments. Following ligation of adapters to the ends of the DNA fragments, PCR was used to enrich the adapter-modified DNA fragments to obtain a DNA library suitable for high-throughput sequencing using Illumina Genome Analyzer. The concentration of DNA library was obtained by RT-PCR. Libraries for single-end sequencing were prepared for G3.2, G3.6, and X3.5 and a paired-end sequencing library was prepared for G3.266.7. Single-end libraries were sequenced with the Illumina Genome Analyzer 2 using a 36 cycle run, while the G3.266.7 paired-end library was sequenced on the same platform using 2 x 36 cycle runs.

4.6.5 Sequence analysis

A custom perl script (available upon request) was written to automate the sequence analysis described below. Raw illumina reads were aligned to the *E. coli* EcNR1 reference sequence using Novoalign v2.04.02 [129]. Novoalign output was converted to MAQ (Mapping and Assembly with Qualities) map format, and MAQ v0.7.0 was used to build a consensus sequence and call SNPs and short indels [130]. Annotations and descriptions of mutated genes were downloaded from KEGG (Kyoto Encyclopedia of Genes and Genomes), and amino acid changes due to SNPs were automatically computed [184]. Large deletions were detected by tabulating coverage gaps in the alignments. To detect structural variation (SV) breakpoints with single-end reads, unmappable reads were filtered to eliminate long homopolymer runs (since erroneous homopolymer runs at tile edges are common with the Illumina platform) and de novo assembled with Velvet v0.7.51 [131]. For each set of reads, assemblies were done over a range of k-mer values. The assemblies were aligned to the *E. coli* EcNR1 reference sequence with BLAST, and were inspected manually to detect breakpoints [185]. For paired-end sequencing data, structural variations were detected using BreakDancer v0.0.1, using mean paired-end insert sizes calculated by MAQ [186]. Primer3 v2.0.0 was utilized to automatically design primers (with amplicon sizes of 800-1000 bp and melting temperatures of 60°C - 65°C) flanking detected mutations for Sanger sequencing verification

[187]. For clones G3.2 and G3.6, only mutations shared between these two clones were verified by Sanger sequencing. In other sequenced clones, SNPs with consensus quality > 200 (as computed by MAQ) or indels with frequency > 0.4 were verified; we have found empirically that lower quality/frequency mutations are almost always false positives [130].

4.6.6 Sanger sequencing

Mutations identified in the genome sequence analysis, as well as mutations reconstructed in *E. coli* EcHW24, were verified using Sanger sequencing. Sanger sequencing was also used to search for mutations in *marC*, *acrAB*, *tolC*, *mdh*, and *hfq* in evolved lineages that were not characterized by genome resequencing. See Additional file 5 for primer list. For mutation verification, 200 to 1000 bp regions containing putative mutations were amplified by PCR, using Phusion HotStart polymerase (Finnzymes USA, Woburn, MA) with manufacturers recommended cycling conditions and a 6:00 min initial denaturation at 95°C. Cell suspensions were prepared by resuspending colony material in 250 μL sterile water; 1 μL cell suspension was used as template per 50 μL PCR reaction. For mutation searches, the entire locus of interest plus at least 200 bp upstream/downstream was amplified, using Phusion HotStart polymerase for PCR. Agarose gel electrophoresis (50 mL 0.7% agarose gel, tris-acetate-EDTA (TAE) or tris-borate-EDTA (TBE) buffer, 98 V for 45 minutes running time) was used to verify PCR product size and reaction specificity. PCR products were purified with a QIAquick spin column kit (Qiagen), as per manufacturers protocol. DNA concentration in purified PCR products was quantified with a Thermo Scientific Nanodrop spectrophotometer. Purified PCR products were diluted to 3 ng/μL/kb product size, and were Sanger sequenced with the same primers used for PCR. For amplicons > 1200 bp in size, internal primers were designed and also used for sequencing, such that sequencing primers were spaced every 600 bp. Sanger sequencing was completed by the University of Michigan DNA Sequencing Core. Returned sequences were aligned to the reference *E. coli* EcNR1 sequence using BLAST, and chromatograms were manually inspected to verify mutations.

4.6.7 Allele specific PCR

Primers were designed such that one primer in a pair was complementary to a mutation of interest at its 3 end [188]. Under stringent PCR conditions, non-proofreading DNA polymerases are unable to extend from 3 mismatches, allowing genotype discrimination based on the presence or absence of a PCR product [188]. Allele specific primers for SNPs were designed using

BatchPrimer3 v1.0, and allele specific primers for indels were designed manually [188]. See Additional file 5 for primer list. Platinum Taq polymerase (Invitrogen, Carlsbad, CA, USA) was used for PCR, with 0.2 μM of each primer and other reagent concentrations as per manufacturers recommendations. Cell suspensions were used as template as described for Sanger sequencing above. Cycling conditions used were initial denaturation at 94°C for 6:00, followed by 28 cycles of 94°C denaturation for 30 seconds, optimal annealing temperature for 30s, and 72°C elongation for 1:00, followed by a final 72°C extension for 5:00. Optimal annealing temperature for each PCR reaction was determined via annealing temperature gradient with WT *E. coli* EcNR1 and appropriate mutants as controls. Agarose gel electrophoresis (50 mL 1% agarose gel, TBE buffer, 110 V for 30 minutes running time) was used to examine PCR products.

Multiplex allele specific PCR was done with a Qiagen Multiplex PCR kit, using manufacturers recommended reagent concentrations and cell suspensions for template. Cycling conditions used were initial denaturation at 94°C for 15:00, followed by 28 cycles of 94°C denaturation for 30 seconds, optimal annealing temperature for 1:30, and 72°C elongation for 1:30, followed by a final 72°C extension for 10:00. Optimal annealing temperature for each PCR reaction was determined via annealing temperature gradient with *E. coli* EcNR1 and appropriate mutants as controls. Agarose gel electrophoresis (50 mL 3% agarose gel, TBE buffer, 110 V for 1 h 25 minutes running time) was used to examine PCR products.

4.6.8 Genetic manipulation methods

4.6.8.1 High efficiency ssDNA mediated homologous recombination

High efficiency ssDNA mediated mutagensis in *E. coli* EcHW24 was used to engineer chromosomal SNPs and short indels, as per previously described procedures [59]. 90-mer oligonucleotides containing mutations of interest flanked by homologous genomic sequences were designed such that they targeted the lagging strand of the replication fork during DNA replication and had $\Delta G > 12.5$ kcal/mol for secondary structures (evaluated with mfold v3.2) [189]. Oligos were synthesized by IDT (Integrated DNA Technologies, Coralville, IA) with four 5 phosphorothioated bases to enhance in vivo stability. See Additional file 5 for oligo list.

Homologous recombination was done by heat shocking *E. coli* EcHW24 to induce expression of λ-Red genes, preparing electrocompetent cells from the induced cultures, and electroporating the competent cells with oligonucleotide at 5 μM concentration. Overnight cultures of *E. coli* EcHW24 were inoculated 1:70 into fresh LB medium and incubated with shaking at 30°C until

reaching OD_{600}=0.7. Cultures were then heat shocked at 42°C for 15 minutes in a water bath with 200 rpm shaking. Immediately after heat shocking, the cultures were chilled on ice for 10 minutes. All subsequent manipulations were done at 4°C, which is vital for maximum recombination efficiency. For each electroporation, 1 mL of induced cells was centrifuged at 16,000 g x 1 minute. Cells were resuspended in 1 mL chilled ultrapure dH_2O. The centrifugation and washing process was repeated twice. A final centrifugation was performed and the cell pellet was resuspended in 50 μL ultrapure dH_2O. Oligonucleotide solution was mixed with the cell suspension such that the final DNA concentration was 5 μM. Immediately after adding oligonucleotide, cell/DNA mixture was transferred to 0.1 cm gap electroporation cuvette and electroporated at 1.8 kV, using an Eppendorf Electroporator 2510. Cell mix was immediately resuspended in 1 mL room temperature LB medium, added directly to the electroporation cuvette. The resuspended cell mix was transferred to a Falcon tube, 1-5 mL LB medium was further added, and cells were allowed to recover at 30°C with shaking.

After 3-5 hours of incubation, 100 μL of $1{:}10^4$ and $1{:}10^5$ diluted recovery mix were plated on LB agar and incubated at 30°C overnight. The described recombination approach has efficiencies ranging from 1-40%, so selection is not required; mutants were recovered by direct genetic screens of colonies. For each recombination, 20 to 300 colonies were screened for desired mutations using Sanger sequencing or allele specific PCR (described above).

Multiplex recursive ssDNA mediated mutagenesis was used to generate a mutant set containing all combinations of *miaA-hfq*, *rph*, *mdh*, and *groL* mutations identified in the G3 lineage. The above homologous recombination procedure was carried out, using 5 μM oligonucleotide for each mutation (*miaA-hfq*, *rph*, *mdh*, and *groL*, 20 μM total) in electroporation mixes. *E. coli* EcHW24 was electroporated without oligonucleotides (oligo-) as a control. Recovery mixes were allowed to grow to OD_{600}=0.7, and the homologous recombination procedure was repeated. This recursive homologous recombination process was repeated for a total of six cycles. $1{:}10^2$, $1{:}10^3$, and $1{:}10^4$ dilutions of the final recovery mix were prepared and 100 μL aliquots were spread on NG50 plates supplemented with isobutanol at 0.7% (w/v). Plates were wrapped in parafilm to prevent isobutanol evaporation and incubated at 30°C for 48 hours. Clones showing improved isobutanol tolerance, judged by colony size and comparison with the oligo- control plate, were screened using allele specific PCR for *rph* and *groL* and Sanger sequencing for *miaA-hfq* and *mdh*. Putative *rph* and *groL* mutations were verified by Sanger sequencing.

4.6.8.2 Construction of gene knockouts with P1 transduction

E. coli EcNR1 gene knockouts were constructed via P1vir transduction using Keio single gene knockout strains as donors, as described previously [190, 145]. P1vir lysates were prepared for each Keio mutant (Δ*acrA::kan*, Δ*acrB::kan*, and Δ*mdh::kan*), and used to transduce gene knockouts to *E. coli* EcNR1 using LB agar with 100 μg/mL kanamycin as selective medium. Transductants were purified from residual P1vir phage by isolation streaking on LB agar supplemented with 0.8 mM sodium citrate and 100 μg/mL kanamycin, and verified by PCR as per published procedures [145].

4.6.8.3 Construction of Δ*marC::kan* knockouts with homologous recombination

We were unable to generate Δ*marC::kan* knockouts using P1 transduction; dsDNA homologous recombination was used as an alternative procedure. A dsDNA cassette containing *kan* flanked by 50-100 bp of *marC* homologous sequence was produced via PCR (as per PCR procedure described for Sanger sequencing), using the Δ*marC::kan* Keio mutant as template (see Additional file 5 for primers). Homologous recombination was carried out using *E. coli* EcHW24 or progeny host strains. Procedures for heat shock, preparation of electrocompetent cells, and electroporation were identical to those described for high efficiency ssDNA mediated recombination. 50 ng purified Δ*marC::kan* PCR product was used for each electroporation. Electroporation recovery mixes were incubated at 30°C for two hours, then centrifuged at 12,000 rpm x 1 minute. After discarding supernatant, cell pellets were spread on LB agar plates supplemented with 100 μg/mL kanamycin and incubated overnight at 30°C. Colonies were PCR screened to verify Δ*marC::kan* genotype.

4.6.9 Microarray sample processing and data generation

4.6.9.1 Probe and microarray design

Microarray probes targeting coding sequences (CDSs) from the *Escherichia coli* str. K12 substr. MG1655 genome (Genbank NC_000913) were designed using the OligoArray software [191]. In a first pass using the following parameters (Probe length, 45 - 47 nucleotides; GC content 41 - 57% and T_m 83 - 91°C; Up to 3 probes per gene), we obtained 12037 probes targeting 4253 out of the 4292 CDSs annotated for this strain (99% of all CDSs). In a second pass, the GC content and T_m parameters were relaxed to 30 - 70% and 78 - 97°C, respectively. This led to the design of an additional 41 probes targeting 27 genes missed from the first pass for a total of 12078

probes targeting 4280 CDSs. Twelve genes (*thrL*, *ylbI*, *trpL*, *pheM*, *yojO*, *ypaB*, *ypdJ*, *pheL*, *yhaL*, *yrhD*, *yifL* and *pyrL*) failed the second pass and were excluded from this study. Each probe was replicated 6 times on the array, each replicate being randomly distributed across the whole array area for a total of 72,468 spots. Oligonucleotide probes were synthesized by MYcroarray (Ann Arbor, MI, USA) on an 80K array format, one array per slide.

4.6.9.2 Sample Preparation

Total RNA was extracted from mid log phase cultures using an RNeasy kit (Qiagen). 100 μL RNAase free water was used for final elution; a 90 μL aliquot was immediately ethanol precipitated and stored at -80°C. mRNA was enriched from 10 μg of total RNA using a MICROBExpressTM Bacterial mRNA Purification Kit (#1905, Ambion, Austin, TX, USA) by the removal of 16S and 23S ribosomal RNAs. The enriched mRNA (200 ng) was converted to cRNA containing aminoallyl-UTP with a Message Amp II - Bacteria Prokaryotic RNA amplification kit (#1790, Ambion) following the manufacturers instructions. Amino-allyl modified cRNA (45 μg) was coupled with amine reactive fluorescent dye (Alexa Fluor-555, #32756, Invitrogen) in a 10 μL reaction following the manufacturer's instructions. After fluorescent dye coupling, unincorporated dye was removed with RNEasy mini columns (Qiagen) following the manufacturers instructions. Labeled cRNA was eluted from RNeasy mini columns with RNAse/DNAse free water. The extent of dye incorporation was determined by the Microarray function on a NanoDrop 1000 spectrophotometer (Thermo Scientific). The dye incorporation calculations were performed as described by Invitrogen/Molecular Probes for Alexa Fluor dye products. All samples had dye incorporation between 35 to 43 bases per dye. Fluorescent dye coupled cRNA (20 μg) was fragmented by exposure to zinc sulphate (5 mM final concentration in a 60 μL reaction) at 75°C for 10 min and the reaction was stopped by the addition of 500 mM EDTA to a final concentration of 20 mM. The extent of fragmentation was visualized with a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) using a RNA Nano Chip (Agilent Technologies, #5067-1511). Samples with a mean fragment size of 100-200 nucleotides were qualified for hybridization.

4.6.9.3 Hybridization

Each labeled and fragmented cRNA sample was hybridized individually to one custom 80K microarray by dynamic hybridization as follows: 20 μg of sample was added to hybridization solution (600 μL final volume) and incubated at 65°C for 5 min and then placed on ice. Hybridization solution contained 6X SSPE (1M NaCl, 6.7mM EDTA, 40mM NaH2PO4 and

27.3mM Na2HPO4), 0.01 μg/μL acetylated BSA (#R3961, Promega, Madison, WI, USA), 0.01% Tween-20 (#P9416, Sigma, Saint Louis, MO, USA), and 10% deionized formamide (#P9037, Sigma). A large volume of hybridization solution master mix was prepared from which aliquots were removed to prepare each sample. Hybridizations were performed using a hybridization gasket slide (#G2534-60003, Agilent Technologies) and a Microarray Hybridization Chamber assembly (#G2534A, Agilent Technologies) following the manufacturers instructions except that 585 μL of hybridization solution was used per gasket slide and both the gasket slide and microarray slide were preheated to 65°C prior to Hybridization Chamber assembly. The final Hybridization Chamber assembly was incubated at 50°C for 20 hrs while rotating at 5 rpm (to assure free movement of the mixing bubble) in a hybridization oven (#G2545A, Agilent Technologies).

4.6.9.4 Washing and Scanning

Following hybridization, unbound material was removed as follows: The hybridization chamber assembly was quickly removed from the hybridization oven and the microarray/gasket slide sandwich was immediately submerged in 1X SSPE at room temperature. The slide was quickly transferred to fresh wash solution (1X SSPE) at room temperature and incubated for three minutes with gentle agitation. This wash was repeated twice, however, the first repeat was performed at 50°C. Finally, the slide was rinsed for 30 sec in 0.25X SSPE at room temperature and immediately spun dry in a Microarray Minifuge (ArrayIt, Sunnyvale, CA, USA). Slides were immediately scanned using an Axon 4000B scanner (Molecular Devices) at 5 micron resolution and 100% laser power. The PMT gain in the 532 nm channel was adjusted to appreciate the full dynamic range (0 - 65,000) of signal intensity such that only a few pixels were saturated in a few spots.

4.6.9.5 Data Extraction

A signal intensity value for each probe on the array was extracted from the scanned image using Axon GenePixPro 6.1 software (version 6.1.0.4, Molecular Devices). Fixed diameter (35 microns) circular feature indicators were placed over the centre of each spot (probe) and median pixel intensity was calculated for each feature.

4.6.10 Microarray data analysis

4.6.10.1 Pre-processing

First, background fluorescence intensities of individual spots were subtracted directly from foreground intensities for background adjustment. Then variance-stabilizing normalization (vsn) and quantile methods without weight were applied sequentially for normalization with software R (2.11.0, "vs" and "aroma.light" packages [192, 193, 194, 195, 196]). For each strain/isobutanol condition, spots with acceptable signals in at least two out of three biological replicates were chosen for further analysis, which resulted in 4-6 replicates for each probe. Two or three probes for each gene were included on the array. For simplicity, we chose the probe with the smallest overall variation across replicates (measured by the sum of standard deviations across 12-18 technical/biological replicates over all the four strain/isobutanol conditions) to represent the gene in further analysis. For each chosen probe, the median of the technical replicates on each array was calculated to represent the expression level of the corresponding gene for each sample. The above pre-processing procedure provided normalized expression data for 4235 genes (out of 4280 included on the array). The complete dataset has been deposited in the GEO database (www.ncbi.nlm.nih.gov/geo) under accession number GSE23526.

Identification of differentially expressed/regulated genes. We first applied two filters to select genes that showed notable expression changes, namely by choosing genes with expression levels higher than 100 units in at least 25% of the 12 samples (i.e. 3 samples) and interquartile ranges (IQR) larger than 0.5 [192, 197]. This led to a set of 2026 genes for further analysis and two-sample students t-test was carried out to identify genes that showed differential expression. In particular, three t-tests were conducted: a t-test between WT/0% isobutanol and WT/0.5% isobutanol; a t-test between G3.2/0% isobutanol and G3.2/0.5% isobutanol; and a t-test between changes in G3.2 upon isobutanol treatment and changes in WT upon isobutanol treatment. The first two t-tests identified differentially expressed genes in each strain and the last t-test identified genes that responded to isobutanol differently between G3.2 and WT. Empirical Bayes statistics were applied to remove chip effects with software R (2.11.0, "limma" package [198], parameters: assumed proportion of differentially expressed genes equal to 0.01, assumed lower and upper limits for the standard deviation of log2 fold changes for differentially expressed genes equal to 0.1 and 4 respectively [199, 200]). In addition, the p-values of the t-test were adjusted by controlling false discovery rate (FDR, BH procedure) [201]. Genes with adjusted p-values less than 0.001 were considered as differentially expressed in the first two t-tests or responding to isobutanol differently in the last t-test.

4.6.10.2 Regulatory network analysis

Network Component Analysis (NCA) was carried out with NCA toolbox (v2.3 for Matlab, released Feb. 19, 2007) [202, 203] to predict transcription factor activities using microarray data. Based on previous study of the isobutanol response network in *E. coli* [137] and preliminary examination of our microarray data, we selected 16 transcription factors that are potentially involved in isobutanol tolerance (ArcA, PdhR, Fnr, Fur, FlhDC, OmpR, CRP, GadE, MarA, Nac, LexA, PurR, Fis, IHF, PhoB and PhoP) for this analysis. The corresponding transcriptional network was constructed based on connectivity data in RegulonDB (v6.7, MAR.2, 2010) [204]. A total of 998 genes from the above pre-processed expression dataset were regulated by this set of TFs, with each gene regulated by 1 to 5 TFs (1.5 on average). Due to limited data (i.e. four strain/isobutanol conditions), we used a subset of four TFs in each NCA analysis and repeated the analysis for different combinations of the TFs. Only TFs that were consistently predicted to change significantly across different combinations of TFs and different replicates were kept for further rounds of analysis; at the end, four TFs that showed the most significant activity changes were obtained (FlhDC, GadE, MarA and PhoP). In the study described in [137], it was concluded that activities of ArcA, PhoB, and Fur were significantly perturbed by isobutanol due to quinone/quinol malfunction. We performed NCA for various combinations of ArcA, PhoB, and Fur with FlhDC, GadE, MarA, and PhoP (using up to six TFs in each NCA) to determine whether these results were recapitulated in our study. For each strain/condition, TF activities were calculated for all the biological replicates and we report their average value and 95% confidence interval in Figure 4.8D.

4.6.11 Analytical methods

4.6.11.1 qRT-PCR validation of gene expression changes

Total RNA extraction and mRNA enrichment were carried out as described for microarray sample processing, with three biological replicates per strain/condition. Biological replicates were pooled and 100 ng enriched mRNA was reverse transcribed using a QuantiTect Reverse Transcription kit (Qiagen) as per manufacturers protocol. Quantitative PCR assays were performed in 25 μL samples on an MJ Research (BioRad, Hercules, CA, USA) Chromo4 thermal cycler with a QuantiTect SYBR Green RT-PCR kit (Qiagen), using primer pairs designed as per manufacturers recommendations for genes *gadA*, *fimI*, *fabA*, *rfaJ*, and *rpoD* (Additional file 5). 5 ng reverse transcribed cDNA was used as template for each reaction, with primer and other reagent concentrations as per manufacturers protocol. Cycling conditions used were initial denaturation at 95°C for 15 minutes, followed by 45 cycles of 94°C denaturation for 15 seconds, 60°C annealing for 30

seconds, 72°C elongation for 30 seconds, and fluorescence measurement at 72°C. After the final cycle, reaction specificity was verified by determining melting profiles over a temperature range of 65°C to 95°C in 0.2°C increments. All assays were run in triplicate. qRT-PCR data was analyzed by fitting parameters of the MAK2 model (initial target concentration D_0 and characteristic PCR constant k) , using a custom Mathematica script [135]. Initial target concentrations were normalized to *rpoD*.

4.6.11.2 Fatty acid analysis

Duplicate 100 mL cultures of *E. coli* EcNR1 and 50 mL cultures of G3.2 were grown to mid log phase in NG50 medium supplemented with 0.5% (w/v) isobutanol. Cultures were harvested by centrifugation at 3,200 g x 10 minutes. Cell pellets were washed with Phosphate Buffered Saline (PBS) and dried at 70°C for 24 hours. The dry cell pellets were subjected to ethanolysis at 90°C for 2 h in 5% HCl in anhydrous ethanol. The resulting Fatty Acid Ethyl Esters (FAEEs) were quantified with an Agilent 6890 GC equipped with a 50 m x 0.2 mm x 0.33 mm HPx5 capillary column, flame ionization detector (FID), and autoinjector. 2 μL FAEE samples were analyzed after split injection (1:10). Helium was used as a carrier gas with a constant flow rate of 1.9 mL/min. The temperatures of the injector and detector were 325 and 350°C, respectively. The following temperature program was applied: 50°C for 3 min, increase 10°C/min to 300°C, then 300°C for 10 min.

4.6.11.3 Cell envelope protein analysis

50 mL cultures of *E. coli* EcNR1 and G3.2 were grown to mid log phase in NG50 medium supplemented with 0.5% (w/v) isobutanol. $5x10^9$ cells (estimated by OD_{600}) were harvested by centrifugation at 5,000 g x 10 minutes. Cell pellets were washed with PBS and resuspended in 3 mL 10 mM NaH_2PO_4-NaOH (pH 7.2) buffer. Cell suspensions were placed on ice and lysed by sonication (15 W continuous output; 30 seconds sonication followed by 30 seconds cooling; repeated six times total). Unbroken cells were removed by centrifugation at 1,500 g x 20 minutes. Supernatant was ultracentrifuged at 100,000 g x1 hour at 4°C to pellet cell envelopes. Cell envelope pellets were washed with 10 mM NaH_2PO_4-NaOH (pH 7.2), resuspended in 20 μL dH_2O plus 20 μL 2x Laemmli buffer (BioRad), and then incubated at 95°C for 5 minutes. Cell envelope samples were then analyzed with SDS-PAGE (12.5% gel; 25 μL loading volume; 200 V for 55 minutes). After electrophoresis, gel was incubated in Coomassie staining solution (50% v/v methanol, 10% v/v acetic acid, and 0.25% w/v Coomassie brilliant blue R-250 in dH_2O)

for 40 minutes with gentle shaking and then destained by boiling in 1 L dH_2O for 20 minutes. Protein band intensity was quantified by densitometry analysis (using ImageJ software) and normalized to the sum of intensities of the major protein bands (i.e. the 95 kDa, 72 kDa, 55 kDa, 44 kDa, OmpC/OmpF, OmpA, 29 kDa, and 26 kDa bands). Experiment was repeated to verify results.

4.6.11.4 RpoS Western blot

Total cellular protein was extracted from early log phase cultures using B-PER lysis buffer (Thermo Fisher Scientific, Waltham, MA, USA) as per manufacturers protocol. Protein concentration was quantified using Bradford assay [96]. Samples were diluted with PBS to 4 μg/μL protein and mixed 1:1 with 2x Laemmli buffer (BioRad) and then incubated at 95°C for 5 minutes. Protein samples were separated with SDS-PAGE (12.5% gel; 10 μL loading volume, corresponding to 20 μg protein; 100 V for 55 minutes) and then electroblotted onto a 0.22 μm nitrocellulose membrane (BioRad). Membrane was blocked by incubating for one hour with gentle shaking in 15 mL Tris-Buffered Saline Tween-20 (TBST) buffer (BioRad) supplemented with 5% (w/v) nonfat dry milk (BioRad). After blocking, membrane was incubated for two hours with gentle shaking in 15 mL blocking buffer (TBST with 5% w/v milk) supplemented 1:1000 with primary mouse anti-RpoS antibody (NeoClone, Madison, WI, USA). Membrane was then washed three times with TBST and incubated for one hour with gentle shaking in blocking buffer supplemented 1:2000 with secondary goat anti-mouse IgG horseradish peroxidase conjugated antibody (Jackson ImmunoResearch, West Grove, PA, USA). After incubation with secondary antibody, membrane was washed three times with TBST and treated with Immun-Star horseradish peroxidase substrate (BioRad) for detection, as per manufacturers protocol. Experiment was repeated to verify results. Purified *E. coli* RpoS (NeoClone) was used as a positive control.

4.6.11.5 Ethidium bromide accumulation assay

Ethidium bromide accumulation assays were carried as described previously, with minor modifications [142]. For each strain tested, LB medium was inoculated 1:70 with saturated overnight culture and incubated at 30°C with shaking. When cultures reached OD_{600} = 0.6, 600 μL aliquots were withdrawn and centrifuged at 13,000 rpm x 3 minutes. Supernatant was discarded and cell pellets were resuspended in 1000 μL sterile phosphate buffer saline (PBS). The centrifugation and washing process was repeated once more. After a final centrifugation, cell pellets were resuspended in PBS such that OD_{600} = 0.3 in the final cell suspension. Costar black/clear bottom 96-well microplates (Thermo Fisher Scientific) were filled with 2 μL 0.1 mg/mL ethidium bromide

per well and 198 μL OD_{600} = 0.3 cell suspension for 200 μL total volume. Fluorescence (518 nm excitation / 605 nm emission) was measured every minute for 60 minutes using a Molecular Devices Spectramax M5 plate reader, with 25°C incubation temperature and agitation between reads. Specific fluorescence was calculated by dividing measure fluorescence by OD_{600}.

4.6.11.6 Malate dehydrogenase assay

Malate dehydrogenase assays were carried out as described previously, with modifications [205]. For each strain tested, LB medium was inoculated 1:70 with saturated overnight culture and incubated at 30°C with shaking. When cultures reached OD_{600} = 0.6, 5 mL aliquots were withdrawn from each culture and centrifuged at 5000 g x 10 minutes. Supernatant was discarded. Cell pellets were suspended in 50 μL B-PER (Thermo Fisher Scientific) lysis buffer and incubated for 15 minutes at room temperature. After incubation, cell lysates were centrifuged at 15,000 g x 5 minutes at 4°C. Lysis supernatant was reserved and kept on ice for the remainder of the procedure. Standard 96-well microplates were filled with 185.6 μL PBS, 4.45 μL 10 mM NADH, and 10 μL 1:10 diluted lysate supernatant per well. Absorbance at 340 nm was measured every 30 seconds for 10 minutes at 30°C with shaking to determined background NADH oxidation rate. 22.2 μL 20 mM oxaloacetate was then added to each well. Absorbance at 340 nm was measured every 30 seconds for 15 minutes at 30°C with shaking to determine rate of NADH oxidation. Malate dehydrogenase activity was calculated as Units/g-wet-cells $= \frac{\Delta A_{340}}{\Delta t \varepsilon_{340} l c}$, where ΔA_{340} = rate of change in absorbance at 340 nm wavelength (AU), Δt = time interval (15 minutes), ε_{340} = NADH molar extinction coefficient at 340 nm wavelength (6.22 1/mM/cm), l = 1 cm, and c = 0.0045 g-wet cells per reaction volume.

4.6.11.7 Iodine staining assay

Iodine staining assay was adapted from a previously described procedure [144]. Overnight cultures of selected strains were streaked on NG50 agar spiked with 0.35% (w/v) isobutanol. The plate was tightly wrapped in parafilm, incubated at 30°C for 48 hours, and then flooded with 5 mL USP tincture of iodine. After 2 minutes of incubation at room temperature, excess iodine was poured off the plate, and iodine was allowed to evaporate until media was translucent.

CHAPTER 5

Elucidating and improving the genetic architecture of isobutanol tolerance in *Escherichia coli* through targeted genome engineering

5.1 Summary

Due to the complexity of microbial stress tolerance, there is a great need for phenotype improvement methods that expand accessible genotype search spaces and which can account for the multigenic nature this phenotype. Guided by the evolutionary-genomic studies described in the previous chapter, in this work we generate combinatorial libraries of 38 mutations associated with isobutanol tolerance in *E. coli* using Multiplex Automated Genome Engineering (MAGE), a recently developed technology that entails repeated cycles of high efficiency recombination using libraries of mutagenic DNA oligonucleotides. This strategy enables rapid exploration of vast genotype space without being constrained to adaptive walks. Variants with improved isobutanol tolerance are isolated and then further characterized via quantitative phenotype and genotype analysis, allowing for systematic mapping of isobutanol tolerance phenotypes and genotypes. We identified numerous novel isobutanol tolerant genotypes, yielding additional insights into biochemical mechanisms of tolerance. Additionally, the improved strains generated in this study may be immediately useful for isobutanol production.

The work presented in this chapter was done in collaboration with the labs of George Church (Harvard Medical School) and Harris Wang (Pathology and Cell Biology, Columbia University), and will be reported in a future manuscript: J. Minty, L. Lai, J. Park, S. Wang, B. Johnson, T. Zaroff, L. Kennedy, D. Boyer, H. Wang, M. Burns, G. Church, and X. Lin. Elucidating the genetic architecture of isobutanol tolerance in *Escherichia coli* through targeted genome engineering and high throughput screening.

5.2 Introduction and background

5.2.1 Integrating genome mapping and genome engineering for phenotype improvement

Microbial stress tolerance is a complex multigenic trait that is intractable to traditional genetic study and rational engineering efforts, thus most methods for stress tolerance phenotype improvement involve combinatorial or random-search approaches. In the past, limitations in genetic manipulation techniques hindered efforts to improve tolerance phenotypes; thus strategies were restricted to random mutagenesis, adaptive evolution, and breeding (for sexual organisms). These limitations motivated the development of genome-scale techniques for improving and studying complex phenotypes, including whole genome shuffling, transcription factor mutagenesis (e.g. global transcription machinery engineering; gTME), and screening of gene deletion libraries, genomic fragment overexpression libraries, and promoter and intergenic region libraries (as discussed in Chapter 1). Advances in DNA sequencing, DNA synthesis, and genome-scale genetic manipulation techniques have radically expanded our ability to read and write genetic information, enabling unprecedented opportunities for both genetically decoding and programming complex phenotypes.

Integrating methods for identifying genes/mutations conferring desired phenotypes with genome-scale combinatorial engineering represents a promising strategy for improving microbial chemical tolerance phenotypes. This general approach was recently tried by Sandoval *et al.* to identify combinations of genetic loci correlated with improved tolerance of *E. coli* to acetate, low pH, and cellulosic hydrolysate [206] (overview shown in Figure 5.1). Trackable Multiplex Recombineering (TRMR; described in Chapter 1 and [56]) was used to map the effect of increased or decreased expression of every gene in the *E. coli* genome onto each of the aforementioned phenotypes, resulting in identification of gene subsets associated with each trait (Figure 5.1A&B) [206]. Ribosome Binding Sites (RBS) of identified target genes were mutagenized using Multiplex Automated Genome Engineering (MAGE; described in Chapter 1 and [59]) to generate combinatorial libraries of gene expression variants, which were then subjected to growth selections (Figure 5.1C&D) [206]. A total of 27 genes were targeted for hydrolysate tolerance, 14 genes for low pH tolerance, and 8 genes for acetate tolerance [206]. However, despite extensive growth selections, all variants isolated from the combinatorial libraries were either single or double mutants; furthermore, few of the double mutants had improved growth relative to single mutants, and some actually showed overall antagonistic interactions (i.e. the relative growth rates of double mutants were less than either of the corresponding single mutants) [206]. While this study

demonstrates the feasibility of integrating genome mapping and genome engineering methods for improving microbial phenotypes, it also illustrates inherent challenges that need to be addressed.

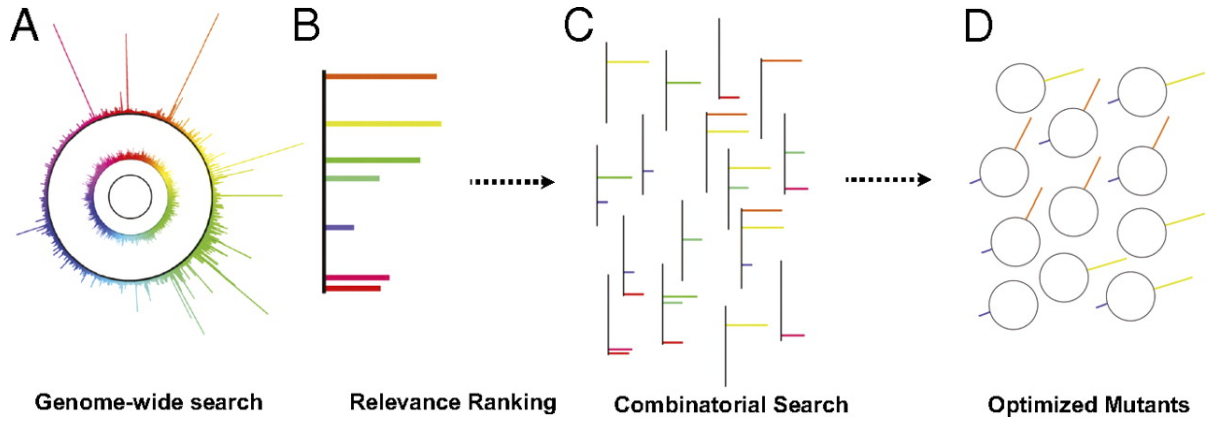

Figure 5.1: Overview of strategy used by Sandoval *et al*. Adapted from [206] **(A)** Trackable Multiplex Recombineering (TRMR; [56]) is used to map fitness effects of genetic loci. **(B)** Targets conferring high fitness are selected for further study. **(C)** Recursive multiplex recombineering is used to generate a combinatorial library of the selected targets. **(D)** Growth selection is performed on the library to yield clones with high tolerance.

One key issue with integrating genome mapping and genome engineering methods is that most genome-scale trait mapping techniques cannot capture interactions between distal genetic loci, thus limiting detection of potential synergistic combinations. Since most mapping techniques identify fitness effects of single genes, a common strategy is to identify single genes with high fitness effects and create combinatorial variants with top candidates, as exemplified by Sandoval *et al* [206]. From an intuitive standpoint this may seem like a good strategy, but recent evolutionary biology studies [207, 117] have reported widespread negative epistasis (i.e. less than additive effects) between beneficial mutations, thus reducing the potential utility of this approach. Prevalent negative epistasis between beneficial mutations was recently reported for *E. coli* evolved on glucose minimal media [117] and for an engineered *Methylobacterium extorquens* strain evolved with methanol as the sole carbon source [207, 117]. In our work on isobutanol tolerance (Chapter 4), we observed negative epistasis between beneficial mutations in *hfq* and *marC*, but there was evidence for positive epistasis for each of *hfq* or *marC* with subsequent mutations (*groL*, *mdh*, and *rph*). Collectively, these results suggest that negative epistasis between beneficial mutations may be a statistically common phenomena which needs to be considered when devising genome-scale phenotype improvement methods.

5.2.2 Overview

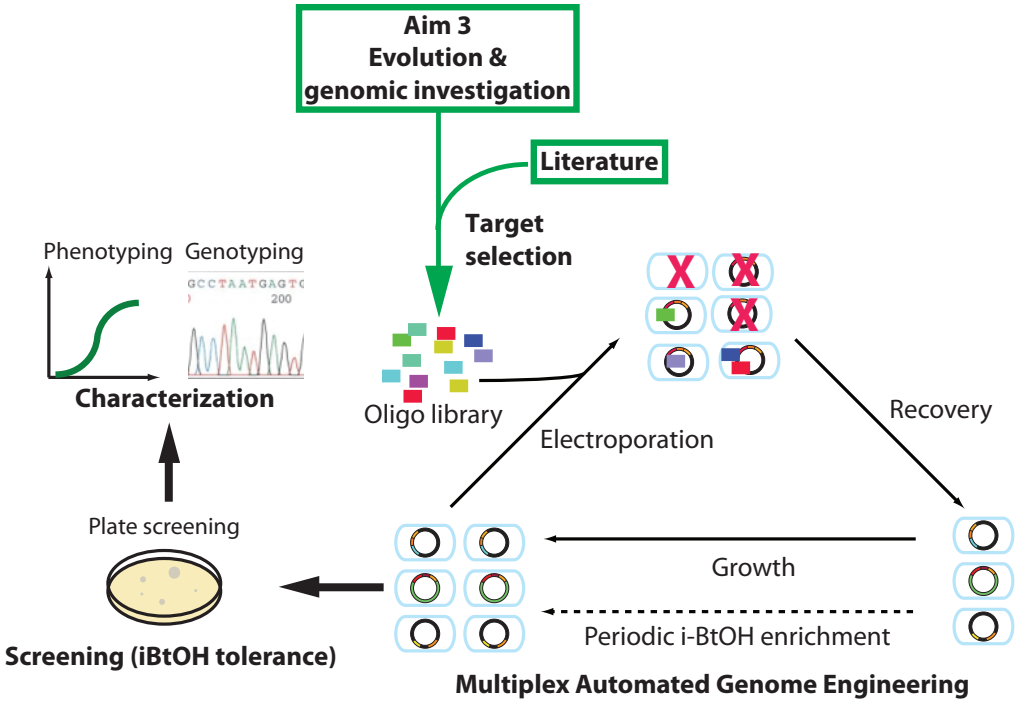

Figure 5.2: Genome engineering for further characterization and improvement of isobutanol tolerance.

Guided by our evolutionary-genomics studies (Chapter 4), in this work we devise and implement a genome engineering strategy to further investigate and improve isobutanol tolerance in *E. coli* (Figure 5.2). In collaboration with the labs of George Church (Harvard Medical School) and Harris Wang (Pathology and Cell Biology, Columbia University), we performed targeted mutagenesis of candidate genetic loci using Multiplex Automated Genome Engineering (MAGE) [59]. We utilize the following strategy: i) create mutagenic ssDNA oligo libraries, ii) perform MAGE, with periodic selection cycles to enrich highly tolerant genotypes, iii) isolate hypertolerant variants, and iv) characterize isolates for systematic genotype-phenotype correlation (Figure 5.2). A key feature of this work is that we employ a holistic and comprehensive approach for target gene selection, utilizing the results from our evolutionary-genomics studies (Chapter 4) as well as leveraging investigations of *E. coli* alcohol tolerance reported in literature. We suggest that through judicious selection of target genes, directed genome engineering can be used to successfully improve multigenic phenotypes with complex underlying epistatic interactions. In the next section, we provide background on the selection of target mutations in this study.

5.3 Results

5.3.1 Selection of target mutations

Due to the importance of microbial chemical tolerance in bioprocessing, there is a large body of research on the mechanisms and genetic bases of chemical toxicity and tolerance. Amongst biofuels, short-chain alcohols (i.e. ethanol, n-butanol, and isobutanol) are at the most advanced development stage and thus there is a disproportionate amount of research concerning the tolerance and toxicity of short chain alcohols in model organisms such as *S. cerevisiae* and *E. coli*. In June 2011, we conducted a literature and patent review of genetic studies of ethanol, n-butanol, and isobutanol tolerance in *E. coli*. Methods used in these studies included evolution/genome resequencing [124, 208] as well as screening of knockout/overexpression libraries for alcohol tolerant variants [209, 210, 121, 165, 211]. Since mechanisms of toxicity/tolerance for short chain alcohols are closely related, we expect that a subset of mutations conferring tolerance to ethanol or n-butanol may also provide some degree of cross-tolerance to isobutanol; thus we include these loci as potential genome engineering targets. From this review we compiled 103 genetic loci correlated with short-chain alcohol tolerance in *E. coli*, which combined with the 144 loci[1] identified in our evolutionary-genomics studies of isobutanol tolerance (Chapter 4) yields a total of 247 potential genome engineering targets for improving isobutanol tolerance in *E. coli*.

As shown in Figure 5.3, genetic loci correlated with alcohol tolerance are widely distributed across many different biochemical functions and genomic locations, consistent with the notion that tolerance to alcohols is a complex trait that involves a diversity of cellular adaptations and responses. Many different types of mutations were reported, including SNPs, indels, large-scale chromosomal rearrangements, transposon insertions, gene knockouts, and gene overexpression (i.e. cloning genomic fragments into inducible high-copy vectors). While there was a diversity of different types of mutations, it appears that many of the reported mutations simply lead to loss-of-function (i.e. indels leading to frameshifts, transposon insertions, and knockouts).

To maintain a tractable combinatorial search space and level of experimental complexity, we decided to select a reduced subset of target mutations from the 247 identified in our review. We selected targets based on the following criteria: i) extent to which alcohol tolerance can be attributed target (i.e. significance rating in Figure 5.3), ii) oligo recombination efficiency for the required genetic modification (i.e. using MAGE; [59]), and iii) likelihood of synergistic interactions between targets. Targets with demonstrated or known effects on alcohol tolerance

[1] This number is larger than the total of 131 given Chapter 4 since it includes mutations identified in our parallel evolution studies and X2.5 genome resequencing

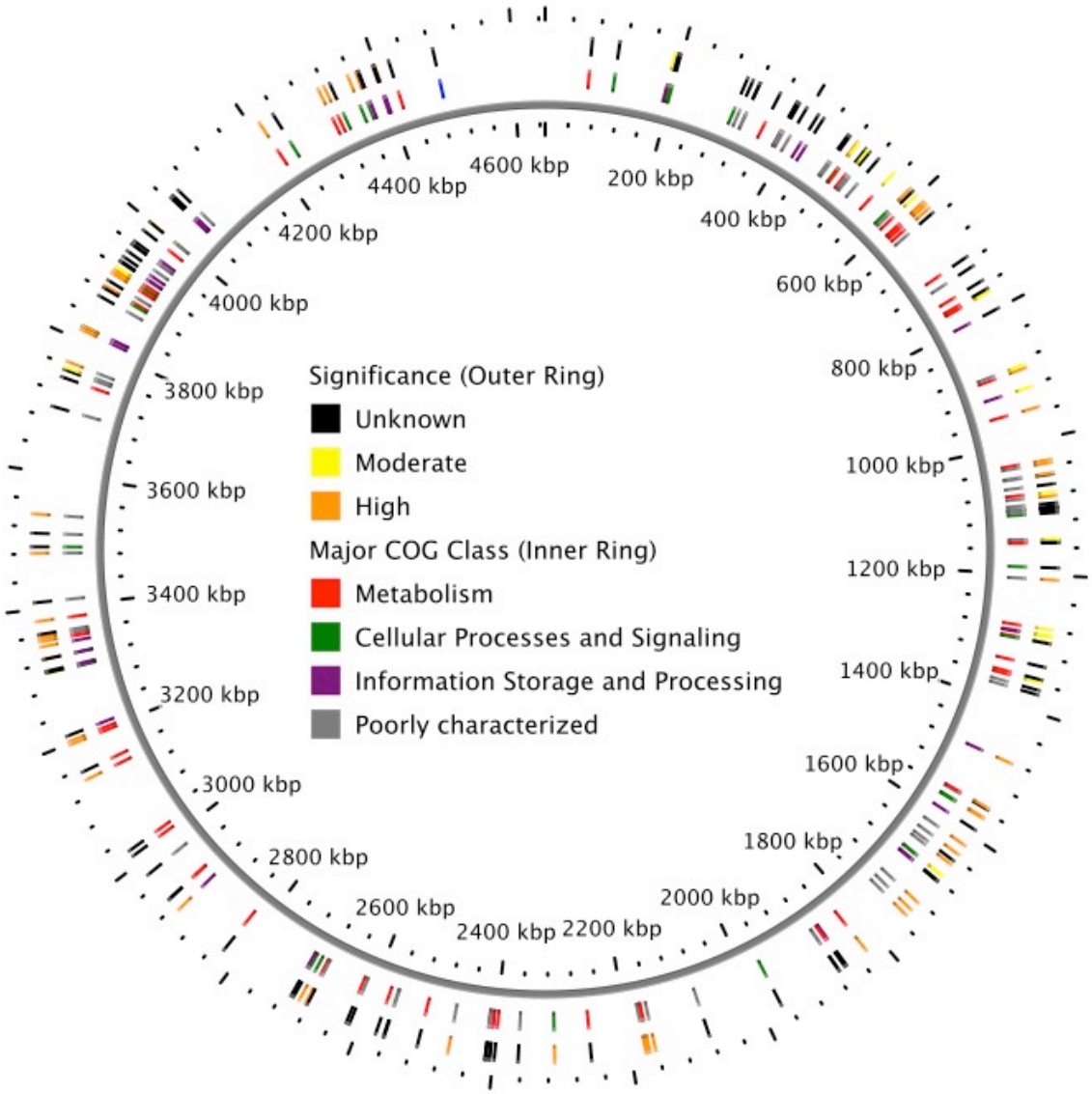

Figure 5.3: Chromosome map (*E. coli* EcNR1 reference sequence) of genetic loci correlated with short-chain (C2-C4) alcohol tolerance in *E. coli*. Phenotype significance and major COG (Clusters of Orthologous Groups) class for each locus is given according to color code in legends. "Significance" is a qualitative metric that captures the extent to which alcohol tolerance can be attributed to a particular locus. High significance indicates that there is direct experimental evidence linking alcohol tolerance to a locus; medium significance indicates that direct experimental evidence is lacking, but the locus has been identified in more than one study and/or is functionally related to known mechanisms of toxicity; and unknown significance indicates that the locus has an unknown role in alcohol tolerance. Data compiled from the following sources: [124], [121], [208], [209], [210], [165], and [211].

are obviously preferred, thus the rationale for the first criterion. For the second criterion, oligo recombination efficiency is correlated to the amount of homologous sequence between an oligonucelotide and the chromosomal target; point mutations such as SNPs and short indels can be produced at the highest efficiencies (>10%; [59]). Recombination efficiencies are much lower for large mutations, though specialized techniques for such mutations are currently under

development (such as co-selection MAGE; [61]). Therefore, we give preference to targets that comprise small point mutations, while larger mutations (such as promoter replacements) could be explored in future work. For our final criterion, it is quite difficult to predict epistasis between mutations a priori. However, general patterns of epistasis can be used to provide guiding principles for target selection. Negative epistasis between genetic loci frequently arises due to functional overlap. For deleterious mutations, if two genes can individually perform a common function, then each can partially compensate for the loss of the other; however, loss of both genes will completely eliminate the function, which would be lethal if the function is essential [154]. Similar effects may occur for beneficial mutations; for example, if two different beneficial mutations have the same functional effect, then the combined effect of both mutations may be less than additive due to saturation of the underlying biochemical process [117]. Thus we can minimize the likelihood of negative epistasis in our genome engineering work by avoiding targets with potential functional redundancy. For positive epistasis, patterns can be quite complex. For instance, positive epistasis is prevalent between metabolic genes with non-related functions [154]. On the other hand, positive epistasis is frequently observed between related or interacting functions; a common example occurs when mutations enhance the activity/function of a protein but also reduce structural stability, thus driving the evolution of compensatory mutations that improve stability of the mutant protein but have little effect in a WT background [212]. Thus we may be able to increase the likelihood of synergistic interactions in our genome engineering work by selecting functionally diverse targets, selecting targets with possible functional interactions (e.g. *lepB* and *secA*, components of the Sec protein translocation apparatus; Chapter 4), and most obviously by selecting targets with known positive interactions (i.e. *hfq* or *marC* with subsequent mutations *groL*, *mdh*, and *rph*; Chapter 4).

