Genome Engineering for Improved Furfural and Product Tolerance in *Escherichia coli* for Renewable Biofuel Applications

By

Tirzah Ya'el Glebes

B.S., West Virginia University, 2007

M.S., University of Colorado Boulder, 2009

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Chemical Engineering 2014

This thesis entitled:

Genome Engineering for Improved Furfural and Product Tolerance in *Escherichia coli* for Renewable Biofuel Applications

written by Tirzah Ya'el Glebes

has been approved for the Department of Chemical Engineering

_____

Ryan T. Gill

_____

Theodore W. Randolph

Date _____

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Glebes, Tirzah Ya'el (Ph.D., Chemical Engineering)

Genome Engineering for Improved Furfural and Product Tolerance in *Escherichia coli* for Renewable Biofuel Applications

Thesis directed by Associate Professor Ryan T. Gill

Abstract

As engineers, we are interested in designing controlled, predictable, and maintainable strategies for performing or improving tasks. Genome engineering aims to use these same principles to design or re-design biological systems for targeted purposes. Strategies for genome engineering are no longer primarily limited by the cost of DNA synthesis or sequencing as they have been in the past. Instead, strategies are limited by not having methods to inform efficient and directed design. In these studies, we present an example of overcoming this current limitation by using various tools to identify genetic manipulations of interest, and then subsequently use these findings to motivate the directed design of cells for novel phenotypes. Initial studies are focused on elucidating genetic manipulations that confer furfural tolerance. Furfural is a key microbial inhibitor found in lignocellulosic hydrolysate, which is the proposed renewable sugar source for fermentation of sustainable biofuels that do not rely on food-based feedstocks. We transition towards engineering biofuel tolerance based upon identifications made in the furfural studies.

Using libraries of 10^3 to 10^5 unique members with defined and trackable mutations, we tested, in parallel, their effect on growth in the presence of furfural. We used two different search strategies (multiSCale Analysis of Library Enrichments and TRackable Multiplex Recombineering) to map genotype-to-phenotype relationships for furfural tolerance. Improved growth was confirmed for six novel furfural tolerance alleles: *lpcA* (lipopolysaccharide biosynthesis), *groESL* (chaperonin), *dicA* (regulator of cell division proteins), *rna* (ribonuclease), *ahpC* (alkylhydroperoxide reductase subunit), and *yhjH* (involved in flagellar motility regulation). The diversity of beneficial mutations found here highlights the breadth of changes that can be made to confer the same phenotype.

Building upon one of the most tolerant genes elucidated for furfural tolerance (*lpcA*), we informed the directed design of mutants with altered lipopolysaccharide biosynthesis to confer tolerance to hydrophobic biofuels, like *n*-butanol. Using a recursive recombineering approach to create libraries of increasingly mutated strains, we isolated clones capable of up to 50% growth improvements in *n*-butanol. We also initiated use of a new method for tracking multiple mutations across the genome, which has the potential to further reduce DNA sequencing costs by an order of magnitude.

Together, these studies identify novel mutations which confer industrially relevant phenotypes that can be used in future cellulosic biofuel production efforts. We show mutations identified for furfural tolerance can be redirected to improve biofuel tolerant phenotypes, suggesting a unified approach towards engineering both feedstock and product tolerance. Our findings also discuss broader applications to genome engineering, both in the importance of library and selection design, and the propagation of random mutations during commonly used engineering strategies that convolute the mapping of genotype-to-phenotype relationships.

Acknowledgements

I consider myself blessed to have been supported by mentors, coworkers, and friends during my graduate studies that are all due gracious appreciation. Firstly, I would like to thank my advisor, Prof. Ryan Gill, for his initial belief in me as an REU student many years ago. Thank you for trusting me to teach your class in Spring 2012—it was truly a highlight of my career at CU. Thank you for the members of my committee, Prof. Detweiler, Prof. Medlin, Prof. Randolph, and Dr. Zhang, for your support, comments, and opinions geared at shaping this work into what it was able to become. Corrie, I appreciate you hosting the microbiology journal club when I first started in grad school, as the course got me excited about the breadth of the field. Will, thank you for your continuing efforts supporting C2B2. The Center is a unique feature to the Front Range and benefits many students through its project funding and REU program, including me. Ted, I appreciate your work in keeping the Biotechnology Training Grant running. Completing my internship requirement for it was such an educational and eye-opening experience as to what industry is like. I count the encouragement from the program to complete an internship as a prime example of something that CU is doing right. And Min, thank you for your feedback regarding my project from a very early date. The NREL lab tours your gave our team after meetings were above and beyond the call of duty.