Based on the above considerations, we selected a subset of 38 target mutations for genome engineering (with a combinatorial search space size of $2^{38} \approx 2.75\text{x}10^{11}$ genotypes) for improving isobutanol tolerance in *E. coli*. Selected targets are given in Table 5.1, with a summary chromosome map shown in Figure 5.4. For each target, we designed a 90 bp mutagenic oligonucleotide for use in MAGE (see section 5.5.3 for more details). Approximately 2/3 of the selected targets were derived from our evolutionary-genomics studies (Chapter 4) while the remainder were obtained from various literature and patent sources (Table 5.1). Due to the difficulty in producing mutations $>$10bp with MAGE, we avoided targets that involve promoter replacement (i.e. gene overexpression). Most of the selected mutations are thus point mutations, indels, or gene deletions (produced with SNPs that introduce premature stop codons) (Table 5.1). All of the significant mutations identified in our evolutionary-genomics studies (Chapter 4) are included, as well as a number of uncharacterized mutations (Table 5.1). Overall, our library includes a large number

targets with confirmed effects on alcohol tolerance, and represents a diversity of different cellular processes (Table 5.1). Having selected our target mutations, in the next section we discuss considerations in the design of our genome engineering strategy.

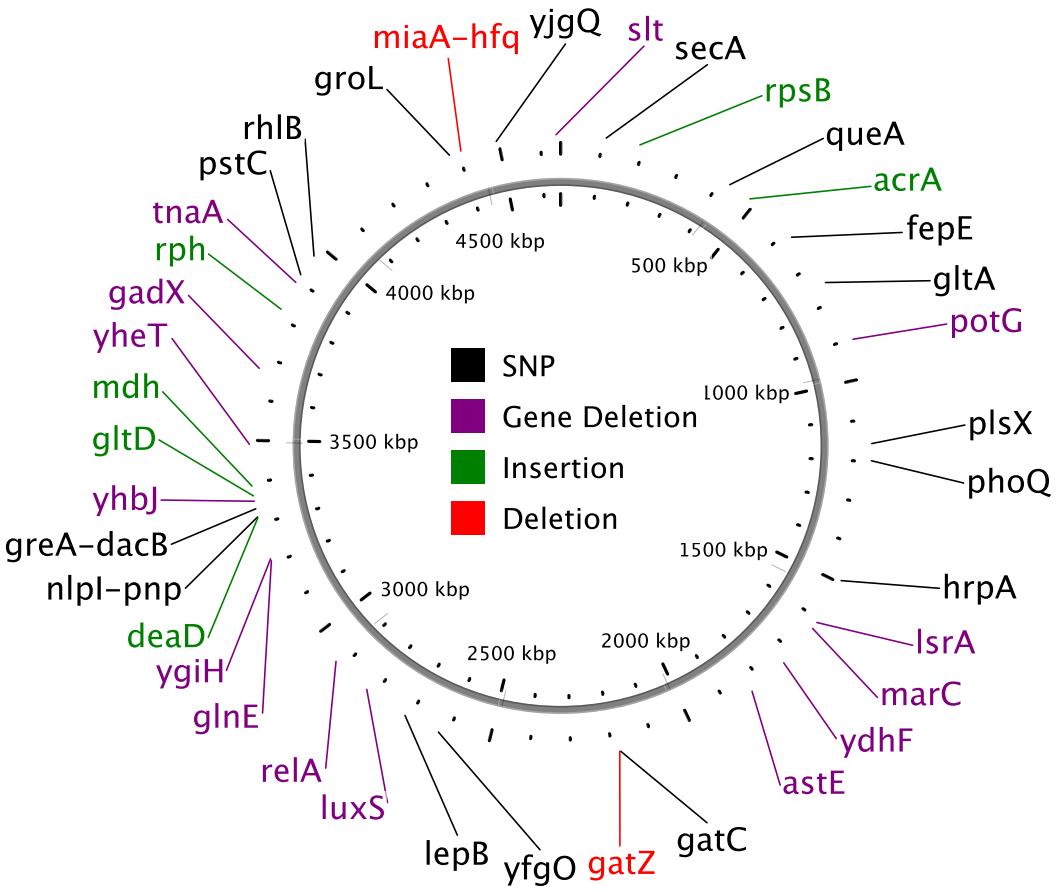

Figure 5.4: Chromosome map (*E. coli* EcNR1 reference sequence) of target loci selected for combinatorial genome engineering with MAGE. Mutation types given according to color code in legend. See Table 5.1 for more details.

Table 5.1: Target mutations selected for combinatorial genome engineering. Mutation positions are given as absolute genomic coordinates in the *E. coli* EcNR1 reference sequence. SNPs are indicated by base transition/transversion. Small insertions are indicated by a "+", with the size (number of bp) of the insertion and sequence of inserted bases. Small deletions are designated by "−" with a format similar to that for small insertions. "DELETION" indicates gene deletion and/or loss-of-function.

Locus	Position	Mutation	Locus description	Source
fepE	618531	T→C	Regulates length of O-antigen; component of the ferric enterobactin transport system	This work (Chapter 4); G3 lineage
gatZ	2182915	-1:C	D-tagatose 1,6-bisphosphate aldolase 2, subunit (EC:4.1.2.40)	This work (Chapter 4); G3 lineage
gltD	3367270	1:G	Glutamate synthase, 4Fe-4S protein, small subunit (EC:1.4.1.13)	This work (Chapter 4); G3 lineage
glnE	3206870	DELETION	Fused deadenylyltransferase/adenylyltransferase for glutamine	This work (Chapter 4); G3 lineage
greA-dacB	3336132	A→G	Promoter region of D-alanyl-D-alanine endopeptidase *dacB*	This work (Chapter 4); G3 lineage
groL	4378650	A→C	Cpn60 chaperonin GroEL, large subunit of GroESL	This work (Chapter 4); G3 lineage
lepB	2711902	G→A	Leader peptidase (signal peptidase I) (EC:3.4.21.89)	This work (Chapter 4); G3 lineage
miaA-hfq	4407505	-7:AGGAAAA	Ribosome binding site of pleiotroic regulator *hfq*, involved with many RNA-based regulators	This work (Chapter 4); G3 lineage
nlpI-pnp	3316213	T→C	Lipoprotein involved in osmotic sensitivity and filamentation	This work (Chapter 4); G3 lineage
phoQ	1197581	A→G	Sensory histidine kinase in two-component regulatory system with PhoP	This work (Chapter 4); G3 lineage
pstC	3917582	T→C	Phosphate transporter subunit	This work (Chapter 4); G3 lineage
queA	425270	A→G	S-adenosylmethionine:tRNA ribosyltransferase-isomerase (EC:5.-.-.-)	This work (Chapter 4); G3 lineage
rhlB	3972054	A→G	ATP-dependent RNA helicase	This work (Chapter 4); G3 lineage
secA	108975	T→C	Preprotein translocase subunit, ATPase	This work (Chapter 4); G3 lineage
yjgQ	4494710	T→C	Lipopolysaccharide export ABC permease of the LptBFGC export complex	This work (Chapter 4); G3 lineage
mdh	3390936	5:AACCT	Malate dehydrogenase, NAD(P)-binding (EC:1.1.1.37)	This work (Chapter 4); X3 lineage
gatC	2180640	C→T	Galactitol-specific enzyme IIC component of PTS	This work (Chapter 4); X3 lineage
gltA	753345	T→C	Citrate synthase (EC:2.3.3.1)	This work (Chapter 4); X2 lineage
rph	3823229	4:GGTC	Defective ribonuclease PH	This work (Chapter 4); G3 lineage
acrA	483735	1:A	Multidrug efflux system	This work (Chapter 4); X3 lineage
deaD	3314027	4:AGAC	ATP-dependent RNA helicase	This work (Chapter 4); X3 lineage
hrpA	1493514	C→T	Predicted ATP-dependent helicase	This work (Chapter 4); X3 lineage
plsX	1156697	A→G	Probable phosphate acyltransferase	This work (Chapter 4); X3 lineage
rpsB	190557	1:A	30S ribosomal subunit protein S2	This work (Chapter 4); X3 lineage

Table 5.1: Target mutations selected for combinatorial genome engineering.

Locus	Position	Mutation	Locus description	Source
yfgO	2623022	C→T	Predicted inner membrane permease, UPF0118 family	This work (Chapter 4); X3 lineage
astE	1834154	DELETION	Succinylglutamate desuccinylase	n-butanol tolerance study; [209]
ygiH	3211923	DELETION	Conserved inner membrane protein	n-butanol tolerance study; [209]
yhbJ	3354344	DELETION	Predicted P-loop containing ATPase; involved in *glmS* regulation	Isobutanol tolerance study [124]
tnaA	3895960	DELETION	L-cysteine desulfhydrase / tryptophanase	Isobutanol tolerance study [124]
marC	1626139	DELETION	Conserved protein; predicted transporter	Common mutation in isobutanol tolerant strains; Chapter 4, [208], [124]
ydhF	1732863	DELETION	Predicted oxidoreductase	n-butanol tolerance study; [210]
potG	903421	DELETION	PotFGHI ATP-dependent putrescine transporter	n-butanol tolerance study; [210]
yheT	3490431	DELETION	Predicted hydrolase, member of the UPF0017 protein family	n-butanol tolerance study; [210]
gadX	3673040	DELETION	GadX DNA-binding transcriptional dual regulator; controls the transcription of pH-inducible genes, including the principal acid resistance system	ethanol tolerance study; [121]
slt	4637963	DELETION	Soluble lytic murein transglycosylase	ethanol tolerance study; [121]
relA	2920880	DELETION	GDP pyrophosphokinase / GTP pyrophosphokinase; ppGpp synthetase	n-butanol tolerance study; [211]
luxS	2821962	DELETION	S-ribosylhomocysteine lyase; biosynthesis of quorum sensing signal AI-2	Reported to increase isobutanol production; [213]
lsrA	1608721	DELETION	Predicted ATP-binding component of the AI-2 ABC transporter	Reported to increase isobutanol production; [213]

5.3.2 Design considerations for combinatorial genome engineering

The basic genome engineering strategy that we devised entails performing repeated cycles of MAGE using an oligo library targeting the above described mutations (Table 5.1), with periodic selection steps (passaging populations on isobutanol spiked media) to enrich genotypes with improved fitness. Important cycling parameters include total number of cycles, frequency and duration (i.e. growth time) of selection steps, and strength of selection (i.e. isobutanol concentration). To gain insights into how cycling parameters affect population dynamics and genotype distributions, we constructed a simple model of our genome engineering strategy. For each cycle of multiplex oligonucleotide recombination, mutations are distributed binomially; thus assuming equal probability (i.e. recombination efficiency) for each mutation, the frequency of cells with recombination events at one locus, two loci, ... to n loci can be predicted with the binomial probability

density function [59][2]. For repeated cycles of MAGE, we can then iteratively model mutation distributions as follows:

$$f_{j,k} = f_{j,k-1} + \sum_{i=0}^{j-1} f_{i,k-1} B(n, j-i, p) - \sum_{i=1}^{n-j} f_{j,k-1} B(n, i, p) \tag{5.1}$$

$$B(n, x, p) = \binom{n}{x} p^x (1-p)^{n-x} \tag{5.2}$$

Where $f_{j,k}$ is the frequency of genotypes with j loci modified after k cycles, p is the recombination probability per locus, n is the total number of targeted loci, and $B(x, y, z)$ is the probability density function for a binomial distribution. The first summation term in Equation 5.1 represents generation of genotypes with j modified loci due to mutations added to genotypes with $< j$ modified loci, while the second summation term represents transitions from genotype j to higher-order mutants.

Since we do not know the relative fitness of different genetic variants a priori, we cannot precisely predict the effect of enrichment steps on population dynamics. To qualitatively explore the effect enrichment cycles on genotype distributions and population dynamics, we can assume a simple of model of relative fitness as a function of number of genomic loci modified. As discussed previously, recent studies suggest prevalent negative epistasis between beneficial mutations in evolving bacterial populations [207, 117]. In contrast, we observe both positive and negative epistasis between mutations acquired in the evolution of isobutanol tolerance in *E. coli* (Chapter 4). As a conservative estimate, we will assume a predominance of negative epistatic effects between mutations in our library, leading to eventual saturation of relative fitness. This can be represented by a simple phenomenological model:

$$R_{f,j} = \frac{\mu_{max,MUT}}{\mu_{max,WT}} = R_{f,max} \left(\frac{j}{j + K_m} \right) + 1 \tag{5.3}$$

Where $R_{f,j}$ is the relative fitness of genotypes with j loci modified, $R_{f,max}$ is the maximum relative fitness, and K_m is a pseudo affinity. Parameter values will vary with selective pressure; in the absence of selection, $R_{f,j} = 0$ for all loci. Assuming exponential growth, population dynamics during selection can be modeled as:

[2]If probabilities of each mutation are not equal, then predicting genotype distributions requires Monte Carlo simulations

$$f_{j,k} = \frac{f_{j,k-1} e^{\left(R_{f,j}-1\right)t_*}}{f_{0,k-1} + \sum_{i=1}^{n} f_{i,k-1} e^{\left(R_{f,i}-1\right)t_*}} \tag{5.4}$$

Where $t_* = \mu_{max,WT} t$ is dimensionless time, and other terms are as described previously.

Parameter	Value	Notes
p	0.0066	Assuming 25% overall recombination efficiency (divided by 38 loci)
n	38	Total number of targeted loci
$R_{f,max}$	≈ 5	Maximum relative fitness at (0.5% to 1% w/v isobutanol selection); estimated from results in Chapter 4
K_m	≈ 10	Order of magnitude estimate

Table 5.2: Parameters used in genome engineering model.

Using the parameter values given in Table 5.2, we modeled mutation distributions and overall population fitness for a total of 50 MAGE cycles, without and with selection steps every 10th cycle (Figure 5.5). Even with a conservative fitness model, selection steps have a dramatic effect on the distribution of genetic variants, increasing frequencies of high fitness genotypes by multiple orders of magnitude within a single enrichment step (Figure 5.5A vs. C). Thus selection steps may decrease the total number of cycles required to reach a given genotype frequency, reducing labor and accumulation of off-target mutations. While we could further increase the frequency, duration, and intensity of selection steps, our model suggests that our current scheme of isobutanol selection (0.5 to 1% w/v) every 10th cycle should yield highly modified variants (10 to 20 mutations) at selectable frequencies ($> 10^{-7}$) within a manageable number of cycles (20 to 30 total).

In addition to the cycling parameters discussed above, there are other factors that may affect mutation distributions and population dynamics. In general, epistatic effects and strength of selection could have major impacts on genotype distributions. Epistatic effects can dramatically alter fitness landscapes, especially in the case of global regulators (i.e. *hfq*) [214]. Thus while we wish to generate synergistic mutation combinations, if epistatic mutations become fixed in populations at early MAGE cycles this could constrain accessible evolutionary trajectories and limit exploration of combinatorial search spaces. For selection, increasing intensity favors enrichment of highly fit genotypes while purging less fit variants; however, if selective pressure is too strong, population diversity will be drastically reduced, again limiting search space exploration. Beyond the effects of epistasis / selection, there are additional caveats that must be considered. In prior

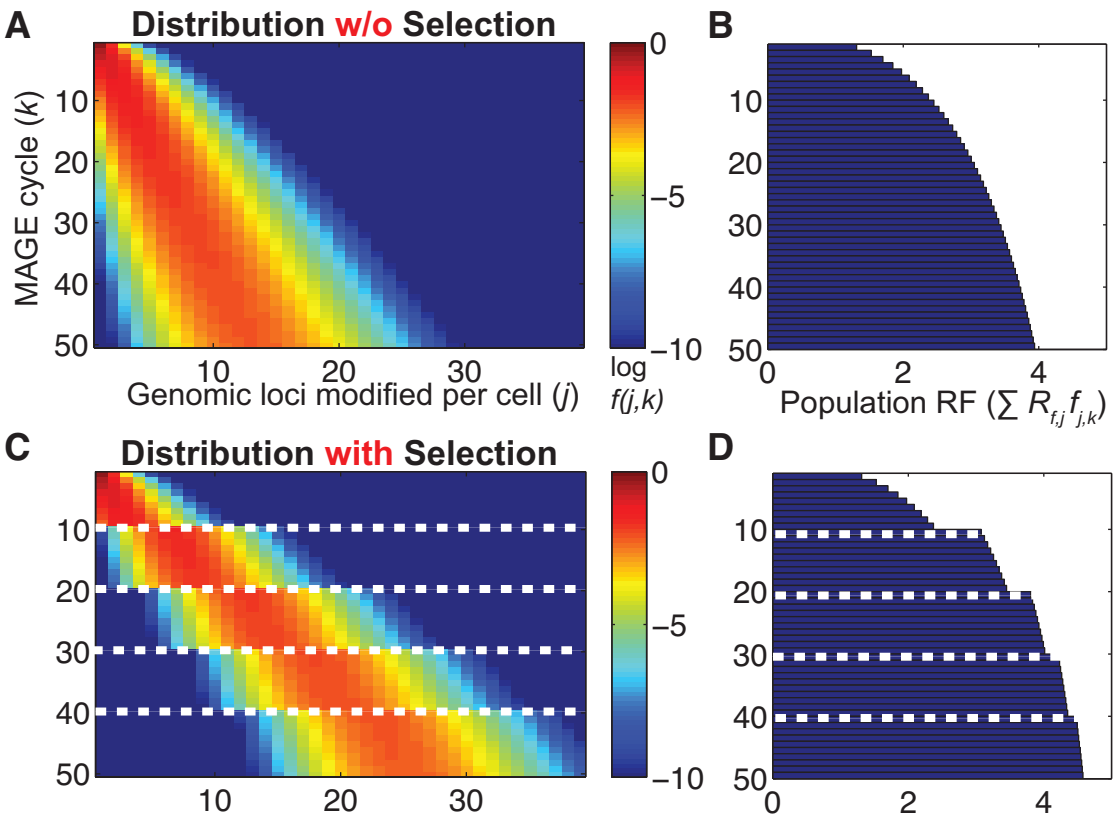

Figure 5.5: Theoretical modeling of genome engineering strategies, with parameter values as described in Table 5.2. **(A)** Mutation distribution as a function of MAGE cycle number with no selection steps (i.e. $t_* = 0$). **(B)** Overall population relative fitness as a function of MAGE cycle number with no selection steps. **(C)** Mutation distribution with selection every 10th cycle; total growth time of $t_* = 4.8$ per selection. **(D)** Overall population relative fitness with selection every 10th cycle; $t_* = 4.8$ per selection.

work (Chapter4) we have established that i) certain mutations reduce subsequent recombination efficiency by ≈100x or more and ii) certain mutations may exhibit antagonistic pleiotropy between rich and minimal media, causing negative selection during MAGE (since rich media is required for optimal recombination efficiency). We have specifically observed both effects (recombination efficiency / antagonistic pleiotropy) with mutation *miaA-hfq* 4407505 -7:AGGAAAA. Beyond *hfq*, it seems likely that mutations which alter cell envelope properties could affect electroporation efficiency (and thus overall recombination efficiency), and such cell envelope targeting mutations are prevalent in our library (since the cell envelope is a primary target of isobutanol toxicity).

5.3.3 Multiplex genome engineering experiments and phenotypic characterization of populations

Based on the above considerations, we decided to test several different MAGE schemes. In our first series of experiments, we performed MAGE cycles with a full (38x target) oligonucleotide library ("38*" lineages; see Figure 5.6 for details). In a second set of experiments, we performed MAGE with reduced dosage of the *miaA-hfq* 4407505 -7:AGGAAAA oligonucleodtide ("37" and "38" lineages; see Figure 5.7 for details). The *miaA-hfq* 4407505 -7:AGGAAAA mutation has confirmed epistasis with several other loci (*groL*, *mdh*, and *rph*) and potentially many more uncharacterized interactions. Additionally, *hfq* causes reduced recombination efficiency / antagonistic pleiotropy in rich media, as discussed above. Due to these issues, fixation of *hfq* in MAGE populations may constrain accessible evolutionary trajectories and limit evolvability, thus motivating us to examine the effects of reduced *hfq* oligonucleotide dosage (Figure 5.6).

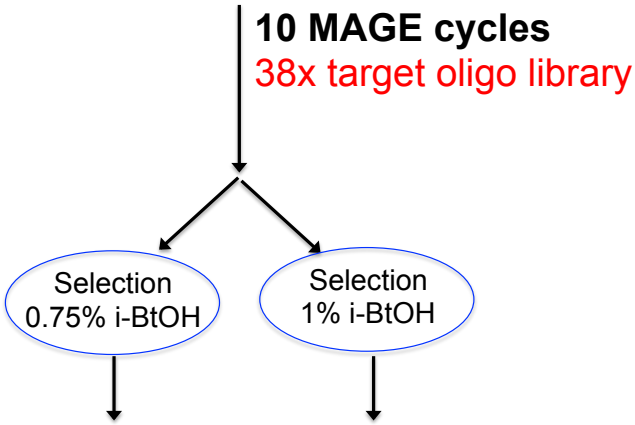

Figure 5.6: Genome engineering strategy: full (38x target) library. 10 cycles of MAGE are performed with full oligonucleotide library. After 10 cycles, population is split into two fractions subject to selection on NG50 minimal media spiked with 0.75% and 1% (w/v) isobutanol, respectively. Lineages are referred to as "38* / 0.75% sel" and '38* / 1% sel", respectively.

MAGE experiments were performed according to the above described cycling schemes in collaboration with the labs of George Church (Harvard Medical School) and Harris Wang (Pathology and Cell Biology, Columbia University), using *E. coli* EcHW24 (K12 λ-Red ΔmutS) *intC::yfp* as a host strain. We performed a total of 10 MAGE cycles for the 38* lineages (full oligonucelotide library; see Figure 5.6 for details), 20 cycles for the 38 lineage (reduced *hfq*; see Figure 5.7A & B for details), and 30 cycles for the 37 lineage (reduced *hfq*; see Figure 5.7A & B for details), with selection steps after every 10th cycle (see Figure 5.6, and Figure 5.7B & C for details). For selection steps, cell populations were washed and diluted 1:10 to 1:100 in NG50 minimal media supplemented with isobutanol (concentrations indicated in Figure 5.6 & 5.7) and

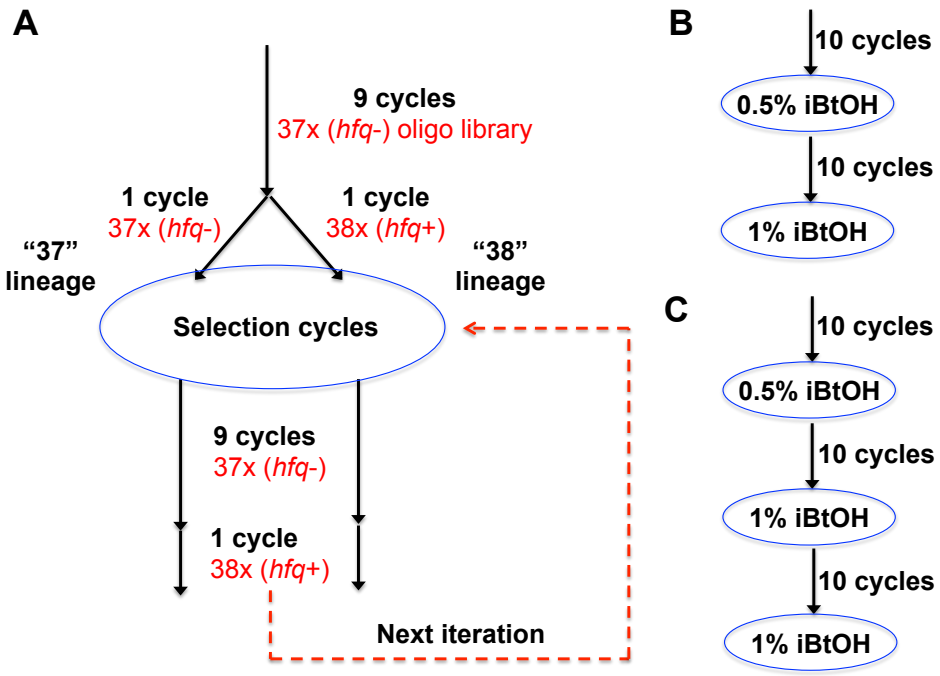

Figure 5.7: Genome engineering strategy: reduced *hfq* dosage. **(A)** 9 cycles of MAGE with reduced (37x target; *hfq* omitted) oligonucleotide library. After 10th cycle, population is split into two fractions, one of which is subject to recombination with the reduced (37x target / *hfq*-) library (termed "37" lineage) and the other with the full (38x target / *hfq*+) library (termed "38 lineage"), respectively. Populations are subject to selection cycles at various isobutanol concentrations. MAGE is continued using a scheme of 9 cycles with reduced (*hfq*-) library and 1 cycle with full (+*hfq*) library per 10 cycles. Thus the main difference between the 37 and 38 lineages is that the 37 lineage received 19 MAGE cycles before first *hfq* dose vs. 9 cycles for the 38 lineage. **(B)** Selection steps used in 38 lineage. 20 cycles total. **(C)** Selection steps used in 37 lineage. 30 cycles total.

then grown to saturation. We use a notation of lineage-cycle # to refer to MAGE populations (e.g. "37-20" corresponds to lineage 37, cycle 20 population).

As an initial characterization, we examined isobutanol tolerance phenotypes of whole population samples from each MAGE lineage by measuring growth rates (μ_{max}; 1/h) and maximum OD_{600} in NG50 media with 1% (w/v) isobutanol (Figure 5.8). We calculated relative fitness (RF; mutant phenotype divided by parental *E. coli* EcHW24 *intC::yfp* phenotype) in terms of both μ_{max} (RF_{μ}; Figure 5.8A) and maximum OD_{600} (RF_{OD}; Figure 5.8B). There was little variation in maximum OD_{600} between the different populations, with the exception of 37-30 (Figure 5.8B); RF_{μ} values varied more between populations and number of cycles. After 10 MAGE cycles and a selection step, RF_{μ} for all lineages was increased substantially, indicating that isobutanol tolerant variants were present at high frequency in all populations (Figure 5.8A & B). As expected, population RF_{μ} tended to increase with MAGE cycle number, with the exception of population 37-20 (Figure 5.8A). Overall, the 37-30 had the highest fitness at 1% (w/v) isobutanol out of the populations examined, with $RF_{\mu} = 3.1 \pm 0.1$ and $RF_{OD} = 1.48 \pm 0.06$. Interestingly, the 38*-10

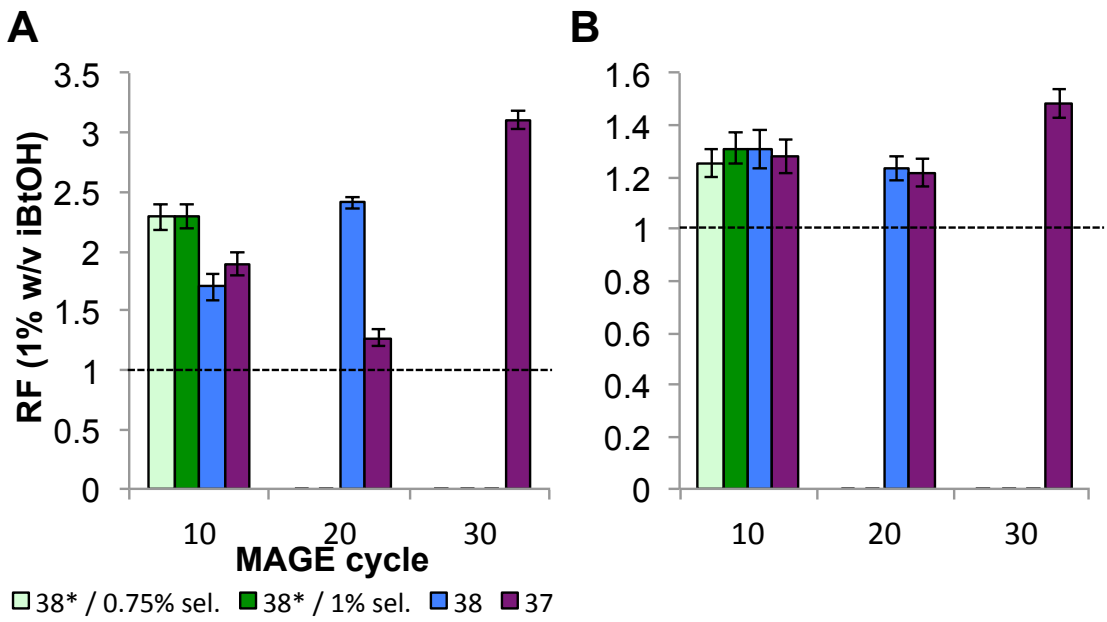

Figure 5.8: Population relative fitness (RF) after 10, 20, and 30 MAGE cycles (post isobutanol selection step). Lineages are noted in the legend, and correspond to those described in Figure 5.6 & 5.7. **(A)** Relative fitness at 1% (w/v) isobutanol, calculated from μ_{max} values: $\text{RF}_{\mu} = \mu_{max,MUT}/\mu_{max,WT}$. **(B)** Relative fitness at 1% (w/v) isobutanol, calculated from maximum OD_{600} values: $\text{RF}_{OD} = OD_{600,MUT}/OD_{600,WT}$.

populations (full target library) populations had higher RF_{μ} values compared to 37-10 and 38-10 (reduced *hfq* dosage). However, 38*-10 was subjected to stronger selection (0.75% or 1% w/v isobutanol; Figure 5.6) compared to 37-10 and 38-10 (0.5% w/v isobutanol; Figure 5.7B & C), so we cannot ascertain whether these fitness differences are simply due to selection effects or arise from genotype distribution differences between different *hfq* dosing strategies.

To further examine the effects of selection intensity, we subjected MAGE populations to additional selections at various isobutanol concentrations and then measured population RF_{μ} in 1% (w/v) isobutanol NG50 (Figure 5.9). Intuitively we would expect that population RF_{μ} should increase with selection pressure. However, several populations show a trend wherein population RF_{μ} increases with isobutanol concentration up to a certain point, and then declines (Figure 5.9B & C). Surprisingly, in some cases population RF_{μ} after high selection pressure (1% to 1.5% w/v isobutanol) is actually lower than pre-selection RF_{μ} (e.g. 38-10, 37-10, and 37-20; Figure 5.9B & C). There are various possible explanations for this unusual trend. At high isobutanol concentrations ($> 1\%$ w/v) growth saturates at a relatively low OD_{600}, and so reduction in population doublings may correspondingly reduce enrichment of tolerant genotypes. In extreme cases, if isobutanol levels exceed the minimum inhibitory concentration (MIC) of all variants present in the population, then selection changes to favor retention of viability as opposed to maximum growth rate. Since there is often a tradeoff between survival and growth rate (e.g. allocating resources to re-

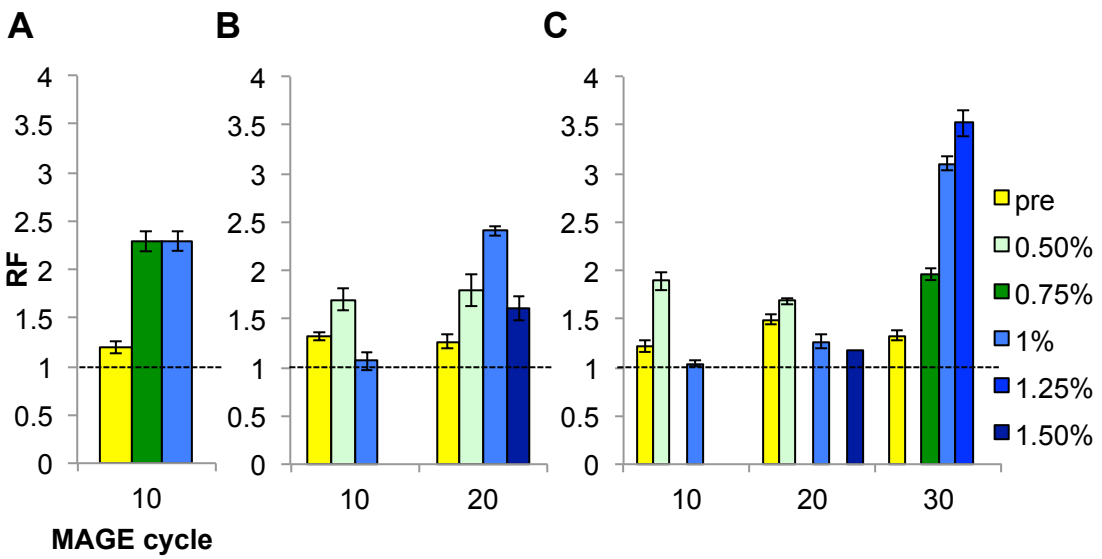

Figure 5.9: Population RF at 1% (w/v) isobutanol (RF = $RF_\mu = \mu_{max,MUT}/\mu_{max,WT}$) after selection at indicated isobutanol concentrations. Legend is shown to right; "pre" indicates pre-selection populations. **(A)** 38* lineage. **(B)** 38 lineage. **(C)** 37 lineage.

production and growth vs. cellular maintenance and repair) [215], variants with improved survival at high isobutanol concentrations may actually have reduced growth rates at non-inhibitory levels. Additionally, at isobutanol concentrations high enough to reduce cell viability, population bottlenecking effects could occur. Overall, the 37-30 population selected at 1.25% (w/v) isobutanol had the highest fitness out of the populations examined, with $RF_\mu = 3.5 \pm 0.1$ and $RF_{OD} = 1.53 \pm 0.07$.

5.3.4 Phenotype/genotype characterization of highly tolerant isolates

For each of the populations depicted in Figures 5.6 & 5.7, we screened for isolates with high isobutanol tolerance by plating population samples (10^4 to 10^5 cells) on NG50 agar spiked with 1% (w/v) isobutanol[3]. We selected a total of 10 colonies from each of the 38*-10 populations (0.75% and 1% isobutanol selection cycles; Figure 5.6), 37-10 and 38-10 populations (0.5% w/v isobutanol selection cycles; Figure 5.7), and 37-20 and 38-20 populations (1% w/v isobutanol selection cycles; Figure 5.7), and 15 colonies from the 37-30 population (1% w/v isobutanol selection cycle; Figure 5.7). Clonal isolates were genotyped at each targeted locus using PCR based assays, and phenotyped by measuring μ_{max} and maximum OD_{600} in 1% (w/v) isobutanol NG50 media. We use a notation of lineage-cycle-colony # to refer to isolates (e.g. "37-20-10" corresponds to lineage 37, cycle 20 population, colony # 10).

[3]Isobutanol toxicity varies between liquid and agar media; 1% (w/v) isobutanol is completely inhibitory to the parental *E. coli* EcHW24 *intC::yfp* on NG50 agar media.

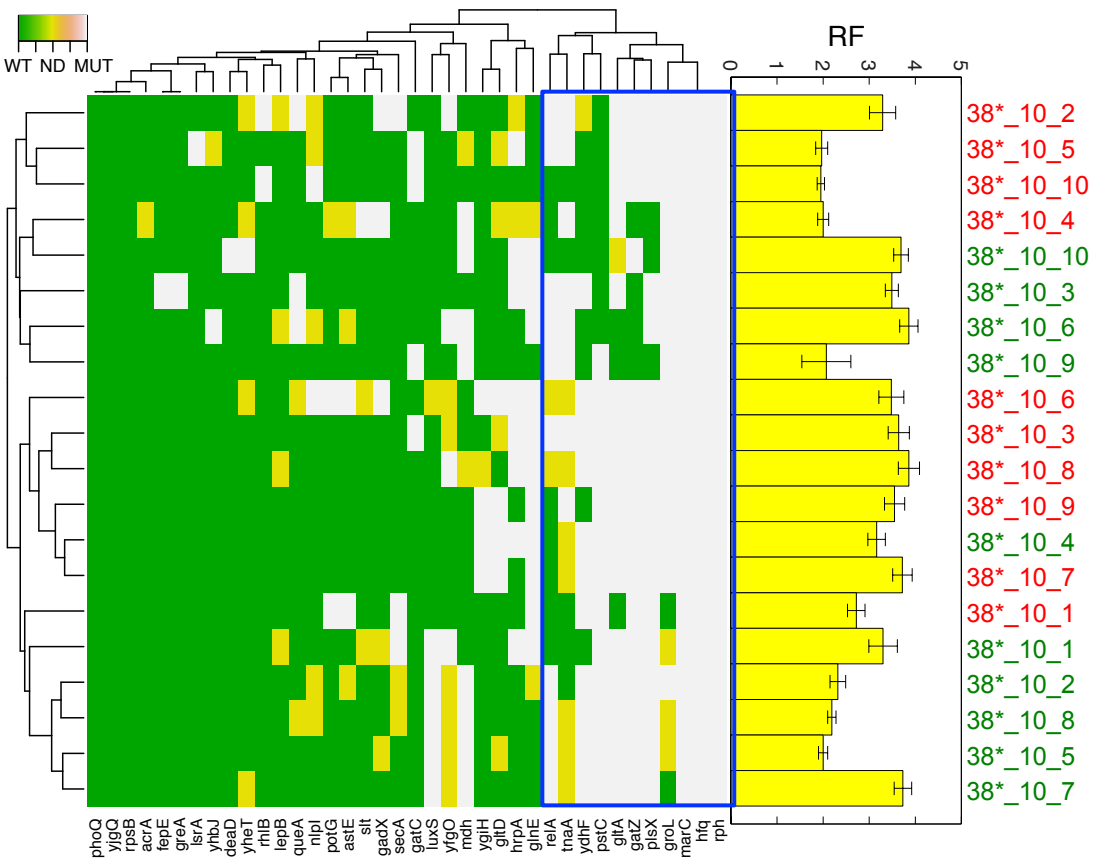

Figure 5.10: Genotype/phenotype map for clonal isolates from 38* lineages. Genotypes for each clonal isolate ($N = 20$ total) are shown with hierarchical clustering (Ward's method; Euclidean distance). Genotypes are coded according to legend in upper left corner (green, wild type; yellow, not determined; grey, mutant). Relative fitness ($\text{RF} = \text{RF}_\mu = \mu_{max,MUT}/\mu_{max,WT}$) was measured in 1% (w/v) isobutanol NG50 media. Clone labels (right hand side) are color coded by isobutanol selection level (red, 1% w/v isobutanol; green, 0.75%).

A genotype/phenotype map for isolates from the 38* lineage is shown in Figure 5.10. Number of mutations per isolate ranged from 9 to 17 mutations total (with a mean of 13), compared to a statistically expected mean of 2.4 mutations per isolate after 10 cycles of MAGE without selection (see Figure 5.12A); this represents an $\approx 10^5$ fold increase over expected frequency. All isolates had RF_μ values between 1.95 and 3.86 (Figure 5.10). While we observe a diversity of phenotypes and genotypes in the 38* populations, hierarchical clustering reveals that mutations are particularly prevalent for a group of loci including *rph*, *hfq*, *marC*, *groL*, *plsX*, *gatZ*, *gltA*, *pstC*, *ydhF*, *tnaA*, and *relA* (area boxed in blue in Figure 5.10). Most notably, all genotyped clones have *hfq*, *rph*, and *marC* mutations (i.e. 100% allele frequency). We characterized phenotypic effects and epistasis for *hfq*, *marC*, *rph*, and *groL* mutations from the G3 lineage in our evolutionary-genomics work (Chapter 4); given the known importance of these mutations, the observed high frequencies are not surprising. Interestingly, in our previous work *mdh* was also shown to be epistatic with *hfq* /

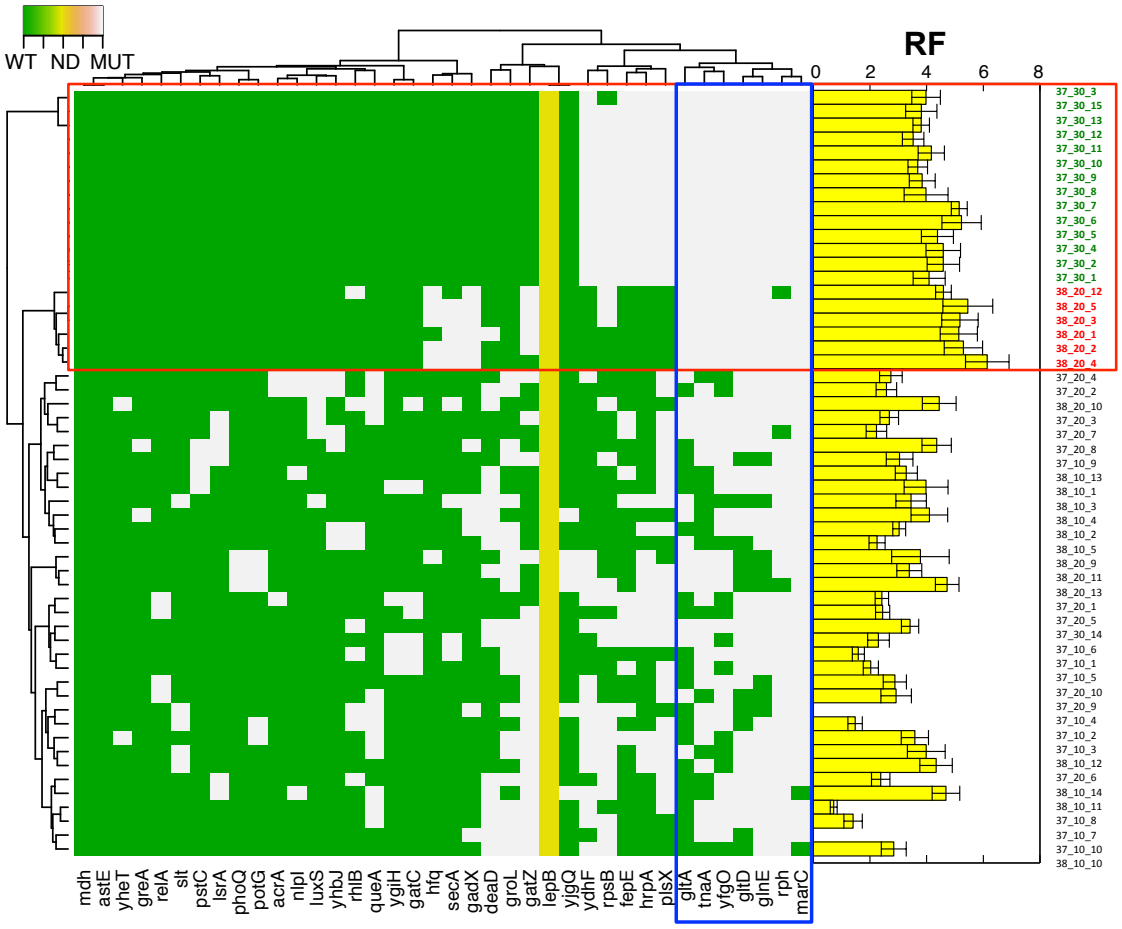

Figure 5.11: Genotype/phenotype map for clonal isolates from 37 and 38 lineages. Genotypes for each clonal isolate ($N = 55$ isolates total) are shown with hierarchical clustering (Ward's method; Euclidean distance). Genotypes are coded according to legend in upper left corner (green, wild type; yellow, not determined; grey, mutant). Relative fitness ($\text{RF} = \text{RF}_\mu = \mu_{max,MUT}/\mu_{max,WT}$) was measured in 1% (w/v) isobutanol NG50 media. The boxed regions indicate clusters of isolates with high RF and shared mutations; isolates with green labels (right hand side) were derived from population 37-30, with red labels denoting 38-20.

marC (Chapter 4), but is less prevalent (40% frequency).

In addition to 38*, we also characterized clonal isolates from the 37 and 38 lineages (Figure 5.11). Number of mutations per isolate ranged from 7 to 16 mutations total, comparable to the range observed for the 38* lineage. Interestingly, there was little variation in the number of mutations between different lineages or different numbers of MAGE cycles (Figure 5.12), suggesting that fitness saturates within the observed mutation ranges. Isolates from the aggregate 37/38 set had a broad range of RF_μ values, ranging between 0.7 and 6.1, and a diversity of genotypes (Figure 5.11). In contrast to other populations, isolates from 37-30 were nearly isogenic (generally containing *marC*, *rph*, *glnE*, *gltD*, *yfgO*, *tnaA*, *gltA*, *plsX*, *hrpA*, *fepE*, *rpsB*, and *ydhF* mutations). Phenotypes of 37-30 isolates were less uniform, possibly due to accumulation of

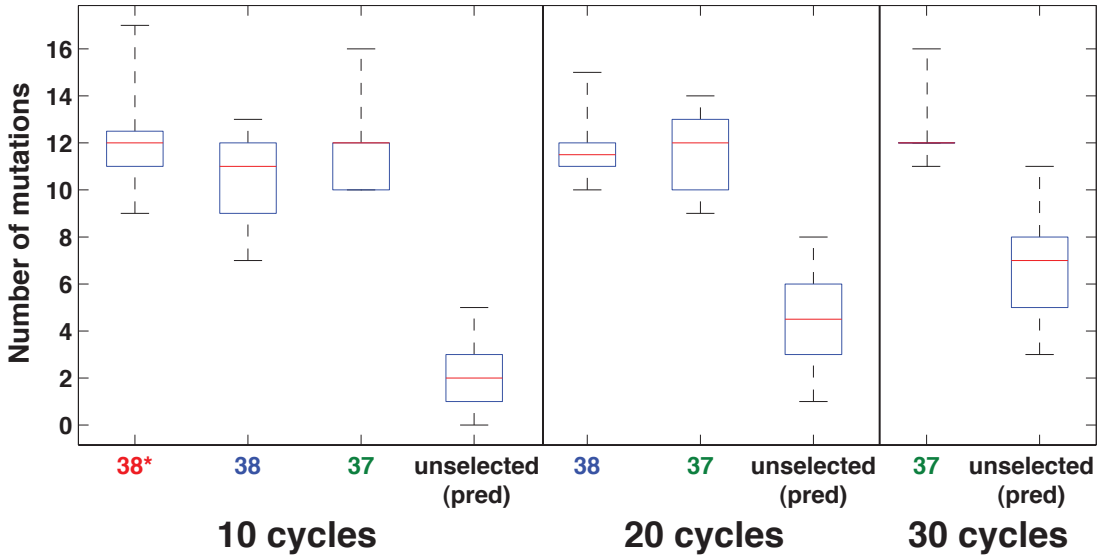

Figure 5.12: Distribution of genetic variants for different MAGE lineages / cycle numbers. Central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points. The series "unselected (pred)" corresponds to predicted distribution (based on model (Equation 5.1) with $p = 0.0066$ and $n = 38$) without selection (i.e. $R_{f,j} = 0$).

off-target mutations (Figure 5.11). Overall 37-30 had the highest population RF_{μ} (Figure 5.8A), suggesting that the observed genotype is present at high frequency in the population. Hierarchical clustering reveals two groups of clones (corresponding to isolates from 37-30 and 38-20) with high RF_{μ} that share a set of common mutations including *gltA*, *tnaA*, *yfgO*, *gltD*, *glnE*, *rph*, and *marC*, suggesting that these mutations may be important determinants of high isobutanol tolerance. However, we note that several isolates with markedly different genotypes also had high fitness. Out of all isolates characterized, 38-20-4 had the highest isobutanol tolerance, with $RF_{\mu} = 6.1 \pm 0.8$ and mutations in *hfq*, *secA*, *gadX*, *gltA*, *tnaA*, *yfgO*, *gltD*, *glnE*, *rph*, and *marC*.

5.3.5 Cross-lineage comparison of phenotypes and genotypes

Comparison of phenotypes and genotypes between the 38* and 37/38 lineages reveals several key patterns. In terms of phenotypes, 37/38 lineages showed a much greater range of RF_{μ} and RF_{OD} values compared to 38* (Figure 5.13). For 37/38 isolates, there is visible correlation between RF_{μ} and RF_{OD}, while there is little or no correlation between these metrics for the 38* isolates. However, lack of correlation for 38* isolates may simply be due to reduced sample size ($N = 20$ vs. $N = 55$) and/or narrower phenotype range. In general, both high RF_{μ} and RF_{OD} are desirable characteristics for isobutanol production strains, thus isolates that have high tolerance by both metrics are of greatest interest (i.e. 34-20-4 and other points in the upper right side of Figure 5.13).

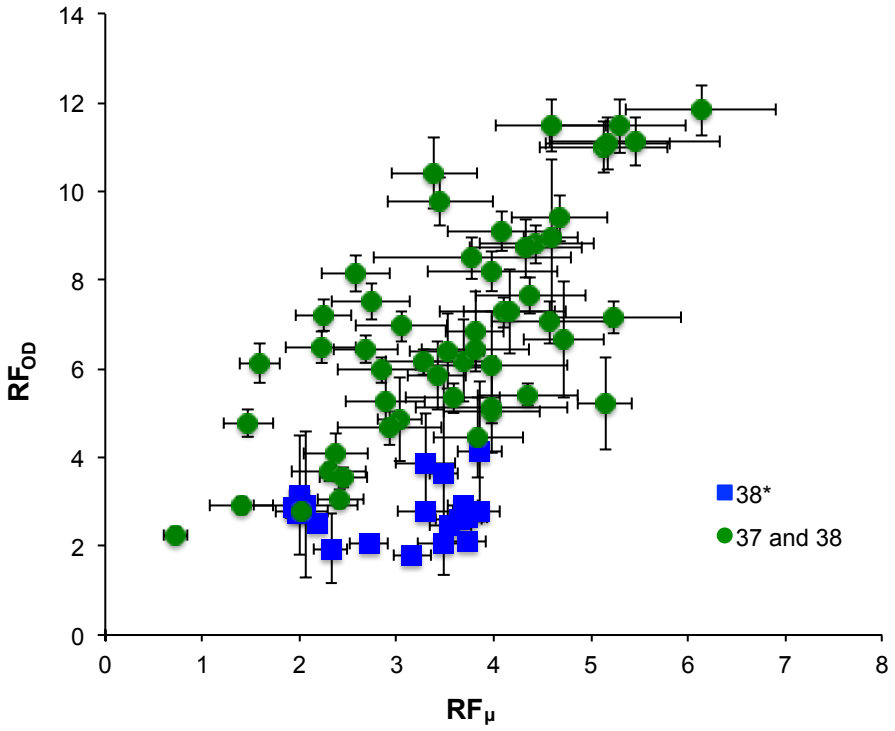

Figure 5.13: Comparison of observed phenotypes in clonal isolates from 38* lineage ($N = 20$ isolates total) with 37 and 38 lineages (reduced *hfq* dosage; $N = 55$ total isolates). For each clonal isolate, RF$_{\mu} = \mu_{max,MUT}/\mu_{max,WT}$ and RF$_{OD} = OD_{600,MUT}/OD_{600,WT}$ in 1% (w/v) isobutanol NG50 are plotted.

In addition to phenotypic differences, isolates from 38* have a distinctly different mutation signature compared to 37/38 isolates. In general, there is greater genotypic diversity among 37/38 isolates compared to 38*, which may underlie the greater range of phenotypes observed for these lineages (i.e. compare Figures 5.10 & 5.11). As discussed above, we hypothesized that *hfq* may constrain accessible evolutionary trajectories and limit evolvability; however, we cannot ascertain whether reduced *hfq* dosage is the origin of greater diversity in the 37/38 lineages, since this set includes two MAGE lineages (vs. one 38*), more MAGE cycles, different selection levels, and a greater number of isolates compared to 38*. Direct comparison of mutation frequencies between 38* and 37/38 provides additional insights into the genotypic patterns in these lineages (Figure 5.14). As expected, *hfq* is present at higher frequency in 38* compared to 37/38; additionally, frequencies of *groL*, *mdh*, *relA*, and *pstC* are also increased in 38* relative to 37/38 (Figure 5.14). The increased frequencies of *groL* and *mdh* in 38* are likely due to epistasis with *hfq*. Interestingly, the uncharacterized *pstC* mutation originates in the same evolved lineage as *hfq* (G3; Chapter 4); thus the increased frequency in 38* suggests potential epistasis with *hfq*. The *relA* mutation was obtained from our literature review and is correlated with n-butanol tolerance; however, a recent study demonstrated functional interaction between *hfq* and *relA* wherein RelA

enhances multimerization of Hfq and stimulates Hfq binding to sRNAS, thus suggesting a possible link between these genes [216].

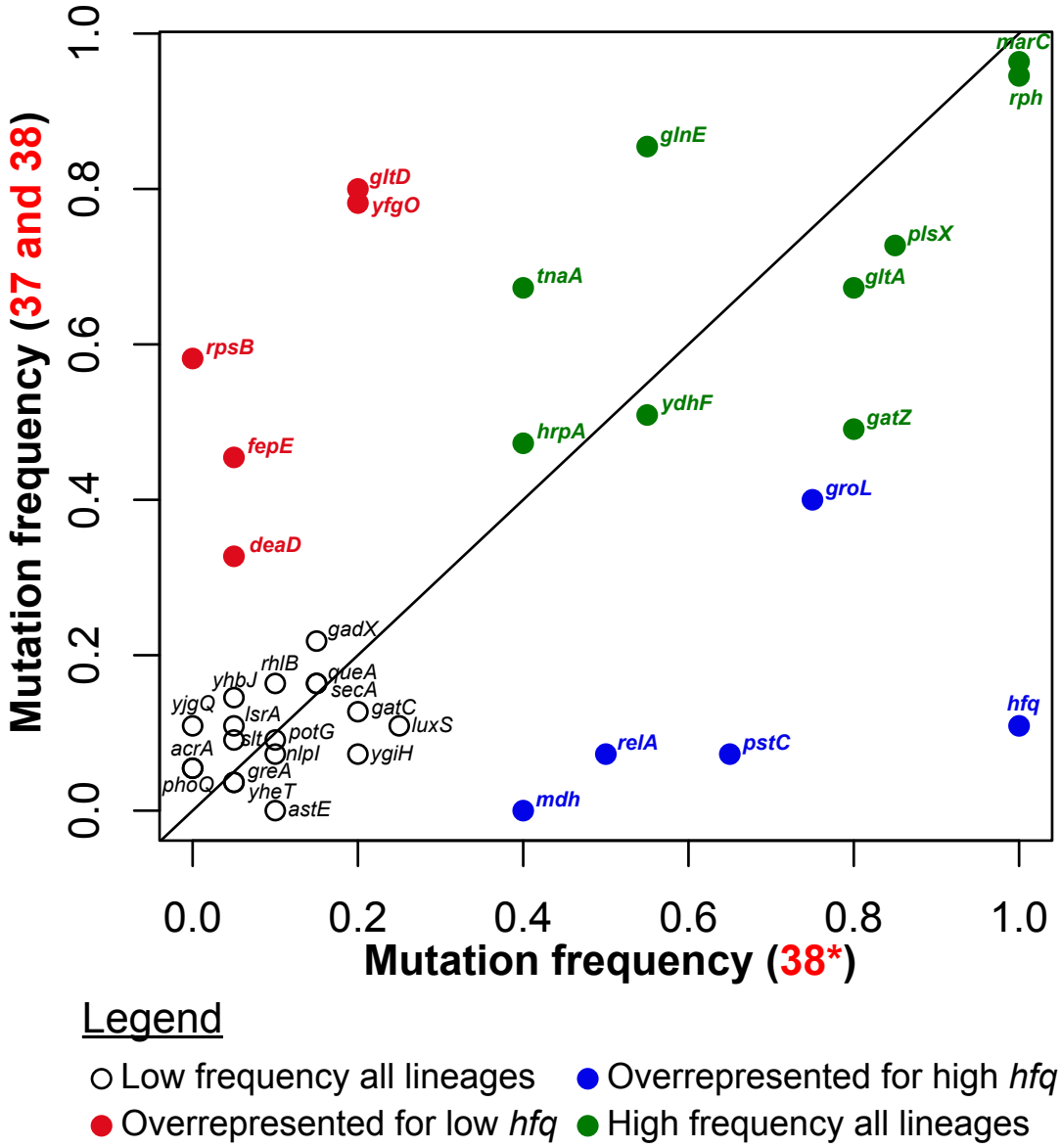

Figure 5.14: Comparison of mean mutation frequencies in 38* lineage with 37 and 38 lineages. The data set for 37 and 38 lineages represents an aggregate of all populations (i.e. 38-10, 38-20, 37-10, 37-20, and 37-30). Open symbols denote mutations that occur at low frequency ($\lesssim 0.25$) in both lineages; green symbols denote mutations that occur at high frequency ($\gtrsim 0.4$) in both lineages; red symbols denote mutations that occur more frequently in 37 and 38 lineages compared to 38*; and blue symbols denote mutations that occur more frequently in 38* compared to 37 and 38.

There are several mutations which occur at higher frequency in 37/38 compared to 38*, including *rpsB*, *deaD*, *fepE*, *yfgO*, and *gltD* (Figure 5.14). All of these mutations derive from

the G3 and X3 lineages of our evolutionary-genomics work and are uncharacterized (Chapter 4). *yfgO*, *deaD*, and *rpsB* all derive from lineage X3 (which has WT *hfq*), and *deaD*/*rpsB* may interact functionally in ribosome assembly, suggesting possible epistasis between these mutations. Possible relationships between other mutations enriched in 37/38 vs. 38* are less obvious, and it is not clear whether frequencies of these mutations are higher by chance or due to reduced *hfq* dosage (e.g. due to negative epistasis with *hfq*). In addition to differences in mutation frequencies between the 38* and 37/38 lineages, there are also notable similarities. There is a large group of mutations that occur at relatively low frequency ($\lesssim 0.25$) in both 38* and 37/38 (Figure 5.14). These low-frequency mutations may be of lesser importance to isobutanol tolerance, or alternately may only confer fitness benefits in certain genotypic backgrounds (e.g. due to epistasis). Interestingly, *acrA* mutations occur at very low frequency in both 38* and 37/38 lineages, despite that *acrAB* mutations were prevalent in our evolutionary-genomics studies and were demonstrated to confer improved isobutanol tolerance (Chapter 4). In addition to low-frequency mutations, there are also mutations that occur at high frequency ($\gtrsim 0.4$) in both 38* and 37/38 lineages, including *marC*, *rph*, *glnE*, *plsX*, *gltA*, *gatZ*, *ydhF*, *hrpA*, and *tnaA*. These higher frequency mutations could be key determinants of isobutanol tolerance, and may act independently of *hfq*.

5.4 Discussion and conclusion

5.4.1 Rapid generation of highly tolerant novel variants with directed genome engineering

The genome engineering strategy that we devised successfully generated highly modified variants (ranging from 6 to 17 mutations per genome) with improved isobutanol tolerance (RF_{μ} up to 6.1 ± 0.8 and RF_{OD} up to 11.8 ± 0.6 at 1% w/v isobutanol). To put these results into perspective, we can compare the characteristics of tolerant strains generated through genome engineering vs. experimental evolution (Chapter 4). As shown in Table 5.3, the top performing isolate from our genome engineering work (38-20-4) has ≈60% higher RF_{μ} at 1% isobutanol compared to the best evolved strain (G3.2; Chapter 4). To comprehensively compare isobutanol tolerance between these two strains we would need to characterize 38-20-4 at isobutanol concentrations higher than 1%; however, these preliminary results suggest that 38-20-4 is more tolerant than G3.2. Furthermore, 38-20-4 has far fewer mutations than G3.2 (Table 5.3). Thus through genome engineering we have refactored isobutanol tolerance into simpler and higher fitness genotypes, which facilitates further investigation into mechanisms and genetic bases of adaptation. In addition to generating highly tolerant strains, genome engineering is a rapid technique - we were able to generate highly

tolerant variants in a few weeks time, compared to approximately six months for the evolution (Table 5.3). However, the genome engineering strategy we devised requires a priori selection of target loci. Since target selection was partially based on our evolutionary-genomics studies of isobutanol tolerance, our genome engineering strategy was necessarily dependent on the outcomes of these studies and thus we must exercise prudence in comparing these methods for phenotype improvement. Overall, with the high speed of MAGE, methods for target gene identification are likely to be a rate limiting factor for phenotype improvement methods integrating both approaches.

Method	**Best strain**	**RF$^*_\mu$**	**# mutations**	**Time**
Evolution	G3.2	3.9 ± 0.5	47	≈6 months
Genome engineering	38-20-4	6.2 ± 0.8	10	≈2 weeks

Table 5.3: Comparison of top-performing strains generated with genome engineering vs. experimental evolution. $^*\text{RF}_\mu$ values determined in 1% (w/v) isobutanol NG50.

Approximately 2/3 of the genome engineering targets in this work were derived from our evolutionary-genomics studies (Chapter 4) while the remainder were obtained from various literature and patent sources (Table 5.1). We experimentally characterized isobutanol tolerance effects for certain mutations from our evolutionary-genomics studies (*marC*, *hfq*, *groL*, *mdh*, *rph*, and *acrA*), but the majority remain uncharacterized. Many of the highly tolerant isolates in this library include mutations that are uncharacterized (e.g. *plsX*, *gltA*, *pstC*, *glnE*, *gltD*, *yfgO*, *hrpA*, *gatZ*, *rpsB*, and *deaD*), or which were that identified from literature (e.g. *ydhF*, *tnaA*, and *relA*), thus suggesting that these loci may be important for isobutanol tolerance and should be examined in greater depth. Furthermore, many isolates feature combinations of mutations derived from several different evolved lineages or literature sources. For example, 38-20-4 (genotype shown in Table 5.4) includes mutations derived from three different evolved lineages (G3, X3, and X2; Chapter 4) as well as two literature reports ([124] and [121]). Recombination of mutations from diverse origins represents a prime advantage of our directed genome engineering method, as it allows exploration of larger and more complex genotype search spaces compared to strategies that rely on one particular method for target gene identification (e.g. TRMR in Sandoval *et al* [206]). While recombination techniques have been developed for asexual organisms (e.g. genome shuffling via protoplast fusion), our genome engineering strategy is more powerful due to the high efficiency/speed of MAGE, and does not require pre-existing *in vivo* mutations [217]. Overall, through our genome engineering work we have correlated additional genetic loci (beyond those investigated in Chapter 4) with isobutanol tolerance and generated novel highly isobutanol tolerant genotypes that provide new insights into mechanisms of adaptation, which are discussed in the

next section.

Locus	Position	Mutation	Locus description	Source
marC	1626139	DELETION	Conserved protein; predicted transporter	Common mutation in isobutanol tolerant strains; Chapter 4, [208], [124]
rph	3823229	4:GGTC	Defective ribonuclease PH	Chapter 4; G3 lineage (*rph* mutations are common in K12 strains evolved on minimal media).
glnE	3206870	DELETION	Fused deadenylyltransferase/adenylyltransferase for glutamine	Chapter 4; G3 lineage
gltD	3367270	1:G	Glutamate synthase, 4Fe-4S protein, small subunit (EC:1.4.1.13)	Chapter 4; G3 lineage
yfgO	2623022	C→T	Predicted inner membrane permease, UPF0118 family	Chapter 4; X3 lineage
tnaA	3895960	DELETION	L-cysteine desulfhydrase / tryptophanase	Isobutanol tolerance study; [124]
gltA	753345	T→C	Citrate synthase (EC:2.3.3.1)	Chapter 4; X2 lineage
gadX	3673040	DELETION	GadX DNA-binding transcriptional dual regulator; controls the transcription of pH-inducible genes, including the principal acid resistance system	Ethanol tolerance study; [121]
secA	108975	T→C	Preprotein translocase subunit, ATPase	Chapter 4; G3 lineage
hfq	4407505	-7:AGGAAAA	Ribosome binding site of pleiotroic regulator *hfq*, involved with many RNA-based regulators	Chapter 4; G3 lineage

Table 5.4: Genotype of isolate 38-20-4. Bold gene names indicate that these loci belong to the high RF_{μ} cluster shown in Figure 5.11. Mutation positions are given as absolute genomic coordinates in the *E. coli* EcNR1 reference sequence. SNPs are indicated by base transition/transversion. Small insertions are indicated by a "+", with the size (number of bp) of the insertion and sequence of inserted bases. Small deletions are designated by "−" with a format similar to that for small insertions. "DELETION" indicates gene deletion and/or loss-of-function. Source names are color-coded to highlight the different origins of the mutations.

5.4.2 Possible mechanisms of adaptation in highly tolerant variants

Mechanisms of adaptation to isobutanol toxicity were extensively investigated and discussed in Chapter 4. However, the mutation patterns generated by genome engineering lend additional insights into possible mechanisms of tolerance. Of particular interest, we identified two groups of clones with high RF_{μ} from the 37 and 38 lineages that share a set of common mutations, including *gltA*, *tnaA*, *yfgO*, *gltD*, *glnE*, *rph*, and *marC*. *marC* and *rph* mutations were previously characterized in our evolutionary-genomics work (Chapter 4), while *tnaA* was confirmed to play a role in isobutanol tolerance in *E. coli* in another study [124]. However, the significance of *yfgO*, *gltA*, *gltD*, and *glnE* was not apparent in our previous studies. These mutations may represent important core adaptations to isobutanol toxicity, and merit further examination.

yfgO has not been characterized yet (no information available in literature as of April 2013), so the role of this protein in isobutanol tolerance is not clear. However, based on sequence homology *yfgO* is predicted to be an inner membrane protein and is thus a possible target of

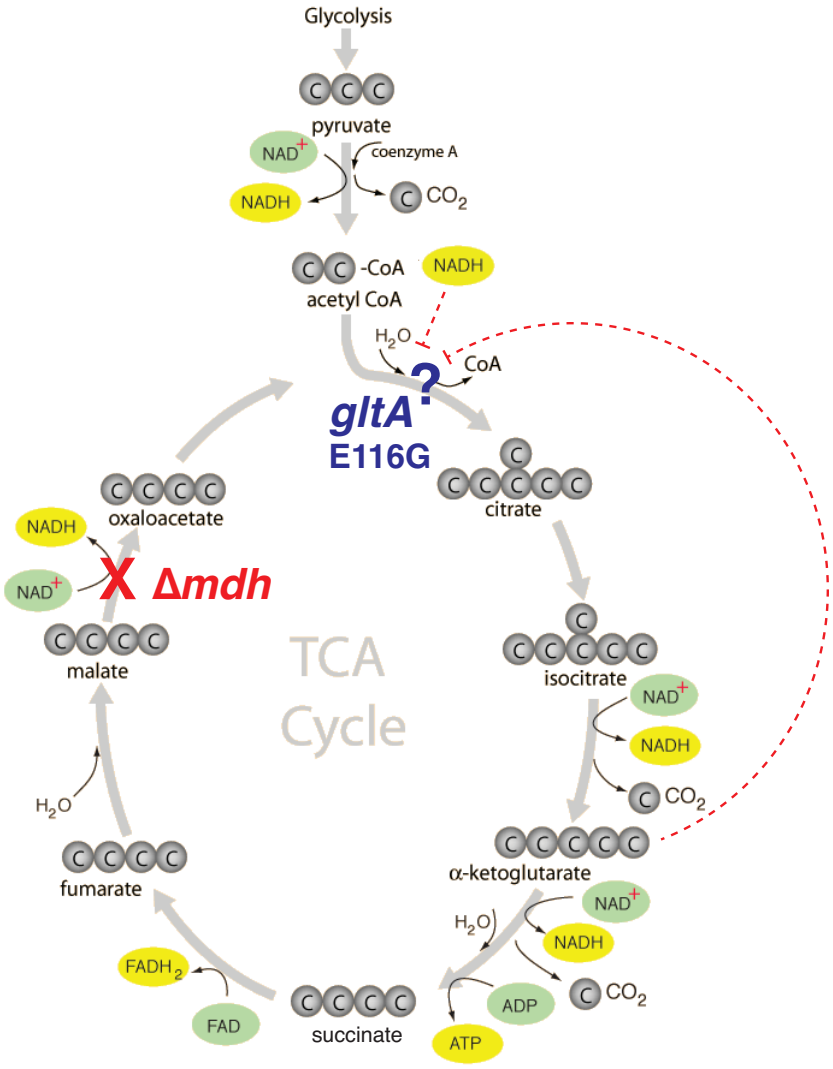

Figure 5.15: *mdh* and *gltA* mutations may rewire central metabolism around the oxaloacetate node. *mdh* mutation leads to loss-of-function (verified in Chapter 4). Functional effects of *gltA* E116G mutation have not yet been elucidated. Portions of figure adapted from [218].

isobutanol toxicity [150]. *gltA* (citrate synthase), *gltD* (glutamate synthase, subunit), and *glnE* (deadenylyltransferase/adenylyltransferase; regulator of glutamine synthase activity) mutations are all of great interest since these genes represent high centrality metabolic network nodes. *gltA* is a key metabolic enzyme and represents a potential rate-controlling step of the TCA cycle (Figure 5.15) [9]. Our previous work (Chapter 4) and other studies [124] demonstrate that many adaptive mutations in isobutanol tolerance simply involve loss-of-function. However, *gltA* is an essential gene so the mutation (753345 T→C; E116G) likely has some other effect; possibilities include changes in substrate specificity, catalytic activity, or modulation of NADH / α-ketoglutarate inhibition (Figure 5.15). In our previous studies loss-of-function mutations in TCA cycle gene *mdh* (malate dehydrogenase) were prevalent in evolved isobutanol tolerant

lineages and shown to be epistatic with *hfq* (Chapter 4). In the TCA cycle, Mdh and GltA operate sequentially, with Mdh catalyzing the reversible oxidation of malate to oxaloacetate (with concomitant generation of NADH) and GltA catalyzing the condensation of acetyl-CoA and oxaloacetate to form citrate (Figure 5.15). Collectively, the prevalence of *mdh* and/or *gltA* mutations in isobutanol tolerant variants suggests that rerouting TCA cycle metabolism may be an important mechanism of adaption to isobutanol toxicity.

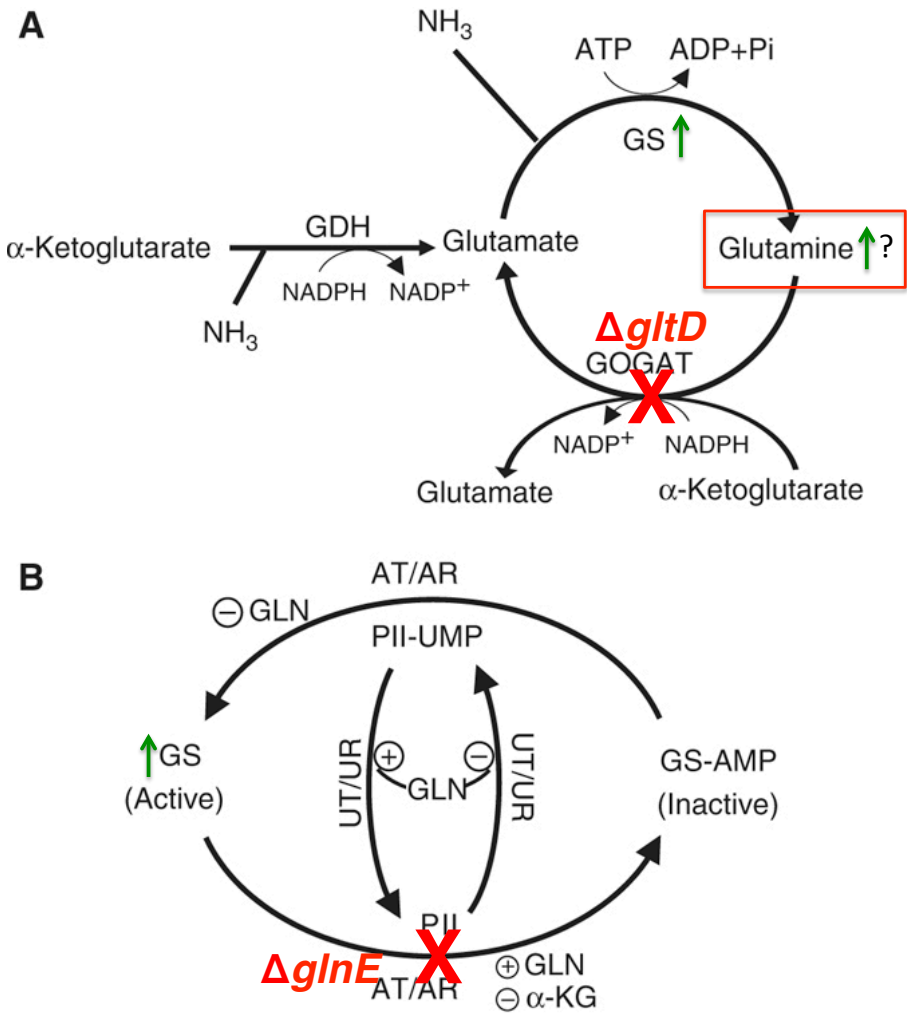

Figure 5.16: Potential effects of *gltD* and *glnE* mutations on nitrogen metabolism. Portions of figure adapted from [219]. **(A)** Loss-of-function of glutamate synthase (GOGAT; subunit *gltD*) is expected to disrupt conversion of glutamine to glutamate, possibly increasing intracellular glutamine pool. **(B)** Glutamine synthase (GS) activity is regulated by adenylyltransferase/adenylyl-removing enzyme (AT/AR), with adenylylation inactivating GS. Loss-of-function of AT/AR (*glnE*) is expected to increase GS activity via elimination of the adenylylation reaction, potentially further increasing glutamine biosynthesis.

In addition to the TCA cycle, nitrogen metabolism may also be significant to isobutanol tolerance. In our previous work, we identified putative loss-of-function mutations in *gltD* and

glnE in the G3 lineage (Chapter 4). *glnE* and *gltD* mutations occurred at high frequency in the 37/38 lineages of our genome engineering studies, and interestingly also clustered closely, suggesting correlation between these mutations (Figure 5.11 & 5.14). We hypothesize that these mutations may act in concert to increase glutamine biosynthesis via elimination of glutamate synthase (Figure 5.16A) and upregulation of glutamine synthase activity (Figure 5.16B). We have not yet fully elucidated how TCA cycle rerouting via *mdh*/*gltA* mutations or nitrogen metabolism rerouting via *gltD*/*glnE* mutations may contribute to improved isobutanol tolerance. We speculate that *mdh*/*gltA* and *gltD*/*glnE* may be convergent metabolic adaptations in glutamine biosynthesis, since *mdh*/*gltA* mutations may increase the intracellular pool of α-ketoglutarate, which is a precursor to glutamine (Figure 5.17).

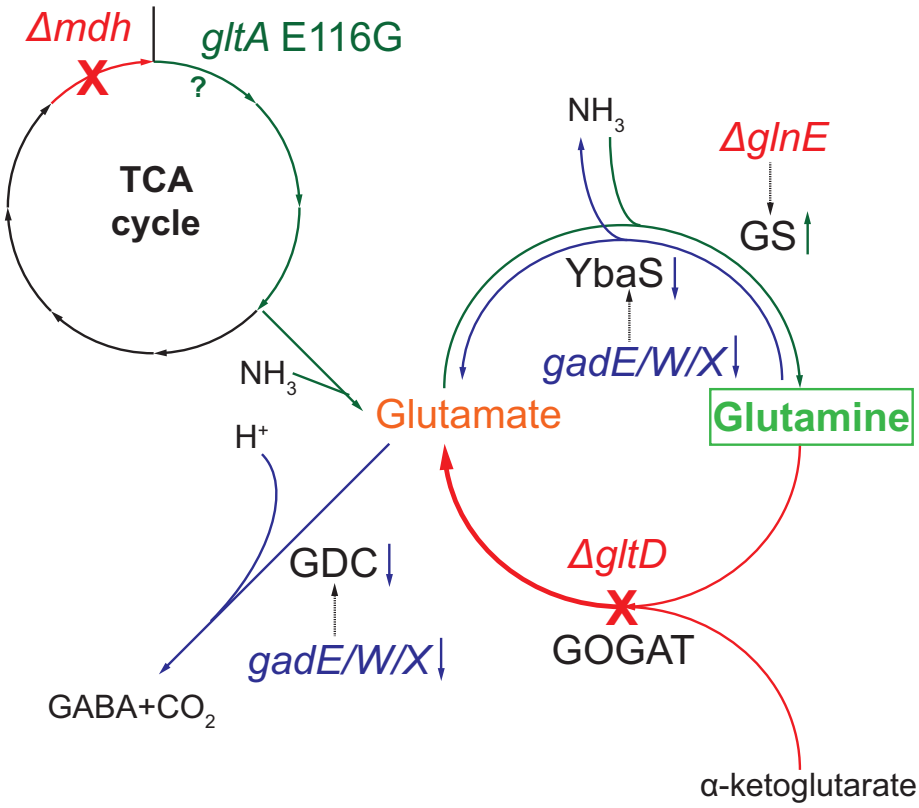

Figure 5.17: Convergent adaptations in glutamine biosynthesis. Mutations in TCA cycle genes may increase metabolic flux to glutamate (though the functional effect of *gltA* E116G has not been characterized). Δ*glnE* is expected to increase glutamine synthase (GS) activity, while glutamate synthase activity is eliminated by Δ*gltD*, which may increase glutamine production. Several glutamate/glutamine catabolic genes are part of the GadE/GadW/GadX acid stress response regulon, including glutaminase (YbaS) and glutamate decarboxylase (GDC; GadAB). Thus attenuation or loss-of-function of *gadE*/*gadW*/*gadX* may reduce glutamate/glutamine catabolism. Color code: red denotes loss-of-function for genes/reactions; green denotes possible increased expression/activity for genes/reactions; blue denotes possible decreased expression/activity for genes/reactions.

Beyond these mutations, our evolutionary-genomics studies (Chapter 4) suggest additional

possible adaptations in glutamate/glutamine metabolism. We observed that the acid tolerance GadE/GadX/GadW regulon is strongly repressed in G3.2 (Chapter 4), and additionally acid tolerance genes were negatively correlated with ethanol tolerance in another study [121]. Glutamate decarboxylase (GadAB) and glutaminase (YbaS) are key mechanisms of acid tolerance (via consumption of H^+ and generation of basic NH_3, respectively; see Figure 5.17) and are regulated by GadE/GadX/GadW; indeed, *gadB* and *ybaS* were among the most differentially regulated genes in G3.2 compared to WT (Chapter 4). Downregulation of glutamate/glutamine catabolic genes is expected to increase glutamine availability, thus representing further potential convergent adaptations (Figure 5.17). Collectively, our results suggest a link between glutamine biosynthesis and isobutanol tolerance, but the mechanistic basis of improved tolerance in not clear. Glutamine is a key protein component and acts as a nitrogen donor for many important reactions, including the biosynthesis of amino acids, nucleotides, glucosamime and carbamoyl phosphate [150]. Glucosamine is a peptidoglycan precursor, thus increased glutamine availability may ultimately lead to increased peptidoglycan biosynthesis, which has previously been implicated in ethanol [121] and isobutanol [124] tolerance. We conducted a preliminary study to determine whether supplementing culture media with glutamine or glutamate improves isobutanol tolerance in *E. coli* EcNR1; however, there was no difference in growth between the supplemented and unsupplemented cultures (data not shown). Glutamine is unstable in aqueous solutions, so the results of this experiment are not conclusive and further study is warranted.

5.4.3 Future work

Our genome engineering studies have generated a rich set of data and isobutanol tolerant isolates that could provide many additional insights into mechanisms and genetic bases of tolerance in future work. Our results suggest that metabolic adaptations may be more important to isobutanol tolerance than previously thought. We suggest further investigation of possible links between glutamine metabolism and isobutanol tolerance. Potential followup studies could include reconstructing metabolic mutations (*mdh*, *gltA*, *gltD*, and *glnE*) singly and in various combinations to determine functional and phenotypic effects and investigating epistasis between metabolic genes and other mutations (e.g. *hfq* or *marC*). At a broader level, metabolomic and ^{13}C flux analysis studies of reconstructed mutants and highly tolerant genome engineering isolates could provide insights into the global metabolic basis of isobutanol tolerance. In studies of *E. coli* evolved for ethanol tolerance under aerobic conditions, ethanol degradation emerged as an important mechanism of adaptation [121]; thus isobutanol catabolism should be assessed in the isolates from this study, especially in light of the observed metabolic mutations. In addition to metabolic

adaptations, it would also be fruitful to examine additional aspects of isobutanol tolerance through transcriptomic studies and cell envelope profiling. With the high genetic diversity in our library of genome engineering isolates, it would be particularly interesting to perform comparative studies between different isolates to explore whether adaptations are convergent or vary between different genotypes. While we have a large number ($N = 75$) of isobutanol tolerant strains available for study, isolates/genotypes of interest could be narrowed down by performing additional phenotyping studies at isobutanol concentrations $> 1\%$ (w/v) to characterize minimum inhibitory concentration (MIC).

In addition to the above suggested investigations, it would also be fruitful to conduct further genetic studies with our genome engineering isolates. We suspect that isobutanol tolerance arises from complex epistatic interactions between many genetic loci. Total epistasis could be determined by constructing and phenotyping single mutants for each target in our library (or a subset of mutations of interest, such as those from a highly tolerant isolate), and then comparing single-mutant fitness to the fitness of isolates with multiple mutations. Additional insights could be gained by "knocking-in" WT alleles in highly-tolerant isolates and measuring residual fitness. As suggested above, additional phenotyping studies would facilitate selection of isolates / sets of genetic loci of interest. Since our genome engineering host strain is MutS- (mutation rate ≈100x higher than WT), it would also be beneficial to perform whole-genome resequencing on selected isolates to identify off-target mutations that accumulate during genome engineering. Off-target mutations could potentially contribute to isobutanol tolerance and thus represent an important confounding factor in this study.

5.5 Materials and methods

5.5.1 Model analysis

The model described in section 5.3.2 was implemented and iteratively solved in MATLAB (MathWorks Inc), using parameter values given in Table 5.2.

5.5.2 Base strains, media, growth conditions, and cryopreservation

E. coli EcHW24 *intC::yfp* (*E. coli* K12 λ-Red+ *bla* Δ*mutS intC::*P_L*lacO*$_1$*-yfp-cat*) was used as a host strain for genome engineering and general ssDNA mediated recombination. *E. coli* EcHW24 is a derivative of *E. coli* K12 MG1655 containing a modified λ prophage integrated at

the *bioA*/*bioB* locus [59], and 2864887 T→G and 2864892 G→T SNPs that produce premature stop codons in *mutS*. *intC::*P_L*lacO*$_1$*-yfp-cat* (cassette for IPTG inducible *yfp* expression) was transferred from strain RP22 [220] to EcHW24 to create the final strain used in this study. NG50 minimal medium, consisting of M9 salts at 1x concentration, 50 g/L glucose, and 0.25 mg/L biotin (identical to that used in Chapter 4) was used for selection cycles and phenotyping assays. NG50 agar plates were prepared by supplementing NG50 with 15 g/L agar. Isobutanol was added to media as indicated; to prevent evaporation, liquid cultures were conducted in screw-cap tubes or flasks and plates were tightly sealed with parafilm. LB Lennox broth (10 g/L tryptone, 5 g/L yeast extract, and 5 g/L NaCl) and LB agar (10 g/L tryptone, 5 g/L yeast extract, 10 g/L NaCl, and 15 g/L agar) were used for propagating strains during genetic manipulations. Growth conditions were 30°C with 150 to 200 rpm shaking for liquid cultures. When needed, culture samples were cryopreserved by centrifuging 1 mL aliquots at 16,000 g x 1 to 2 minutes, removing supernatant, resuspending cell pellets in 500 μL LB broth and 500 μL 50% (v/v) glycerol, and then transferring to cryovials and storing at -80°C.

5.5.3 Design of mutagenic ssDNA oligonucleotides

Mutagenic oligoucleotides were designed as per specifications given in Wang *et al*. 2009 [59]. For each mutation in Table 5.1, we designed a 90-mer oligonucleotide containing the mutation flanked by homologous genomic sequences. Oligonucleotides were designed such that they targeted the lagging strand of the replication fork during DNA replication and had predicted ΔG $>$ -12.5 kcal/mol for secondary structures (evaluated with mfold v3.2) [189]. For sequences with a high propensity for folding, we disrupted base pairing in predicted secondary structures by introducing additional mutations in the form of synonymous SNPs, which are expected to be neutral. Oligos were synthesized by IDT (Integrated DNA Technologies, Coralville, IA) with four 5' phosphorothioated bases to enhance *in vivo* stability. Designed oligonucleotide sequences are shown in Table 5.5.

Oligo name	Locus	Position	Mutation	Oligo sequence
fepE_g3.2_mut	*fepE*	618,531	T→C	t⋆t⋆t⋆c⋆tggatgccatcgcatagcgcaacagcacgccaccacaagccGccatcccgccgatcaacgcggaaaggatcacaataatcgcctta
gatZ_g4.2_mut	*gatZ*	2,182,915	-1:C	c⋆a⋆a⋆a⋆ttccacacccggctgcaccacgatggcaatcacacgtgttaa-gcctctgtcagcccacgggcaataaaggccttttgatgcgtacg
gltD_g3.2_mut	*gltD*	3,367,270	1:G	c⋆a⋆a⋆T⋆acTaaacagttaatgggctttggtgaaaccccgTgacgaaccGgttcgtcagcatggaaggcaaacgcgtggtggtccttggcggtg
glnE_SNP_del	*glnE*	3,206,870	Δ: $55, $56, $57; 946	t⋆g⋆c⋆g⋆gcggttggctttAcTAttAcgtcagccactctggatgcgcaatcacgctgtcctgcacaaaatcactaaaagtaagtactgacttc
greA_dacB_g3.2_mut	*greA-dacB*	3,336,132	A→G	t⋆t⋆t⋆g⋆atagtcattttaccctgaagttcccgaagggtcatcgtttGctttatagggcgttgcgccgtagtatgacggctcgattccaggtt
groL_g4.2_mut	*groL*	4,378,650	A→C	a⋆a⋆c⋆c⋆tgagcaatcgctttagagtcagagcatggtacggacagcgcGttcagttcttcaactgcagcggtaaccgctttgtcgataccacg
lepB_g3.2_mut	*lepB*	2,711,902	G→A	a⋆c⋆g⋆c⋆tcggaaagacgaattccatttttctttggtttcgtttttcgAcacttcaaagaatccgctggtcgcttccccaccattacggcgtga
miaA-hfq_g4.2_mut	*miaA-hfq*	4,407,505	-7:AGGAAAA	g⋆t⋆t⋆c⋆aggaacggatcttgtaaagattgccccttagccattctctc-------tatatgcttatttgtactttgaacctttcgattctgaaaaaattg
nlpI_pnp_g3.2_mut	*nlpI-pnp*	3,316,213	T→C	a⋆a⋆a⋆a⋆ggcttcatttcccactcccgaagaccacggttgaatgaacgtcctgttcccggttgctaacaaggcgCcctgcccggttaaaagcc
phoQ_g3.2_mut	*phoQ*	1,197,581	A→G	g⋆g⋆t⋆t⋆tatctatgtgctctcagccaatctgctgttagtgatcccgcCgctgtgggtcgccgcctggtggagtttacgccccatcgaagccc
pstC_g3.2_mut	*pstC*	3,917,582	T→C	g⋆a⋆g⋆a⋆gacaataatgccacccaacatcaatagcacaatcagcgccgccagttttaccagcacgctgaaaattatgCcgccctttttacccg
queA_g3.2_mut	*queA*	425,270	A→G	c⋆g⋆c⋆c⋆gcggaattactccccgacgcgctcattaattgcctgcggatCgtacgtgataaacatcgcatcaccgtaactaaaaaagcgatatt
rhlB_g3.2_mut	*rhlB*	3,972,054	A→G	t⋆g⋆c⋆t⋆gcaaacgctgatcgaagaagagtggccagacagagcgattatttttcgccaacaccaaacaccgttgCgaagagatctggggccacc
secA_g3.2_mut	*secA*	108,975	T→C	g⋆g⋆a⋆a⋆ggtttcggagtcttctttttcctgacggatcaggtgcggaataatttttattcacgcgtttatacatttccgGgctgtcttctgccg
yjgQ_g3.2_mut	*yjgQ*	4,494,710	T→C	t⋆c⋆g⋆g⋆gaagaagatctgcacatctttcggcacgctcagcaaggtatGcattcctgcgcctaacgcgtcgtaactcccctgcccggctttttt
mdh_x3.5_mut	*mdh*	3,390,936	5:AACCT	g⋆g⋆t⋆c⋆agatcagccacttcctgctcggtaaaactaacgccaggAACCTaacctgtgacagcagcggcagaatggtaacaccagagtgaccg
gatC_x3.5_mut	*gatC*	2,180,640	C→T	g⋆c⋆c⋆g⋆tatgtcccagcagcaacgccggatcaaggccaatcaggaactTctgaccgccgaacttcgcctgtaaacgactacgagcctgctta
gltA_x2.5_mut	*gltA*	753,345	T→C	a⋆t⋆g⋆a⋆cgaatttaaaactacggtgacccgtcataccatgatccacgGgcagattacccgtctgttccatgctttccgtcgcgactcgcatc
rph_g4.2_mut	*rph*	3,823,229	4:GGTC	g⋆c⋆a⋆c⋆ttcaatgatgcgcccgtcttcggtcatcactacgttcatGTCGgtcggtctctgcggcagagtcttcaacgtattccagatcgcaa
acrA_x3.5_mut	*acrA*	483,735	1:A	a⋆g⋆g⋆c⋆gatcgcgtagtaataagtgggctgcagaaagtgcgtcctggTtgtccaggtaaaagcacaagaagttaccgctgataataaccagc
deaD_x3.5_mut	*deaD*	3,314,027	4:AGAC	c⋆g⋆t⋆t⋆cggcagttctacttccggaatagtcagcttcatagtacgttcaatgttgcgcagcagacAGACgacgctcgcggttctcaacgaac
hrpA_x3.5_mut	*hrpA*	1,493,514	C→T	t⋆a⋆a⋆c⋆ttttcatgtaaatatttgattggcgatggaatattcagcaAcagtagacggcgaagaccgttccacattgcctgcttttgctccag
plsX_x3.5_mut	*plsX*	1,156,697	A→G	t⋆c⋆a⋆c⋆aaaccagcacatctgtcttgccagttaacaactcattggctCcaagatagccgatataattgatagaagggattgttttaagcacc
rpsB_x3.5_mut	*rpsB*	190,557	1:A	c⋆t⋆a⋆t⋆tttgggggagttatcaagccttattactcagcttctacgaagctttcttccgcctgggaagccagaTtcctgagaacggccttcac
yfgO_x3.5_mut	*yfgO*	2,623,022	C→T	g⋆c⋆c⋆a⋆ccagcaacggagcaagcaggccactaaagaaaaagataatgTcaaatccggcaactaaaatcaccagcaaggcaatcgcttccggg
astE_SNP_del	*astE*	1,834,154	Δ: $49, $51, $56; 322	a⋆g⋆c⋆a⋆acgcgtccagcatctccacaggtgccgtctAattaccgtgtattcAcgctCaaatcaccagtgcgccttgcggtggcgttaatggc
ygiH_SNP_del	*ygiH*	3,211,923	Δ: $25, $27, $28; 205	c⋆g⋆t⋆a⋆cctctgcggctccatttccagtgccattctggtttgAcgctAgtgAgggctgcccgatccgcgaaccagcggctccggcaatccag
yhbJ_SNP_del	*yhbJ*	3,354,344	Δ: $22, $27, $28; 284	a⋆g⋆g⋆t⋆tcaggtaaatctgtcgccctgcgtgcgctgTaagatatgggtttttaGtgAgtggataaccttcccgtagtgttgttacccgatct
tnaA_SNP_del	*tnaA*	3,895,960	Δ: $26, $28, $29; 471	a⋆t⋆t⋆c⋆gtgttattgagccagtaaaacgtaccactcgcgcttaGcgtTaaTaggcaattattaaatccggtatgaacccgttcctgctggat
marC_SNP_del	*marC*	1,626,139	Δ: $43, $44, $45; 221	a⋆c⋆g⋆c⋆caccatcatgatggcaaatacgtacaccgaggccatcTacTactAacgattacgttcggcactgttcatgttgcctgccaggccaa
ydhF_SNP_del	*ydhF*	1,732,863	Δ: $21, $22, $23; 298	a⋆a⋆a⋆c⋆tgaccagctggcgggcggacatattccagtccatcaatcAcTaCtagcccatcacaaaacgggaaaactccgggccttgcggcgca
potG_SNP_del	*potG*	903,421	Δ: $43, $44, $46; 377	a⋆a⋆a⋆c⋆ctgccagcatacgcagcagcgtggacttgccacagccggatgAgcccagcagcgcgaagatttAaccttACtagatggtcaggctg
yheT_SNP_del	*yheT*	3,490,431	Δ: $46, $47, $48; 340	t⋆g⋆c⋆t⋆gccgcgtctgtttcgtcgtcaggtgaaattcaccccgtaGtgATagcggctggagttgcccgacggcgattttgtcgatcttgcgt
gadX_SNP_del	*gadX*	3,673,040	Δ: $25, $26, $28; 274	t⋆g⋆g⋆c⋆ttgcatccgcaaaaaccaggtcaccgccattaaaCtagcgCtattAaccattaaccatggtgagaatatatttatgtcttgcatac
slt_SNP_del	*slt*	4,637,963	Δ: $29, $32, $33; 645	t⋆g⋆t⋆t⋆cgaccacatccatttgtcgattatcccaggcctgcttgatttgAgcgtaacgactacgctActAatccagtCagtcggctcgcgcc
relA_SNP_del	*relA*	2,920,880	Δ: $29, $30, $31; 744	g⋆t⋆t⋆g⋆cagacaatacgcccaggtttcggctaagcactcacacTactActAgctggtaatacccagacttgcgatccatttttccggatcaa
luxS_SNP_del	*luxS*	2,821,962	Δ: $3, $4, $5; 171	a⋆c⋆t⋆g⋆caggcgcttccatccgggtatgatcgactgtgaagctCtAtTacTacggcatttagccacctccggtaatttttttaaaaattttc
lsrA_SNP_del	*lsrA*	1,608,721	Δ: $20, $21, $22; 511	c⋆c⋆c⋆g⋆cgcgttaccgctactttgcgcccgctcggtttataaaTagtaAtGaggggtcaatgtcctgaaaggcatcgattttacgttgcatc

Table 5.5: Oligonucleotide library used for genome engineering. Mutagenic oligonucleotides were designed for each of the target mutations listed in Table 5.1. For deletion mutations ("Δ"), position denotes start of gene. We use the following notation for deletion mutations: $X, $Y,.....; N where X, Y, = stop codon residue position and N= total number residues in peptide. Notation for other mutations is the same as in Table 5.1. For oligo sequences, "⋆" denotes phosphorothioate bond; "–" denotes deleted nucleotide; and N denotes inserted nucleotides or SNPs.

5.5.4 Multiplex ssDNA mediated homologous recombination

ssDNA mediated homologous recombination was done by heat shocking *E. coli* EcHW24 *intC::yfp* to induce expression of λ-Red genes, preparing electrocompetent cells from the induced cultures, and electroporating competent cells with oligonucleotides at 1 μM/oligo concentration. Overnight cultures of *E. coli* EcHW24 *intC::yfp* were inoculated 1:70 into fresh LB medium and incubated with shaking at 30°C until reaching mid-log phase (OD_{600} ≈0.7). Cultures were then heat shocked at 42°C for 15 minutes in a water bath with 200 rpm shaking. Immediately after heat shocking, the cultures were chilled on ice for 10 minutes. All subsequent manipulations were done at 4°C, which is vital for maximum recombination efficiency. For each electroporation, 1 mL of induced cells was centrifuged at 16,000 g x 1 minute. Cells were resuspended in 1 mL chilled ultrapure dH_2O. The centrifugation and washing process was repeated twice. A final centrifugation was performed and the cell pellet was resuspended in 50 μL ultrapure dH_2O. Oligonucleotide solution was mixed with the cell suspension such that final DNA concentration was 1 μM/oligo (38 μM total for full library or 37 μM for *hfq-* library). Immediately after adding oligonucleotide solution, cell/DNA mixture was transferred to a 0.1 cm gap electroporation cuvette and electroporated at 1.8 kV. Cell mix was immediately resuspended in 1 mL room temperature LB medium, added directly to the electroporation cuvette. The resuspended cell mix was transferred to a Falcon tube, diluted with 1-5 mL additional LB medium, and allowed to recover at 30°C with shaking for 2 to 3 h. Once cells reached mid-log phase they were used for additional genome engineering cycles, selection cycles, or isolated and subjected to genotype/phenotype analysis (all described below).

5.5.5 Multiplex genome engineering

This work was performed in collaboration with the labs of Professor George Church (Harvard Medical School) and Harris Wang (Pathology and Cell Biology, Columbia University), with technical assistance from John Harding (Wyss Institute). Genome engineering was performed via recursive cycles of multiplex ssDNA mediated homologous recombination (described above), using cycling schemes described in section 5.3.3. As a control/reference, a lineage of *E. coli* EcHW24 *intC::yfp* was cycled similarly to those described in section 5.3.3, but with a *lacZ-* oligo (g*g*a*a*acagctatgaccatgattacggattcactggccgtcgtttGacaacgtcgtgactgggaaaaccctggcgttacccaacttaatc) instead of the above described oligonucleotide library. Recombination efficiency during genome engineering was monitored by differential plating of population samples from the *lacZ-* lineage on LB agar supplemented with 0.1 mM IPTG and 40 μg/mL Xgal. 1 mL aliquots of each population were cryopreserved before and after selection cycles. For selection cycles, inocula were prepared by centrifuging 1 mL population samples (with OD_{600} ≈0.7) at 16,000 g x 1 minute, discarding

supernatant, and resuspending cell pellets in 1 mL NG50 media. Resuspended pellets were inoculated into 100 mL NG50 media with isobutanol at indicated concentration, and incubated at 30°C with shaking until saturation.

5.5.6 Plate screening for tolerant isolates

Isobutanol tolerant variants were isolated by plating population samples on 1% (w/v) isobutanol NG50 agar plates (on solid media, 1% isobutanol is completely inhibitory to WT *E. coli* EcHW24 *intC::yfp*). To prepare media, NG50 agar plates were weighed to measure the volume of media (weight of empty plate subtracted; $\rho_{NG50} = 0.98$ g/mL) and determine volume of isobutanol to add to each plate. The required volume of isobutanol was spread onto each plate using glass beads. After adding isobutanol, plates were tightly wrapped with parafilm to prevent evaporation and incubated for 12 hours before use to ensure uniform diffusion of isobutanol through the media. Serial dilutions (prepared in NG50 media) of population samples were plated at a density of 10^4 to 10^5 cells per plate, estimated by OD_{600} of the undiluted culture (with 1 $OD_{600} \approx 1x10^9$ cells/mL). Plates were incubated at 30°C until colonies were visible (up to 5 days). For each plate, NG50 cultures were inoculated from 10 to 15 randomly selected colonies, incubated at 30°C until saturation, and then cryopreserved. Cryostocks were used for subsequent phenotype/genotype analysis (described below).

5.5.7 Phenotype analysis

Isobutanol tolerance was quantified by measuring the maximum specific growth rate (μ_{max}, 1/h) and saturating OD_{600} in 1% (w/v) isobutanol NG50, using a microplate spectrophotometer (similar to procedure described in Chapter 4). Inoculum was prepared by centrifuging 1 mL of overnight culture at 16,000 g x 2 min and resuspending cell pellets in a volume of fresh medium such that OD_{600} = 1. Standard 96-well microplates were filled with 200 μL medium per well and seeded with 2 μL of prepared inoculum per well. Microplates were covered with adhesive film to prevent isobutanol evaporation and microplate lids were affixed with tape. OD_{600} was measured every 10 minutes for 48 hours using Molecular Devices Spectramax M5 or Molecular Devices Versamax plate readers, with 30°C incubation temperature and agitation between reads. μ_{max} was calculated via linear regression of ln(OD_{600}) vs. time (h) after subtracting blank values; regression was done over the time intervals corresponding to log growth phase. To provide cross-experiment references, WT *E. coli* EcHW24 *intC::yfp* and arbitrarily selected genome engineering isolate 37-30-15 were included in each microplate study. Relative fitness (RF) was calculated as $\mu_{max,MUT}/\mu_{max,WT}$ or

maximum $OD_{600,MUT}$ / $OD_{600,WT}$, were MUT denotes genome engineering isolates.

5.5.8 Genotype analysis

5.5.8.1 Multiplex allele specific PCR (MASPCR) genotyping assays

We designed allele specific PCR genotyping assays for each mutation in our library, except for *hfq*. Primers were designed such that one primer in a pair was complementary to a mutation of interest at it's 3' end [188]. Under stringent PCR conditions, non-proofreading DNA polymerases are unable to extend from 3' mismatches, allowing genotype discrimination based on the presence or absence of a PCR product [188]. Allele specific primers for SNPs were designed using BatchPrimer3 v1.0, and allele specific primers for indels were designed manually [188]. See Table 5.6 for list of primers and sequences. To increase genotyping throughput, we multiplexed allele specific PCR reactions according to the reaction sets listed in Table 5.6. Multiplex allele specific PCR was done with a Qiagen Multiplex PCR kit, using manufacturer's recommended protocol with 0.2 μM each primer in the final reaction mix. Cell suspensions were prepared by resuspending cryostock material in 250 μL sterile water; 0.4 μL cell suspension was used as template per 20 μL PCR reaction.

For each MASPCR reaction set, we optimized reaction conditions by performing annealing temperature gradient PCR over varying numbers of cycles, using *E. coli* EcHW24 *intC::yfp* cells as a negative control and whole population samples from the 38*-10 lineage as a positive control[4]; optimal annealing temperatures and cycle numbers for each reaction set are shown in Table 5.6. Cycling conditions used for final PCR reactions were initial denaturation at 94°C for 15:00, followed by cycles of 94°C denaturation for 30 seconds, optimal annealing temperature for 1:30, and 72°C elongation for 1:30, followed by a final 72°C extension for 10:00. Agarose gel electrophoresis (Invitrogen E-gel 48 or 96 system; 2% or 4% agarose; 15-25 minutes total run time) was used according to manufacturer's recommended protocol to examine PCR products.

5.5.8.2 PCR genotyping assay for *hfq* mutation

We were unable to develop a reliable allele specific PCR assay for the *miaA-hfq* 4407505 -7:AGGAAAA mutation in this study. Instead, we designed primers flanking *miaA-hfq* 4407505 -7:AGGAAAA (Table 5.6) and genotype this mutation via a short amplicon PCR reaction

[4]All targeted mutations are expected to be present in MAGE populations; 38*-10 was arbitrarily chosen for use as a positive control

that allows discernment of the -7:AGGAAAA mutation based on product size. Platinum Taq polymerase (Invitrogen, Carlsbad, CA, USA) was used for PCR, with 0.4 μM of each primer and other reagent concentrations as per manufacturer's recommendations. Cell suspensions were used as template as described above. Cycling conditions used were initial denaturation at 94°C for 6:00, followed by 32 cycles of 94°C denaturation for 30 seconds, 60°C annealing for 30s, and 72°C elongation for 1:00, followed by a final 72°C extension for 5:00. Agarose gel electrophoresis (Invitrogen E-gel 48 system; 4% agarose; 17-23 minutes total run time) was used according to manufacturer's recommended protocol to examine PCR products; WT *miaA-hfq* yields a 75 bp product, while *miaA-hfq* 4407505 -7:AGGAAAA yields a 68 bp product.

5.5.8.3 PCR genotyping assays for *hrpA* mutation

We were unable to develop a reliable allele specific PCR assay for the *hrpA* 1493514 C→T mutation in this study. All allele specific PCR reactions tested for *hrpA* appear to produce an off-target product of similar size to the expected product, thus obfuscating the results. However, we discovered that the off-target product has a lower melting temperature (73.5°C) compared to the expected product (79°C). Thus we can discriminate between these two PCR products using quantitative PCR with melt curve analysis. Quantitative PCR (qPCR) assays were performed in 25 μL samples on an MJ Research (BioRad, Hercules, CA, USA) Chromo4 thermal cycler with a QuantiTect SYBR Green RT-PCR kit (Qiagen), using primer pairs for *hrpA* given in Table 5.6. Cell suspensions were used as template as described above, with primer and other reagent concentrations as per manufacturer's protocol. Cycling conditions used were initial denaturation at 95°C for 15 minutes, followed by 55 cycles of 94°C denaturation for 15 seconds, 67.5°C annealing for 30 seconds, 72°C elongation for 30 seconds, and fluorescence measurement at 72°C. Melting profile was determined over a temperature range of 55°C to 95°C in 0.2°C increments. qPCR data was analyzed by fitting parameters of the MAK2 model (initial target concentration D_0 and characteristic PCR constant k) with a custom R script [135]. *E. coli* EcHW24 *intC::yfp* and *E. coli* X3.5 (origin of *hrpA* mutation; Chapter 4) were used as negative and positive controls, respectively.

5.5.8.4 PCR primers used in this study

Table 5.6: Primers and MASPCR cycling conditions for PCR genotyping reactions. See corresponding sections for *hfq* and *hrpA* PCR procedure and cycling conditions.

Set #	Locus	Primer names	Sequence	Size	Cycles	Anneal. T
1	*gatC*	gatC_x3.5_fwd_mut	gatcaaggccaatcaggaaatt	65	**29x**	**61.3°C**
		gatC_x3.5_rev_out	ctaagcaggctcgtagtcgtttaca			
	yfgO	yfgO_x3.5_fwd_out	gtagcaatataccgacgaaaaccac	182		
		yfgO_x3.5_rev_mut1	gttattttagttgccggattgga			
	plsX	plsX_x3.5_fwd_mut1	ttctatcaattatatcggctatctggg	303		
		plsX_x3.5_rev_out	ggctgcaccatgacttttttatcac			
	groL	groL_g4.2_fwd_out	gcacgacactgaacatacgaattta	453		
		groL_g4.2_rev_mut	gcatggtacggacagcacg			
2	*mdh*	mdh_x3.5_fwd_mut	aaaactaacgccaggaacctaacctgtg	101	**29x**	**68.3°C**
		mdh_x3.5_rev_out	aggcaaacagccaggcgaagt			
	gatZ	gatZ_g4.2_fwd_out3	gtctggtaatcggtagaatgtgctt	168		
		gatZ_g4.2_rev_mut3	cgtgggctgacagaggctta			
	rph	Rph_g4.2_fwd_mut	tcatcactacgttcatgtcggtcggtc	193		
		Rph_g4.2_rev_out	tcgattacgggtgcctgcgtgg			
	acrA	acrA_x3.5_fwd_mut2	acttcttgtgcttttacctggacaacc	211		
		acrA_x3.5_rev_out	agtcccgcaacagggcgtaacc			
	deaD	deaD_x3.5_fwd_mut1	ttgcgcagcagacagacg	362		
		deaD_x3.5_rev_out1	ctggaagcggaagattttgatg			
	rpsB	rpsB_x3.5_fwd_out1	cggtatgctgactaactggaaaac	406		
		rpsB_x3.5_rev_mut2	cgcctgggaagccagatt			
3	*glnE*	glnE_4SNP_fwd	tacagagcgcctcctgcaacc	92	**29x**	**68°C**
		glnE_4SNP_AS_rev	gcatccagagtggctgacgtaatagt			
	potG	potG_1SNP_AS_fwd2	ttcgcgctgctgggct	144		
		potG_1SNP_rev2	atattgatggggcgcaggtaag			
	astE	astE_2SNP_fwd	ccaacgaccgccaaacattc	199		
		astE_2SNP_AS_rev	cgcactggtgatttgagcgt			
	yheT	yheT_3SNP_fwd	cgaactgtatgaccagagccgtaaa	308		
		yheT_3SNP_AS_rev	ggcaactccagccgctatcac			
	gadX	gadX_2SNP_AS_fwd	gtcaccgccattaaactagcgc	386		
		gadX_2SNP_rev	ttggctgtctattcatcgtgttgat			
	slt	slt_2SNP_AS_fwd	gcgagccgactgactggatt	512		
		slt_2SNP_rev	tacccgctttcatcgccagac			

Table 5.6: Primers and MASPCR cycling conditions for PCR genotyping reactions.

Set #	Locus	Primer names	Sequence	Size	Cycles	Anneal. T
	relA	relA_2SNP_AS_fwd2	gtttcggctaagcactcacactacta	596		
		relA_2SNP_rev2	gtactggatctgttctgcggtatg			
	luxS	luxS_4SNP_fwd	cgccaatgagatcttctggatgt	703		
		luxS_4SNP_AS_rev	ggaggtggctaaatgccgtagtaatag			
	lsrA	lsrA_3SNP_fwd	cgcaacggagccggaataata	862		
		lsrA_3SNP_AS_rev	tttcaggacattgacccctcattacta			
4	*fepE*	fepE_g3.2_fwd_mut	gttgatcggcgggatggc	120	**28x**	**72°C**
		fepE_g3.2_rev_out	acggcaccttcaaccgcaaatagt			
	greA-dacB	greA_dacB_g3.2_fwd_mut	gaagttcccgaagggtcatcgtttg	156		
		greA_dacB_g3.2_rev_out	atcaacatttgcggcctgaacact			
	lepB	lepB_ASv3_fwd	catcatgaagtattgtcccggagga	203		
		lepB_ASv3_mut3bp_rev	gaccagcggattctttgaagggt			
	nlpI-pnp	nlpI_pnp_g3.2_fwd_out	tgccctcagaccgagactatcataca	316		
		nlpI_pnp_g3.2_rev_mut	gggcttttaaccgggcaggg			
	phoQ	phoQ_g3.2_fwd_mut	cggcgacccacagcg	408		
		phoQ_g3.2_rev_out	caagcaaagccccaccatga			
	pstC	pstC_g3.2_fwd_mut	cagttttaccagcacgctgaaaattatgc	513		
		pstC_g3.2_rev_out	gtacaccaaactgatctccgctgatg			
	rhlB	rhlB_g3.2_fwd_mut	caggtggccccagatctcttcg	618		
		rhlB_g3.2_rev_out	cttctctctcatcctgcgattgcc			
	secA	secA_g3.2_fwd_mut	ggcccggcagaagacagcc	715		
		secA_g3.2_rev_out	ccagctccgatttttcgatggagat			
	yjgQ	yjgQ_g3.2_fwd2	agtgttagcagaagcgaaagcgaac	869		
		yjgQ_g3.2_mut3bp_rev	ggcacgctcagcaaggtctg			
5	*ygiH*	ygiH_fwd_AS_mut	gccattctggtttgacgctagtga	97	**29x**	**67.6°C**
		ygiH_rev_out	ccttgccaccgatacgtaacacatt			
	gltA	gltA_x2.5_fwd_mut4	gaacagacgggtaatctgcc	140		
		gltA_x2.5_rev_out4	agctggcgaccgattctaa			
	yhbJ	yhbJ_fwd_AS_mut	tgcgctgtaagatatgggtttttagtga	180		
		yhbJ_rev_out	gagaaagcgtcaggcaggttactca			
	queA	queA_g3.2_fwd_out3	gttggcaccacttcagtacgtt	300		
		queA_g3.2_rev_mut3	ctcattaattgcctgcggatc			
	gltD	gltD_g3.2_fwd_out3	aggtctggcgtgtgcgga	393		

Table 5.6: Primers and MASPCR cycling conditions for PCR genotyping reactions.

Set #	Locus	Primer names	Sequence	Size	Cycles	Anneal. T
		gltD_g3.2_rev_mut2	tgccttccatgctgacgaacc			
	tnaA	tnaA_fwd_out	tgctccccgaacgattgtga	510		
		tnaA_rev_AS_mut	tcataccggatttaataattgcctattaacgc			
	marC	marC_fwd_out	tggccagtgacactcgaagaaca	619		
		marC_rev_AS_mut	gaacagtgccgaacgtaatcgttagtagta			
	ydhF	ydhF_fwd_AS_mut	cggacatattccagtccatcaatcactac	704		
		ydhF_rev_out	gcaactcagtcagtcgctggtgata			
Misc	*hfq*	hfq_75_up	ttcagaatcgaaaggttcaaagtaca	-	-	-
		hfq_75_down	ttgtaaagattgccccttagcc	-	-	-
	hrpA	hrpA_x3.5_fwd_mut	aacggtcttcgccgtctactgt	-	-	-
		hrpA_x3.5_rev_out	ttatccacaccgcaggagatacagt	-	-	-

CHAPTER 6

Concluding remarks and perspectives

6.1 Part I: Synthetic fungal-bacterial consortia for production of cellulosic biofuels and commodity chemicals

6.1.1 Concluding remarks and perspectives

Inspired by power of natural microbial ecosystems, in this work we explored engineering synthetic microbial consortia for flexible conversion of cellulosic feedstocks to biofuels. The required biochemical functions are divided between two specialist organisms: the fungus *Trichoderma reesei*, which secretes cellulases to hydrolyze lignocellulose into soluble saccharides, and the bacterium *Escherichia coli*, which metabolizes soluble saccharides into desired products. We developed and experimentally validated a comprehensive modeling framework for *T. reesei* / *E. coli* (TrEc) consortia, allowing us to elucidate key ecological interactions, as well as mechanisms that permit stable and tunable population compositions. To illustrate bioprocessing applications, we demonstrated direct conversion of cellulosic feedstocks to isobutanol, achieving titers up to 1.86 g/L and 62% theoretical yield. These titers and yields represent the highest reported to date (April 2013) for production of advanced biofuels with consolidated bioprocessing, and were achieved without detailed optimization of culture conditions or nutrient supplementation other than minimal salts. Beyond isobutanol, the modularity of our system will allow it to be readily adapted to the large portfolio of existing *E. coli* strains metabolically engineered to produce biofuels or commodity chemicals, representing a key advantage over the conventional approach of using a single microbe, or "super-bug".

While the notion of using saccharolytic / fermentative (S/F) consortia to produce fuels and chemicals from biomass is not new and has been tried a number of times, our comprehensive theoretical and experimental studies of the TrEc model system have yielded new insights into the underlying ecological interactions and properties of S/F consortia. For example, our results

suggest that growth and substrate uptake kinetics play a major role in determining carbon flow partition between *T. reesei* and *E. coli*, which in turn determines the tradeoff between substrate hydrolysis rate and potential product yields. As another example, we show that cooperator-cheater dynamics within the TrEc consortium lead to stable coexistence between *T. reesei* and *E. coli*, and that equilibrium population composition can be tuned by modulating relative cooperation/cheating benefits, and possibly other ecological parameters. More broadly, cooperator-cheater tuning could serve as a general tool for designing and implementing microbial consortia, and could be accomplished via manipulation of culture conditions as we have done or through genetically programmed cellular behavior. Overall, by demonstrating a promising strategy for direct conversion of cellulosic biomass to biofuels and engineering microbial consortia with stable and controllable population compositions, this work makes substantive contributions to the state-of-the-art in biofuels and microbial engineering. Ultimately, we envision that synthetic microbial consortia will become increasingly important in industrial bioprocessing, and this work provides a valuable foundation for future studies. As our understanding of the complex interactions and dynamics in microbial communities improves, it is likely that engineered microbial consortia will become the most cost-effective route for biosynthesis of many different fuels and chemicals from biomass, and may prove useful for other biotechnology applications, such bioremediation or medicine.

As a final perspective, we would like to point out that our results illustrate the importance of engineering and testing microbial strains under conditions representative of their intended usage (e.g. industrial scale fermentation). In the course of this study, we examined three different *E. coli* strains engineered for isobutanol production (JCL260 [1], NV3 [3], and NV3r1 [3], all with plasmids pSA55 and pSA69), generously provided to us by the lab of James Liao (Chemical and Biomolecular Engineering, UCLA). Through the use of rich media (i.e. yeast extract supplementation) and glucose/nutrient feeding cycles, the Liao lab was able to reach isobutanol titers $>$20 g/L with strains JCL260 [1] and NV3r1 [3], while titers were lower for NV3 [3]. However, when we tested these strains on minimal media (more representative of industrial biofuel fermentations), titers were much lower; for example strain NV3 produced only 3.5$\pm$0.1 g/L isobutanol in minimal media vs. 13.6 g/L in rich media with feeding cycles. We also noted that performance of these strains varied greatly between monoculture and co-culture with *T. reesei*. Under co-culture conditions, all three strains showed evidence of instability, including plasmid loss and high variance in isobutanol titers/yield between replicate cultures (see Figure 2.7). Interestingly, NV3 achieved the highest and most consistent titers under co-culture conditions, despite that this strain had the poorest titers/yields under monoculture conditions. Further investigation revealed that fermentation product (succinate, acetate, ethanol, and isobutanol) distributions varied between monoculture and co-culture, with greatly reduced

isobutanol selectivity under co-culture conditions. These results thus demonstrate the dependence of strain performance on environmental and physiological context. This issue is also of consequence in our genome evolution and engineering work, and will be touched upon again in 6.2.2.2.

6.1.2 Future directions

6.1.2.1 Extending the TrEc consortium to include C5/C6 specialists

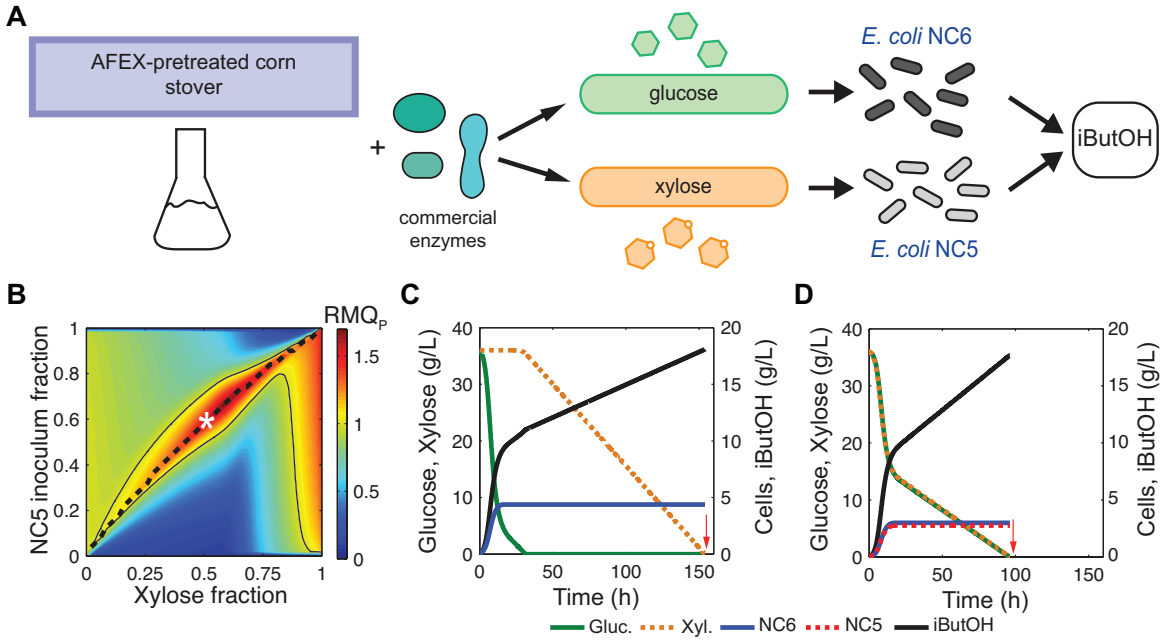

Figure 6.1: General scheme and model analysis of two member *E. coli* consortium for co-fermentation of hexose and pentose sugars into isobutanol. Adapted from [31]. **(A)** Process for converting AFEX-pretreated corn stover into isobutanol using *E. coli* specialist strains. The enzyme mixture breaks down the AFEX-pretreated corn stover, producing an enzymatic hydrolysate. The *E. coli* specialists then convert the sugars into isobutanol. Although the enzymes can hydrolyze the corn stover into glucose and xylose mono and oligosaccharides, the glucose oligosaccharides are currently not utilized by the consortium. **(B)** Predicted consortium productivities at 72 g/L total initial sugars. Numerical solutions to our ODE model were found over a range of NC5 inoculum fractions and xylose fractions, and used to calculate relative mean isobutanol volumetric productivity (RMQ_P; ratio of consortium productivity to diauxic monoculture productivity). The region where $RMQ_P > 1$ (i.e. consortium outperforms diauxic monoculture) is enclosed by solid black line; dashed line corresponds to optimal NC5 inoculum fractions. Sample numerical solutions for point marked by white * are shown in panels C and D. **(C)** Sample numerical solution for diauxic monoculture, corresponding to the white * in panel B (72 g/L total sugars; 0.5 g/g initial xylose fraction) **(D)** Sample numerical solution for two-member consortium, corresponding to the white * in panel B (72 g/L total sugars; 0.5 g/g initial xylose fraction; 0.58 g/g NC5 inoculum fraction).

Isobutanol production with *E. coli* C5/C6 specialists In conjunction with the TrEc consortium presented in this dissertation, we have also developed a consortium of pentose (C5) and hexose

(C6) *E. coli* specialists engineered for isobutanol production (Figure 6.1A) [31]; this work was done in collaboration with Alissa Kerner, another member of Xiaoxia Nina Lin's laboratory. Achieving efficient utilization of mixed hexose/pentose saccharides remains an important challenge in biofuel production from lignocellulosic biomass [16] [14]. Engineering hexose and pentose specialized microbes, each exclusively metabolizing it's respective carbon source, represents a promising strategy for co-utilization of mixtures of hexoses/pentoses. Co-cultures of hexose and pentose specialists would be expected to utilize both types of sugar simultaneously (in contrast to diauxic sequential utilization, observed in most natural species), thus improving the conversion rate of mixed sugars derived from lignocellulosic biomass. We created an isobutanol producing C5 specialist by knocking out key genes involved in glucose transport and assimilation ($\Delta ptsG$, $\Delta manX$, and Δglk) in *E. coli* NV3 pSA55/69, thus eliminating glucose metabolism and making this strain obligately dependent on pentose sugars (NC5 in Figure 6.1A). Unmodified *E. coli* NV3 pSA55/69 serves as the C6 specialist (NC6 in Figure 6.1A). Modeling and theoretical analysis suggest that the NC5/NC6 biculture will outperform diaxuic monocultures under many different conditions (see Figure 6.1B-D). Specifically, our model predicts that for a given sugar composition, there is a range of consortia population compositions over which which the two-member consortium will achieve higher mean volumetric isobutanol productivity than a diauxic monoculture (Figure 6.1B); as would be expected, the optimal NC5 inoculum fraction increases with pentose fraction (Figure 6.1B).

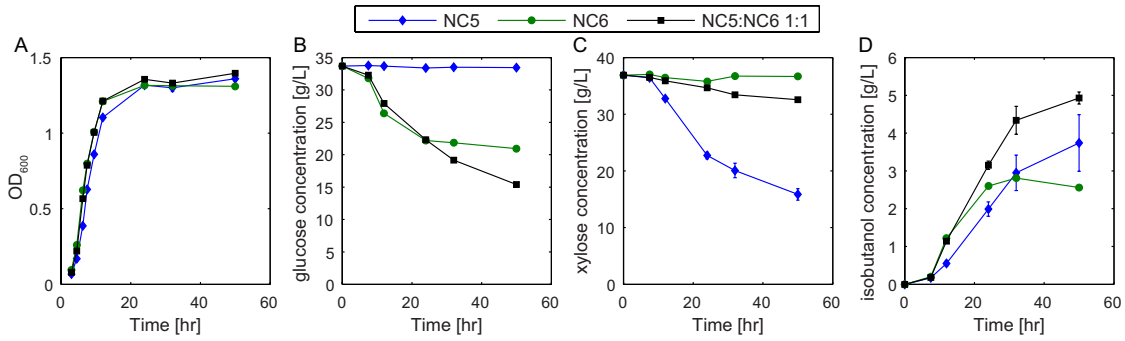

Figure 6.2: Mono and co-culture growth profiles **(A)**, glucose **(B)** and xylose **(C)** consumption, and isobutanol production **(D)** on defined media containing both glucose and xylose. Concentration of each sugar is 36 g/L, 72 g/L total, data shown for 72 hours. Error bars represent the ±S.D. of two biological replicates (with only one biological replicate for NC6). Adapted from [31].

Experiments performed by Alissa Kerner have substantiated that the NC5/NC6 biculture can outperform a diauxic NC6 monoculture. For example, NC5/NC6 mono and biculture studies on defined media demonstrated that the biculture utilized glucose and xylose simultaneously, and achieved higher xylose conversion and isobutanol titers compared to monocultures of either strain (sample data shown in Figure 6.2). More importantly, the biculture grows well

on real lignocellulosic biomass (AFEX pretreated corn stover) hydrolysate, producing up to 2.95 g/L isobutanol at ≈65% of theoretical yield without detoxification or additional nutrient supplementation. These results are detailed in a manuscript that is presently under preparation [31].

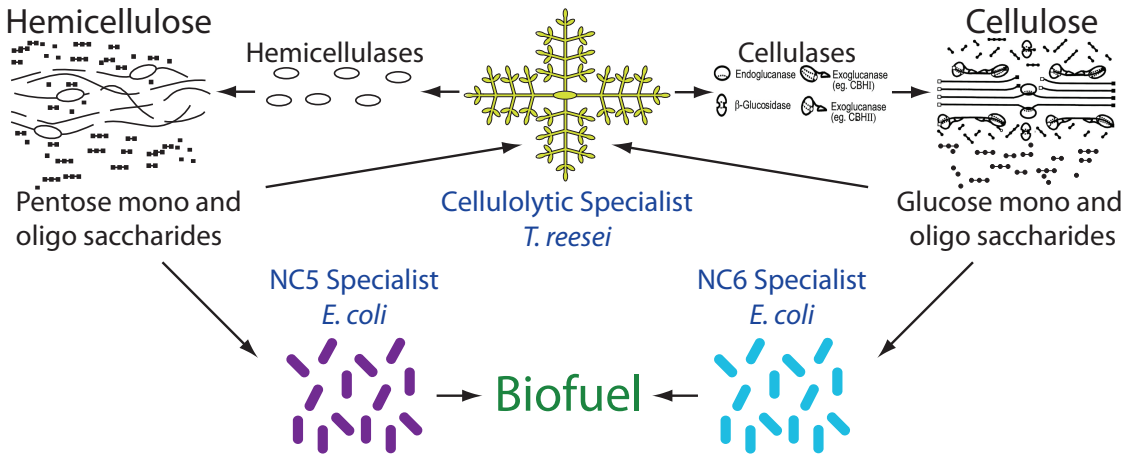

Figure 6.3: TrEc consortium with NC5/NC6 *E. coli* specialists. The *T. reesei* cellulolytic specialist hydrolyzes hemicellulose and cellulose into pentose and glucose mono and oligosaccharides, which are cofermented by the respective *E. coli* specialists into isobutanol.

Extending the TrEc consortium by combining C5/C6 specialists with *T. reesei* The ultimate goal for the NC5/NC6 consortium is to integrate it with *T. reesei* to form a three-specialist system for direct conversion of lignocellulosic biomass to isobutanol (Figure 6.3). The combined *T. reesei*/NC5/NC6 consortium may be able to achieve better performance compared to *T. reesei* / diauxic *E. coli* bicultures, especially at higher substrate loadings. To implement the three-member consortium, *T. reesei*/NC5/NC6 tricultures will need to be characterized on model substrates (defined mixtures of microcrystalline cellulose and purified xylan) and real lignocellulosic biomass over a range of different population compositions. In conjunction with experimental work, our modeling framework could be extended to include hemicellulose hydrolysis, pentose consumption, and kinetics for the NC5 specialist, allowing us to perform theoretical analyses to gain insights into consortium performance and behavior. While our theoretical and experimental results for NC5/NC6 bicultures suggest that the three-member consortium is a promising approach, several caveats must be mentioned. Co-cultures of *T. reesei* and diauxic *E. coli* NV3 pSA55/69 (NC6) on AFEX pre-treated corn stover achieved high conversion of pentose sugars while conversion of cellulose was incomplete (see Figure 2.9A & B), in contrast to results with NC5/NC6 bicultures on hydrolyzed AFEX pre-treated corn stover (e.g. Figure 6.2). Additionally, our theoretical analysis of the TrEc consortium suggests that substrate uptake and growth rates will be very low

during co-culture growth on cellulose. Low glucose flux through phosphoenolpyruvate-dependent phosphotransferase systems (PTSs) is expected to increase intracellular cyclic AMP (cAMP) levels, resulting in transcriptional activation of catabolite repressed operons [150]. Thus WT diauxic *E. coli* strains may actually co-utilize hexose and pentose sugars under co-culture conditions due to low substrate uptake and growth rates. This phenomena could be investigated by using transcriptional analysis to determine whether catabolite repressed genes are expressed in *E. coli* under co-culture conditions, or through ^{13}C flux analysis studies.

6.1.2.2 Spatial structure and cross-feeding: additional mechanisms for stability and tunability

In this work, we exploited cooperator-cheater dynamics as a mechanism for stabilizing and tuning population composition in TrEc consortia. However, it is important to consider the long-term evolutionary stability of cooperator-cheater systems. While ecological motifs such as metabolic cross-feeding [42, 106] and cooperator-cheater dynamics may exhibit stable short-term behavior, these systems are not necessarily stable over long evolutionary timescales, especially in homogenous environments. For example, consortia based on metabolic cross-feeding are prone to invasion by cheating strains which consume nutrients secreted by the partner strain without reciprocation [221]. Additionally, in experimental evolution studies with cross-feeding *E. coli* auxotrophs, the cross-feeding strains eventually evolved to become metabolically independent (Xiaoxia Nina Lin, personal communication). However, numerous studies have shown that spatially structured environments promote the evolution and persistence of cooperation in microbes [221]. In contrast to homogenous (e.g. well-mixed) environments, in spatially structured environments each individual microbe interacts with a limited number of neighbors, creating local feedback that can reward cooperators and punish cheaters [221]. The effect of spatial structure on cooperation was recently demonstrated with a symbiotic microbial system consisting of a methionine secreting *Salmonella enterica* mutant (met+ *S. enterica*) and an *E. coli* methionine auxotroph (met- *E. coli*) [221]. On lactose minimal medium, the two strains are interdependent: met- *E. coli* ferments lactose to acetate, which is metabolized by met+ *S. enterica*, while met+ *S. enterica* secretes methionine to support met- *E. coli* [221]. Methionine secretion is costly for met+ *S. enterica*, and when met- *E. coli* and met+ *S. enterica* are cultured in a well mixed environment, the population is subject to invasion by non-methionine secreting *S. enterica*, which has a higher growth rate than met+ *S. enterica* [221]. In contrast, a spatially structured 2-D environment with isolated patches of *E. coli* / *S. enterica* favored the persistence of cooperation; patches dominated by cooperators engender more growth than cheater-dominated patches, thus increasing the proportion of cooperators in the total population [221].

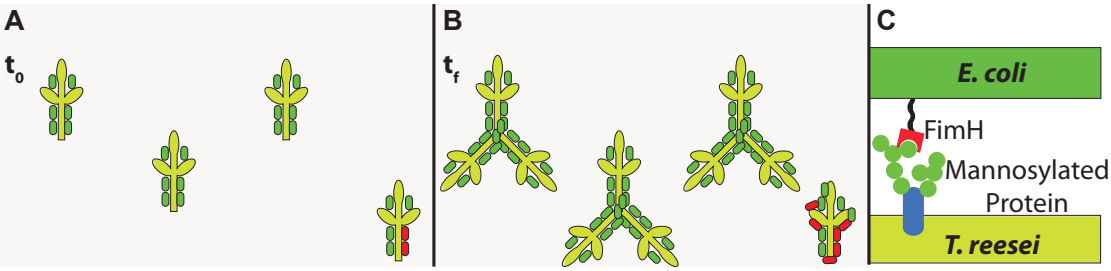

Figure 6.4: Spatially structured *T. reesei* / *E. coli* consortium featuring substrate surface colonization and adhesion between *T. reesei* and *E. coli*. **(A)** Consortium at time t_0. WT *E. coli* cells are indicated by green, while red cells represent a fast growing mutant (i.e. $\mu_{max,MUT} > \mu_{max,WT}$) **(B)** Consortium at time t_f. **(C)** Proposed system for fimbrial adhesion of *E. coli* to *T. reesei*.

We suggest that spatial structure may also be beneficial for maintaining stability of TrEc consortia and other systems with cooperator-cheater dynamics. Over long time-scales, we expect that *E. coli* and *T. reesei* will evolve in the co-culture environment, ultimately leading to changes in ecological parameters (i.e. Table 3.1) that shift population equilibria. Since mutation rates are generally higher in bacteria compared to filamentous fungi [222], it is likely that *E. coli* will evolve faster than *T. reesei*, which in the case of a well-mixed homogenous environment may lead to increased equilibrium X_{Ec} over time. We may be able to improve the stability of *E. coli* growth rate by introducing spatial structure to the consortium. Figure 6.4 depicts a proposed spatial structure scheme featuring substrate surface colonization and *E. coli* adhesion to *T. reesei* mycelium, and illustrates how this structure might maintain growth rate stability in *E. coli*. At time t_0 the substrate surface is colonized by patches of *T. reesei* mycelia with adherent WT *E. coli* (green) and a fast growing mutant (red) (Figure 6.4A). The fast growing mutant *E. coli* will utilize a greater proportion of cellulose hydrolysis products, reducing growth rate and cellulase production in adherent mycelia. After a period of time t_f, mycelia with WT *E. coli* have outgrown mycelia with fast growing mutant *E. coli*, and the overall proportion of WT *E. coli* has increased despite the higher growth rate of the mutant; thus spatial structure could mitigate invasion by faster growing mutants (Figure 6.4B).

Promoting surface colonization by *T. reesei* and engineering fimbrial adhesion of *E. coli* specialists to *T. reesei* mycelium (Figure 6.4C) could create the spatial structure depicted in Figure 6.4A&B. *T. reesei* can be encouraged to colonize substrate surfaces by choosing appropriate culture conditions, such as low shear submerged culture or solid state fermentation [108, 223]. Many strains of *E. coli* naturally produce type I fimbriae, which facilitate binding to host epithelial cells [224]. Fimbrial binding is mediated by mannose-specific FimH adhesins, located at fimbriae tips (Figure 6.4C) [224]. Since *T. reesei* has a high proportion of mannosylated

proteins in the cell wall, fimbrial adhesion of *E. coli* to *T. reesei* mycelium seems highly plausible (Figure 6.4C) [225]. Unfortunately, most non-pathogenic laboratory strains of *E. coli* weakly express fimbriae. Fimbriae expression can be improved by knocking out *fimE* and overexpressing *fimB*, which are negative and dual regulators of fimbriae genes, respectively [226]. Thus we could construct Δ*fimE* / *fimB* overexpressing strains and test for *T. reesei* adhesion using agglutination and microscopy assays, and if warranted investigate the effects of *fimB* expression on adhesion. To test the effects of spatial structure on the consortium, we could perform coculture studies with *T. reesei* and mixtures of a slow/fast growing adherent (i.e. Δ*fimE* / *fimB* overexpressing) strains or slow/fast growing non-adherent strains. If spatial structure creates the anticipated local-feedback, we would expect co-cultures with adherent *E. coli* to be resistant to invasion by faster growing strains. We expect that *E. coli* adhesion to *T. reesei* mycelium would alter local glucose concentration gradients, thus changing cooperator-cheater dynamics and shifting population equilibria. Additionally, fimbriae expression may be deleterious to isobutanol tolerance in *E. coli* (see Chapter 4). These potential issues would need to be carefully considered in the design and implementation of the proposed spatial structure, as well as choice of culture conditions.

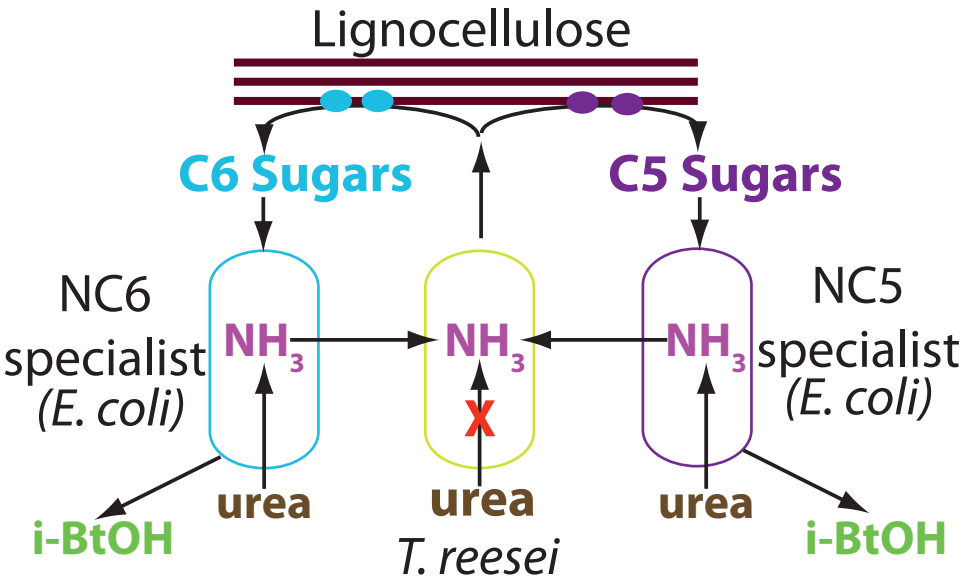

Figure 6.5: Proposed nitrogen/carbon metabolic cross-feeding circuit for TrEc consortium with NC5 and NC6 specialists.

An alternative or additional approach for achieving stability/tunability in TrEc consortia is engineering a symbiotic relationship between *T. reesei* and *E. coli*. Engineering a synthetic nitrogen/carbon metabolic cross-feeding circuit between *E. coli* and *T. reesei* is one possibility that we suggest exploring (Figure 6.5). In this system, fermentation media containing urea as the sole nitrogen source would be utilized. The *T. reesei* urea assimilation pathway will be disabled by

deleting the gene(s) for urease, while *E. coli* will be engineered to express a heterologous urease (such as *ureABCEFG* from *Klebsiella aerogenes* [227]) that hydrolyzes urea to NH_3 and CO_2 (Figure 6.5). NH_3 can then be secreted out of *E. coli* cells (via membrane diffusion and the Amt facilitated transport system) into the medium where it is available to *T. reesei*. Thus the relationship between *T. reesei* and *E. coli* would be symbiotic: *T. reesei* secretes cellulases and hemicellulases that hydrolyze lignocellulose to soluble saccharides, while *E. coli* hydrolyzes urea, providing both organisms with nitrogen (Figure 6.5). The proposed symbiosis is expected to be stable in a spatially structured environment, as discussed above [221]. Besides creating a stable interdependence between *T. reesei* and *E. coli*, the suggested cross-feeding circuit could also provide mechanisms for tuning population composition. A key drawback to this design is that it requires urea as the sole nitrogen source. If the cellulosic substrate contains other nitrogen sources, then the cross-feeding relationship will be obviated. Some cellulosic feedstocks such as waste paper or cardboard have very low nitrogen levels, and thus nitrogen would be supplied by the fermentation media. However, most types of lignocellulosic biomass will have significant nitrogen content (e.g. proteins, amino acids, etc.), and biomass pre-treated by AFEX may contain residual NH_3. These types of issues would need to be considered in the design and deployment of any kind of symbiotic cross-feeding circuit.

6.1.2.3 The TrEc consortium as modular and flexible platform for consolidated bioprocessing

While we demonstrated isobutanol production as a proof-of-concept application of the TrEc consortium, this system represents a modular and flexible CBP platform that can be readily adapted to biosynthesize a variety of products from lignocellulosic feedstocks. The modularity of our system allows it to leverage the vast existing portfolio of *E. coli* strains metabolically engineered to produce valuable products, many of which could be swapped directly into the TrEc consortium in a "plug-and-play" fashion. As a caveat, if desired products are excessively toxic, tolerance of both *T. reesei* and *E. coli* will need to be improved to achieve efficient fermentation. Additionally, *E. coli* strains optimized for monoculture conditions may require additional metabolic engineering and other improvements to produce desired products at high yields/titers under co-culture conditions (e.g. as discussed in section 2.4.1).

The TrEc consortium requires aerobic metabolism of desired products, but this constraint is not excessively prohibitive. For example, *E. coli* can be readily engineered to produce anaerobic metabolites under aerobic conditions via elimination of respiratory chain components, as discussed in the introduction to Chapter 2 [75]. Furthermore, many next-generation biofuels and commodity chemicals are produced through biosynthetic non-fermentative pathways with substantial ATP

and NADPH requirements, and thus it may actually be more efficient to produce these products aerobically or microaerobically. Examples include products of isoprenoid pathways (such as farnesene and isopentenol), products of fatty acid pathways (such as alkanes, alkenes, fatty alcohols, and fatty acid ethyl esters), and 1,3-propanediol[1] [228]. Indeed, the majority of promising value-added building block chemicals that can be produced biologically will likely require aerobic metabolic pathways (Figure 6.6) [229] . Many fine and specialty chemical products, such as amino acids and vitamins, are also produced via aerobic fermentation. While we developed the TrEc consortium with an eye towards consolidated bioprocessing of lignocellulosic biomass into fuels and commodity chemicals, it may also prove useful for production of higher-value products from biomass. Beyond the TrEc consortium, other fermentation specialists could be readily used in place of *E. coli* (e.g. *S. cerevisiae*), and other cellulolytic specialists are possible as well. Since many cellulolytic microbes have cell-wall localized (i.e. complexed) cellulase systems [14], cooperator-cheater dynamics could be preserved; additionally, use of cellulolytic microbes with different levels of privileged access to hydrolysis products represents a possible means of tuning/optimization.

[1]Glycerol can be fermented to 1,3-propanediol anaerobically, but production from glucose requires aerobic metabolism

Identification of Actual and Potential Pathways to Building Blocks from Sugars						
Building Blocks	Yeast or Fungal	Bacterial	Yeast or Fungal	Bacterial	Chemical & Catalytic Processes	Biotrans-formation
	AEROBIC FERMENTATIONS		ANAEROBIC FERMENTATIONS		CHEM-Enzyme TRANSFORMATIONS	
3 Carbon	Commercial Product - C	Commercial Product - C	Commercial Product - C	Commercial Product - C	Commercial Product - C	Commercial Product - C
3-Hydroxy propionic acid	X	X				
Glycerol	X	X	X	X	C	
Lactic acid	X		X	C		
Malonic acid	X				X	
Propionic acid				X		
Serine	X	C				C
4 Carbon						
3-Hydroxy butyrolactone					X	
Acetoin	X	X		X		
Aspartic Acid	X	X				X
Fumaric Acid	X	X			X	
Malic acid	X	X				
Succinic acid	X	X		X		X
Threonine	X	C				
5 Carbon						
Arabitol	X		X		C	X
Furfural					C	
Glutamic	X	C				
Itaconic Acid	C					
Levulinic acid					X	
Xylitol	X		X	X		C
6 Carbon						
2,5 Furan dicarboxylic acid					X	
Aconitic acid	X					
Citric acid	C					
Glucaric acid	X	X			X	
Gluconic acid	C	X				X
Levoglucosan					X	
Lysine	X	C				
Sorbitol	X	X			C	X
Number in each Pathway category*	21	14	4	6	11	7
Commercial processes	3	4	0	1	4	2

Figure 6.6: Summary of pathways for conversion of sugars to building block chemicals. Adapted from [229].

6.2 Part II: Genome evolution and engineering for elucidating the genetic architecture of isobutanol tolerance in *Escherichia coli*

6.2.1 Concluding remarks

In this work, we studied and improved isobutanol tolerance in *E. coli* by integrating experimental evolution, genome re-sequencing to identify acquired mutations, genomic and functional dissection to reverse engineer mechanisms of tolerance, and multiplex genome engineering to explore a large combinatorial genotype space for improved variants. Consistent with the complex nature of solvent tolerance, our experimental evolution and genomic studies revealed adaptations in a diversity of cellular processes. By examining genotypic adaptation in multiple independent lineages, we find evidence of parallel evolution in genes *marC*, *hfq*, *mdh*, *acrAB*, *gatYZABCD*, and *rph*. Many isobutanol tolerant lineages show reduced RpoS activity, perhaps related to mutations in *hfq* or *acrAB*. Collectively, the genotypic adaptations we observe suggest mechanisms of adaptation to isobutanol stress based on remodeling the cell envelope and, surprisingly, stress response attenuation. Guided by these results and findings from literature, we selected a set of 38 target mutations for multiplex genome engineering. Populations generated through genome engineering contained a diversity of novel genotypes and phenotypes. We identified a large cluster of highly tolerant genotypes that share a set of common mutations, including *gltA*, *tnaA*, *yfgO*, *gltD*, *glnE*, *rph*, and *marC*. These core mutations are derived from four different origins (three independent evolutionary lineages plus one from literature), and the significance of some of these loci in isobutanol tolerance was uncertain prior to this study. The *gltA*, *gltD*, and *glnE* mutations appear to be convergent metabolic adaptations that may increase intracellular levels of glutamine (a precursor for key cell envelope components), representing a potential new mechanism of adaptation. The best strain generated through our genome engineering work had higher fitness at 1% (w/v) isobutanol and fewer mutations compared to the best evolved strain, and thus through genome engineering we have effectively refactored isobutanol tolerance into simpler and higher fitness genotypes.

This work makes a number of important contributions to the current state-of-the-art in microbial alcohol tolerance and phenotype improvement methods. We generated a large library of isobutanol tolerant variants that could potentially serve as chassis strains for isobutanol production. Additionally, we have identified the underlying genetic bases of several important mechanisms of isobutanol tolerance, providing a foundation for further efforts to improve this phenotype. Beyond our contributions to isobutanol tolerance, this work provides valuable fundamental insights into

engineering complex phenotypes. In our evolutionary-genomics studies of isobutanol tolerance, we observed a trend of evolution targeting post-transcriptional regulation and high centrality nodes of biochemical networks. Phenotype improvement methods based on mutagenizing global transcription factors have been developed, but thus far perturbation of other regulatory or biochemical network hubs has not been explored. We suggest that post-transcriptional regulators (such as Hfq or RNA helicases) and perhaps protein interaction network hubs (such as GroEL and components of the Sec apparatus) are interesting mutagenesis targets for future global phenotype engineering.

Our results also demonstrate the importance and impact of epistasis in phenotype engineering. Mutation reconstruction and phenotyping studies revealed a striking predominance of positive epistatic effects in isobutanol tolerance. We observed evidence of positive epistasis between *hfq* or *marC* and each of *groL*, *mdh*, and *rph* mutations. Interestingly, individual *groL*, *mdh*, and *rph* mutations had neutral effects on isobutanol tolerance, suggesting that these mutations may be compensating adaptations to maladaptive side effects of global regulatory changes cause by *hfq* or possibly *marC* mutations. Such evolutionary patterns (i.e. global effect mutations and subsequent compensatory mutations) are not unique to isobutanol tolerance, and have been observed for a variety of phenotypes [123]. Many standard target gene identification strategies (such as knockout or overexpression libraries) are unable to capture interactions between distal genetic loci, and are therefore unable identify these types of adaptations. Thus while experimental evolution and genome resequencing studies are laborious and require relatively long time scales, these studies are of great value for studying highly epistatic phenotypes.

In addition to target gene identification, epistasis is also an important factor to consider in directed genome engineering studies. Several recent studies have demonstrated that negative epistasis between beneficial mutations may be a statistically common phenomena, both in evolution [117, 207] as well as integrated trait mapping / directed genome engineering studies [206]. In our work, we used general patterns of epistasis to guide selection of target genes so as to minimize the likelihood of negative epistatic effects, namely by avoiding combinations of genetic loci with functional redundancy and seeking loci with functional interactions. We suggest that this strategy could be of broad utility in future genome engineering studies. Epistasis can also affect genome engineering by altering fitness landscapes, especially in the case of global regulators [214]. Thus while we wish to generate synergistic mutation combinations, if epistatic mutations become fixed in populations too early this could constrain accessible evolutionary trajectories and limit exploration of combinatorial search spaces. To explore this notion, we conducted genome engineering studies with low and high dosage of the epistatic *hfq* mutation.

Our results demonstrate distinctly different mutation profiles between low *hfq* and high *hfq* dosage lineages, with *hfq* and associated epistatic mutations reaching much greater frequencies in the high dosage populations. Qualitatively, our results suggest that low *hfq* dosage populations have greater genotype and phenotype diversity than the high *hfq* populations, and we ultimately identified isolates with the highest isobutanol tolerance from low *hfq* populations. The effects of epistatic interactions on combinatorial genome engineering will need to be further studied in future work.

6.2.2 Future directions and perspectives

6.2.2.1 Increasing scope, scale, and throughput of our integrated phenotype engineering methodology

In this section, we describe potential ways to leverage new technologies in DNA synthesis, next-generation sequencing, and microfluidics to expand the scope, scale, and throughput of our methodology for studying and improving complex phenotypes. The use of experimental evolution represents a key limiting step in our current approach. We suggest that integrating rapid target gene identification methods (knockout/overexpression libraries, TRMR, etc.) with experimental evolution may be a good strategy for improving scope and productivity. For example, TRMR could be used to generate a library of single gene upregulation/downregulation variants with improved fitness. The single gene variant library could then be subjected to a short round of experimental evolution (e.g. 25-50 generations) and genome resequencing to identify subsequent beneficial mutations. Throughput of this approach could be increased by use of liquid handling robotics systems for culturing and colony screening and highly multiplexed libraries for genome resequencing. In addition to this proposed approach, it would be of great value to develop rapid target gene identification methods that can capture interactions between distal genetic loci. Conducting TRMR recursively (described in Chapter 1) is a particularly promising approach, but has not yet been reported in the literature as of this writing (April 2013).

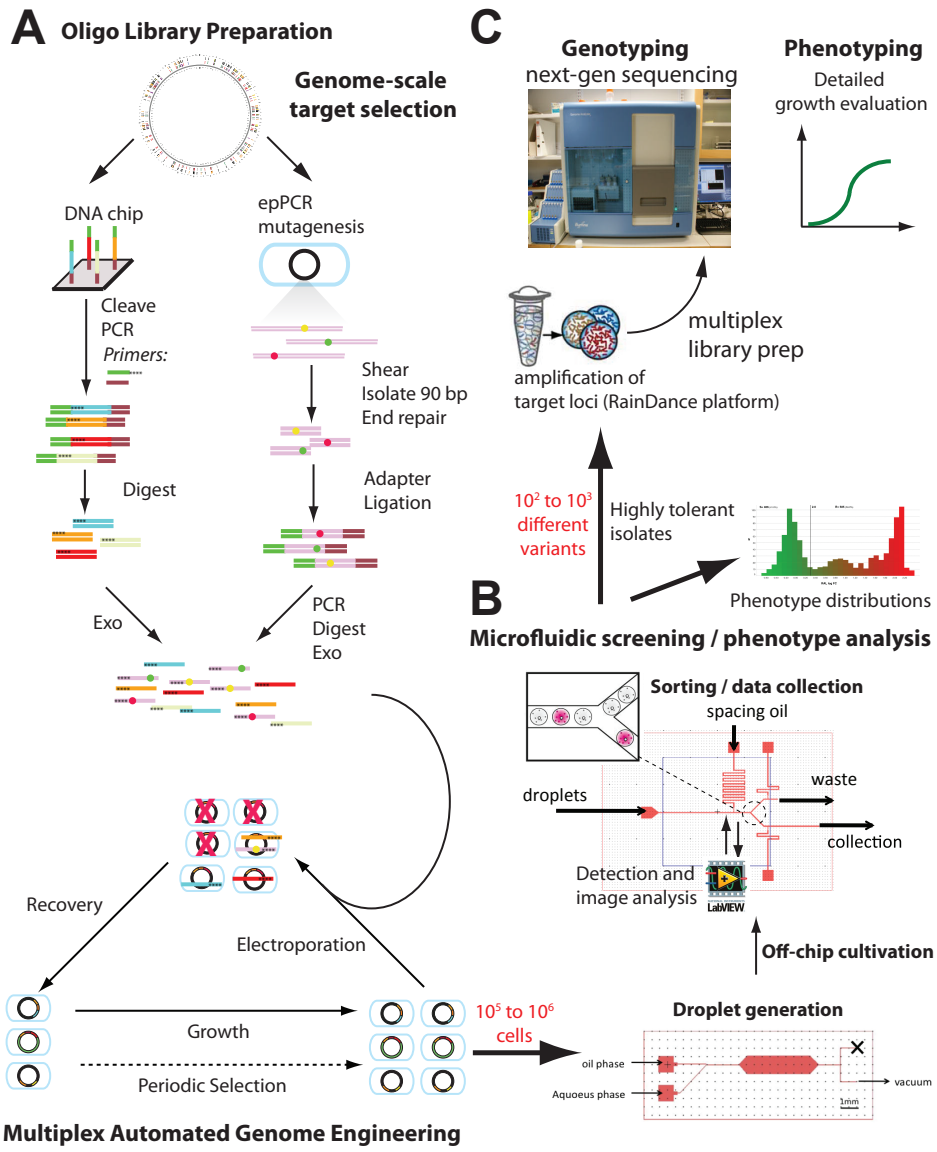

Figure 6.7: Future directions in genome engineering: expanding scale and throughput. **(A)** Proposed methods for synthesis of high complexity oligonucleotide libraries. For DNA chip based libraries, 90 bp MAGE oligo sequences will be flanked by ≈20 bp PCR priming sites incorporating type IIs restriction sites, such that the priming sites can be scarlessly cleaved from PCR products. The designed oligo library will be synthesized on a DNA microchip and then PCR amplified. PCR primers in the same strand as the required ssDNA product will have four degenerate, phosphorothioate 3 nucleotides. After amplification and restriction digestion to remove priming sites, the 90 bp dsDNA library will be treated with a 5'→3' exonuclease to selectively degrade the unwanted strand yielding the desired lagging-strand targeted 90-mer ssDNA library with 5' phosphorothioate nucleotides. For random mutagenesis of genetic loci, we will employ an epPCR based approach for generating an oligo library. Selected genetic loci will be PCR amplified and then epPCR will be used to generate mutagenized PCR products. The epPCR products will be sheared, and ≈90 bp fragments will be isolated and end repaired. The products will then be ligated to dsDNA adapters containing PCR priming sites similar to those above, and the library will then be PCR amplified, restriction digested, and exonuclease digested as above. **(B)** High throughput microfluidic phenotyping and sorting system. Portions of figure adapted from [230] and [231]. **(C)** Genotyping and phenotyping highly tolerant isolates. Portions of figure adapted from [232].

We used a relatively small library (38x targets; combinatorial search space size of $2^{38} \approx 2.75\text{x}10^{11}$ genotypes) for the multiplex genome engineering work described in Chapter 5. Expanding genotype search spaces to larger scales would be of great benefit for engineering complex phenotypes such as isobutanol tolerance, but requires more sophisticated methods for designing and synthesizing oligonucleotide libraries, and for screening, phenotyping, and genotyping. Due to the small library size used in Chapter 5, we were able to use conventionally synthesized oligonucleotides. For larger libraries, conventionally synthesized oligonucleotides will be prohibitively expensive, and more scalable synthesis methods must be employed. The homologous recombination procedure used in MAGE requires 90-mer ssDNA (containing a mutation of interest with flanking homologous sequence) targeted to the lagging strand of the DNA replication fork and having four 5' phosphorothioate modified nucleotides [59]. Generating a complex oligonucleotide library conforming to MAGE requirements is not trivial and requires novel design strategies. For constructing precise point mutations (e.g. reconstructing specific mutations found from genome resequencing), oligonucleotide libraries on the order of 10^4 to 10^5 sequences could be synthesize on DNA chips, with subsequent amplification and processing to produce ssDNA with the required 5' phosphorothioate modifications (see Figure 6.7A for more details). For random mutagenesis of genetic loci (e.g. focused mutagenesis of global transcription factors), we suggest using error-prone PCR (epPCR) to generate mutant sequence libraries, with subsequent processing to produce the required ≈90-mer ssDNA with 5' phosphorothioate modifications (see Figure 6.7A for more details). For the epPCR method, library complexity will depend on the number of targeted loci and epPCR mutation rate, which will need to be modulated to maintain reasonable library sizes (e.g. 10^7 to 10^8 sequences). Both of these proposed methods will require extensive testing and optimization, and the final oligo libraries will need to be validated. Beyond point mutations (i.e. SNPs and small indels), it would also be desirable to create larger mutations (e.g. promoter replacements to generate overexpression libraries). MAGE has recently been modified to permit construction of longer mutations (i.e. insertions $>$10 bp) via the co-selection MAGE (CoS-MAGE) method [61] and also for selection of highly modified cells via co-operative oligonucleotide co-selection [62]. These new methods could be incorporated into future genome engineering work.

In conjunction with expanding genotype search spaces, higher throughput methods for screening, phenotyping, and genotyping combinatorial libraries are also needed. We are currently working with the lab of Mark Burns (Chemical Engineering, University of Michigan) to develop a high throughput microfluidic phenotyping and sorting system for screening combinatorial libraries. The system consists of microfluidic devices for encapsulating cells in microdroplets (developed by Jihyang Park; [230]), and for sorting microdroplets based on fluorescence intensity

(originally designed by Jihyang Park and currently under development by Mathieu Rossion and Sida Wang as of April 2013) (Figure 6.7B). The first device generates aqueous droplets (consisting of media with cells) dispersed in a continuous oil phase; by adjusting the concentration of cells in the aqueous phase, we can achieve a density of ≈1 cell/droplet (Figure 6.7B). The droplets can then be removed from the device and incubated off-chip in a microtube. After cultivation, the droplet library will be sorted on a second microfluidic device based on their level of growth (i.e. fluorescence); if selection is applied (e.g. isobutanol added to media used for droplets), then we would expect higher growth for more tolerant cells. The sorting device consists of a split junction, and the flow of droplets to the collection or waste channels can be controlled via pressure (Figure 6.7B). Droplets will be electronically imaged with fluorescence microscopy, and then sorted based on threshold fluorescence[2]. In addition to selecting droplets with high growth (i.e. tolerance), droplet images can be saved and analyzed, yielding insights into growth phenotype distributions in combinatorial populations (Figure 6.7B). Currently the sorting device can operate at a rate of ≈1 droplet/s, but we hope to increase the throughput in future work so that it will be possible to screen large libraries (i.e. 10^6 or more droplets).

After droplet sorting, the isolated highly tolerant variants can be further characterized off-chip. For the genome engineering work described in Chapter 5, we designed and optimized allele specific PCR (ASPCR) assays to genotype target mutations. However, due to the cost and labor involved this approach is intractable for larger numbers of mutations; additionally, ASPCR would not be applicable to sequences subjected to random mutagenesis. Therefore different methods will need to be developed and optimized for genotyping large numbers of target loci with potentially unknown mutations. Recently developed technologies for target enrichment and next-generation sequencing could be deployed for this work, which we describe briefly here. For target enrichment, the RainDance[3] platform could be used for multiplex amplification of target loci from genome engineering isolates (Figure 6.7C). The RainDance technology is similar to emulsion PCR but uses a microfluidic system to create droplets that contain different pairs of preselected primers, with up 4,000 primer pairs in single reaction set. The droplets are collected in a single tube and then cycled similarly to emulsion PCR. Enriched target loci can then be used to prepare barcoded libraries for multiplex next-generation sequencing (Figure 6.7C). Using this general approach, up to ≈100 isolates can be genotyped in a single experiment. For phenotyping, we have previously developed 96 and 384 well microplate based assays (Chapter 4) that allow detailed characterization of growth kinetics at different isobutanol concentrations; these assays should

[2]This method requires fluorescent markers; we have modified standard genome engineering host strain *E. coli* EcHW24 to express Yfp, as described in section 5.5.2

[3]http://www.raindancetech.com

be sufficient for phenotyping sample libraries of ≈100 isolates. Detailed phenotype/genotype characterization of greater than 10^2 or 10^3 isolates will require the development of more advanced methods.

Continued advances in DNA synthesis, genome engineering, next-generation sequencing, and microfluidics technologies will enable further increases in the scale and throughput of phenotype engineering work. We envision that in the future, much of the experimental work required for phenotype engineering could be miniaturized and automated, allowing for rapid hands-off engineering and study of complex phenotypes. Additionally, genome engineering technologies may eventually be extended beyond *E. coli* to allow for integrated phenotype engineering in other bacteria or fungi hosts.

6.2.2.2 Perspectives

Due to the complex multigenic nature of chemical tolerance phenotypes, most phenotype improvement approaches developed to date follow a "top-down" combinatorial approach of generating diversity in a population and then studying isolates with desired traits. Recent advances in DNA synthesis technologies have enabled de novo synthesis and assembly of multigene constructs, chromosomes, and even entire microbial genomes [63]. These new synthesis technologies will allow for bottom up design and construction of large-scale synthetic gene networks and perhaps even synthetic organisms, offering new opportunities for engineering of complex cellular behaviors and phenotypes. Advances in the mechanistic understanding of chemical tolerance phenotypes combined with new technologies for large-scale editing and synthesis of genetic information may eventually permit a synthetic biology paradigm of "bottom-up" rational programming of complex chemical tolerance phenotypes. However, this will require a more complete and quantitative understanding of mechanisms of tolerance, as well as expanded tools for the design and construction of complex genetic circuits and signal transduction pathways. *In silico* modeling may prove useful in this regard. As a recent example, modeling of synthetic control systems for efflux pumps suggested that feed-forward control of efflux pump expression in response to biofuel concentration might improve fermentation performance [233]. Genome-scale metabolic network models could be used in conjunction with constraint-based optimization techniques to predict genetic manipulations to maximize/minimize a tolerance or toxicity objective function, provided that tolerance/toxicity mechanisms that can be represented in terms of metabolic fluxes. Examples of such mechanisms could include maintenance ATP requirements and changes in the synthesis and composition of cell envelope components (i.e. LPS, peptidogolycan, and lipids).

Hybrid rational-combinatorial methods for phenotype improvement are possible (i.e. rational engineering of a phenotype and then applying combinatorial approaches for optimization), and perhaps represent the best approach at present. Many recent studies demonstrate rational engineering approaches to improve chemical tolerance phenotypes based on established mechanisms of toxicity or tolerance. Examples from literature include overexpressing heat-shock proteins to improve alcohol tolerance in *E. coli* [234], engineering trehalose accumulation [235] and lipid composition [236] in *S. cerevisiae* to improve ethanol tolerance, overexpression of efflux pumps to improve tolerance of *E. coli* [237] to hydrophobic biofuel molecules, and engineering mechanisms for detoxification of alcohol induced reactive oxygen species (ROS) in *E. coli* [238]. However, all of these approaches focus on improving tolerance via a single mechanism; future work needs to integrate multiple mechanisms of tolerance to achieve maximum phenotype improvement.

Research in microbial chemical toxicity and tolerance is usually motivated by the need to improve fermentation titers and productivity. However, tolerance to a toxic product does not necessarily confer improved production, as exemplified by the Atsumi *et al.* study of isobutanol tolerance in *E. coli* [124]. Indeed, metabolic degradation is a common mechanism of tolerance to toxic chemicals [50]. As a striking example, Goodarzi *et al.* revealed that ethanol degradation and assimilation was a dominant component of adaptation in an ethanol tolerant *E. coli* mutant [121]. These studies highlight the need for new methods of co-selecting tolerance and production phenotypes. Given that adaptations can be highly specific to a particular environment or physiological context, it is also important to perform screens/selections under conditions that mimic those in industrial scale production. For example, many studies perform screens and/or selections in rich media, even though such media is not likely to be economically viable for commercial production of biofuels. In our work we discovered that adaptations acquired in evolved isobutanol tolerant strains exhibited antagonistic pleiotropy between minimal and rich medium, demonstrating the importance of considering environmental and physiological context (Chapter 4). Tolerance adaptations may also vary between different growth phases (i.e. exponential vs. stationary phase), thus the growth phase in which biofuel production occurs must also be considered in selection/screening. As an example, in *E. coli* RpoS attenuation was found to be improve growth [3], isobutanol production [3], and isobutanol tolerance (Chapter 4) in exponential phase, but decreased isobutanol production and survival in stationary phase, leading to overall reduced titers and yields [3]. Finally, it is also possible that tolerance mechanisms may differ between exogenous exposure to toxic chemicals compared to endogenous production. For example, cell envelope modifications may protect cells from exogenous alcohol toxicity by improving envelope integrity and/or reducing diffusion of exogenous alcohols into the cell. However, if such mechanisms likewise prevent diffusion of endogenously produced alcohols out of the cells, they may not

improve tolerance in production strains and may even be detrimental under such conditions.

As a case in point, Woodruff *et al*. recently developed a chassis ethanol producing *E. coli* strain (LW06) for phenotype engineering [239]. Several genes previously identified to improve tolerance to exogenous ethanol in a WT non-production strain were overexpressed in LW06 [239]. None of the tested genes improved tolerance or ethanol titer under production conditions, underscoring the importance of the caveats discussed above [239]. In a followup study, Woodruff *et al*. applied genomic library enrichment under ethanol production conditions to identify genes improving tolerance and production in LW06, leading to identification of a different set of genetic loci compared to a similar study with WT strains [167, 240]. In our own work, we attempted to construct an isobutanol producing chassis strain for genome engineering by transducing λ-Red to *E. coli* JCL260 [1] and using oligo recombination to inactivate MutS. However, λ-Red produces a severe growth defect in the JCL260 background (data not shown), perhaps due to interactions between the JCL260 genotype and λ-Red or an unknown mutation co-transduced with λ-Red. For future work, we suggest developing an isobutanol producing strain for genome engineering using a more robust base strain such as NV3 or NV3r1 [3].

APPENDIX A

Modeling the *T. reesei-E. coli* (TrEc) consortium

A.1 Introduction

To gain insights into the behavior and ecology of the *T. reesei* / *E. coli* (TrEc) consortium, we developed a comprehensive ordinary differential equation (ODE) modeling framework that captures salient features of the system. We derived rate expressions for microbial growth, uptake of soluble saccharides, production of cellulase enzymes (endoglucanase, exoglucanase, and β-glucosidase) by *T. reesei*, enzymatic cellulose hydrolysis (based on novel mechanistic models for each type of enzyme), isobutanol production by *E. coli*, and isobutanol toxicity. The model was developed by writing differential mole/mass balances (for a batch reactor) for each species of interest, including microbial biomass (*T. reesei* and *E. coli*), cellulases, insoluble cellulose polysaccharides, soluble oligo and monosaccharides, and isobutanol. We give a detailed description of our model in the following sections.

A.2 Microbial growth and substrate utilization

A.2.1 Maintenance model

Substrate consumption by microbes has two components: substrate consumed for growth, and substrate consumed for non-growth associated maintenance. We expect that maintenance substrate consumption will be important in the *T. reesei* / *E. coli* consortium due to low growth rates. Maintenance substrate uptake is usually empirically modeled as follows [83]:

$$r_{sm} = mC_c \tag{A.1}$$

Where r_{sm} is maintenance substrate uptake rate (g L^{-1} h^{-1}), m is the maintenance coefficient (g-substrate g-cells^{-1} h^{-1}), and C_c is cell concentration (g L^{-1}). We find this model to be unrealistic,

however, as it supposes that substrate uptake rate is completely independent of substrate concentration. This assumption may be reasonable in situations involving high substrate concentrations (e.g. batch culture growth on soluble substrates), but for growth on cellulosic substrates, concentrations of soluble saccharides are likely to be very low, requiring us to consider the actual kinetics of substrate uptake. We thus propose an alternative maintenance model centered on substrate uptake kinetics. We assume that total substrate uptake follows Michaelis-Menten kinetics, and allow for the uptake of multiple substrates simultaneously [241]. Uptake rate of substrate S_i can then be modeled as:

$$r_{S_i} = p_{S_i} \frac{K_{max,S_i} S_i}{K_{S_i} + S_i} C_c \text{ with } p_{S_i} = \frac{S_i}{\sum_j S_j} \tag{A.2}$$

Where r_{S_i} is uptake rate of substrate i (g L^{-1} h^{-1}), p_{S_i} is the fraction of substrate i out of total soluble substrate, K_{max,S_i} is maximum specific uptake rate of substrate i (g-substrate g-cells^{-1} h^{-1}), K_{S_i} is affinity for substrate i (g L^{-1}), S_i is concentration of substrate i (g L^{-1}), and C_c is cell concentration (g L^{-1}). We assume that substrate is partitioned between growth and maintenance uses. We can thus write a mass balance for substrate uptake as follows:

$$r_{S_i} = p_{S_i} \frac{K_{max,S_i} S_i}{K_{S_i} + S_i} C_c = Y_{S_i/C_c} \mu_{S_i} C_c + m p_{S_i} C_c \tag{A.3}$$

Where Y_{S_i/C_c} is the substrate-biomass yield coefficient (g-substrate g-cells^{-1}), μ_{S_i} is specific growth rate on substrate i (h^{-1}), and other terms are as described previously. The term $Y_{S_i/C_c} \mu_{S_i} C_c$ represents substrate consumption for growth, while the $m p_{S_i} C_c$ term represents maintenance consumption. If we hold m constant (i.e. assume constant maintenance requirement), then we can rearrange and write μ_{S_i} in terms of the other parameters and variables:

$$\mu_{S_i} = \frac{p_{S_i}}{Y_{S_i/C_c}} \left(\frac{K_{max,S_i} S_i}{K_{S_i} + S_i} - m \right) \tag{A.4}$$

The maximum specific growth rate $\mu_{S_i,max}$ corresponds to $S_i \gg K_{S_i}$ with i as the sole substrate:

$$\mu_{S_i,max} = \frac{1}{Y_{S_i/C_c}} \left(K_{max,S_i} - m \right) \tag{A.5}$$

Since $\mu_{S_i,max}$ is readily available from experimental data, it makes sense to reformulate our model in terms of this parameter:

$$K_{max,S_i} = Y_{S_i/C_c} \mu_{S_i,max} + m \tag{A.6}$$

Microbial growth rate can then be written as:

$$r_{g,S_i} = \mu_{S_i} C_c = p_{S_i} \left[\left(\mu_{S_i,max} + \frac{m}{Y_{S_i/C_c}} \right) \frac{S_i}{K_{S_i} + S_i} - \frac{m}{Y_{S_i/C_c}} \right] C_c \tag{A.7}$$

Where r_{g,S_i} is growth rate on substrate i (g L^{-1} h^{-1}) and all other terms are as described previously. Substrate consumption can be written as:

$$r_{S_i} = p_{S_i} \frac{K_{max,S_i} S_i}{K_{S_i} + S_i} C_c = p_{S_i} \left(Y_{S_i/C_c} \mu_{S_i,max} + m \right) \frac{S_i}{K_{S_i} + S_i} C_c \tag{A.8}$$

We can write the total growth rate as:

$$r_g = \sum_i r_{g,S_i} = \sum_i \left[p_{S_i} \left(\left(\mu_{S_i,max} + \frac{m}{Y_{S_i/C_c}} \right) \frac{S_i}{K_{S_i} + S_i} - \frac{m}{Y_{S_i/C_c}} \right) C_c \right] \tag{A.9}$$

Where all terms are as described previously

A.2.2 *T. reesei* privileged access to substrate

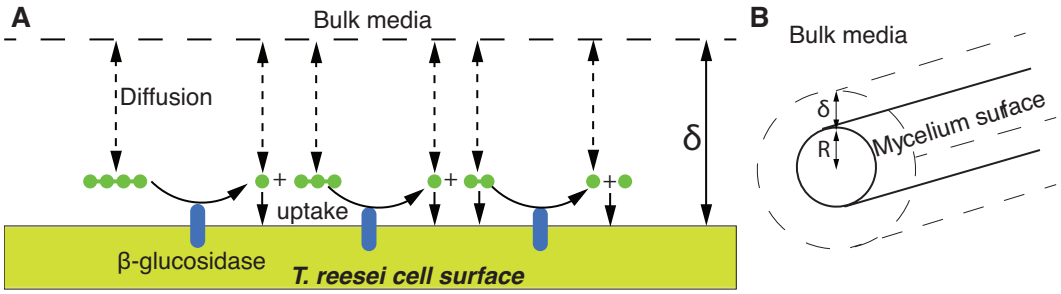

Figure A.1: Mass transfer model for oligosaccharide hydrolysis by cell-wall bound β-glucosidase of *T. reesei* **(A)** Hydrolysis and diffusion of oligosaccharides at *T. reesei* cell surface **(B)** Approximation of mycelium as cylindrical surface.

Accounting for the hydrolysis of cellulosic feedstocks to soluble bioavailable saccharides is a crucial aspect of modeling the *T. reesei* / *E. coli* consortium. An important subtlety in this process is that soluble oligosaccharides (i.e. saccharides 2 to 4 glucose monomers in size) are hydrolyzed to glucose via β-glucosidase bound to the cell wall of *T. reesei* [81], as depicted in Figure A.1A. This leads to locally increased concentration of glucose at the cell surface relative to the bulk media, thus affording privileged access to *T. reesei*. We perform a mass-transfer analysis to estimate the concentration of glucose at the cell surface, making the following assumptions:

1. Soluble oligosaccharides are hydrolyzed by β-glucosidase via a heterogenous Michaelis-Menten reaction at the cell surface.

2. Concentration of soluble saccharides is lower than β-glucosidase affinity ($S_{G_i} < K_{M,G_i}^{BGL}$) — reasonable for co-culture conditions.

3. No homogenous hydrolysis reactions in the bulk media.

4. Pseudo steady state conditions — reasonable due to large timescale for changes in bulk saccharide concentrations.

5. Most of the glucose produced at the *T. reesei* cell surface is lost to diffusion (verified in [107] for *S. cerevisiae* / sucrose hydrolysis).

6. The cell surface is surrounded by a stagnant boundary layer of depth δ which provides the primary resistance to mass transfer to the bulk media (Figure A.1B).

7. *T. reesei* mycelial geometry can be approximated as cylindrical (Figure A.1B).

Based on assumption 2, we can simplify β-glucosidase kinetics to yield the following rate law for hydrolysis of soluble saccharides:

$$r_{S_{G_i}}^{BGL} \approx -\frac{k_{cat,BGL,G_i}\rho_E}{K_{M,G_i}^{BGL}} S_{G_i} = -k_{BGL,G_i} S_{G_i} \tag{A.10}$$

$$k_{BGL,G_i} = \frac{k_{cat,BGL,G_i}\rho_E}{K_{M,G_i}^{BGL}} \tag{A.11}$$

Where $r_{S_{G_i}}^{BGL}$ is the cellobiose hydrolysis rate per unit area (mmol dm^{-2} h^{-1}), k_{cat,BGL,G_i} is the specific activity of β-glucosidase for substrate i (mmol g-BGL^{-1} h^{-1}), ρ_E is the density of β-glucosidase on the cell surface (g-BGL dm^{-2}), K_{M,G_i}^{BGL} is affinity of β-glucosidase for substrate i (mM), S_{G_i} is concentration of substrate i (mM; i is equivalent to degree of polymerization DP, i.e. number of glucose monomers), and k_{BGL,G_i} is an apparent first-order rate constant (dm h^{-1}). Applying conservation equations and simplifying yields the following:

$$\nabla \cdot N_{G_i} = 0 \tag{A.12}$$

$$\nabla^2 S_{G_i} = 0 \tag{A.13}$$

Where N_{G_i} is the molar flux of substrate i (mmol dm^{-2} h^{-1}) and other terms are as defined previously. By argument of symmetry (Figure A.1B) we can neglect all components except for the radial direction, allowing us to simply further:

$$\frac{1}{r}\frac{d}{dr}\left(r\frac{dS_{G_i}}{dr}\right) = 0 \tag{A.14}$$

The above equation can be solved to yield a concentration profile of S_{G_i} as a function of position by applying the following boundary conditions:

$$S_{G_i} = S_{G_i,0} \text{ at } r = R + \delta \tag{A.15}$$

$$S_{G_i} = S_{G_i,Tr} \text{ at } r = R \tag{A.16}$$

Where $S_{G_i,0}$ is the bulk concentration substrate i (mM), $S_{G_i,Tr}$ is the concentration of substrate i at the *T. reesei* mycelium surface (mM), R is the hyphal radius (dm), and δ is the boundary layer thickness (dm). Integrating and applying the above BCs yields:

$$S_{G_i} = (S_{G_i,Tr} - S_{G_i,0}) \frac{\ln r - \ln(R+\delta)}{\ln R - \ln(R+\delta)} + S_{G_i,0} \tag{A.17}$$

Next we apply a mole balance to the interface between the boundary layer and the mycelium surface, equating the diffusive flux at the surface to the rates of hydrolysis of substrate i and production of substrate i from hydrolysis of $i+1$ saccharides:

$$-D_{G_i} \left.\frac{dS_{G_i}}{dr}\right|_{r=R} = r^{BGL}_{S_{G_i}} - r^{BGL}_{S_{G_{i+1}}} = -k_{BGL,G_i} S_{G_i,Tr} + k_{BGL,G_{i+1}} S_{G_{i+1},Tr} \tag{A.18}$$

Where D_{G_i} is the diffusion coefficient of substrate i (dm^2 h^{-1}), $-D_{G_i}\frac{dS_{G_i}}{dr}$ is the diffusive flux of substrate i (mmol dm^{-2} h^{-1}), $r^{BGL}_{S_{G_i}}$ is the hydrolysis rate of substrate i on the mycelium surface (mmol dm^{-2} h^{-1}), and $r^{BGL}_{S_{G_{i+1}}}$ is the hydrolysis rate of substrate $i+1$ at the mycelium surface (mmol dm^{-2} h^{-1}). Substituting the radial concentration profile S_{G_i} in the derivative above and rearranging yields the following expression for $S_{G_i,Tr}$:

$$S_{G_i,Tr} = \frac{R \ln\left(\frac{R}{R+\delta}\right)}{-D_{G_i}} \left[-k_{BGL,G_i} S_{G_i,Tr} + k_{BGL,G_{i+1}} S_{G_{i+1},Tr}\right] + S_{G_i,0} \tag{A.19}$$

To estimate a value for δ, we can assume that the stagnant boundary layer is the primary barrier to mass transfer. The mass transfer coefficient is then simply given by Sherwood number $Sh = 2.0$ and therefore $\delta \approx R$. The above equation then simplifies to:

$$S_{G_i,Tr} = \frac{-k_{BGL,G_i} S_{G_i,Tr} + k_{BGL,G_{i+1}} S_{G_{i+1},Tr}}{D_{G_i}} R \ln 2 + S_{G_i,0} \tag{A.20}$$

We calculate the surface concentrations of each soluble saccharide (DP=1 to 4) as follows:

A.2.2.1 Cellotetraose: DP=4

$$S_{G_4,Tr} = \frac{-k_{BGL,G_4} S_{G_4,Tr}}{D_{G_4}} R \ln 2 + S_{G_4,0} \tag{A.21}$$

Rearranging,

$$\frac{S_{G_4,Tr}}{S_{G_4,0}} = \frac{1}{1 + \frac{k_{BGL,G_4} R \ln 2}{D_{G_4}}} = \eta_4 \tag{A.22}$$

with parameters defined for substrate i as:

$$\eta_i = \frac{1}{1 + k_{BGL,G_i} \phi_i} \tag{A.23}$$

$$\phi_i = \frac{R \ln 2}{D_{G_i}} \tag{A.24}$$

Where η_i is the ratio of mycelium surface concentration to bulk concentration for substrate i, and ϕ_i is the ratio of characteristic boundary layer length to the diffusion coefficient for substrate i (h dm^{-1})

A.2.2.2 Cellotriose: DP=3

$$S_{G_3,Tr} = \frac{-k_{BGL,G_3} S_{G_3,Tr} + k_{BGL,G_4} S_{G_4,Tr}}{D_{G_3}} R \ln 2 + S_{G_3,0} \tag{A.25}$$

Which can be rearranged to yield:

$$S_{G_3,Tr} = k_{BGL,G_4} \phi_3 \eta_3 S_{G_4,Tr} + \eta_3 S_{G_3,0} \tag{A.26}$$

$$= k_{BGL,G_4} \phi_3 \eta_3 \eta_4 S_{G_4,0} + \eta_3 S_{G_3,0} \tag{A.27}$$

With parameters defined as above.

A.2.2.3 Cellobiose: DP=2

Proceeding as for cellotetraose and cellotriose:

$$S_{G_2,Tr} = k_{BGL,G_3} \phi_2 \eta_2 S_{G_3,Tr} + \eta_2 S_{G_2,0} \tag{A.28}$$

$$= \frac{-k_{BGL,G_2} S_{G_2,Tr} + k_{BGL,G_3} S_{G_3,Tr}}{D_{G_2}} (R \ln 2) + S_{G_2,0} \tag{A.29}$$

$$= k_{BGL,G_3} \phi_2 \eta_2 \left(k_{BGL,G_4} \phi_3 \eta_3 \eta_4 S_{G_4,0} + \eta_3 S_{G_3,0}\right) + \eta_2 S_{G_2,0} \tag{A.30}$$

With parameters defined as above.

A.2.2.4 Glucose

For glucose, we modify the interfacial mole balance since glucose is the product of all β-glucosidase hydrolysis reactions:

$$-D_{G_1}\frac{dS_{G_1}}{dr}\bigg|_{r=R} = 2r_{S_{G_2}}^{BGL} + r_{S_{G_3}}^{BGL} + r_{S_{G_4}}^{BGL} \tag{A.31}$$

$$= -2k_{BGL,G_2}S_{G_2,Tr} - k_{BGL,G_3}S_{G_3,Tr} - k_{BGL,G_4}S_{G_4,Tr} \tag{A.32}$$

Where all terms are as defined previously. Proceeding as above, $S_{G_1,Tr}$ can be expressed as:

$$S_{G_1,Tr} = \left(2k_{BGL,G_2}S_{G_2,Tr} + k_{BGL,G_3}S_{G_3,Tr} + k_{BGL,G_4}S_{G_4,Tr}\right)\phi_1 + S_{G_1,0} \tag{A.33}$$

Which can be simplified to the following expression by substituting in the above expressions for $S_{G_2,Tr}$, $S_{G_3,Tr}$, and $S_{G_4,Tr}$ and making some rearrangements:

$$S_{G_1,Tr} = S_{G_1,0} + \theta_2 S_{G_2,0} + \theta_3 S_{G_3,0} + \theta_4 S_{G_4,0} \tag{A.34}$$

Where the θ coefficients are recursively defined as follows:

$$\theta_2 = \frac{2k_{BGL,G_2}\phi_1}{1 + k_{BGL,G_2}\phi_2} \tag{A.35}$$

$$\theta_3 = \theta_2\frac{k_{BGL,G_3}\phi_2}{1 + k_{BGL,G_3}\phi_3} + \frac{k_{BGL,G_3}\phi_1}{1 + k_{BGL,G_3}\phi_3} = \left(\theta_2\phi_2 + \phi_1\right)\frac{k_{BGL,G_3}}{1 + k_{BGL,G_3}\phi_3} \tag{A.36}$$

$$\theta_4 = \left(\theta_3\phi_3 + \phi_1\right)\frac{k_{BGL,G_4}}{1 + k_{BGL,G_4}\phi_4} \tag{A.37}$$

$$\tag{A.38}$$

A.2.2.5 Proposed general framework

In general, we can describe the surface concentration of any soluble saccharide in terms of coefficients similar to those defined above for glucose. A proposed framework:

$$S_{G_i,Tr} = S_{G_i,0} + \sum_{k=i+1}^{4}\theta_{k\to i}S_{G_k,0} \tag{A.39}$$

Where $\theta_{k\to i}$ is the contribution of substrate i concentration at the mycelium surface due to hydrolysis of substrate k (mM mM^{-1} or any other ratio of consistent concentration units)

A.3 *T. reesei* model

A.3.1 *T. reesei* growth

T. reesei is a multicellular filamentous fungus that has different mycelial growth states. Vegetative growth and enzyme secretion are highly active at hyphal tips, while senescent mycelium is relatively dormant [82]. Assuming that growth at hyphal tips follows Monod kinetics and that *T. reesei* is capable of simultaneous utilization of multiple soluble sugars (i.e. glucose and soluble glucose oligosaccharides), *T. reesei* growth in the presence of isobutanol can be described with a segregated kinetic model:

$$\frac{dC_{Tr,v}}{dt} = \mu_{Tr} C_{Tr,v} - k_{v\to s} C_{Tr,v} \tag{A.40}$$

$$\frac{dC_{Tr,s}}{dt} = k_{v\to s} C_{Tr,v} - k_{Tr,d} C_{Tr,s} \tag{A.41}$$

In the first expression, $C_{Tr,v}$ is the vegetative mycelium concentration (g L^{-1}), μ_{Tr} is a generalized Monod function (h^{-1}) depending on isobutanol concentration I (g L^{-1}) and concentration of soluble glucose saccharides S_{G_i} (g L^{-1}; i is the degree of polymerization), and $k_{v\to s}$ is the specific rate of conversion of vegetative mycelium to senescent mycelium (h^{-1}). In the second expression, $C_{Tr,s}$ is the concentration of senescent mycelium (g L^{-1}) and $k_{Tr,d}$ is the specific death rate of the senescent mycelium (h^{-1}). We formulate μ_{Tr} as follows:

$$K_{Tr}^{I} = \begin{cases} \left(1 - \frac{I}{I_{Tr}^{*}}\right)^{n_{Tr}} & \text{if } I \le I_{Tr}^{*} \\ 0 & \text{if } I > I_{Tr}^{*} \end{cases} \tag{A.42}$$

$$\mu_{Tr} = K_{Tr}^{I} \sum_{i} p_{S_{G_i}} \left[\left(\mu_{max,Tr,S_{G_i}} + \frac{m_{Tr}}{Y_{S_{G_i}/C_{Tr}}} \right) \frac{S_{G_i} + \sum_{k=i+1}^{4} \theta_{k\to i} S_{G_k}}{K_{Tr,S_{G_i}} + S_{G_i} + \sum_{k=i+1}^{4} \theta_{k\to i} S_{G_k}} - \frac{m_{Tr}}{Y_{S_{G_i}/C_{Tr}}} \right] \tag{A.43}$$

$$\text{with } p_{S_{G_i}} = \frac{S_{G_i}}{\sum_{j} S_{G_j}}$$

Where K_{Tr}^{I} is an empirical inhibition function (dimensionless) [83], I is isobutanol concentration (g L^{-1}), I_{Tr}^{*} is the growth inhibiting concentration of isobutanol (g L^{-1}) for *T. reesei*, n_{Tr} is an empirically determined exponent, $\mu_{max,Tr,S_{G_i}}$ is maximum specific growth rate on substrate i

(h^{-1}), $p_{S_{G_i}}$ is the proportion of substrate i in the total substrate concentration ($S_{G_i}/\sum S_{G_k}$), S_{G_i} is substrate i concentration (g L^{-1}), $K_{Tr,S_{G_i}}$ is substrate i affinity (g L^{-1}), the coefficients $\theta_{k\to i}$ are as described in section A.2.2, m_{Tr} is the maintenance coefficient (g-substrate g-biomass^{-1} h^{-1}), and $Y_{S_{G_i}/C_{Tr}}$ is the substrate/biomass yield coefficient for substrate i (g-substrate g-biomass^{-1}). We assume that growth occurs via utilization of multiple substrates simultaneously, as opposed to diauxic substrate utilization. Available experimental data suggests that this is a reasonable assumption for *T. reesei*, especially the RUTC30 strain, which contains a loss-of-function mutation in catabolite repression gene *cre1* [84]. Our model assumes a total substrate maintenance requirement m_{Tr} rather than an individual maintenance term for each substrate i; this is reasonable for substrates with similar metabolism / energy yields (e.g. glucose and cellobiose), but could be revised for more diverse substrates.

A.3.2 *T. reesei* enzyme secretion

Assuming that enzyme secretion is stoichiometrically coupled to growth and that composition of secreted enzymes is constant, the following expression can be derived for cellulase production:

$$\frac{dE_T}{dt} = Y_{E_T/C_{Tr}}\mu_{Tr}C_{Tr,v} + k_{E_T}C_{Tr,s} \tag{A.44}$$

$$\frac{dE_i}{dt} = x_{E_i}\frac{dE_T}{dt} = x_{E_i}\left[Y_{E_T/C_{Tr}}\mu_{Tr}C_{Tr,v} + k_{E_T}C_{Tr,s}\right] \tag{A.45}$$

Where E_T is the total concentration of secreted enzymes (g L^{-1}), $Y_{E_T/C_{Tr}}$ is the enzyme/biomass yield coefficient (g-protein g-biomass^{-1}), k_{E_T} is the specific enzyme production rate of senescent mycelium (g-protein g-biomass^{-1} h^{-1}), E_i is concentration of enzyme i (g L^{-1}), x_{E_i} is the fraction of enzyme i in the total secretome (E_i/E_T), and the other terms are as described in previous sections. *T. reesei* produces a large suite of biomass degrading enzymes, but for the purpose of our cellulose hydrolysis model, we consider the most important enzymes [14]:

- cellobiohydrolase 1 ($CBH1$) and 2 ($CBH2$)
- endoglucanase 1 ($EG1$)
- β-glucosidase 1 (BGL)

A.3.3 *T. reesei* RUTC30 saccharide uptake

Assuming that saccharide uptake is stoichiometrically coupled to growth of vegetative mycelium and that both vegetative and senescent mycelia consume saccharides for maintenance, the follow-

ing expression for saccharide uptake by *T. reesei* can be derived:

$$r_{S_{G_i}}^{Tr} = \left[Y_{S_{G_i}/C_{Tr}} K_{Tr}^{I} p_{S_{G_i}} \mu_{max,Tr,S_{G_i}} + m_{Tr}\right] \frac{S_{G_i}}{K_{Tr,S_{G_i}} + S_i} C_{Tr,v} + m_{Tr} p_{S_{G_i}} \frac{S_{G_i}}{K_{Tr,S_{G_i}} + S_{G_i}} C_{Tr,s} \tag{A.46}$$

Where $r_{S_{G_i}}^{Tr}$ is the total rate of saccharide i uptake by *T. reesei* (g L^{-1} h^{-1}), and all other terms are as described in previous sections.

A.4 *E. coli* model

A.4.1 *E. coli* growth

We model *E. coli* growth with Monod kinetics [83], assuming that only glucose is utilized for growth (i.e. glucose oligosaccharides cannot be metabolized):

$$\frac{dC_{Ec}}{dt} = \left(\mu_{Ec} - k_{Ec,d}\right) C_{Ec} \tag{A.47}$$

Where C_{Ec} is *E. coli* concentration (g L^{-1}), μ_{Ec} is specific growth rate (h^{-1}), and $k_{Ec,d}$ is the specific cell death rate (h^{-1}). μ_{Ec} is assumed to be a function of glucose concentration S_{G_1}, with concentration-dependent inhibition from isobutanol:

$$K_{Ec,S_{G_1}}^{I} = \begin{cases} \left(1 - \frac{I}{I_{Ec,S_{G_1}}^{*}}\right)^{n_{Ec,S_{G_1}}} & \text{if } I \leq I_{Ec,S_{G_1}}^{*} \\ 0 & \text{if } I > I_{Ec,S_{G_1}}^{*} \end{cases} \tag{A.48}$$

$$\mu_{Ec} = K_{Ec,S_{G_1}}^{I} \left[\left(\mu_{max,Ec,S_{G_1}} + \frac{m_{Ec,S_{G_1}}}{Y_{S_{G_1}/C_{Ec}}}\right) \frac{S_{G_1}}{K_{Ec,S_{G_1}} + S_{G_1}} - \frac{m_{Ec,S_{G_1}}}{Y_{S_{G_1}/C_{Ec}}}\right] \tag{A.49}$$

Where $K_{Ec,S_{G_1}}^{I}$ is an empirical inhibition function (dimensionless) [83], I is isobutanol concentration (g L^{-1}), $I_{Ec,S_{G_1}}^{*}$ is the growth inhibiting concentration of isobutanol (g L^{-1}) for *E. coli*, $n_{Ec,S_{G_1}}$ is an empirically determined exponent, $\mu_{max,Ec,S_{G_1}}$ is maximum specific growth rate of *E. coli* on glucose (h^{-1}), $K_{Ec,S_{G_1}}$ is glucose affinity (g L^{-1}), $Y_{S_{G_1}/C_{Ec}}$ is the glucose/biomass yield coefficient (g-substrate g-biomass^{-1}), and $m_{Ec,S_{G_1}}$ is the maintenance coefficient (g-substrate g-biomass^{-1} h^{-1}).

A.4.2 *E. coli* saccharide uptake

Substrate uptake is assumed to be stoichiometrically coupled to growth. Additionally, experimental data for *E. coli* demonstrates non-growth associated substrate uptake (i.e. during stationary phase) for maintenance / isobutanol production [1]. We then model uptake of glucose as follows:

$$r^{Ec}_{S_{G_1}} = Y_{S_{G_1}/C_{Ec}} K^I_{Ec,S_{G_1}} \frac{\mu_{max,Ec,S_{G_1}} S_{G_1}}{K_{Ec,S_{G_1}} + S_{G_1}} C_{Ec} + m_{Ec,S_{G_1}} \frac{S_{G_1}}{K_{Ec,S_{G_1}} + S_{G_1}} C_{Ec} \tag{A.50}$$

Where $r^{Ec}_{S_{G_1}}$ is the rate of total glucose uptake by *E. coli* (g L^{-1} h^{-1}), and the other terms are as described in previous sections.

A.4.3 *E. coli* isobutanol production

Unlike many metabolic products, isobutanol production is not stoichiometrically coupled to growth, since substantial isobutanol production is observed during stationary phase [1]. To account for this, all consumed substrates, both for growth and maintenance, will be assumed to be converted to isobutanol. For generality, we allow yield coefficients to vary between growth and non-growth associated substrate uptake:

$$\frac{dI}{dt} = Y^{growth}_{I/S_{G_1}} K^I_{Ec,S_{G_1}} \frac{\mu_{max,Ec,S_{G_1}} S_{G_1}}{K_{Ec,S_{G_1}} + S_{G_1}} C_{Ec} + Y^{maint}_{I/S_{G_1}} m_{Ec,S_{G_1}} \frac{S_{G_1}}{K_{Ec,S_{G_1}} + S_{G_1}} C_{Ec} \tag{A.51}$$

Where $Y^{growth}_{I/S_{G_1}}$ is the growth associated isobutanol/glucose yield coefficient (g-iBtOH g-substrate^{-1}) and $Y^{maint}_{I/S_{G_1}}$ is the non-growth (maintenance) isobutanol/glucose yield coefficient (g-iBtOH g-substrate^{-1}). In the case of *E. coli* K12, both yield coefficients would be 0 (i.e. no isobutanol production).

A.5 Enzymatic cellulose hydrolysis: general framework

There are numerous models reported in literature for microbial growth on cellulose [14]. However, few of these models accounts for the hydrolysis of cellulose to soluble saccharides. Competition between the *E. coli* and *T. reesei* for soluble saccharides is a crucial ecological interaction that needs to be accounted for to accurately predict population dynamics, behavior, and isobutanol production in the TrEc consortium. Enzymatic cellulose hydrolysis is a complex process that is poorly understood, and remains an active area of research [14]. There are two main classes of cellulase: endoglucanases and cellobiohydrolases (also known as exoglucanases) [14]. Most

cellulolytic organisms produce multiple, even dozens of cellulases of each type [14]. As a starting point for developing mechanistic models of cellulose hydrolysis, we utilize the general framework for enzymatic cellulose hydrolysis proposed by [14] and [85], which we describe in the following sections. Additionally, we also include generalized soluble saccharide mole balances that describe rate of production/consumption due to enzymatic hydrolysis and microbial uptake.

A.5.1 Endoglucanase

Endoglucanases adsorb at random to cellulose molecules and cleave them to release two shorter chain polysaccharides [85]. This mechanism can be represented as [85]:

$$S_{G_i} + E_{EGm} \overset{K_{dis}^{EGm}}{\longleftrightarrow} S_{G_i} \cdot E_{EGm} \xrightarrow{k_{EGm}} E_{EGm} + S_{G_{i-j}} + S_{G_j} \tag{A.52}$$

Where E_{EGm} is endoglucanase m, K_{dis}^{EGm} is the dissociation constant for endoglucanase m (mM bonds), k_{EGm} is the rate constant of adsorbed E_{EGm} (mmol bonds g-EGm·$S_{G_i}{}^{-1}$ h^{-1}), i and j are cellulose chain lengths, with $1 \leq j < i$, $S_{G_i} \cdot E_{EGm}$ is adsorbed E_{EGm}, and other terms are as previously described. The rate of hydrolysis of saccharide S_{G_i} by endoglucanase m is then [85]:

$$r_{S_{G_i}}^{EGm} = -k_{EGm} \left[S_{G_i} \cdot E_{EGm}\right] \tag{A.53}$$

Where $r_{S_{G_i}}^{EGm}$ is hydrolysis rate (mM bonds h^{-1}) and $[S_{G_i} \cdot E_{EGm}]$ is the mass concentration of E_{EGm} adsorbed to S_{G_i} (g L^{-1}). Cellulose saccharides S_{G_i} can be formed from endoglucanase hydrolysis of longer cellulose molecules S_{G_k}, with $k > i$. The rate of hydrolysis of S_{G_k} to S_{G_i} is equal to the overall rate of hydrolysis of S_{G_k} times the fraction of hydrolysis events that lead to a chain length S_{G_i}, $f_{G_k \to G_i}$. If all glycosidic bonds are cleaved at an equal rate, then $f_{G_k \to G_i} = 2/\left(k-1\right)$, leading to the following [85]:

$$r_{S_{G_k} \to S_{G_i}}^{EGm} = f_{G_k \to G_i} r_{S_{G_k}}^{EGm} = -\frac{2}{k-1} k_{EGm} \left[S_{G_k} \cdot E_{EGm}\right] \tag{A.54}$$

The overall rate of formation of S_{G_i} by endoglucanase is then the sum of the rate of hydrolysis of S_{G_i} and the rate of formation of S_{G_i} from S_{G_k} with $k > i$ [85]:

$$r_{S_{G_i}}^{EGm} = -k_{EGm} \left[S_{G_i} \cdot E_{EGm}\right] - \sum_{k>i} f_{G_k \to G_i} r_{S_{G_k}}^{EGm} \tag{A.55}$$

$$= -k_{EGm} \left[S_{G_i} \cdot E_{EGm}\right] + \sum_{k>i} \frac{2}{k-1} k_{EGm} \left[S_{G_k} \cdot E_{EGm}\right] \tag{A.56}$$

Where the upper limit of the summation is implicitly understood as DP_{max} (maximum polysaccharide length i for given type of cellulose) and other terms are as described previously.

A.5.2 Exoglucanase

In contrast to endoglucanases, exoglucanases (often referred to as cellobiohydrolases) bind to the ends of cellulose chains and processively hydrolyze cellobiose units. Mechanistically, this can be represented as [85]:

$$S_{G_i} + E_{CBHm} \overset{K_{dis}^{CBHm}}{\longleftrightarrow} S_{G_i} \cdot E_{CBHm} \xrightarrow{k_{CBHm}} E_{CBHm} + S_{G_{i-2}} + S_{G_2} \tag{A.57}$$

Where E_{CBHm} represents cellobiohydrolase m, K_{dis}^{CBHm} is the dissociation constant for cellobiohydrolase m (mM bonds), k_{CBHm} is the rate constant of adsorbed E_{CBHm} (mmol bonds g-CBHm·$S_{G_i}{}^{-1}$ h^{-1}), $S_{G_i} \cdot E_{CBHm}$ is adsorbed E_{EGm}, and the other terms are as described previously. The rate of hydrolysis of saccharide S_{G_i} by cellobiohydrolase is then [85]:

$$r_{S_{G_i} \to S_{G_2}}^{CBHm} = -k_{CBHm} \left[S_{G_i} \cdot E_{CBHm} \right] \tag{A.58}$$

Where terms are similar to those described for endoglucanase. S_{G_i} can also be formed from cellobiohydrolase hydrolysis of $i+2$ chain length cellulose molecules. The overall rate of formation of S_{G_i} by cellobiohydrolase is then the sum of the rate of hydrolysis of S_{G_i} and the rate of formation of S_{G_i} from $i+2$ saccharides [85]:

$$r_{S_{G_i}}^{CBHm} = -k_{CBHm} \left[S_{G_i} \cdot E_{CBHm} \right] + k_{CBHm} \left[S_{G_{i+2}} \cdot E_{CBHm} \right] \tag{A.59}$$

Where terms are as described in previous sections. The overall rate of formation of S_{G_2} is the sum of cellobiohydrolase hydrolysis rates for all saccharides S_{G_i} for $i \geq 3$:

$$r_{S_{G_2}}^{CBHm} = k_{CBHm} \sum_{i \geq 3} \left[S_{G_i} \cdot E_{CBHm} \right] \tag{A.60}$$

Where the upper limit of the summation is implicitly understood as DP_{max} and other terms are as described previously.

A.5.3 β-glucosidase

β-glucosidase hydrolyzes cellobiose and other soluble cellulose oligosaccharides to glucose [86]. For soluble saccharides of DP $i = 2..4$ this can be mechanistically represented as:

$$S_{G_i} + E_{BGLm} \xleftrightarrow{K_{G_i,dis}^{BGLm}} S_{G_i} \cdot E_{BGLm} \xrightarrow{k_{BGLm,G_i}} E_{BGLm} + S_{G_{i-1}} + S_{G_1} \tag{A.61}$$

Where E_{BGLm} represents β-glucosidase m, $K_{G_i,dis}^{BGLm}$ is the dissocation constant for β-glucosidase m (mM substrate), k_{BGLm,G_i} is the rate constant of adsorbed E_{BGLm} (mmol g-BGLm·$S_{G_i}{}^{-1}$ h^{-1}), $S_{G_i} \cdot E_{BGLm}$ is the β-glucosidase / substrate complex, and the other terms are as described in previous sections. The rate of hydrolysis of S_{G_i} or formation of S_{G_1} is then:

$$r_{S_{G_i}}^{BGLm} = -r_{S_{G_1}}^{BGLm} = -k_{BGLm,G_i} \left[S_{G_i} \cdot E_{BGLm} \right] \tag{A.62}$$

For cellobiose, the coefficient before $r_{S_{G_1}}^{BGLm}$ would be -2 instead -1 since two glucoses are produced per cellobiose.

A.5.4 Saccharide mass balances

In general, saccharide mass balances must account for both enzymatic cellulose hydrolysis and microbial saccharide uptake. However, insoluble cellulose molecules (chain length $i > 4$) are not utilized biologically. Thus for chain length $i > 4$ cellulose molecules, net rates of formation from endoglucanase and cellobiohydrolase need only be considered. Writing a mass balance for each cellulose molecule of chain length i with $i > 4$ yields [85]:

$$\frac{dS_{G_i}}{dt} = \sum_m r_{S_{G_i}}^{EGm} + \sum_n r_{S_{G_i}}^{CBHn} \text{ for } i > 4 \tag{A.63}$$

Where all terms are as described in previous sections. For the case of the *T. reesei* cellulase system, consisting of endoglucanase 1 ($EG1$), cellobiohydrolase 1 ($CBH1$), and cellobiohydrolase 2 ($CBH2$), the mass balances reduce to:

$$\frac{dS_{G_i}}{dt} = r_{S_{G_i}}^{EG1} + r_{S_{G_i}}^{CBH1} + r_{S_{G_i}}^{CBH2} \text{ for } i > 4 \tag{A.64}$$

Where all terms are as described in previous sections. For soluble saccharides, microbial saccharide uptake and β-glucosidase hydrolysis must also be considered. For a co-culture of *T. reesei* and *E. coli*, writing a mass balance on cellulose molecules of chain length $1 \leq i \leq 4$ yields:

$$\frac{dS_{G_i}}{dt} = r_{S_{G_i}}^{EG1} + r_{S_{G_i}}^{CBH1} + r_{S_{G_i}}^{CBH2} + r_{S_{G_i}}^{BGL} - \frac{1}{MW_{S_{G_i}}} \left(r_{S_{G_i}}^{Tr} + r_{S_{G_i}}^{Ec} \right) \text{ for } 1 \leq i \leq 4 \tag{A.65}$$

Where $MW_{S_{G_i}}$ is the molecular weight of S_{G_i} (g $mmol^{-1}$), and all other terms are as described in previous sections. Most *E. coli* strains cannot metabolize cellulose oligosaccharides and are thus

only able to use S_{G_1}; additionally, while it seems biologically plausible, there is little evidence to support significant uptake and metabolism of $i > 2$ glucose saccharides by *T. reesei*. We thus reduce the $i \leq 4$ saccharide balances to:

$$\frac{dS_{G_i}}{dt} = r^{EG1}_{S_{G_i}} + r^{CBH1}_{S_{G_i}} + r^{CBH2}_{S_{G_i}} + r^{BGL}_{S_{G_i}} \text{ for } 3 \leq i \leq 4 \tag{A.66}$$

$$\frac{dS_{G_2}}{dt} = r^{EG1}_{S_{G_2}} + r^{CBH1}_{S_{G_2}} + r^{CBH2}_{S_{G_2}} + r^{BGL}_{S_{G_2}} - \frac{1}{MW_{S_{G_2}}} r^{Tr}_{S_{G_2}} \tag{A.67}$$

$$\frac{dS_{G_1}}{dt} = r^{EG1}_{S_{G_1}} + r^{CBH1}_{S_{G_1}} + r^{CBH2}_{S_{G_1}} + r^{BGL}_{S_{G_1}} - \frac{1}{MW_{S_{G_1}}} \left(r^{Tr}_{S_{G_1}} + r^{Ec}_{S_{G_1}} \right) \tag{A.68}$$

Where all terms are as described in previous sections.

A.6 Enzymatic cellulose hydrolysis: kinetics and rate laws

Deriving tractable kinetic expressions for enzymatic cellulose hydrolysis requires making simplifying assumptions, many of which are idealizations that do not apply to real systems. Zhang and Lynd [85] derived rate laws for endoglucanase and exoglucanase by incorporating enzyme mass balances with adsorption equilibria and making the following simplifying assumptions:

1. Random adsorption.
2. Continuous equilibrium between adsorbed and free components.
3. No interactions between adsorbing components / affinity does not vary with fractional coverage.
4. Substrate binding sites are in excess of enzyme.
5. Constant substrate reactivity.
6. Negligible inhibition from hydrolysis products (e.g. S_{G_1} and S_{G_2}).

Some of the assumptions made in [85] are clearly not applicable to enzymatic cellulose hydrolysis in the TrEc consortium. In particular, the assumption of excess substrate binding sites is valid only in the early stages of growth; during later stages, when enzyme concentrations are maximal and cellulose concentrations low, substrate binding sites are clearly not in excess of enzyme binding sites [14]. Additionally, in real systems, substrate reactivity (defined in terms of apparent rate constants k) is found to decrease by one to two orders of magnitude as cellulose conversion approaches 100% [14].

To better model enzymatic cellulose hydrolysis with the TrEc consortium, we advanced Zhang and Lynd's model [85] by deriving a new set of kinetics for endoglucanase and exoglucanase. Our derivation incorporates substrate site balances and empirical correlations for declining reactivity, described below. As a secondary consideration, we also include empirical non-competitive product (S_{G_1} and S_{G_2}) inhibition expressions in our kinetics, though these terms are likely to be unimportant since soluble saccharide concentrations are generally low during TrEc consortium growth on cellulosic substrates.

A.6.1 Endoglucanase

As described in section A.5.1, the overall rate of formation of S_{G_i} by endoglucanase is the sum of the hydrolysis rate of S_{G_i} and the rate of formation of S_{G_i} from S_{G_k} hydrolysis with $k > i$ [85]:

$$r_{S_{G_i}}^{EGm} = -k_{EGm}\left[S_{G_i} \cdot E_{EGm}\right] + \sum_{k>i} \frac{2}{k-1} k_{EGm}\left[S_{G_k} \cdot E_{EGm}\right] \tag{A.56}$$

Where the upper limit of the summation is implicitly understood as DP_{max} (maximum polysaccharide length i for given type of cellulose) and terms are as described previously. Computing $r_{S_{G_i}}^{EGm}$ requires us to express $[S_{G_i} \cdot E_{EGm}]$ in terms of known or measurable variables. Zhang and Lynd [85] derive such expressions by incorporating E_{EGm} mass balances with the E_{EGm} adsorption equilibrium, given below:

Adsorption equilibrium [85]:

$$K_{dis}^{EGm} = \frac{E_{EGm,f} C_f}{\sum_{i\geq 2}\left[S_{G_i} \cdot E_{EGm}\right]} \tag{A.69}$$

Where $E_{EGm,f}$ is concentration of free (i.e. unadsorbed) endoglucanase m (g-EGm L^{-1}), C_f is concentration of free substrate binding sites (mM bonds), the upper limit of the summation is implicitly understood as DP_{max}, and other terms are as described previously.

Enzyme balance [85]:

$$E_{EGm} = E_{EGm,f} + \sum_{i\geq 2}\left[S_{G_i} \cdot E_{EGm}\right] \tag{A.70}$$

Where the upper limit of the summation is implicitly understood as DP_{max} and other terms are as described previously.

We modify the derivation described in [85] by incorporating a balance on substrate binding sites [85]:

$$\sum_{i\geq 2} F_a\,(i-1)\,S_{G_i} = \sum_{i\geq 2} 2\alpha_{EGm}\frac{[S_{G_i}\cdot E_{EGm}]}{MW_{EGm}} + C_f \tag{A.71}$$

Where F_a is the fraction of enzyme accessible β-glycosidic bonds, α_{EGm} is the number of cellobiose lattice units occupied by a single molecule of endoglucanase m, MW_{EGm} is the molecular weight of endoglucanase m (g-EGm mmol^{-1}), C_f is the concentration of free β-glycosidic bonds accessible to cellulase (mM bonds), and the upper limit of the summation is implicitly understood as DP_{max}. We then derive expressions for $[S_{G_i}\cdot E_{EGm}]$:

$$[S_{G_i}\cdot E_{EGm}] = \frac{(i-1)\,S_{G_i}}{\sum_{i\geq 2}(i-1)\,S_{G_i}}Y \tag{A.72}$$

$$Y = \frac{1}{2}b_Y \pm \frac{1}{2}\sqrt{b_Y^2 - \frac{4}{\beta_{EGm}}E_{EGm}\sum_{i\geq 2}F_a\,(i-1)\,S_{G_i}} \tag{A.73}$$

$$b_Y = E_{EGm} + \frac{K_{dis}^{EGm}}{\beta_{EGm}} + \frac{1}{\beta_{EGm}}\sum_{i\geq 2}F_a\,(i-1)\,S_{G_i} \tag{A.74}$$

$$\beta_{EGm} = \frac{2\alpha_{EGm}}{MW_{EGm}} \tag{A.75}$$

Where all terms are as described previously. Incorporating balances on both enzyme binding sites and substrate binding sites into the adsorption equilibrium expression results in a quadratic equation in $[S_{G_i}\cdot E_{EGm}]$; the physically meaningful root of Y is the one for which binding site balances are satisfied: $0 < [S_{G_i}\cdot E_{EGm}] < E_{EGm}$ and $0 < [S_{G_i}\cdot E_{EGm}] < (1/\beta_{EGm})\sum_{i\geq 2}F_a\,(i-1)\,S_{G_i}$ [14].

The overall rate of formation of S_{G_i} due to endoglucanase is then:

$$r_{S_{G_i}}^{EGm} = \frac{k_{EGm}Y}{\sum_{i\geq 2}(i-1)\,S_{G_i}}\left(2\sum_{k>i}S_{G_k} - (i-1)\,S_{G_i}\right) \tag{A.76}$$

Where all terms are as described previously.

A.6.2 Exoglucanase

As described in section A.5.2, the overall rate of formation of S_{G_i} by cellobiohydrolase is the sum of the hydrolysis rate of S_{G_i} and the rate of formation of S_{G_i} from hydrolysis of $i+2$ saccharides [85]:

$$r_{S_{G_i}}^{CBHm} = -k_{CBHm}\left[S_{G_i} \cdot E_{CBHm}\right] + k_{CBHm}\left[S_{G_{i+2}} \cdot E_{CBHm}\right] \tag{A.59}$$

Computing $r_{S_{G_i}}^{CBHm}$ requires us to express $[S_{G_i} \cdot E_{CBHm}]$ in terms of known or measurable variables. Zhang and Lynd [85] derive expressions for $[S_{G_i} \cdot E_{CBHm}]$ analogous to the above described procedure for endoglucanase, that is by incorporating E_{CBHm} mass balances with E_{CBHm} adsorption equilibrium, shown below:

Adsorption equilibrium [85]:

$$K_{dis}^{CBHm} = \frac{E_{CBHm,f} C_f}{\sum\limits_{i\geq 3}\left[S_{G_i} \cdot E_{CBHm}\right]} \tag{A.77}$$

Where $E_{CBHm,f}$ is concentration of free (i.e. unbound) cellobiohydrolase m (g-CBHm L^{-1}), C_f is concentration of free substrate sites (mM bonds), the upper limit of the summation is implicitly understood as DP_{max}, and other terms are as described previously.

Enzyme balance [85]:

$$E_{CBHm} = E_{CBHm,f} + \sum_{i\geq 3}\left[S_{G_i} \cdot E_{CBHm}\right] \tag{A.78}$$

Where terms are as described previously.
We modify the approach described in [85] by including a balance on substrate binding sites [85]:

$$\sum_{i\geq 3} 2F_a S_{G_i} = \sum_{i\geq 3} \frac{\left[S_{G_i} \cdot E_{CBHm}\right]}{MW_{CBHm}} + C_f \tag{A.79}$$

Where MW_{CBHm} is the molecular weight of cellobiohydrolase m (g-CBHm $mmol^{-1}$), the upper limit of the summation is implicitly understood as DP_{max}, and other terms are as described previously. Note that we assume cellobiohydrolases adsorbs only at cellulose chain ends. We then derive expressions for $[S_{G_i} \cdot E_{CBHm}]$:

$$[S_{G_i} \cdot E_{CBHm}] = \frac{S_{G_i}}{\sum_{i\geq 3} S_{G_i}} Z \tag{A.80}$$

$$Z = \frac{1}{2} b_Z \pm \frac{1}{2} \sqrt{b_Z^2 - \frac{8 F_a E_{CBHm}}{\beta_{CBHm}} \sum_{i\geq 3} S_{G_i}} \tag{A.81}$$

$$b_Z = E_{CBHm} + \frac{K_{dis}^{CBHm}}{\beta_{CBHm}} + \frac{2F_a}{\beta_{CBHm}} \sum_{i\geq 3} S_{G_i} \tag{A.82}$$

Where $\beta_{CBHm} = 1/MW_{CBHm}$ and all other all terms are as described in previous sections. Incorporating balances on both enzyme binding sites and substrate binding sites into the adsorption equilibrium expression results in a quadratic equation in $[S_{G_i} \cdot E_{CBHm}]$; the physically meaningful root of Z is the one for which binding site balances are satisfied: $0 < [S_{G_i} \cdot E_{CBHm}] < E_{CBHm}$ and $0 < [S_{G_i} \cdot E_{CBHm}] < MW_{CBHm} \sum_{i=n} 2F_a S_{G_i}$ [14]. The overall rate of formation of S_{G_i} by cellobiohydrolase is the sum of the hydrolysis rate of S_{G_i} and the rate of formation of S_{G_i} from hydrolysis of $i+2$ saccharides:

$$r_{S_{G_i}}^{CBHm} = \frac{k_{EGm} Z}{\sum_{i\geq 3} S_{G_i}} \left(S_{G_{i+2}} - S_{G_i} \right) \tag{A.83}$$

Where all terms are as described previously. The overall rate of formation of S_{G_2} is the sum of cellobiohydrolase hydrolysis rates for all saccharides S_{G_i} for $i \geq 3$:

$$r_{S_{G_2}}^{CBHm} = k_{CBHm} Z \tag{A.84}$$

Where all terms are as described previously.

A.6.3 β-glucosidase

We adopt multisubstrate Michaelis-Menten kinetics for β-glucosidase [86]:

$$r_{S_{G_i}}^{BGLm} = -\frac{k_{BGLm,G_i} E_{BGLm} S_{G_i}}{K_{M,G_i}^{BGLm} \left(1 + \frac{S_{G_1}}{K_{G_1}^{BGLm}} \sum_{i=2}^{4} \frac{S_{G_i}}{K_{M,G_i}^{BGLm}} \right)} \tag{A.85}$$

Where K_{M,G_i}^{BGLm} is the Michaelis constant for S_{G_i} (mM), $K_{G_1}^{BGLm}$ is the glucose inhibition term (mM), and other terms are as described in previous sections. The total rate of glucose production via β-glucosidase is then:

$$r_{S_{G_1}}^{BGLm} = -2r_{S_{G_2}}^{BGLm} - \sum_{i=2}^{4} r_{S_{G_i}}^{BGLm} \tag{A.86}$$

A.6.4 Empirical relations for substrate reactivity and product inhibition

A.6.4.1 Declining substrate reactivity

It has long been observed that specific rates of cellulose hydrolysis (i.e. rate per adsorbed cellulase) decline by one to two orders of magnitude as the reaction proceeds to completion [14]. Numerous explanations for this phenomenon have been proposed, but presently the most widely accepted theory is declining substrate reactivity [14]. Whatever the cause, the phenomenon can be empirically modeled. We adopt the following model which describes rate constants for cellulose hydrolysis as a function of substrate conversion [242]:

$$k_m = k_{m,max}\left[f_{deact}\left(1-X\right)^{n_{deact,m}} + \left(1-f_{deact}\right)\right] \tag{A.87}$$

$$X = \frac{\sum\limits_{i>4} MW_{G_i}\left(S_{G_i}|_{t_0} - S_{G_i}|_t\right)}{\sum\limits_{i>4} MW_{G_i} S_{G_i}|_{t_0}} \tag{A.88}$$

Where k_m is the apparent rate constant for enzyme m ($m = EG1$, $CBH1$, or $CBH2$ in the case of the *T. reesei* cellulase system; mmol bonds g-m^{-1} h^{-1}), $k_{m,max}$ the maximum rate constant for enzyme m (i.e. at 0% conversion; mmol bonds g-m^{-1} h^{-1}), $1 - f_{deact}$ is fractional residual activity, $n_{deact,m}$ is an empirically determined exponent for cellulase m, X is substrate conversion (g-consumed g-initial^{-1}), MW_{G_i} is the molecular weight of saccharide G_i (g mmol^{-1}), $S_{G_i}|_{t_0}$ is the initial concentration of saccharide G_i (mM), $S_{G_i}|_t$ is the concentration of saccharide G_i at time t (mM), and the upper limit of the summation is implicitly understood as DP_{max}.

A.6.4.2 Product inhibition

Glucose and cellobiose have been observed to non-competitively inhibit most cellulase enzymes [14]. Expressions for inhibition can be formally derived from reaction mechanisms; rather than performing a formal derivation for each enzyme, we empirically describe non-competitive inhibition in terms of an apparent reaction rate constant [243]:

$$k_m = \frac{k_{m,max}}{1 + \frac{S_{G_1}}{K_{G_1}^m} + \frac{S_{G_2}}{K_{G_2}^m}} \tag{A.89}$$

Where $K_{G_1}^m$ and $K_{G_2}^m$ are dissociation constants between enzyme m ($m = EG1, CBH1$, or $CBH2$ in the case of the *T. reesei* cellulase system; g L^{-1} or mM) and S_{G_1} and S_{G_2} respectively, and all other terms are as described previously. During microbial growth on cellulose, soluble saccharide concentrations are likely to be very low, so accounting for product inhibition is probably unimportant. However, for completeness we include inhibition in our model.

A.6.4.3 Combined expression for declining reactivity and substrate inhibition

The above empirical expressions can be combined to describe effects of declining substrate reactivity and product inhibition on reaction rate constants:

$$k_m = k_{m,max} \left[\frac{f_{deact} \left(1 - X\right)^{n_{deact,m}} + \left(1 - f_{deact}\right)}{1 + \frac{S_{G_1}}{K_{G_1}^m} + \frac{S_{G_2}}{K_{G_2}^m}} \right] \tag{A.90}$$

Where all terms are as described previously.

APPENDIX B

Model parameters

Table B.1: List of all parameters used in global sensitivity analysis, sampling distribution (normal or uniform) for each parameter, values of μ_j, $c_{v,j}$ (normal distribution) or $a_{j,min}$, $a_{j,max}$ (uniform distribution), description/units for each parameter, and additional notes/references.

IC or parameter	$\mu_j/a_{j,min}$	$c_{v,j}/a_{j,max}$	Distribution	Description	Notes / source
$\log(F_a)$	-3	-1	Uniform	Fractional accessible bonds	Range from PASC to Avicel [85]
$\overline{DP}$	1.6	2.25	Uniform	Cellulose mean degree of polymerization (DP; log scale)	Median value corresponds to Avicel [85]
$c_{v,DP}$	0	0.1	Uniform	c_v of cellulose DP distribution	Est. from [91]
$\sum S_{G_i}(t_0)$	10	40	Uniform	Initial cellulose concentration (g L^{-1})	–
$\sum C_i(t_0)$	0.1	1	Uniform	Total initial microbial biomass (g L^{-1})	–
$X_{Tr,v}(t_0)$	0.72	0.2	Normal	Initial fraction vegetative mycelium	Biologically reasonable range
$X_P(t_0)$	0.5	0.2	Normal	Initial protein concentration (fraction of total biomass)	–
$\log(X_{Ec})(t_0)$	-3	0	Uniform	Initial Ec population (fraction of total biomass)	–
$I(t_0)$	0	5	Uniform	Initial isobutanol concentration (g L^{-1})	–
x_{CBH1}	0.6	0.2	Normal	Fraction of CBH1 in total protein	[85]
x_{CBH2}	0.2	0.2	Normal	Fraction of CBH2 in total protein	[85]
x_{EG1}	0.12	0.2	Normal	Fraction of EG1 in total protein	[85]
x_{BGL}	0.03	0.2	Normal	Fraction of BGL in total protein	Est. from [244] and [81]
n_{deact}	5	0.2	Normal	Declining substrate reactivity exponent	Est. from [242]
f_{deact}	0.8	0.99	Uniform	$1 - f_{deact}$ = residual cellulase activity at 100% conversion	Est. from [242]
k_{EG1}	24	0.2	Normal	EG1 rate constant (mmol g-EG1^{-1} h^{-1})	Est. from [85]
K_{dis}^{EG1}	0.33	0.2	Normal	EG1 dissociation constant (mM)	[85]
β_{EG1}	0.166	0.2	Normal	$2\alpha_{EG1}/MW_{EG1}$ (mmol substrate sites g-EG1^{-1})	Est. from [245] and [14]

Table B.1: List of all parameters used in global sensitivity analysis, sampling distribution (normal or uniform) for each parameter,values of μ_j, $c_{v,j}$ (normal distribution) or $a_{j,min}$, $a_{j,max}$ (uniform distribution), description/units for each parameter, and additional notes/references.

IC or parameter	$\mu_j/a_{j,min}$	$c_{v,j}/a_{j,max}$	Distribution	Description	Notes / source
k_{CBH1}	4.8	0.2	Normal	CBH1 rate constant (mmol g-CBH1^{-1} h^{-1})	Est. from [85]
K_{dis}^{CBH1}	0.25	0.2	Normal	CBH1 dissociation constant (mM)	[85]
β_{CBH1}	0.019	0.2	Normal	$1/MW_{CBH1}$ (g-CBH1 mmol^{-1})	Est. from [245] and [14]
k_{CBH2}	9.6	0.2	Normal	CBH2 rate constant (mmol g-EG1^{-1} h^{-1})	Est. from [85]
K_{dis}^{CBH2}	0.25	0.2	Normal	CBH2 dissociation constant (mM)	[85]
β_{CBH2}	0.022	0.2	Normal	$1/MW_{CBH2}$ (g-CBH2 mmol^{-1})	Est. from [245] and [14]
k_{BGL,G_2}	339	0.2	Normal	BGL rate constant for cellobiose (mmol g-BGL^{-1} h^{-1})	[86]
K_{M,G_2}^{BGL}	1.36	0.2	Normal	BGL affinity for cellobiose (mM)	[86]
$K_{G_1}^{BGL}$	11.33	0.2	Normal	BGL competitive glucose inhibition constant (mM)	[86]
k_{BGL,G_3}	758	0.2	Normal	BGL rate constant for cellotriose (mmol g-BGL^{-1} h^{-1})	
K_{M,G_3}^{BGL}	0.2	0.2	Normal	BGL affinity for cellotriose (mM)	[86]
k_{BGL,G_4}	1810	0.2	Normal	BGL rate constant for cellotetraose (mmol g-BGL^{-1} h^{-1})	
K_{M,G_4}^{BGL}	0.38	0.2	Normal	BGL affinity for cellotetraose (mM)	[86]
$K_{G_2}^{i}$	17.2	0.2	Normal	Cellulase (EG1, CBH1, CBH2) non-competitive cellobiose inhibition constant (mM)	Est. from [243]
$K_{G_1}^{i}$	66.6	0.2	Normal	Cellulase (EG1, CBH1, CBH2) non-competitive glucose inhibition constant (mM)	Est. from [243]
$k_{Tr,d}$	0.005	0.2	Normal	Death rate of senescent mycelium (h^{-1})	Est. from [82]
$K_{Tr,S_{G_1}}$	0.02	0.5	Normal	Tr glucose affinity (g L^{-1})	Est. from [246]
$K_{Tr,S_{G_2}}$	0.0001	0.5	Normal	Tr cellobiose affinity (g L^{-1})	Est. from [247]
$\theta_{G_2 \to G_1}$	1.7	0.2	Normal	Glucose contribution at mycelium surface due to cellobiose hydrolysis (mM mM^{-1})	Calculated from [248] and [249]
$\theta_{G_3 \to G_1}$	2.2	0.2	Normal	Glucose contribution at mycelium surface due to cellotriose hydrolysis (mM mM^{-1})	Calculated from [248] and [249]
$\theta_{G_4 \to G_1}$	2.5	0.2	Normal	Glucose contribution at mycelium surface due to cellotetraose hydrolysis (mM mM^{-1})	Calculated from [248] and [249]
$Y_{S_{G_1}/C_{Tr}}$	2	0.2	Normal	Glucose/biomass yield coefficient for Tr (g-glucose g-biomass^{-1})	Est. from [246]

Table B.1: List of all parameters used in global sensitivity analysis, sampling distribution (normal or uniform) for each parameter,values of μ_j, $c_{v,j}$ (normal distribution) or $a_{j,min}$, $a_{j,max}$ (uniform distribution), description/units for each parameter, and additional notes/references.

IC or parameter	$\mu_j/a_{j,min}$	$c_{v,j}/a_{j,max}$	Distribution	Description	Notes / source
$Y_{S_{G_1}/C_{Tr}}$	2	0.2	Normal	Cellobiose/biomass yield coefficient for Tr (g-cellobiose g-biomass^{-1})	Est. ≈glucose
$Y_{E_T/C_{Tr}}$	0.5	0.2	Normal	Enzyme/biomass yield coefficient for Tr (g-protein g-biomass^{-1})	Est. from [14] and [250]
$\mu_{max,Tr,S_{G_1}}$	0.087	0.2	Normal	Maximum specific growth rate of Tr on glucose (h^{-1})	Est. from [82], [251], and [246]
$\mu_{max,Tr,S_{G_2}}$	0.025	0.5	Normal	Maximum specific growth rate of Tr on cellobiose (h^{-1})	Est. from [247]
$k_{v\rightarrow s}$	0.01	0.2	Normal	Specific rate of conversion of vegetative mycelium to aged mycelium (h^{-1})	Est. from [82]
m_{Tr}	0.027	0.2	Normal	Maintenance coefficient for Tr (g-substrate g-biomass^{-1} h^{-1})	Est. from [246]
$I_{Tr,*}$	0.2	0.2	Normal	Isobutanol inhibition exponent for Tr (exponential inhibition model; g-iBtOH L^{-1})	Est. ≈Ec
k_{E_T}	0.008	0.2	Normal	Specific enzyme production rate of senescent mycelium (g-protein g-biomass^{-1} h^{-1})	Est. as $\approx 1/5$ rate of vegetative mycelium
$k_{Ec,d}$	0.001	0.2	Normal	Ec death rate (h^{-1})	Est. from [2]
$K_{Ec,S_{G_1}}$	0.001	0.5	Normal	Ec glucose affinity (g L^{-1})	Est. from [2]
$Y_{S_{G_1}/C_{Ec}}$	6.4	0.5	Normal	Glucose/biomass yield coefficient for Ec (g-glucose g-biomass^{-1})	Est. from [2]
$m_{Ec,S_{G_1}}$	0.23	0.5	Normal	Maintenance coefficient for Ec (g-glucose g-biomass^{-1} h^{-1})	Est. from [1] and [2]
$\mu_{max,Ec,S_{G_1}}$	0.08	0.2	Normal	Maximum specific growth rate of Ec on glucose (h^{-1})	Est. from experimental data in this study
$Y^{growth}_{I/S_{G_1}}$	0	0.41	Uniform	Growth associated isobutanol/glucose yield coefficient for Ec (g-iBtOH g-glucose^{-1})	Est. from [1] and [2]; [3]
$Y^{maint}_{I/S_{G_1}}$	0	0.41	Uniform	Maintenance associated isobutanol/glucose yield coefficient for Ec (g-iBtOH g-glucose^{-1})	Est. from [1] and [2]; [3]
$I_{Ec,*}$	0.2	0.2	Normal	Isobutanol inhibition exponent for Ec (exponential inhibition model; g-iBtOH L^{-1})	Est. from experimental data in this study

		Strain	
Parameter	**Units**	**NV3 pSA55/69**	**JCL260 pSA55/69**
$\mu_{max,Ec,S_{G_1}}$	h^{-1}	0.14 ± 0.02	0.11 ± 0.01
$k_{Ec,d}$	h^{-1}	0.001	0.001
$K_{Ec,S_{G_1}}$	g L^{-1}	0.001	0.001
$Y_{S_{G_1}/C_{Ec}}$	g-glucose g-biomass^{-1}	8.3	6.4
$m_{Ec,S_{G_1}}$	g-glucose g-biomass^{-1} h^{-1}	0.1	0.2
$Y^{growth}_{I/S_{G_1}}$	g-iBtOH g-glucose^{-1}	0.25	0.3
$Y^{maint}_{I/S_{G_1}}$	g-iBtOH g-glucose^{-1}	0.17	0.41

Table B.2: Model parameter values for *E. coli* JCL260 pSA55/69 and *E. coli* NV3 pSA55/69. $\mu_{max,Ec,S_{G_1}}$ was measured experimentally; remaining parameter values were estimated from data presented in [1], [2] and [3].

Species	Experiment notes	Parameter	Value	Units
E. coli K12	Microplate	$\mu_{max,Ec,S_{G_1}}$	0.41 ± 0.02	h^{-1}
E. coli NV3	Microplate	$\mu_{max,Ec,S_{G_1}}$	0.43 ± 0.03	h^{-1}
E. coli JCL260	Microplate	$\mu_{max,Ec,S_{G_1}}$	0.34 ± 0.01	h^{-1}
E. coli NV3 pSA55/69*	Microplate	$\mu_{max,Ec,S_{G_1}}$	0.14 ± 0.02	h^{-1}
E. coli JCL260 pSA55/69	Microplate	$\mu_{max,Ec,S_{G_1}}$	0.11 ± 0.01	h^{-1}
E. coli JCL260 pSA55/69	Microplate; NG50; 0% (w/v) i-BtOH	$\mu_{max,Ec,S_{G_1}}$	0.11 ± 0.005	h^{-1}
E. coli JCL260 pSA55/69	Microplate; NG50; 0.25% (w/v) i-BtOH	$\mu_{max,Ec,S_{G_1}}$	0.071 ± 0.003	h^{-1}
E. coli JCL260 pSA55/69	Microplate; NG50; 0.5% (w/v) i-BtOH	$\mu_{max,Ec,S_{G_1}}$	0.006 ± 0.002	h^{-1}
T. reesei RUTC30	Flask	$\mu_{max,Tr,S_{G_1}}$	0.092 ± 0.01	h^{-1}
T. reesei RUTC30	Flask; 0.2% (w/v) i-BtOH	$\mu_{max,Tr,S_{G_1}}$	0.08 ± 0.01	h^{-1}
E. coli K12	Chemostat	$K_{Ec,S_{G_1}}$	0.24 ± 0.19	mg L^{-1}
E. coli K12	Chemostat	$Y_{S_{G_1}/C_{Ec}}$	1.4 ± 0.7	g-glucose g-biomass^{-1}
E. coli K12	Chemostat	$m_{Ec,S_{G_1}}$	0.4 ± 0.2	g-glucose g-biomass^{-1} h^{-1}
T. reesei RUTC30	Chemostat	$K_{Tr,S_{G_1}}$	24 ± 18	mg L^{-1}
T. reesei RUTC30	Chemostat	$Y_{S_{G_1}/C_{Tr}}$	1.4 ± 0.3	g-glucose g-biomass^{-1}
T. reesei RUTC30	Chemostat	m_{Tr}	0.0175 ± 0.005	g-substrate g-biomass^{-1} h^{-1}
T. reesei RUTC30	Bioreactor; RUTC30 on Avicel	$Y_{S_{G_1}/C_{Tr}}$	3.95	g-glucose g-biomass^{-1}
T. reesei RUTC30	Bioreactor; K12/RUTC30 on Avicel	$K_{Tr,S_{G_1}}$	0.1	mg L^{-1}
T. reesei RUTC30	Bioreactor; K12/RUTC30 on Avicel	$Y_{S_{G_1}/C_{Tr}}$	1.75	g-glucose g-biomass^{-1}
T. reesei RUTC30	Bioreactor; K12/RUTC30 on Avicel	$Y_{S_{G_2}/C_{Tr}}$	1.75	g-cellobiose g-biomass^{-1}
T. reesei RUTC30	Bioreactor; K12/RUTC30 on Avicel	$Y_{E_T/C_{Tr}}$	0.1	g-protein g-biomass^{-1}
T. reesei RUTC30	Bioreactor; K12/RUTC30 on Avicel	k_{E_T}	0.001	g-protein g-biomass^{-1} h^{-1}
T. reesei RUTC30	Bioreactor; K12/RUTC30 on Avicel	$\mu_{max,Tr,S_{G_1}}$	0.08	h^{-1}
Endoglucanase I	Bioreactor; K12/RUTC30 on Avicel	k_{EG1}	67.2	mmol g-EG1^{-1} h^{-1}
Cellobiohydrolase I	Bioreactor; K12/RUTC30 on Avicel	k_{CBH1}	13.44	mmol g-CBH1^{-1} h^{-1}
Cellobiohydrolase II	Bioreactor; K12/RUTC30 on Avicel	k_{CBH2}	26.88	mmol g-CBH2^{-1} h^{-1}
β-glucosidase I	Bioreactor; K12/RUTC30 on Avicel	k_{BGL,G_2}	949.2	mmol g-BGL^{-1} h^{-1}
β-glucosidase I	Bioreactor; K12/RUTC30 on Avicel	k_{BGL,G_3}	2122.4	mmol g-BGL^{-1} h^{-1}
β-glucosidase I	Bioreactor; K12/RUTC30 on Avicel	k_{BGL,G_3}	5068	mmol g-BGL^{-1} h^{-1}
E. coli K12	Bioreactor; K12/RUTC30 on Avicel	$k_{Ec,d}$	0.00001	h^{-1}
E. coli K12	Bioreactor; K12/RUTC30 on Avicel	$K_{Ec,S_{G_1}}$	0.00001	mg L^{-1}
E. coli K12	Bioreactor; K12/RUTC30 on Avicel	$Y_{S_{G_1}/C_{Ev}}$	3	g-glucose g-biomass^{-1}
E. coli K12	Bioreactor; K12/RUTC30 on Avicel	$m_{Ec,S_{G_1}}$	0.02	g-glucose g-biomass^{-1} h^{-1}
E. coli K12	Bioreactor; K12/RUTC30 on Avicel	$\mu_{max,Ec,S_{G_1}}$	0.0425	h^{-1}

Table B.3: Experimentally measured model parameters. *We estimated $\mu_{max,Ec,S_{G_1}}$ for *E. coli* NV3 pSA55/69 by multiplying $\mu_{max,Ec,S_{G_1}}$ for NV3 (i.e. the plasmid free strain) by the ratio of JCL260 pSA55/69 : JCL260 $\mu_{max,Ec,S_{G_1}}$ values.

	Major carbohydrates (g g^{-1})						
Species	**Ara**	**Xyl**	**Mann**	**Gal**	**Glc**	**Cry cel**	**Total**
T. reesei RUTC30	(1 ± 1)E-3	(5 ± 2)E-3	0.049 ± 0.008	0.021 ± 0.003	0.14 ± 0.02	0	0.21 ± 0.02
E. coli K12	(2 ± 0.4)E-4	(2 ± 0.3)E-4	(2 ± 0.5)E-4	(1 ± 0.2)E-3	0.029 ± 0.001	$0.021 \pm 0.005^{*}$	0.05 ± 0.005
AFEX CS	0.03 ± 0.002	0.20 ± 0.03	(2.6 ± 0.4)E-3	0.01 ± 0.001	0.034 ± 0.007	0.47 ± 0.04	0.74 ± 0.05

Table B.4: Carbohydrate composition of microbial biomass and AFEX pre-treated corn stover (taken from AFEX CS TMM media sample). Abbreviations: Arabinose, Ara; Xylose, Xyl; Mannose, Mann; Galactose, Gal; Hemicellulose glucan, Glc; Crystalline cellulose, Cry cel. *Neither *T. reesei* nor *E. coli* are known to produce crystalline cellulose; since the Updegraff assay is non-specific, the measured value probably reflects some other recalcitrant polysaccharide

pH	$\mu_{max,Tr}$ **(h^{-1})**	**Source**	$\mu_{max,Ec}$ **(h^{-1})**	**Source**	μ^{*}
5.3	0.104 ± 0.01	Est from [109]	0.09 ± 0.01	This study	0.9 ± 0.1
5.5	0.098 ± 0.01	Est from [109]	0.21 ± 0.03	This study	2.1 ± 0.4
6	0.092 ± 0.01	This study	0.39 ± 0.01	This study	4.2 ± 0.5

Table B.5: $\mu_{max,Tr}$, $\mu_{max,Ec}$, and μ^{*} for *E. coli* K12 and *T. reesei* RUTC30 on 20 g L^{-1} glucose TMM media at various pH levels. All $\mu_{max,Ec}$ values were measured experimentally; For *T. reesei*, we measured $\mu_{max,Tr}$ at pH 6 and extrapolated to pH 5.5 and 5.3 using data presented in [109].

APPENDIX C

Supporting figures for Chapter 2

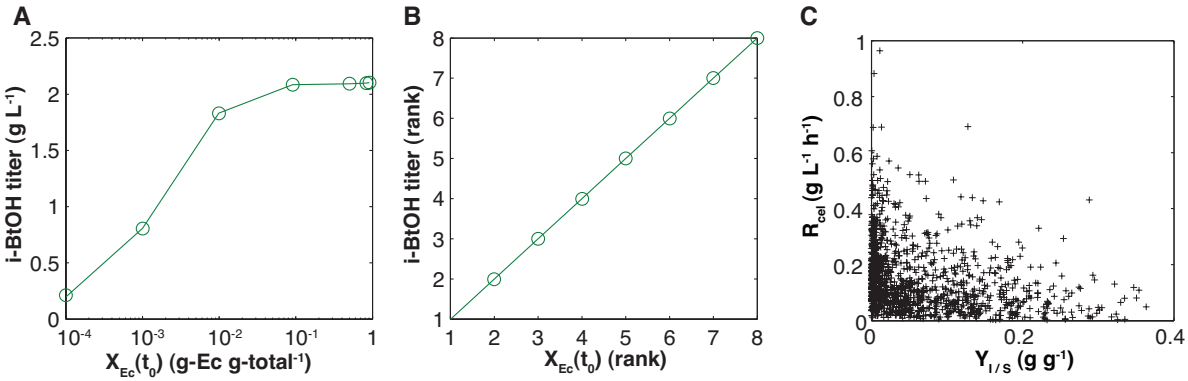

Figure C.1: Model predictions for isobutanol production with TrEc consortium. **(A)** Predicted isobutanol titer vs. $X_{Ec}(t_0)$ for *T. reesei* RUTC30 / *E. coli* NV3 pSA55/69 co-cultures on 20 g L^{-1} Avicel. **(B)** Rank transform of data in panel A (i.e. isobutanol titer rank vs. $X_{Ec}(t_0)$ rank); $R^2 = 1$ **(C)** Mean cellulose hydrolysis rate (R_{cel}; g L^{-1} h^{-1}) vs. isobutanol yield ($Y_{I/S}$; g g-cellulose^{-1}; 0.41 g g^{-1} theoretical maximum) over all parameter and initial condition (IC) values sampled in sensitivity analysis.

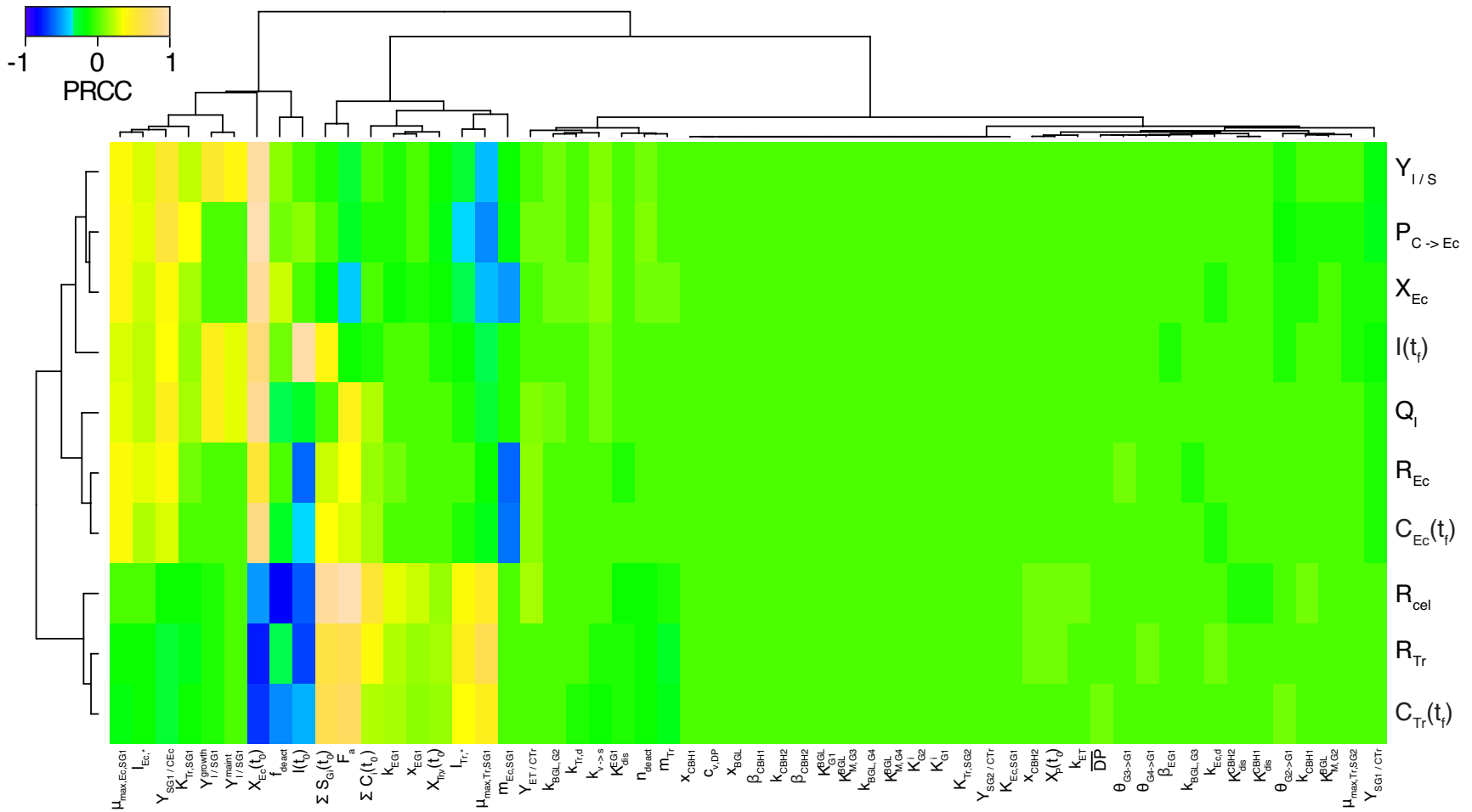

Figure C.2: Global sensitivity analysis of TrEC consortium model - full results. The TrEc consortium model was numerically integrated with 1000 different sets of parameter values and initial conditions (ICs) sampled from appropriate statistical distributions using latin hypercube selection (LHS) [87]. We calculated partial rank correlation coefficients (PRCC) [87] between each parameter or IC and a set of output metrics, including mean *T. reesei* growth rate (R_{Tr}; g L^{-1} h^{-1}), *T. reesei* titer ($C_{Tr}(t_f)$; g L^{-1}), mean *E. coli* growth rate (R_{Ec}; g L^{-1} h^{-1}), *E. coli* titer ($C_{Ec}(t_f)$; g L^{-1}), mean cellulose hydrolysis rate (R_{Cel}; g L^{-1} h^{-1}), *E. coli* population fraction at fermentation endpoint ($X_{Ec}(t_f)$; g g-total^{-1} microbial biomass), fraction of substrate carbon consumed by *E. coli* ($P_{C \to Ec}$; g g-total^{-1}), isobutanol yield ($Y_{I/S}$; g g-cellulose^{-1}), isobutanol titer ($I(t_f)$; g L^{-1}), and isobutanol productivity (Q_I; g g-cellulose^{-1} h^{-1}). Parameter PRCCs are shown with hierarchical clustering (Wards method; Euclidean distance); statistically insignificant PRCCs ($p < 0.05$) set to 0. For description of parameters, parameter values, and statistical distributions used, see Table B.1

APPENDIX D

Supporting information for Chapter 3

Parameter	Value	Description	Notes and source
$\mu_{max,Tr}$	0.087	Maximum specific growth rate of Tr on glucose (h^{-1})	Est. from [82], [251], and [246]
$K_{S,Tr}$	0.02	Tr glucose affinity (g L^{-1})	Est from [246]
$Y_{E_{cel}/C_{Tr}}$	0.5	Enzyme/biomass yield coefficient for Tr (g-protein g-biomass^{-1})	Est. from [14] and [250]
$Y_{E_{BGL}/C_{Tr}}$	0.015	β-glucosidase / biomass yield coefficient for Tr (g-BGL g-biomass^{-1})	Est. from [244] and [81]
$Y_{S_{G_1}/C_{Tr}}$	2	Glucose/biomass yield coefficient for Tr (g-glucose g-biomass^{-1})	Est. from [246]
k_{cel}	0.4	Cellulase rate constant (g-cellobiose g-cellulose^{-1} h^{-1})	Est. from [14]
k_{BGL}	116	β-glucosidase rate constant (g-glucose g-cellobiose^{-1} h^{-1})	Est. from [86]
$K_{M,BGL}$	0.465	β-glucosidase affinity for cellobiose (g L^{-1})	Est. from [86]
$Y_{BGL/cel}$	0.03	β-glucosidase fraction of total cellulase; same as x_{BGL} (g g-total^{-1})	Est. from [14] and [250]
α	8.7	$\frac{Y_{E_{BGL}/C_{Tr}} k_{BGL}}{Y_{E_{cel}/C_{Tr}} k_{cel}}$	–
$Y_{Tr/Ec}$	1.5	$Y_{S_{G_1}/C_{Ec}}/Y_{S_{G_1}/C_{Tr}}$	Biologically reasonable value

Table D.1: Baseline parameters used for steady state analysis of the simplified TrEc consortium model. Other parameters, including $\theta_{G_2 \to G_1}$, β, $K_{S,Ec}$ (K_S^*) and $\mu_{max,Ec}$ (μ^*) were varied to explore population dynamics and steady state composition.

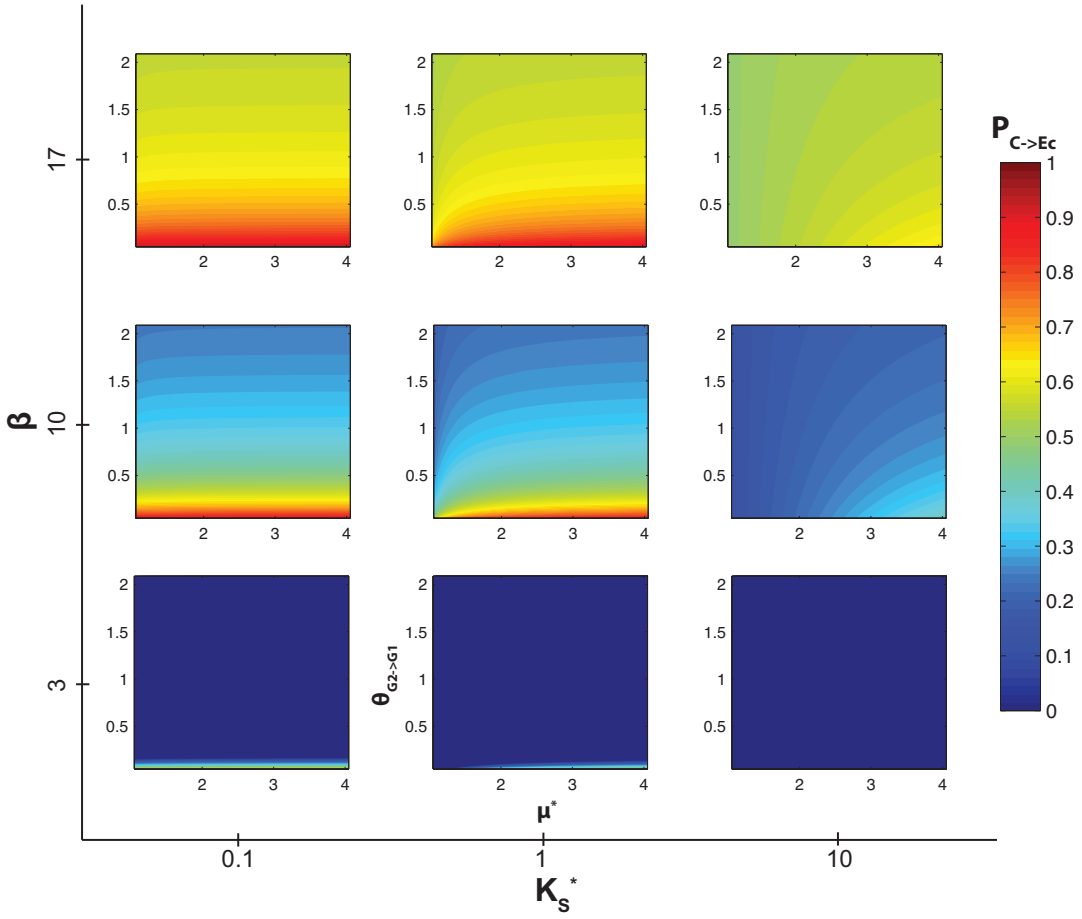

Figure D.1: Steady state analysis of simplified TrEc model. $P_{C\rightarrow Ec}$ (carbon flow partition to *E. coli*) as a function of β, K_S^*, μ^*, and $\theta_{G_2\rightarrow G_1}$ over biologically reasonable ranges for each parameter. Parameter values are as follows: $\alpha = 8.7$, $K_{BGL}^* = 23.25$, $Y_{Tr/Ec} = 1.5$; $\beta = 3, 10, 17$, $K_S^* = 0.1, 1, 10$, $\mu^* = 1.05...4.05$, and $\theta_{G_2\rightarrow G_1} = 0.05...2.05$. For explanation of parameters see section 3.3.1.

APPENDIX E

Supporting information for Chapter 4

E.1 Tables and figures

Table E.1: Mutations discovered in genome resequencing of evolved isobutanol tolerant *E. coli* clones. Mutation entries in bold denote loci that were mutated in all sequenced clones from end populations. Mutation positions are given as absolute genomic coordinates in the *E. coli* EcNR1 reference sequence. SNPs are indicated by base transition/transversion. Small insertions are indicated by a "+", with the size (number of bp) of the insertion and sequence of inserted bases. Small deletions are designated by "−" with a format similar to that for small insertions; for large deletions, the sequence of deleted bases is excluded. Transposons are indicated by the insertion sequence (IS) identity.

Clone	Gene	Gene Description	Biological Process	Cellular Location	Genomic Coordinate	Nucleotide Change	Protein Change
G3.2/G3.6	*phoQ*	sensory histidine kinase in two-component regulatory system with PhoP	Signal transduction	Inner membrane	1197581	A→G	L209P
	acrB	**multidrug efflux system protein**	**Transport**	**Inner and outer membrane**	**480665**	**G→A**	**P988L**
	queA	S-adenosylmethionine:tRNA ribosyltransferase-isomerase (EC:5.-.-.-)	Translation	Cytoplasm	425270	A→G	N346D
	secA	preprotein translocase subunit, ATPase	Protein secretion	Inner membrane	108975	T→C	S233P
	cadA	lysine decarboxylase 1 (EC:4.1.1.18)	Amino acid metabolism	Cytoplasm	4363790	A→G	I686T
	groL	Cpn60 chaperonin GroEL, large subunit of GroESL	Protein folding	Cytoplasm	4378650	A→C	K132N
	mutL	methyl-directed mismatch repair protein	Mismatch repair	Cytoplasm	4405650	-19 bp	Frameshift
	pstC	phosphate transporter subunit	Transport	Inner membrane	3917582	T→C	D16G
	rph	**defective ribonuclease PH**	**RNA processing**	**Cytoplasm**	**3823229**	**+4:GGTC**	**Frameshift**
	yiaK	2,3-diketo-L-gulonate dehydrogenase, NADH-dependent (EC:1.1.1.-)	Carbohydrate metabolism	Cytoplasm	3750540	T→C	L193P
	gltD	glutamate synthase, 4Fe-4S protein, small subunit (EC:1.4.1.13)	Amino acid, Nitrogen metabolism	Cytoplasm	3367270	+1:G	Frameshift
	mdh	**malate dehydrogenase, NAD(P)-binding (EC:1.1.1.37)**	**Tricarboxylic acid cycle**	**Membrane peripheral**	**3390726**	**-1:C**	**Frameshift**
	nlpI-pnp	lipoprotein involved in cell division	Cell cycle	Inner membrane	3316213	T→C	Non-coding region; Possible effect on terminator before nlpI
	glnE	deadenylyltransferase/adenylyltransferase for glutamine	Nitrogen metabolism	Cytoplasm	3205272	IS186 insertion	Disruption
	hybB	predicted hydrogenase 2 cytochrome b type component	Electron transport chain	Inner membrane	3150318	A→G	V359A

Table E.1: Mutations discovered in genome resequencing of evolved isobutanol tolerant *E. coli* clones

Clone	Gene	Gene Description	Biological Process	Cellular Location	Genomic Coordinate	Nucleotide Change	Protein Change
	gatZ	**D-tagatose 1,6-bisphosphate aldolase 2, subunit (EC:4.1.2.40)**	**Carbohydrate metabolism**	**Cytoplasm**	**2182915**	**-1:C**	**Frameshift**
	yeeE	predicted inner membrane protein	-	Inner membrane	2092513	A→G	S333P
	lepB	leader peptidase (signal peptidase I) (EC:3.4.21.89)	Transport	Inner membrane	2711902	G→A	P213S
	hfq	HF-I, host factor for RNA phage Q β replication	Translation	Cytoplasm	4407505	-7:AGGAAAA	Non-coding region; Ribosome binding site deletion
	marC	**conserved protein; predicted transporter**	**-**	**Inner membrane**	**1625925**	**IS1 insertion**	**Disruption**
G3.266.7	*groL*	Cpn60 chaperonin GroEL, large subunit of GroESL	Protein folding	Cytoplasm	4378650	A→C	K132N
	rph	**defective ribonuclease PH**	**RNA processing**	**Cytoplasm**	**3823229**	**+4:GGTC**	**Frameshift**
	gltB-yhcE	-	-	-	-	-9.9kb	ΔgltBDF, ΔyhcADE
	mdh	**malate dehydrogenase, NAD(P)-binding (EC:1.1.1.37)**	**Tricarboxylic acid cycle**	**Membrane peripheral**	**3390726**	**-1:C**	**Frameshift**
	glnE	deadenylyltransferase/adenylyltransferase for glutamine	Nitrogen metabolism	Cytoplasm	3205272	IS186 insertion	Disruption
	hfq	HF-I, host factor for RNA phage Q β replication	Translation	Cytoplasm	4407505	-7:AGGAAAA	Non-coding region; Ribosome binding site deletion
	ycfK	e14 prophage; predicted protein	-	-	1216432	-1688 bp	ΔycfK
	marC	**conserved protein; predicted transporter**	**-**	**Inner membrane**	**1625925**	**IS1 insertion**	**Disruption**
X3.5	***acrA***	**multidrug efflux system protein**	**Transport**	**Inner and outer membrane**	**483735**	**+1:A**	**Frameshift**
	rpsB	30S ribosomal subunit protein S2	Translation	Cytoplasm	190557	+1:A	Frameshift
	rph	**defective ribonuclease PH**	**RNA processing**	**Cytoplasm**	**3823180**	**-1:C**	**Frameshift**
	mdh	**malate dehydrogenase, NAD(P)-binding (EC:1.1.1.37)**	**Tricarboxylic acid cycle**	**Membrane peripheral**	**3390936**	**+5:AACCT**	**Frameshift**
	deaD	DEAD-box RNA helicase	Translation	Cytoplasm	3314027	+4:AGAC	Frameshift
	yfgO	predicted inner membrane protein	-	Inner membrane	2623022	C→T	G30D
	gatC	**galactitol-specific enzyme IIC component of PTS**	**Transport**	**Inner membrane**	**2180640**	**C→T**	**E290K**
	plsX	fatty acid/phospholipid synthesis protein	Fatty acid metabolism	Cytoplasm	1493514	A→G	E216G
	hrpA	ATP-dependent helicase	RNA processing	Cytoplasm	1493514	C→T	L1075L
	mdtJ-tqsA	MdtJ SMR protein; transporter of quorum signal AI-2	Transporter/Transporter	Inner membrane	1681114	IS5 insertion	Non-coding region; mdtJ and tqsA promoter region
	marC	**conserved protein; predicted transporter**	**-**	**Inner membrane**	**1626081**	**IS1 insertion**	**Disruption**
X2.5*	***acrA***	**multidrug efflux system protein**	**Transport**	**Inner and outer membrane**	**484383**	**T→G**	**N154T**
	gltA	citrate synthase	Tricarboxylic acid cycle	Membrane and cytoplasm	753345	T→C	E116G
	marC	**conserved protein; predicted transporter**	**-**	**Inner membrane**	**1626084**	**IS5 insertion**	**Disruption**
	gatC	**galactitol-specific enzyme IIC component of PTS**	**Transport**	**Inner membrane**	**2180721**	**C→A**	**V263L**
	mdh	**malate dehydrogenase, NAD(P)-binding (EC:1.1.1.37)**	**Tricarboxylic acid cycle**	**Membrane peripheral**	**3390659**	**+4:GATT**	**Frameshift**
	pyrE-rph	**defective ribonuclease PH**	**RNA processing**	**Cytoplasm**	**3822991**	**G→A**	**Non-coding region**

Table E.1: *See addendum on X2.5 sequencing (section 4.5).

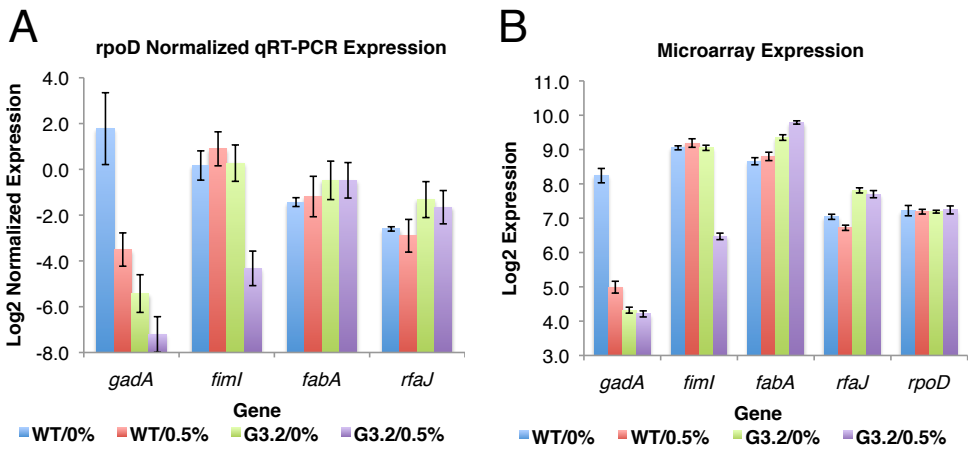

Figure E.1: qRT-PCR validation of *gadA*, *fimI*, *fabA*, and *rfaJ* gene expression changes. qRT-PCR was used to validate gene expression changes measured by DNA microarray. Target concentrations were determined by fitting the MAK2 PCR model to qRT-PCR data [135]. Expression levels were normalized to house keeping gene *rpoD* (sigma factor 70). **(A)** rpoD normalized expression levels determined by qRT-PCR. **(B)** Expression levels from DNA microarray study.

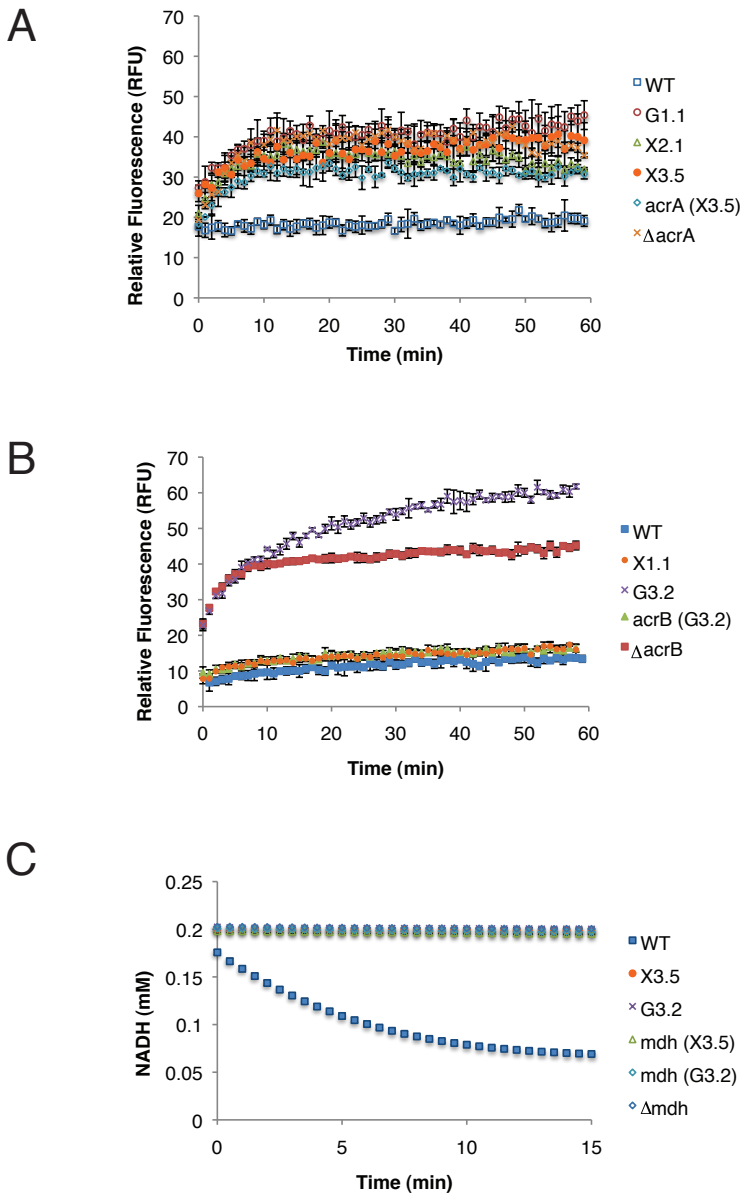

Figure E.2: *acrAB* and *mdh* functional assays. AcrAB-TolC efflux pump activity was measured via ethidium bromide (EtBr) accumulation in reconstructed single mutants and clonal isolates harbouring *acrAB* mutations from evolution end populations. Mid log phase cells were incubated with ethidium bromide and intracellular ethidium bromide was monitored via relative fluorescence (518 nm excitation / 605 nm emission). Mdh (NADH dependent malate dehydrogenase) activity was assayed by incubating cell extracts with oxaloacetate and NADH; disappearance of NADH (due to reduction of oxaloacetate to malate) was monitored by measuring absorbance at 340 nm. **(A)** EtBr accumulation assay for the parent *E. coli* EcNR1 (WT), clonal isolates from evolution end populations harbouring *acrA* mutations (G1.1, X2.1, X3.5), a reconstructed *acrA* single mutant (containing mutation found in X3.5), and Δ*acrA::kan* control. **(B)** EtBr accumulation assay for the parent *E. coli* EcNR1 (WT), clonal isolates from evolution end populations harbouring *acrB* mutations (X1.1, G3.2), a reconstructed *acrB* single mutant (containing mutation found in G3.2), and Δ*acrB::kan* control. **(C)** Mdh assay for the parent *E. coli* EcNR1 (WT), clonal isolates from evolution end populations harbouring *mdh* mutations (G3.2, X3.5), reconstructed *mdh* single mutants (containing mutations found in G3.2 or X3.5), and Δ*mdh::kan control*.

E.2 Additional files

The additional files referenced in Chapter 4 accompany our publication:

J. Minty, A. Lesnefsky, F. Lin, Y. Chen, T. Zaroff, A. Veloso, B. Xie, C. McConnell, R. Ward, D. Schwartz, J.M. Rouillard, Y. Gao, E. Gulari, and X. Lin. Evolution combined with genomic study elucidates genetic bases of isobutanol tolerance in *Escherichia coli*. *Microbial Cell Factories*, 10:18, 2011. doi:10.1186/1475-2859-10-18

Additional file numbers correspond to those used in the publication, except files 5 and 6 are included in this dissertation (Figures E.1 & E.2, respectively) and Additional file 7 in the publication has been renumbered to Additional file 5 in this dissertation. Descriptions of and direct hyperlinks to each additional file (hosted on *Microbial Cell Factories* website) are given below. In that event that the hyperlinks are corrupt, please access the online version of the publication through DOI (e.g. http://dx.doi.org/10.1186/1475-2859-10-18).

Additional file 1: Full mutation list

Full lists of SNP, indel, and SV mutations discovered in G3.2, G3.6, G3.266.7, and X3.5 with Illumina sequencing. Mutation positions are genomic coordinates in the *E. coli* EcNR1 reference sequence, gene descriptions are from the KEGG database, mutation frequency is defined as mutant reads divided by the total number of mapped reads at a position, and consensus quality was computed by MAQ [130]. For G3.2, G3.266.7, and X3.5, SNPs with consensus quality $<$150 or indels with frequency $<$0.4 were discarded; we have found empirically that lower quality/frequency mutations tend to be false positives. G3.6 may have been contaminated with another clonal isolate and thus may be mixed genotype; to reduce false negatives, quality cutoff thresholds were lowered to consensus quality $<$100 for SNPs and frequency $<$0.15 for indels. Mutations that were later discovered to be heterogeneities in the parent *E. coli* EcNR1 strain were discarded. Entries with red text were verified by Sanger sequencing.

Additional file 2: *E. coli* EcNR1 genome reference sequence

E. coli EcNR1 is a derivative of *E. coli* K12 MG1655 containing a modified λ prophage integrated at the *bioA*/*bioB* locus. We created a reference genome sequence for *E. coli* EcNR1 by adding the above genetic modification to the *E. coli K12 MG1655* reference sequence (NC_000913) obtained from the National Center for Biotechnology Information Reference Sequence Collection (NCBI RefSeq). We provide three formats: Lasergene DNA (.seq), FASTA (.fas), and GenBank (.gbk).

Additional file 3: Tracing mutations found in endpoint populations through intermediate generations

We investigated the dynamics of genotypic adaptation in the G3 and X3 lineages by genotyping population samples from intermediate generations for selected mutations identified in the end point populations. Genotyping was conducted by screening whole-population cryopreserved samples for mutations with Sanger sequencing of PCR amplified mutated regions (for G3 *mdh* and *miaA-hfq* mutations), inferred from PCR product size for large insertion mutations (*marC* transposon insertions, X3 *mdtj::IS5::tqsA*, and G3 *glnE::IS186*) or using allele specific PCR (all other genotyped mutations). WT designates wild-type allele (directly detected in Sanger sequencing or inferred from lack of allele specific PCR product), Mut designates mutant allele (directly detected in Sanger sequencing or inferred from of amplification of allele specific PCR product), NT designates not tested. Strength (strong, weak, etc) indicates band intensity on agarose gel electrophoresis of PCR product, and is roughly correlated with allele frequency (allele specific PCR and large insertions). Sanger sequencing and genotyping via PCR product sizes allow discrimination of mixed genotypes, which are reported where applicable.

Additional file 4: Microarray Data and Analyses

Microarray data for gene expression study of G3.2 and the parent *E. coli* EcNR1 (WT) in 0% and 0.5% (w/v) isobutanol glucose minimal medium. Genes that responded to isobutanol the most differently between G3.2 and WT are tabulated with p-values and transcription factors controlling them. Differentially expressed genes, p-values, and related transcription factors are also tabulated for WT / 0% isobutanol and WT/ 0.5% isobutanol, G3.2 / 0% isobutanol and G3.2/ 0.5% isobutanol, and lastly WT / 0.5% isobutanol and G3.2/ 0.5% isobutanol. BiNGO was used to assess any overrepresented GO terms amongst genes that responded to isobutanol the most differently between G3.2 and WT, and for genes differentially expressed between WT / 0.5% isobutanol and G3.2/ 0.5% isobutanol. A summary of NCA results is given as well.

Additional file 5: Primers and oligos used in this study

Sequences of forward and reverse primers used for Sanger sequencing, allele specific PCR, qRT-PCR, and all other PCR reactions described in this study are listed. Sequences of oligonucleotides used for ssDNA mediated homologous recombination are also given.

BIBLIOGRAPHY

[1] Atsumi, S., Hanai, T., and Liao, J. C., "Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels," *Nature*, Vol. 451, No. 7174, 2008, pp. 86–89.

[2] Baez, A., Cho, K. M., and Liao, J. C., "High-flux isobutanol production using engineered *Escherichia coli*: a bioreactor study with in situ product removal," *Applied Microbiology and Biotechnology*, Vol. 90, No. 5, 2011, pp. 1681–1690.

[3] Smith, K. M. and Liao, J. C., "An evolutionary strategy for isobutanol production strain development in *Escherichia coli*," *Metabolic Engineering*, Vol. 13, No. 6, 2011, pp. 674–681.

[4] U.S. Energy Information Administration., "Annual Energy Review 2011," September 2012.

[5] U.S. Department of Energy Genomic Science Program., "DOE Genomics: GTL Roadmap. Systems biology for energy and environment," August 2005.

[6] Rude, M. A. and Schirmer, A., "New microbial fuels: a biotech perspective," *Current Opinion in Microbiology*, Vol. 12, No. 3, 2009, pp. 274–281.

[7] Demirbas, A., "Political, economic and environmental impacts of biofuels: A review," *Applied Energy*, Vol. 86, Supplement 1, No. 0, 2009, pp. S108–S117.

[8] Zhang, F., Rodriguez, S., and Keasling, J. D., "Metabolic engineering of microbial pathways for advanced biofuels production," *Current Opinion in Biotechnology*, Vol. 22, No. 6, 2011, pp. 775–783.

[9] Voet, D. and Voet, J. G., *Biochemistry*, 3rd ed., John Wiley and Sons, Hoboken, NJ, 2004.

[10] Gelfand, I., Sahajpal, R., Zhang, X., Izaurralde, R. C., Gross, K. L., and Robertson, G. P., "Sustainable bioenergy production from marginal lands in the US Midwest," *Nature*, Vol. 493, No. 7433, 2013, pp. 514–517.

[11] Lee, S., Oh, Y., Kim, D., Kwon, D., Lee, C., and Lee, J., "Converting carbohydrates extracted from marine algae into ethanol using various ethanolic *Escherichia coli* strains," *Applied Biochemistry and Biotechnology*, Vol. 164, 2011, pp. 878–888.

[12] Wargacki, A. J., Leonard, E., Win, M. N., Regitsky, D. D., Santos, C. N. S., Kim, P. B., Cooper, S. R., Raisner, R. M., Herman, A., Sivitz, A. B., Lakshmanaswamy, A., Kashiyama, Y., Baker, D., and Yoshikuni, Y., "An engineered microbial platform for direct biofuel production from brown macroalgae," *Science*, Vol. 335, No. 6066, 2012, pp. 308–313.

[13] Perlack, R., Wright, L., Turhollow, A., R.L., G., Stokes, B., and Erbach, D. U.S. Department of Energy and U.S. Department of Agriculture., "Biomass as feedstock for a bioenergy and bioproducts industry: the technical feasibility of a billion-ton annual supply," April 2005.

[14] Lynd, L. R., Weimer, P. J., van Zyl, W. H., and Pretorius, I. S., "Microbial cellulose utilization: fundamentals and biotechnology," *Microbiology and Molecular Biology Reviews*, Vol. 66, No. 3, 2002, pp. 506–577.

[15] Lynd, L. R., van Zyl, W. H., McBride, J. E., and Laser, M., "Consolidated bioprocessing of cellulosic biomass: an update," *Current Opinion in Biotechnology*, Vol. 16, No. 5, 2005, pp. 577–583.

[16] Dellomonaco, C., Fava, F., and Gonzalez, R., "The path to next generation biofuels: successes and challenges in the era of synthetic biology," *Microbial Cell Factories*, Vol. 9, No. 3, 2010.

[17] Joshi, C. P., Nookaraju, A., Sekhon, K. K., Khanna, S., Cameotra, S. S., Binta, M., Mushi, E., Liu, S., Goldfeld, M. G., and Podella, C. W., "New avenues of bioenergy production from plants: green alternatives to petroleum," *Journal of Petroleum and Environmental Biotechnology*, Vol. 3, No. 134, 2012.

[18] Olson, D. G., McBride, J. E., Joe Shaw, A., and Lynd, L. R., "Recent progress in consolidated bioprocessing," *Current Opinion in Biotechnology*, Vol. 23, No. 3, 2012, pp. 396–405.

[19] Argyros, D. A., Tripathi, S. A., Barrett, T. F., Rogers, S. R., Feinberg, L. F., Olson, D. G., Foden, J. M., Miller, B. B., Lynd, L. R., Hogsett, D. A., and Caiazza, N. C., "High ethanol titers from cellulose by using metabolically engineered thermophilic, anaerobic microbes," *Applied and Environmental Microbiology*, Vol. 77, No. 23, 2011, pp. 8288–8294.

[20] Higashide, W., Li, Y., Yang, Y., and Liao, J. C., "Metabolic engineering of *Clostridium cellulolyticum* for production of isobutanol from cellulose," *Applied and Environmental Microbiology*, Vol. 77, No. 8, 2011, pp. 2727–2733.

[21] Desvaux, M., "*Clostridium cellulolyticum*: model organism of mesophilic cellulolytic clostridia," *FEMS Microbiology Reviews*, Vol. 29, No. 4, 2006, pp. 741–764.

[22] Thirumalai Vasan, P., Sobana Piriya, P., Immanual Gilwax Prabhu, D., and John Vennison, S., "Cellulosic ethanol production by *Zymomonas mobilis* harboring an endoglucanase gene from *Enterobacter cloacae*," *Bioresource Technology*, Vol. 102, No. 3, 2011, pp. 2585–2589.

[23] Fierobe, H.-P., Mingardon, F., and Chanal, A., "Engineering cellulase activity into *Clostridium acetobutylicum*," *Methods in Enzymology*, Vol. 510, 2012, pp. 301–316.

[24] Brenner, K., You, L., and Arnold, F. H., "Engineering microbial consortia: a new frontier in synthetic biology," *Trends in Biotechnology*, Vol. 26, No. 9, 2008, pp. 483–9.

[25] Mee, M. T. and Wang, H. H., "Engineering ecosystems and synthetic ecologies," *Molecular BioSystems*, Vol. 8, 2012, pp. 2470–2483.

[26] Sabra, W., Dietz, D., Tjahjasari, D., and Zeng, A.-P., "Biosystems analysis and engineering of microbial consortia for industrial biotechnology," *Engineering in Life Sciences*, Vol. 10, No. 5, 2010, pp. 407–421.

[27] Werner, J. J., Knights, D., Garcia, M. L., Scalfone, N. B., Smith, S., Yarasheski, K., Cummings, T. A., Beers, A. R., Knight, R., and Angenent, L. T., "Bacterial community structures are unique and resilient in full-scale bioenergy systems," *Proceedings of the National Academy of Sciences*, Vol. 108, No. 10, 2011, pp. 4158–4163.

[28] Agapakis, C. M., Boyle, P. M., and Silver, P. A., "Natural strategies for the spatial optimization of metabolism in synthetic biology," *Nature Chemical Biology*, Vol. 8, No. 6, 2012, pp. 527–535.

[29] Ronan, P., William Yeung, C., Schellenberg, J., Sparling, R., Wolfaardt, G. M., and Hausner, M., "A versatile and robust aerotolerant microbial community capable of cellulosic ethanol production," *Bioresource Technology*, Vol. 129, No. 0, 2013, pp. 156–163.

[30] Zuroff, T. and Curtis, W., "Developing symbiotic consortia for lignocellulosic biofuel production," *Applied Microbiology and Biotechnology*, Vol. 93, No. 4, 2012, pp. 1423–1435.

[31] Kerner, A., Minty, J., Singer, M., Faulkner, I., Dale, B., Balan, V., and Lin, X., "A synthetic Escherichia coli consortium for efficient conversion of hexose and pentose monomers and oligomers to isobutanol," *To be submitted*, 2013.

[32] Bayer, T. S., Widmaier, D. M., Temme, K., Mirsky, E. A., Santi, D. V., and Voigt, C. A., "Synthesis of methyl halides from biomass using engineered microbes," *Journal of the American Chemical Society*, Vol. 131, No. 18, 2009, pp. 6508–6515.

[33] Trinh, C. T., Unrean, P., and Srienc, F., "Minimal *Escherichia coli* cell for the most efficient production of ethanol from hexoses and pentoses," *Applied and Environmental Microbiology*, Vol. 74, No. 12, 2008, pp. 3634–3643.

[34] Eiteman, M. A., Lee, S. A., and Altman, E., "A co-fermentation strategy to consume sugar mixtures effectively," *Journal of Biological Engineering*, Vol. 2, No. 3, 2008.

[35] Eiteman, M. A., Lee, S. A., Altman, R., and Altman, E., "A substrate-selective co-fermentation strategy with *Escherichia coli* produces lactate by simultaneously consuming xylose and glucose," *Biotechnology and Bioengineering*, Vol. 102, No. 3, 2009, pp. 822–827.

[36] Xia, T., Eiteman, M. A., and Altman, E., "Simultaneous utilization of glucose, xylose and arabinose in the presence of acetate by a consortium of *Escherichia coli* strains," *Microbial Cell Factories*, Vol. 11, No. 77, 2012.

[37] Tsai, S.-L., Goyal, G., and Chen, W., "Surface display of a functional minicellulosome by intracellular complementation using a synthetic yeast consortium and its application to cellulose hydrolysis and ethanol production," *Applied and Environmental Microbiology*, Vol. 76, No. 22, November 15, 2010, pp. 7514–7520.

[38] Bokinsky, G., Peralta-Yahya, P. P., George, A., Holmes, B. M., Steen, E. J., Dietrich, J., Soon Lee, T., Tullman-Ercek, D., Voigt, C. A., Simmons, B. A., and Keasling, J. D., "Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered *Escherichia coli*," *Proceedings of the National Academy of Sciences*, Vol. 108, No. 50, 2011, pp. 19949–19954.

[39] Arai, T., Matsuoka, S., Cho, H.-Y., Yukawa, H., Inui, M., Wong, S.-L., and Doi, R. H., "Synthesis of *Clostridium cellulovorans* minicellulosomes by intercellular complementation," *Proceedings of the National Academy of Sciences*, Vol. 104, No. 5, 2007, pp. 1456–1460.

[40] Goyal, G., Tsai, S. L., Madan, B., DaSilva, N. A., and Chen, W., "Simultaneous cell growth and ethanol production from cellulose by an engineered yeast consortium displaying a functional mini-cellulosome," *Microbial Cell Factories*, Vol. 10, No. 89, 2011.

[41] Balagaddé, F. K., Song, H., Ozaki, J., Collins, C. H., Barnet, M., Arnold, F. H., Quake, S. R., and You, L., "A synthetic *Escherichia coli* predator–prey ecosystem," *Molecular Systems Biology*, Vol. 4, No. 1, 2008.

[42] Shou, W., Ram, S., and Vilar, J. M. G., "Synthetic cooperation in engineered yeast populations," *Proceedings of the National Academy of Sciences*, Vol. 104, No. 6, 2007, pp. 1877–1882.

[43] Regot, S., Macia, J., Conde, N., Furukawa, K., Kjellén, J., Peeters, T., Hohmann, S., de Nadal, E., Posas, F., and Solé, R., "Distributed biological computation with multicellular engineered networks," *Nature*, Vol. 469, No. 7329, 2010, pp. 207–211.

[44] Prindle, A., Samayoa, P., Razinkov, I., Danino, T., Tsimring, L. S., and Hasty, J., "A sensing array of radically coupled genetic 'biopixels'," *Nature*, Vol. 481, No. 7379, 2012, pp. 39–44.

[45] Kim, H. J., Boedicker, J. Q., Choi, J. W., and Ismagilov, R. F., "Defined spatial structure stabilizes a synthetic multispecies bacterial community," *Proceedings of the National Academy of Sciences*, Vol. 105, No. 47, 2008, pp. 18188–18193.

[46] Arkin, A. P. and Fletcher, D. A., "Fast, cheap and somewhat in control," *Genome Biology*, Vol. 7, No. 114, 2006.

[47] Goldman, R. P. and Brown, S. P., "Making sense of microbial consortia using ecology and evolution," *Trends in Biotechnology*, Vol. 27, No. 1, 2009, pp. 3–4.

[48] Zhang, Y., Zhu, Y., Zhu, Y., and Li, Y., "The importance of engineering physiological functionality into microbes," *Trends in Biotechnology*, Vol. 27, No. 12, 2009, pp. 664–672.

[49] Minty, J. J., Lin, F., Nelson, M. C., and Lin, X. N., "Elucidating chemical toxicity and improving microbial tolerance for biofuel production: an update," *Biotechnology Journal*, in revision, 2013.

[50] Nicolaou, S. A., Gaida, S. M., and Papoutsakis, E. T., "A comparative view of metabolite and substrate stress and tolerance in microbial bioprocessing: from biofuels and chemicals, to biocatalysis and bioremediation," *Metabolic Engineering*, Vol. 12, No. 4, 2010, pp. 307-331.

[51] Warner, J. R., Patnaik, R., and Gill, R. T., "Genomics enabled approaches in strain engineering," *Current Opinion in Microbiology*, Vol. 12, No. 3, 2009, pp. 223–230.

[52] Alper, H. and Stephanopoulos, G., "Global transcription machinery engineering: a new approach for improving cellular phenotype," *Metabolic Engineering*, Vol. 9, No. 3, 2007, pp. 258–267.

[53] Klein-Marcuschamer, D., Santos, C. N., Yu, H., and Stephanopoulos, G., "Mutagenesis of the bacterial RNA polymerase alpha subunit for improvement of complex phenotypes," *Applied and Environmental Microbiology*, Vol. 75, No. 9, 2009, pp. 2705–2711.

[54] Santos, C. N. and Stephanopoulos, G., "Combinatorial engineering of microbes for optimizing cellular phenotype," *Current Opinion in Chemical Biology*, Vol. 12, No. 2, 2008, pp. 168–176.

[55] Nicolaou, S. A., Gaida, S. M., and Papoutsakis, E. T., "Coexisting/Coexpressing Genomic Libraries (CoGeL) identify interactions among distantly located genetic loci for developing complex microbial phenotypes," *Nucleic Acids Research*, Vol. 39, No. 22, 2011, pp. e152–e152.

[56] Warner, J. R., Reeder, P. J., Karimpour-Fard, A., Woodruff, L. B. A., and Gill, R. T., "Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides," *Nature Biotechnology*, Vol. 28, No. 8, 2010, pp. 856–862.

[57] Elena, S. F. and Lenski, R. E., "Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation," *Nature Reviews Genetics*, Vol. 4, No. 6, 2003, pp. 457–469.

[58] Conrad, T. M., Lewis, N. E., and Palsson, B. Ø., "Microbial laboratory evolution in the era of genome-scale science," *Molecular Systems Biology*, Vol. 7, No. 509 2011.

[59] Wang, H. H., Isaacs, F. J., Carr, P. A., Sun, Z. Z., Xu, G., Forest, C. R., and Church, G. M., "Programming cells by multiplex genome engineering and accelerated evolution," *Nature*, Vol. 460, No. 7257, 2009, pp. 894–898.

[60] Isaacs, F. J., Carr, P. A., Wang, H. H., Lajoie, M. J., Sterling, B., Kraal, L., Tolonen, A. C., Gianoulis, T. A., Goodman, D. B., Reppas, N. B., Emig, C. J., Bang, D., Hwang, S. J., Jewett, M. C., Jacobson, J. M., and Church, G. M., "Precise manipulation of chromosomes *in vivo* enables genome-wide codon replacement," *Science*, Vol. 333, No. 6040, 2011, pp. 348–353.

[61] Wang, H. H., Kim, H., Cong, L., Jeong, J., Bang, D., and Church, G. M., "Genome-scale promoter engineering by coselection MAGE," *Nature Methods*, Vol. 9, No. 6, 2012, pp. 591–593.

[62] Carr, P. A., Wang, H. H., Sterling, B., Isaacs, F. J., Lajoie, M. J., Xu, G., Church, G. M., and Jacobson, J. M., "Enhanced multiplex genome engineering through co-operative oligonucleotide co-selection," *Nucleic Acids Research*, Vol. 40, No. 17, 2012, pp. e132–e132.

[63] Gibson, D. G., Glass, J. I., Lartigue, C., Noskov, V. N., Chuang, R.-Y., Algire, M. A., Benders, G. A., Montague, M. G., Ma, L., Moodie, M. M., Merryman, C., Vashee, S., Krishnakumar, R., Assad-Garcia, N., Andrews-Pfannkoch, C., Denisova, E. A., Young, L., Qi, Z.-Q., Segall-Shapiro, T. H., Calvey, C. H., Parmar, P. P., Hutchison, C. A., Smith, H. O., and Venter, J. C., "Creation of a bacterial cell controlled by a chemically synthesized genome," *Science*, Vol. 329, No. 5987, 2010, pp. 52–56.

[64] Esvelt, K. M. and Wang, H. H., "Genome-scale engineering for systems and synthetic biology," *Molecular Systems Biology*, Vol. 9, No. 641, 2013.

[65] Stephanopoulos, G., "Challenges in engineering microbes for biofuels production," *Science*, Vol. 315, No. 5813, 2007, pp. 801–804.

[66] Shong, J., Jimenez Diaz, M. R., and Collins, C. H., "Towards synthetic microbial consortia for bioprocessing," *Current Opinion in Biotechnology*, Vol. 23, No. 5, 2012, pp. 798-802.

[67] Szambelan, K., Nowak, J., and Czarnecki, Z., "Use of *Zymomonas mobilis* and *Saccharomyces cerevisiae* mixed with *Kluyveromyces fragilis* for improved ethanol production from Jerusalem artichoke tubers," *Biotechnology Letters*, Vol. 26, No. 10, 2004, pp. 845–848.

[68] Mamma, D., Koullas, D., Fountoukidis, G., Kekos, D., Macris, B. J., and Koukios, E., "Bioethanol from sweet sorghum: simultaneous saccharification and fermentation of carbohydrates by a mixed microbial culture," *Process Biochemistry*, Vol. 31, No. 4, 1996, pp. 377–381.

[69] Park, E., Naruse, K., and Kato, T., "One-pot bioethanol production from cellulose by co-culture of *Acremonium cellulolyticus* and *Saccharomyces cerevisiae*," *Biotechnology for Biofuels*, Vol. 5, No. 64, 2012.

[70] Hahnhagerdal, B. and Haggstrom, M., "Production of ethanol from cellulose, Solka Floc Bw-200, in a fedbatch mixed culture of *Trichoderma reesei*, C-30, and *Saccharomyces cerevisiae*," *Applied Microbiology and Biotechnology*, Vol. 22, No. 3, 1985, pp. 187–189.

[71] Tran, H. T. M., Cheirsilp, B., Hodgson, B., and Umsakul, K., "Potential use of *Bacillus subtilis* in a co-culture with *Clostridium butylicum* for acetone–butanol–ethanol production from cassava starch," *Biochemical Engineering Journal*, Vol. 48, No. 2, 2010, pp. 260–267.

[72] Nakayama, S., Kiyoshi, K., Kadokura, T., and Nakazato, A., "Butanol production from crystalline cellulose by cocultured *Clostridium thermocellum* and *Clostridium saccharoperbutylacetonicum* N1-4," *Applied and Environmental Microbiology*, Vol. 77, No. 18, 2011, pp. 6470–6475.

[73] Kristensen, E., Ahmed, S. I., and Devol, A. H., "Aerobic and anaerobic decomposition of organic matter in marine sediment: which is fastest?" *Limnology and Oceanography*, 1995, pp. 1430–1437.

[74] Zhang, Y. and Lynd, L. R., "Quantification of cell and cellulase mass concentrations during anaerobic cellulose fermentation: development of an enzyme-linked immunosorbent assay-based method with application to *Clostridium thermocellum* batch cultures," *Analytical Chemistry*, Vol. 75, No. 2, 2002, pp. 219–227.

[75] Portnoy, V. A., Herrgard, M. J., and Palsson, B. Ø., "Aerobic fermentation of D-glucose by an evolved cytochrome oxidase-deficient *Escherichia coli* strain," *Applied and Environmental Microbiology*, Vol. 74, No. 24, 2008, pp. 7561–7569.

[76] "*Trichoderma reesei* v2.0," U.S. Department of Energy Joint Genome Institute [http://genome.jgi-psf.org/Trire2/Trire2.home.html. Accessed 4/20/13.]

[77] Haab, D., Gassner, B., and Kubicek, C. P., "Protein hypersecretory *Trichoderma reesei* mutant RUT C-30 displays increased ethanol and polyene resistance," *Journal of Biotechnology*, Vol. 29, No. 1, 1993, pp. 97–108.

[78] "*Cytophaga hutchinsonii* ATCC 33406," U.S. Department of Energy Joint Genome Institute [http://gp-edge3.jgi-psf.org:1080/Cytophaga/Cytophaga.info.html. Accessed 4/20/13.]

[79] Meyer, V., "Genetic engineering of filamentous fungi: progress, obstacles and future trends," *Biotechnology Advances*, Vol. 26, No. 2, 2008, pp. 177–185.

[80] Seidl, V., Gamauf, C., Druzhinina, I., Seiboth, B., Hartl, L., and Kubicek, C., "The *Hypocrea jecorina* (*Trichoderma reesei*) hypercellulolytic mutant RUT C30 lacks a 85 kb (29 gene-encoding) region of the wild-type genome," *BMC Genomics*, Vol. 9, No. 327, 2008.

[81] Kubicek, C. P., "Release of carboxymethyl-cellulase and beta-glucosidase from cell-walls of *Trichoderma reesei*," *European Journal of Applied Microbiology and Biotechnology*, Vol. 13, No. 4, 1981, pp. 226–231.

[82] Bader, J., Klingspohn, U., Bellgardt, K. H., and Schugerl, K., "Modeling and simulation of the growth and enzyme-production of *Trichoderma reesei* Rut C30," *Journal of Biotechnology*, Vol. 29, No. 1-2, 1993, pp. 121–135.

[83] Fogler, H. S., *Elements of chemical reaction engineering*, 4th ed., Prentice Hall, Upper Saddle River, NJ, 2006.

[84] Mach, R. L., Strauss, J., Zeilinger, S., Schindler, M., and Kubicek, C. P., "Carbon catabolite repression of xylanase I (*xyn1*) gene expression in *Trichoderma reesei*," *Molecular Microbiology*, Vol. 21, No. 6, 1996, pp. 1273–1281.

[85] Zhang, Y. H. and Lynd, L. R., "A functionally based model for hydrolysis of cellulose by fungal cellulase," *Biotechnology and Bioengineering*, Vol. 94, No. 5, 2006, pp. 888–898.

[86] Schmid, G. and Wandrey, C., "Characterization of a cellodextrin glucohydrolase with soluble oligomeric substrates: experimental results and modeling of concentration-time-course data," *Biotechnology and Bioengineering*, Vol. 33, No. 11, 1989, pp. 1445–1460.

[87] Marino, S., Hogue, I. B., Ray, C. J., and Kirschner, D. E., "A methodology for performing global uncertainty and sensitivity analysis in systems biology," *Journal of Theoretical Biology*, Vol. 254, No. 1, 2008, pp. 178–196.

[88] Teymouri, F., Laureano-Perez, L., Alizadeh, H., and Dale, B. E., "Ammonia fiber explosion treatment of corn stover," *Applied Biochemistry and Biotechnology*, Vol. 113-116, 2004, pp. 951–963.

[89] Tweeddale, H., Notley-McRobb, L., and Ferenci, T., "Effect of slow growth on metabolism of *Escherichia coli*, as revealed by global metabolite pool ("metabolome") analysis," *Journal of Bacteriology*, Vol. 180, No. 19, 1998, pp. 5109–5116.

[90] Kayser, A., Weber, J., Hecht, V., and Rinas, U., "Metabolic flux analysis of *Escherichia coli* in glucose-limited continuous culture. I. Growth-rate-dependent metabolic efficiency at steady state," *Microbiology*, Vol. 151, No. 3, 2005, pp. 693–706.

[91] Mittal, A., Katahira, R., Himmel, M. E., and Johnson, D. K., "Effects of alkaline or liquid-ammonia treatment on crystalline cellulose: changes in crystalline structure and effects on enzymatic digestibility," *Biotechnology for Biofuels*, Vol. 4, No. 41, 2011.

[92] Juhasz, T., Szengyel, Z., Reczey, K., Siika-Aho, M., and Viikari, L., "Characterization of cellulases and hemicellulases produced by *Trichoderma reesei* on various carbon sources," *Process Biochemistry*, Vol. 40, No. 11, 2005, pp. 3519–3525.

[93] Elbing, K. and Brent, R., "Media preparation and bacteriological tools", *Current Protocols in Molecular Biology*, Vol. 59, 2001.

[94] Kovarova-Kovar, K. and Egli, T., "Growth kinetics of suspended microbial cells: from single-substrate-controlled growth to mixed-substrate kinetics," *Microbiology and Molecular Biology Reviews*, Vol. 62, No. 3, 1998, pp. 646–666.

[95] Mashego, M. R., van Gulik, W. M., Vinke, J. L., and Heijnen, J. J., "Critical evaluation of sampling techniques for residual glucose determination in carbon-limited chemostat culture of *Saccharomyces cerevisiae*," *Biotechnology and Bioengineering*, Vol. 83, No. 4, 2003, pp. 395–399.

[96] Simonian, M. H. and Smith, J. A., "Spectrophotometric and colorimetric determination of protein Cconcentration", *Current Protocols in Molecular Biology*, Vol. 59, 2001.

[97] Kaur, J., Chadha, B. S., Kumar, B. A., Ghatora, S. K., and Saini, H. S., "Purification and characterization of beta-glucosidase from *Melanocarpus sp*. MTCC 3922," *Electronic Journal of Biotechnology*, Vol. 10, No. 2, 2007.

[98] Xiao, Z., Storms, R., and Tsang, A., "Microplate-based carboxymethylcellulose assay for endoglucanase activity," *Analytical Biochemistry*, Vol. 342, No. 1, 2005, pp. 176–178.

[99] Dashtban, M., Maki, M., Leung, K. T., Mao, C., and Qin, W., "Cellulase activities in biomass conversion: measurement methods and comparison," *Critical Reviews in Biotechnology*, Vol. 30, No. 4, 2010, pp. 302–309.

[100] Foster, C. E., Martin, T. M., and Pauly, M., "Comprehensive compositional analysis of plant cell walls (lignocellulosic biomass) part II: carbohydrates," *Journal of Visualized Experiments*, , No. 37, 2010, pp. e1837.

[101] Wintermute, E. H. and Silver, P. A., "Dynamics in the mixed microbial concourse," *Genes & Development*, Vol. 24, No. 23, 2010, pp. 2603–2614.

[102] Weber, W., Daoud-El Baba, M., and Fussenegger, M., "Synthetic ecosystems based on airborne inter- and intrakingdom communication," *Proceedings of the National Academy of Sciences*, Vol. 104, No. 25, 2007, pp. 10435–10440.

[103] Balagaddé, F. K., You, L., Hansen, C. L., Arnold, F. H., and Quake, S. R., "Long-term monitoring of bacteria undergoing programmed population control in a microchemostat," *Science*, Vol. 309, No. 5731, 2005, pp. 137–140.

[104] Bernstein, H. C., Paulson, S. D., and Carlson, R. P., "Synthetic *Escherichia coli* consortia engineered for syntrophy demonstrate enhanced biomass productivity," *Journal of Biotechnology*, Vol. 157, No. 1, 2012, pp. 159–166.

[105] Bernstein, H. and Carlson, R., "Microbial consortia engineering for cellular factories: *in vitro* to *in silico* systems," *Computational and Structural Biotechnology Journal*, Vol. 3, No. 0, 2012.

[106] Kerner, A., Park, J., Williams, A., and Lin, X. N., "A programmable *Escherichia coli* consortium via tunable symbiosis," *PLoS One*, Vol. 7, No. 3, 2012.

[107] Gore, J., Youk, H., and Van Oudenaarden, A., "Snowdrift game dynamics and facultative cheating in yeast," *Nature*, Vol. 459, No. 7244, 2009, pp. 253–256.

[108] Tengerdy, R. P., Rho, W. H., and Mohagheghi, A. M., "Liquid fluidized-bed starter culture of *Trichoderma reesei* for cellulase production," *Applied Biochemistry and Biotechnology*, Vol. 27, No. 3, 1991, pp. 195–204.

[109] Lejeune, R., Nielsen, J., and Baron, G. V., "Influence of pH on the morphology of *Trichoderma reesei* QM9414 in submerged culture," *Biotechnology Letters*, Vol. 17, No. 3, 1995, pp. 341–344.

[110] Presser, K. A., Ratkowsky, D. A., and Ross, T., "Modelling the growth rate of *Escherichia coli* as a function of pH and lactic acid concentration," *Applied and Environmental Microbiology*, Vol. 63, No. 6, 1997, pp. 2355–2360.

[111] Lenski, R. E., "Phenotypic and genomic evolution during a 20,000-generation experiment with the bacterium *Escherichia coli*", *Plant Breeding Reviews: Long-term Selection: Crops, Animals, and Bacteria, Volume 24, Part 2*. John Wiley and Sons, Hoboken, NJ, 2003, pp. 225–265.

[112] Shendure, J. and Ji, H., "Next-generation DNA sequencing," *Nature Biotechnology*, Vol. 26, No. 10, 2008, pp. 1135–1145.

[113] Herring, C. D., Raghunathan, A., Honisch, C., Patel, T., Applebee, M. K., Joyce, A. R., Albert, T. J., Blattner, F. R., van den Boom, D., Cantor, C. R., and Palsson, B. Ø., "Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale," *Nature Genetics*, Vol. 38, No. 12, 2006, pp. 1406–1412.

[114] Applebee, M. K., Herrgård, M. J., and Palsson, B. Ø., "Impact of individual mutations on increased fitness in adaptively evolved strains of *Escherichia coli*," *Journal of Bacteriology*, Vol. 190, No. 14, 2008, pp. 5087–5094.

[115] Conrad, T. M., Joyce, A. R., Applebee, M. K., Barrett, C. L., Xie, B., Gao, Y., and Palsson, B. Ø., "Whole-genome resequencing of *Escherichia coli* K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations," *Genome Biology*, Vol. 10, No. R118, 2009.

[116] Barrick, J. E., Yu, D. S., Yoon, S. H., Jeong, H., Oh, T. K., Schneider, D., Lenski, R. E., and Kim, J. F., "Genome evolution and adaptation in a long-term experiment with *Escherichia coli*," *Nature*, Vol. 461, No. 7268, 2009, pp. 1243–1247.

[117] Khan, A. I., Dinh, D. M., Schneider, D., Lenski, R. E., and Cooper, T. F., "Negative epistasis between beneficial mutations in an evolving bacterial population," *Science*, Vol. 332, No. 6034, 2011, pp. 1193–1196.

[118] Charusanti, P., Conrad, T. M., Knight, E. M., Venkataraman, K., Fong, N. L., Xie, B., Gao, Y., and Palsson, B. Ø., "Genetic basis of growth adaptation of *Escherichia coli* after deletion of pgi, a major metabolic gene," *PLoS genetics*, Vol. 6, No. 11, 2010.

[119] Kishimoto, T., Iijima, L., Tatsumi, M., Ono, N., Oyake, A., Hashimoto, T., Matsuo, M., Okubo, M., Suzuki, S., and Mori, K., "Transition from positive to neutral in mutation fixation along with continuing rising fitness in thermal adaptive evolution," *PLoS genetics*, Vol. 6, No. 10, 2010.

[120] Lee, D. H. and Palsson, B. Ø., "Adaptive evolution of *Escherichia coli* K-12 MG1655 during growth on a nonnative carbon source, L-1,2-propanediol," *Applied and Environmental Microbiology*, Vol. 76, No. 13, 2010, pp. 4158–4168.

[121] Goodarzi, H., Bennett, B. D., Amini, S., Reaves, M. L., Hottes, A. K., Rabinowitz, J. D., and Tavazoie, S., "Regulatory and metabolic rewiring during laboratory evolution of ethanol tolerance in *E. coli*," *Molecular Systems Biology*, Vol. 6, No. 378, 2010.

[122] Philippe, N., Crozat, E., Lenski, R. E., and Schneider, D., "Evolution of global regulatory networks during a long-term experiment with *Escherichia coli*," *Bioessays*, Vol. 29, No. 9, 2007, pp. 846–860.

[123] Hindré, T., Knibbe, C., Beslon, G., and Schneider, D., "New insights into bacterial adaptation through *in vivo* and *in silico* experimental evolution," *Nature Reviews Microbiology*, Vol. 10, No. 5, 2012, pp. 352–365.

[124] Atsumi, S., Wu, T. Y., Machado, I. M., Huang, W. C., Chen, P. Y., Pellegrini, M., and Liao, J. C., "Evolution, genomic analysis, and reconstruction of isobutanol tolerance in *Escherichia coli*," *Molecular Systems Biology*, Vol. 6, No. 449, 2010.

[125] Reyes, L. H., Almario, M. P., Winkler, J., Orozco, M. M., and Kao, K. C., "Visualizing evolution in real time to determine the molecular mechanisms of n-butanol tolerance in *Escherichia coli*," *Metabolic Engineering*, Vol. 14, No. 5, 2012, pp. 579–590.

[126] Dragosits, M., Mozhayskiy, V., Quinones-Soto, S., Park, J., and Tagkopoulos, I., "Evolutionary potential, cross-stress behavior and the genetic basis of acquired stress resistance in *Escherichia coli*," *Molecular Systems Biology*, Vol. 9, No. 643, 2013.

[127] Kim, Y., Ingram, L. O., and Shanmugam, K. T., "Construction of an *Escherichia coli* K-12 mutant for homoethanologenic fermentation of glucose or xylose without foreign genes," *Applied and Environmental Microbiology*, Vol. 73, No. 6, 2007, pp. 1766–1771.

[128] Griepernau, B., Leis, S., Schneider, M. F., Sikor, M., Steppich, D., and Bockmann, R. A., "1-Alkanols and membranes: a story of attraction," *Biochimica et Biophysica Acta*, Vol. 1768, No. 11, 2007, pp. 2899–913.

[129] Novocraft Technologies., "Novoalign," version 2.04.02, 2009.

[130] Li, H., Ruan, J., and Durbin, R., "Mapping short DNA sequencing reads and calling variants using mapping quality scores," *Genome Research*, Vol. 18, No. 11, 2008, pp. 1851–1858.

[131] Zerbino, D. R. and Birney, E., "Velvet: algorithms for de novo short read assembly using de Bruijn graphs," *Genome Research*, Vol. 18, No. 5, 2008, pp. 821–829.

[132] Maere, S., Heymans, K., and Kuiper, M., "BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks," *Bioinformatics*, Vol. 21, No. 16, 2005, pp. 3448–3449.

[133] Krishnamoorthy, G., Tikhonova, E. B., and Zgurskaya, H. I., "Fitting periplasmic membrane fusion proteins to inner membrane transporters: mutations that enable *Escherichia coli* AcrA to function with *Pseudomonas aeruginosa* MexB," *Journal Bacteriology*, Vol. 190, No. 2, 2008, pp. 691–698.

[134] Link, T. M., Valentin-Hansen, P., and Brennan, R. G., "Structure of *Escherichia coli* Hfq bound to polyriboadenylate RNA," *Proceedings of the National Academy of Sciences*, Vol. 106, No. 46, 2009, pp. 19292–19297.

[135] Boggy, G. J. and Woolf, P. J., "A mechanistic model of PCR for accurate quantification of quantitative PCR data," *PLoS One*, Vol. 5, No. 8, 2010, pp. e12355.

[136] Savli, H., Karadenizli, A., Kolayli, F., Gundes, S., Ozbek, U., and Vahaboglu, H., "Expression stability of six housekeeping genes: A proposal for resistance gene quantification studies of *Pseudomonas aeruginosa* by real-time quantitative RT-PCR," *Journal of Medical Microbiology*, Vol. 52, No. 5, 2003, pp. 403–408.

[137] Brynildsen, M. P. and Liao, J. C., "An integrated network approach identifies the isobutanol response network of *Escherichia coli*," *Molecular Systems Biology*, Vol. 5, No. 277, 2009.

[138] Guisbert, E., Rhodius, V. A., Ahuja, N., Witkin, E., and Gross, C. A., "Hfq modulates the sigmaE-mediated envelope stress response and the sigma32-mediated cytoplasmic stress response in , *Escherichia coli*," *Journal of Bacteriology*, Vol. 189, No. 5, 2007, pp. 1963–1973.

[139] Dong, T. and Schellhorn, H. E., "Control of RpoS in global gene expression of *Escherichia coli* in minimal media," *Molecular Genetics and Genomics*, Vol. 281, No. 1, 2009, pp. 19–33.

[140] Miyashiro, T. and Goulian, M., "Stimulus-dependent differential regulation in the *Escherichia coli* PhoQ PhoP system," *Proceedings of the National Academy of Sciences*, Vol. 104, No. 41, 2007, pp. 16305–16310.

[141] Moon, K. and Gottesman, S., "A PhoQ/P-regulated small RNA regulates sensitivity of *Escherichia coli* to antimicrobial peptides," *Molecular Microbiology*, Vol. 74, No. 6, 2009, pp. 1314–1330.

[142] Viveiros, M., Martins, A., Paixao, L., Rodrigues, L., Martins, M., Couto, I., Fahnrich, E., Kern, W. V., and Amaral, L., "Demonstration of intrinsic efflux activity of *Escherichia coli* K-12 AG100 by an automated ethidium bromide method," *International Journal of Antimicrobial Agents*, Vol. 31, No. 5, 2008, pp. 458–462.

[143] Yang, S., Lopez, C. R., and Zechiedrich, E. L., "Quorum sensing and multidrug transporters in *Escherichia coli*," *Proceedings of the National Academy of Sciences*, Vol. 103, No. 7, 2006, pp. 2386–2391.

[144] Notley-McRobb, L., King, T., and Ferenci, T., "*rpoS* mutations and loss of general stress resistance in *Escherichia coli* populations as a consequence of conflict between competing stress responses," *Journal of Bacteriology*, Vol. 184, No. 3, 2002, pp. 806–811.

[145] Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H., "Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection," *Molecular Systems Biology*, Vol. 2, No. 8, 2006.

[146] Martin, G., Elena, S. F., and Lenormand, T., "Distributions of epistasis in microbes fit predictions from a fitness landscape model," *Nature Genetics*, Vol. 39, No. 4, 2007, pp. 555–560.

[147] Whitlock, M. C., Phillips, P. C., Moore, F. B., and Tonsor, S. J., "Multiple fitness peaks and epistasis," *Annual Review of Ecology and Systematics*, Vol. 26, No. 1, 2003, pp. 601–629.

[148] Driessen, A. J. and Nouwen, N., "Protein translocation across the bacterial cytoplasmic membrane," *Annual Review of Biochemistry*, Vol. 77, 2008, pp. 643–667.

[149] Gelis, I., Bonvin, A. M., Keramisanou, D., Koukaki, M., Gouridis, G., Karamanou, S., Economou, A., and Kalodimos, C. G., "Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR," *Cell*, Vol. 131, No. 4, 2007, pp. 756–769.

[150] Keseler, I. M., Bonavides-Martinez, C., Collado-Vides, J., Gama-Castro, S., Gunsalus, R. P., Johnson, D. A., Krummenacker, M., Nolan, L. M., Paley, S., Paulsen, I. T., Peralta-Gil, M., Santos-Zavaleta, A., Shearer, A. G., and Karp, P. D., "EcoCyc: a comprehensive view of *Escherichia coli* biology," *Nucleic Acids Research*, Vol. 37, Database issue, 2009, pp. D464–D470.

[151] Charollais, J., Dreyfus, M., and Iost, I., "CsdA, a cold-shock RNA helicase from *Escherichia coli*, is involved in the biogenesis of 50S ribosomal subunit," *Nucleic Acids Research*, Vol. 32, No. 9, 2004, pp. 2751–2759.

[152] Borden, J. R., Jones, S. W., Indurthi, D., Chen, Y., and Papoutsakis, E. T., "A genomic-library based discovery of a novel, possibly synthetic, acid-tolerance mechanism in *Clostridium acetobutylicum* involving non-coding RNAs and ribosomal RNA processing," *Metabolic Engineering*, Vol. 12, No. 3, 2010, pp. 268–281.

[153] Ochi, K., "From microbial differentiation to ribosome engineering," *Bioscience, Biotechnology, and Biochemistry*, Vol. 71, No. 6, 2007, pp. 1373–1386.

[154] He, X., Qian, W., Wang, Z., Li, Y., and Zhang, J., "Prevalent positive epistasis in *Escherichia coli* and *Saccharomyces cerevisiae* metabolic networks," *Nature Genetics*, Vol. 42, No. 3, 2010, pp. 272–276.

[155] Yamada, J., Yamasaki, S., Hirakawa, H., Hayashi-Nishino, M., Yamaguchi, A., and Nishino, K., "Impact of the RNA chaperone Hfq on multidrug resistance in *Escherichia coli*," *Journal of Antimicrobial Chemotherapy*, Vol. 65, No. 5, 2010, pp. 853–858.

[156] Salis, H. M., Mirsky, E. A., and Voigt, C. A., "Automated design of synthetic ribosome binding sites to control protein expression," *Nature Biotechnology*, Vol. 27, No. 10, 2009, pp. 946–950.

[157] Crombach, A. and Hogeweg, P., "Evolution of evolvability in gene regulatory networks," *PLoS Computional Biology*, Vol. 4, No. 7, 2008, pp. e1000112.

[158] Jovelin, R. and Phillips, P. C., "Evolutionary rates and centrality in the yeast gene regulatory network," *Genome Biology*, Vol. 10, No. 4, 2009, pp. R35.

[159] Bleichert, F. and Baserga, S. J., "The long unwinding road of RNA helicases," *Molecular Cell*, Vol. 27, No. 3, 2007, pp. 339–352.

[160] Waters, L. S. and Storz, G., "Regulatory RNAs in bacteria," *Cell*, Vol. 136, No. 4, 2009, pp. 615–628.

[161] Maharjan, R., Zhou, Z., Ren, Y., Li, Y., Gaffe, J., Schneider, D., McKenzie, C., Reeves, P. R., Ferenci, T., and Wang, L., "Genomic identification of a novel mutation in *hfq* that provides multiple benefits in evolving glucose-limited populations of *Escherichia coli*," *Journal of Bacteriology*, 2010.

[162] Wang, L., Spira, B., Zhou, Z., Feng, L., Maharjan, R. P., Li, X., Li, F., McKenzie, C., Reeves, P. R., and Ferenci, T., "Divergence involving global regulatory gene mutations in an *E. coli* population evolving under phosphate limitation," *Genome Biology and Evolution*, Vol. 2, 2010, pp. 478-487.

[163] Fraser, H. B., Hirsh, A. E., Steinmetz, L. M., Scharfe, C., and Feldman, M. W., "Evolutionary rate in the protein interaction network," *Science*, Vol. 296, No. 5568, 2002, pp. 750–752.

[164] Lu, C., Zhang, Z., Leach, L., Kearsey, M. J., and Luo, Z. W., "Impacts of yeast metabolic network structure on enzyme evolution," *Genome Biology*, Vol. 8, No. 407, 2007.

[165] Larossa, R. A. and Smulski, D. R., "PRODUCTION OF FOUR CARBON ALCOHOLS USING IMPROVED STRAIN," US Patent Application Pub. No. US 2009/0203139 A1, August 2009.

[166] Horinouchi, T., Tamaoka, K., Furusawa, C., Ono, N., Suzuki, S., Hirasawa, T., Yomo, T., and Shimizu, H., "Transcriptome analysis of parallel-evolved *Escherichia coli* strains under ethanol stress," *BMC Genomics*, Vol. 11, No. 1, 2010.

[167] Woodruff, L., Pandhal, J., Ow, S. Y., Karimpour-Fard, A., Weiss, S. J., Wright, P. C., and Gill, R. T., "Genome-scale identification and characterization of ethanol tolerance genes in *Escherichia coli*," *Metabolic engineering*, Vol. 15, 2012, pp. 124–133.

[168] Potrykus, K. and Cashel, M., "(p)ppGpp: still magical?" *Annual Review of Microbiology*, Vol. 62, 2008, pp. 35–51.

[169] Guillier, M., Gottesman, S., and Storz, G., "Modulating the outer membrane with small RNAs," *Genes and Development*, Vol. 20, No. 17, 2006, pp. 2338–2348.

[170] Tsui, H. C., Leung, H. C., and Winkler, M. E., "Characterization of broadly pleiotropic phenotypes caused by an *hfq* insertion mutation in *Escherichia coli* K-12," *Molecular Microbiology*, Vol. 13, No. 1, 1994, pp. 35–49.

[171] Hobbs, E. C., Astarita, J. L., and Storz, G., "Small RNAs and small proteins involved in resistance to cell envelope stress and acid shock in *Escherichia coli*: analysis of a bar-coded mutant collection," *Journal of Bacteriology*, Vol. 192, No. 1, 2010, pp. 59–67.

[172] Baars, L., Ytterberg, A. J., Drew, D., Wagner, S., Thilo, C., van Wijk, K. J., and de Gier, J. W., "Defining the role of the *Escherichia coli* chaperone SecB using comparative proteomics," *Journal of Biological Chemistry*, Vol. 281, No. 15, 2006, pp. 10024–10034.

[173] Hasona, A., Zuobi-Hasona, K., Crowley, P. J., Abranches, J., Ruelf, M. A., Bleiweis, A. S., and Brady, L. J., "Membrane composition changes and physiological adaptation by *Streptococcus mutans* signal recognition particle pathway mutants," *Journal of Bacteriology*, Vol. 189, No. 4, 2007, pp. 1219–1230.

[174] Gonzalez, R., Tao, H., Purvis, J. E., York, S. W., Shanmugam, K. T., and Ingram, L. O., "Gene array-based identification of changes that contribute to ethanol tolerance in ethanologenic *Escherichia coli*: comparison of KO11 (parent) to LY01 (resistant mutant)," *Biotechnology Progress*, Vol. 19, No. 2, 2003, pp. 612–623.

[175] Koo, J. T., Choe, J., and Moseley, S. L., "HrpA, a DEAH-box RNA helicase, is involved in mRNA processing of a fimbrial operon in *Escherichia coli*," *Molecular Microbiology*, Vol. 52, No. 6, 2004, pp. 1813–1826.

[176] Kazuta, Y., Adachi, J., Matsuura, T., Ono, N., Mori, H., and Yomo, T., "Comprehensive analysis of the effects of *Escherichia coli* ORFs on protein translation reaction," *Molecular and Cellular Proteomics*, Vol. 7, No. 8, 2008, pp. 1530–1540.

[177] Seregina, T., Osipov, G., Shakulov, R., and Mironov, A., "Isolation and phenotypic characteristics of the *Escherichia coli* butanol-tolerant mutants," *Microbiology*, Vol. 81, No. 2, 2012, pp. 208–215.

[178] Aono, R. and Kobayashi, H., "Cell surface properties of organic solvent-tolerant mutants of *Escherichia coli* K-12," *Applied and Environmental Microbiology*, Vol. 63, No. 9, 1997, pp. 3637–3642.

[179] Yoshimura, M., Oshima, T., and Ogasawara, N., "Involvement of the YneS/YgiH and PlsX proteins in phospholipid biosynthesis in both *Bacillus subtilis* and *Escherichia coli*," *BMC Microbiology*, Vol. 7, No. 69, 2007.

[180] Cao, Y., Yang, J., Xian, M., Xu, X., and Liu, W., "Increasing unsaturated fatty acid contents in *Escherichia coli* by coexpression of three different genes," *Applied Microbiology and Biotechnology*, Vol. 87, No. 1, 2010, pp. 271–280.

[181] Jeong, H., Kim, S. H., Han, S. S., Kim, M. H., and Lee, K. C., "Changes in membrane fatty acid composition through proton-induced *fabF* mutation enhancing 1-butanol tolerance in *E. coli*," *Journal of the Korean Physical Society*, Vol. 61, No. 2, 2012, pp. 227–233.

[182] Vanegas, J. M., Contreras, M. F., Faller, R., and Longo, M. L., "Role of unsaturated lipid and ergosterol in ethanol tolerance of model yeast biomembranes," *Biophysical journal*, Vol. 102, No. 3, 2012, pp. 507–516.

[183] Ezeji, T., Milne, C., Price, N. D., and Blaschek, H. P., "Achievements and perspectives to overcome the poor solvent resistance in acetone and butanol-producing microorganisms," *Applied Microbiology and Biotechnology*, Vol. 85, No. 6, 2010, pp. 1697–1712.

[184] Aoki, K. F. and Kanehisa, M., "Using the KEGG database resource," *Current Protocols in Bioinformatics*, Chapter 1, 2005, Unit 1.12.

[185] Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T. L., "NCBI BLAST: a better web interface," *Nucleic Acids Research*, Vol. 36, Web issue, 2008, pp. W5–W9.

[186] Chen, K., Wallis, J. W., McLellan, M. D., Larson, D. E., Kalicki, J. M., Pohl, C. S., McGrath, S. D., Wendl, M. C., Zhang, Q., Locke, D. P., Shi, X., Fulton, R. S., Ley, T. J., Wilson, R. K., Ding, L., and Mardis, E. R., "BreakDancer: an algorithm for high-resolution mapping of genomic structural variation," *Nature Methods*, Vol. 6, No. 9, 2009, pp. 677–681.

[187] Koressaar, T. and Remm, M., "Enhancements and modifications of primer design program Primer3," *Bioinformatics*, Vol. 23, No. 10, 2007, pp. 1289–1291.

[188] You, F. M., Huo, N., Gu, Y. Q., Luo, M. C., Ma, Y., Hane, D., Lazo, G. R., Dvorak, J., and Anderson, O. D., "BatchPrimer3: a high throughput web application for PCR and sequencing primer design," *BMC Bioinformatics*, Vol. 9, No. 253, 2008.

[189] Zuker, M., "Mfold web server for nucleic acid folding and hybridization prediction," *Nucleic Acids Research*, Vol. 31, No. 13, 2003, pp. 3406–3415.

[190] Miller, J. H., *A short course in bacterial genetics : a laboratory manual and handbook for Escherichia coli and related bacteria*, Cold Spring Harbor Laboratory Press, Plainview, N.Y., 1992.

[191] Rouillard, J., Zuker, M., and Gulari, E., "OligoArray 2.0: design of oligonucleotide probes for DNA microarrays using a thermodynamic approach," *Nucleic Acids Research*, Vol. 31, No. 12, 2003, pp. 3057–3062.

[192] R Development Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, Austria, 2010.

[193] Huber, W., von Heydebreck, A., Sultmann, H., Poustka, A., and Vingron, M., "Variance stabilization applied to microarray data calibration and to the quantification of differential expression," *Bioinformatics*, Vol. 18, Suppl. 1, 2002, pp. S96–S104.

[194] Bengtsson, H., Jonsson, G., and Vallon-Christersson, J., "Calibration and assessment of channel-specific biases in microarray data with extended dynamical range," *BMC Bioinformatics*, Vol. 5, No. 177, 2004.

[195] Bengtsson, H. and Hössjer, O., "Methodological study of affine transformations of gene expression data with proposed robust non-parametric multi-dimensional normalization method," *BMC Bioinformatics*, Vol. 7, No. 100, 2006.

[196] Bengtsson, H., Irizarry, R., Carvalho, B., and Speed, T., "Estimation and assessment of raw copy numbers at the single locus level," *Bioinformatics*, Vol. 24, No. 6, 2008, pp. 759–767.

[197] Tukey, J. W., *Exploratory data analysis*, Addison-Wesley series in behavioral science, Addison-Wesley Pub. Co, Reading, 1977.

[198] Smyth, G. K., *Limma: linear models for microarray data*, Springer, New York, 2005, pp. 397–420.

[199] Smyth, G. K., "Linear models and empirical bayes methods for assessing differential expression in microarray experiments," *Statistical Applications in Genetics and Molecular Biology*, Vol. 3, No. 3, 2004.

[200] Loennstedt, L. and Speed, T. P., "Replicated microarray data," *Statistica Sinica*, Vol. 12, 2002, pp. 31–36.

[201] Benjamini, Y. and Hochberg, Y., "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 57, No. 1, 1995, pp. pp. 289–300.

[202] Liao, J. C., Boscolo, R., Yang, Y. L., Tran, L. M., Sabatti, C., and Roychowdhury, V. P., "Network component analysis: reconstruction of regulatory signals in biological systems," *Proceedings of the National Academy of Sciences*, Vol. 100, No. 26, 2003, pp. 15522–15527.

[203] Galbraith, S. J., Tran, L. M., and Liao, J. C., "Transcriptome network component analysis with limited microarray data," *Bioinformatics*, Vol. 22, No. 15, 2006, pp. 1886–1894.

[204] Gama-Castro, S., Jimenez-Jacinto, V., Peralta-Gil, M., Santos-Zavaleta, A., Penaloza-Spinola, M. I., Contreras-Moreira, B., Segura-Salazar, J., Muniz-Rascado, L., Martinez-Flores, I., Salgado, H., Bonavides-Martinez, C., Abreu-Goodger, C., Rodriguez-Penagos, C., Miranda-Rios, J., Morett, E., Merino, E., Huerta, A. M., Trevino-Quintanilla, L., and Collado-Vides, J., "RegulonDB (version 6.0): gene regulation model of *Escherichia coli* K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation," *Nucleic Acids Research*, Vol. 36, No. Database issue, 2008, pp. D120–D124.

[205] Oh, T. J., Kim, I. G., Park, S. Y., Kim, K. C., and Shim, H. W., "NAD-dependent malate dehydrogenase protects against oxidative damage in *Escherichia coli* K-12 through the action of oxaloacetate," *Environmental Toxicology and Pharmacology*, Vol. 11, No. 1, 2002, pp. 9–14.

[206] Sandoval, N. R., Kim, J. Y. H., Glebes, T. Y., Reeder, P. J., Aucoin, H. R., Warner, J. R., and Gill, R. T., "Strategy for directing combinatorial genome engineering in Escherichia coli," *Proceedings of the National Academy of Sciences*, Vol. 109, No. 26, 2012, pp. 10540–10545.

[207] Chou, H.-H., Chiu, H.-C., Delaney, N. F., Segrè, D., and Marx, C. J., "Diminishing returns epistasis among beneficial mutations decelerates adaptation," *Science*, Vol. 332, No. 6034, 2011, pp. 1190–1192.

[208] Harper, M. A., Chen, Z., Toy, T., Machado, I. M. P., Nelson, S. F., Liao, J. C., and Lee, C. J., "Phenotype sequencing: identifying the genes that cause a phenotype directly from pooled sequencing of independent mutants," *PLoS ONE*, Vol. 6, No. 2, 2011.

[209] Reyes, L. H., Almario, M. P., and Kao, K. C., "Genomic library screens for genes involved in n-Butanol tolerance in *Escherichia coli*," *PLoS ONE*, Vol. 6, No. 3, 2011.

[210] Juan, H.-F., Mori, H., Chang, H.-Y., Huang, H.-C., and Liao, J. C., "ALCOHOL TOLERANT ESCHERICHIA COLI AND METHODS OF PREPARATION THEREOF," US Patent Application Pub. No. US 2010/0105103 A1, 4 2010.

[211] Larossa, R. A. and Smulski, D. R., "STRAIN FOR BUTANOL PRODUCTION," US Patent Application Pub. No. US 2009/0162911 A1, June 2009.

[212] Soskine, M. and Tawfik, D. S., "Mutational effects and the evolution of new protein functions," *Nature Reviews Genetics*, Vol. 11, No. 8, 2010, pp. 572–582.

[213] Huo, Y.-X., Cho, K. M., Rivera, J. G. L., Monte, E., Shen, C. R., Yan, Y., and Liao, J. C., "Conversion of proteins into biofuels by engineering nitrogen flux," *Nature Biotechnology*, Vol. 29, No. 4, 2011, pp. 346–351.

[214] Freddolino, P. L., Goodarzi, H., and Tavazoie, S., "Fitness landscape transformation through a single amino acid change in the Rho terminator," *PLoS Genetics*, Vol. 8, No. 5, 2012.

[215] Nyström, T., "Stationary-phase physiology," *Annual Review of Microbiology*, Vol. 58, 2004, pp. 161–181.

[216] Argaman, L., Elgrably-Weiss, M., Hershko, T., Vogel, J., and Altuvia, S., "RelA protein stimulates the activity of RyhB small RNA by acting on RNA-binding protein Hfq," *Proceedings of the National Academy of Sciences*, Vol. 109, No. 12, 2012, pp. 4621–4626.

[217] Winkler, J. and Kao, K. C., "Harnessing recombination to speed adaptive evolution in *Escherichia coli*," *Metabolic Engineering*, Vol. 14, No. 5, 2012, pp. 487–495.

[218] Audesirk, T. and Audesirk, G., *Biology, Life on Earth*, Prentice-Hall, 5th ed., 1999.

[219] Yuan, J., Doucette, C. D., Fowler, W. U., Feng, X.-J., Piazza, M., Rabitz, H. A., Wingreen, N. S., and Rabinowitz, J. D., "Metabolomics-driven quantitative analysis of ammonia assimilation in *E. coli*," *Molecular Systems Biology*, Vol. 5, 2009.

[220] Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S., "Stochastic gene expression in a single cell," *Science*, Vol. 297, No. 5584, 2002, pp. 1183–1186.

[221] Harcombe, W., "Novel cooperation experimentally evolved between species," *Evolution*, Vol. 64, No. 7, 2010, pp. 2166–2172.

[222] Drake, J. W., "The distribution of rates of spontaneous mutation over viruses, prokaryotes, and eukaryotes," *Annals of the New York Academy of Sciences*, Vol. 870, No. 1, 1999, pp. 100–107.

[223] Holker, U., Hofer, M., and Lenz, J., "Biotechnological advantages of laboratory-scale solid-state fermentation with fungi," *Applied Microbiology Biotechnology*, Vol. 64, No. 2, 2004, pp. 175–186.

[224] Aprikian, P., Tchesnokova, V., Kidd, B., Yakovenko, O., Yarov-Yarovoy, V., Trinchina, E., Vogel, V., Thomas, W., and Sokurenko, E., "Interdomain interaction in the FimH adhesin of *Escherichia coli* regulates the affinity to mannose," *Journal of Biological Chemistry*, Vol. 282, No. 32, 2007, pp. 23437–23446.

[225] Perlinska-Lenart, U., Orlowski, J., Laudy, A. E., Zdebska, E., Palamarczyk, G., and Kruszewska, J. S., "Glycoprotein hypersecretion alters the cell wall in *Trichoderma reesei* strains expressing the *Saccharomyces cerevisiae* dolichylphosphate mannose synthase gene," *Applied and Environmental Microbiology*, Vol. 72, No. 12, 2006, pp. 7778–7784.

[226] McCusker, M. P., Turner, E. C., and Dorman, C. J., "DNA sequence heterogeneity in Fim tyrosine-integrase recombinase-binding elements and functional motif asymmetries determine the directionality of the *fim* genetic switch in *Escherichia coli* K-12," *Molecular Microbiology*, Vol. 67, No. 1, 2008, pp. 171–187.

[227] Collins, C. M. and D'Orazio, S. E., "Bacterial ureases: structure, regulation of expression and role in pathogenesis," *Molecular Microbiology*, Vol. 9, No. 5, 1993, pp. 907–913.

[228] Nakamura, C. E. and Whited, G. M., "Metabolic engineering for the microbial production of 1,3-propanediol," *Current Opinion in Biotechnology*, Vol. 14, No. 5, 2003, pp. 454–459.

[229] Holladay, J. E., White, J. F., Bozell, J. J., and Johnson, D., "Top value added chemicals from biomass: volume I – results of screening for potential candidates from sugars and synthesis gas," Tech. rep., U.S. Department of Energy, National Renewable Energy Lab, 2004.

[230] Park, J., Kerner, A., Burns, M. A., and Lin, X. N., "Microdroplet-enabled highly parallel co-cultivation of microbial communities," *PloS One*, Vol. 6, No. 2, 2011.

[231] Van der Borght, K., Verheyen, A., Feyaerts, M., Van Wesenbeeck, L., Verlinden, Y., Van Craenenbroeck, E., and van Vlijmen, H., "Quantitative prediction of integrase inhibitor resistance from genotype through consensus linear regression modeling," *Virology Journal*, Vol. 10, No. 1, 2013.

[232] Mamanova, L., Coffey, A. J., Scott, C. E., Kozarewa, I., Turner, E. H., Kumar, A., Howard, E., Shendure, J., and Turner, D. J., "Target-enrichment strategies for next-generation sequencing," *Nature Methods*, Vol. 7, No. 2, 2010, pp. 111–118.

[233] Dunlop, M. J., Keasling, J. D., and Mukhopadhyay, A., "A model for improving microbial biofuel production using a synthetic feedback loop," *Systems and Synthetic Biology*, Vol. 4, No. 2, 2010, pp. 95–104.

[234] Zingaro, K. A. and Papoutsakis, E. T., "Toward a semisynthetic stress response system to engineer microbial solvent tolerance," *Mbio*, Vol. 3, No. 5, 2012.

[235] Moon, M. H., Ryu, J., Choeng, Y.-H., Hong, S.-K., Kang, H. A., and Chang, Y. K., "Enhancement of stress tolerance and ethanol production in *Saccharomyces cerevisiae* by heterologous expression of a trehalose biosynthetic gene from *Streptomyces albus*," *Biotechnology and Bioprocess Engineering*, Vol. 17, No. 5, 2012, pp. 986–996.

[236] Yazawa, H., Kamisaka, Y., Kimura, K., Yamaoka, M., and Uemura, H., "Efficient accumulation of oleic acid in *Saccharomyces cerevisiae* caused by expression of rat elongase 2 gene (rELO2) and its contribution to tolerance to alcohols," *Applied Microbiology and Biotechnology*, Vol. 91, No. 6, 2011, pp. 1593–1600.

[237] Dunlop, M. J., Dossani, Z. Y., Szmidt, H. L., Chu, H. C., Lee, T. S., Keasling, J. D., Hadi, M. Z., and Mukhopadhyay, A., "Engineering microbial biofuel tolerance and export using efflux pumps," *Molecular Systems Biology*, Vol. 7, No. 1, 2011.

[238] Lee, S. J., Oh, E. K., Oh, Y. H., Won, J. I., Han, S. O., and Lee, J. W., "Increased ethanol resistance in ethanolic *Escherichia coli* by insertion of heat-shock genes BEM1 and SOD2 from *Saccharomyces cerevisiae*," *Biotechnology and Bioprocess Engineering*, Vol. 15, No. 5, 2010, pp. 770–776.

[239] Woodruff, L. B. A., May, B. L., Warner, J. R., and Gill, R. T., "Towards a metabolic engineering strain "commons": an *Escherichia coli* platform strain for ethanol production," *Biotechnology and Bioengineering*, Vol. 110, No. 5, 2013, pp. 1520–1526.

[240] Woodruff, L. B. A., Boyle, N. R., and Gill, R. T., "Engineering improved ethanol production in *Escherichia coli* with a genome-wide approach," *Metabolic Engineering*, Vol. 17, No. 0, 2013, pp. 1–11.

[241] Lendenmann, U. and Egli, T., "Kinetic models for the growth of *Escherichia coli* with mixtures of sugars under carbon-limited conditions," *Biotechnology and Bioengineering*, Vol. 59, No. 1, 1998, pp. 99–107.

[242] South, C. R., Hogsett, D. A. L., and Lynd, L. R., "Modeling simultaneous saccharification and fermentation of lignocellulose to ethanol in batch and continuous reactors," *Enzyme and Microbial Technology*, Vol. 17, No. 9, 1995, pp. 797–803.

[243] Peri, S., Karra, S., Lee, Y. Y., and Karim, M. N., "Modeling intrinsic kinetics of enzymatic cellulose hydrolysis," *Biotechnology Progress*, Vol. 23, No. 3, 2007, pp. 626–37.

[244] Herpoel-Gimbert, I., Margeot, A., Dolla, A., Jan, G., Molle, D., Lignon, S., Mathis, H., Sigoillot, J. C., Monot, F., and Asther, M., "Comparative secretome analyses of two *Trichoderma reesei* RUT-C30 and CL847 hypersecretory strains," *Biotechnology for Biofuels*, Vol. 1, 2008.

[245] Zhang, Y. H. P. and Lynd, L. R., "Toward an aggregated understanding of enzymatic hydrolysis of cellulose: noncomplexed cellulase systems," *Biotechnology and Bioengineering*, Vol. 88, No. 7, 2004, pp. 797–824.

[246] Pakula, T. M., Salonen, K., Uusitalo, J., and Penttila, M., "The effect of specific growth rate on protein synthesis and secretion in the filamentous fungus *Trichoderma reesei*," *Microbiology*, Vol. 151, 2005, pp. 135–143.

[247] Kubicek, C. P., Messner, R., Gruber, F., Mandels, M., and Kubicekpranz, E. M., "Triggering of cellulase biosynthesis by cellulose in *Trichoderma reesei* - involvement of a constitutive, sophorose-inducible, glucose-inhibited beta-diglucoside permease," *Journal of Biological Chemistry*, Vol. 268, No. 26, 1993, pp. 19364–19368.

[248] Huang, C. P., Westman, D., Quirk, K., and Huang, J. P., "The removal of Cadmium(II) from dilute aqueous-solutions by fungal adsorbent," *Water Science and Technology*, Vol. 20, No. 11-12, 1988, pp. 369–376.

[249] Kurath, S. F. and Bump, D. D., "Hydrodynamic friction coefficients for cellodextrins in water," *Journal of Polymer Science Part A-General Papers*, Vol. 3, No. 4PA, 1965, pp. 1515.

[250] Esterbauer, H., Steiner, W., Labudova, I., Hermann, A., and Hayn, M., "Production of *Trichoderma* cellulase in laboratory and pilot scale," *Bioresource Technology*, Vol. 36, No. 1, 1991, pp. 51–65.

[251] Bader, J., Bellgardt, K. H., and Schügerl, K., "Modelling and simulation of the enzymatic hydrolysis of potato pulp by a complex enzyme mixture," *The Chemical Engineering Journal*, Vol. 52, No. 1, 1993, pp. B13–B19.