To members of the Gill Group, past and present, I give my sincere thanks. Nich Sandoval, Nanette Boyle, Steele Reynolds, Joost Groot, Paul Handke, and Pip Reeder—thank you for your daily contributions, being a sounding board for advice, lab help, and editing numerous materials. Nich, you have helped shaped so much of my work and allowed me to be trained while contributing to yours, for which I cannot express enough gratitude. You are truly a great mentor and friend. Sean Lynch, Andrew Garst, Ramsey Zeitoun, and Mike Lynch: your

ideas and scientific creativity astound me—thank you for sharing your expertise over the years. I will look forward to seeing your names in all the fancy journals in the future.

To all of the coauthors contributing to the work herein, I thank you for your effort and time. I am also extremely grateful to the team of students that worked with me—Alejandra Vargas (C2B2 NSF REU, Summer 2009), Katherine Schilling (high school student, 2009-2010), Hanna Aucoin (CU CHBE NSF REU, Summer 2011), and Jacob Gillis (CU student, 2013)—your enthusiasm, willingness to learn, and organizational skills have benefitted me beyond measure. I give special thanks to Jacob, who joined the lab just over half a year ago. In many cases, taking on a student to mentor is not the most prudent method for preparing to defend, but working with you has been a real joy and you helped get so much work done that otherwise would have been unexplored. I am confident that you are going to achieve whatever path you pursue. I will carry the Fantasy Football advice with me in the future and appreciate you teaching me about all of its complexities.

Dominique De Vangel… you are such an asset to every graduate student in this department. It is impossible to count the times that you helped with my confusions over enrollment, pay, deadlines, and general advice. Your positive attitude is infectious and much appreciated. There were times that purposefully walk to your office to ask a question instead of e-mailing it, just because I needed a smile in my day and knew you would have one. Your dedication to this job is so appreciated and is a very large contributing factor to the success of the department.

Numerous funding sources are gratefully acknowledged for supporting this work: the National Science Foundation Graduate Research Fellowship Program, the Department of

Education Graduate Assistance in Areas of National Need Program, the National Institutes of Health Biotechnology Training Grant, the C2B2 Seed Scholar Program, as well as funding through the Department of Energy and National Science Foundation. Thanks is also given to the CU Microarray Facility, which ran the arrays used in these studies. I am also grateful for the open use of resources within the Yin Lab in Biochemistry and the Chatterjee Lab in Chemical and Biological Engineering.

Finally, thank you to my family for your continuing support. I hold dear the memories of studying physics on rollerblades and watching UC Berkeley's biochemistry extension courses on Saturday mornings with you, Dad. To my mother, thank you for helping get my disability diagnosed and giving me the confidence to attempt to achieve despite it. Thank you for helping with every application, remembering every name and deadline, and putting up with all of my silly stories on the road to get here. I would not have attempted this path if not for both of your encouragement. And to my husband, Robin—you have been living in Iraq, South Korea, Hawaii, California, Georgia, Washington, Missouri, and bounced around various other places while I have been here in grad school. Thank you for filling out all of the forms required to explain to the Army why I was not there with you and patiently waiting for me to finish. This one's for you.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The projects contained herein are motivated by recent advances allowing for significant cost reduction for DNA synthesis and sequencing. As we enter an age where bacterial genomes can be chemically synthesized with a user-chosen code and then be "booted up" into recipient cells, a question exists—what should we write? The choice of deciding what to encode is the underlying principle of these studies.

Given the genetic code of adenine (A), thymine (T), cytosine (C), and guanine (G), the breadth of the mutational space is vast. For humans, with 3.2 billion bases of DNA, sequence variants of the same genome length number $4^{3,200,000,000}$. Considering instead a simpler organism, like the bacterium *Escherichia coli* (4.6 million bases), a large mutational space still exists even when only considering altering the expression levels of its 4,600 genes by one additional level (wild-type vs. mutant). This space outnumbers the estimated atoms in the known universe, 10^{1365} vs. 10^{80}, respectively. Presented with such a vast combinatorial space, we must develop tools to help us intelligently sift through the possibilities to find "winners"—sequences that are desirable to write.

Here, we approach engineering *E. coli* through a Design-Build-Test cycle in order to determine "winning" sequences (Fig. 1.1) for engineering cellulosic biofuel feedstock and product tolerance. In the initial Design phase, we target the way in which diversity is generated, for which many Design techniques exist. For a single gene, promoters, ribosomal binding sites (RBS), 5' untranslated regions, and increased dose via plasmid expression are some of the parameters we can vary without affecting the protein-coding open reading frame. Build, the

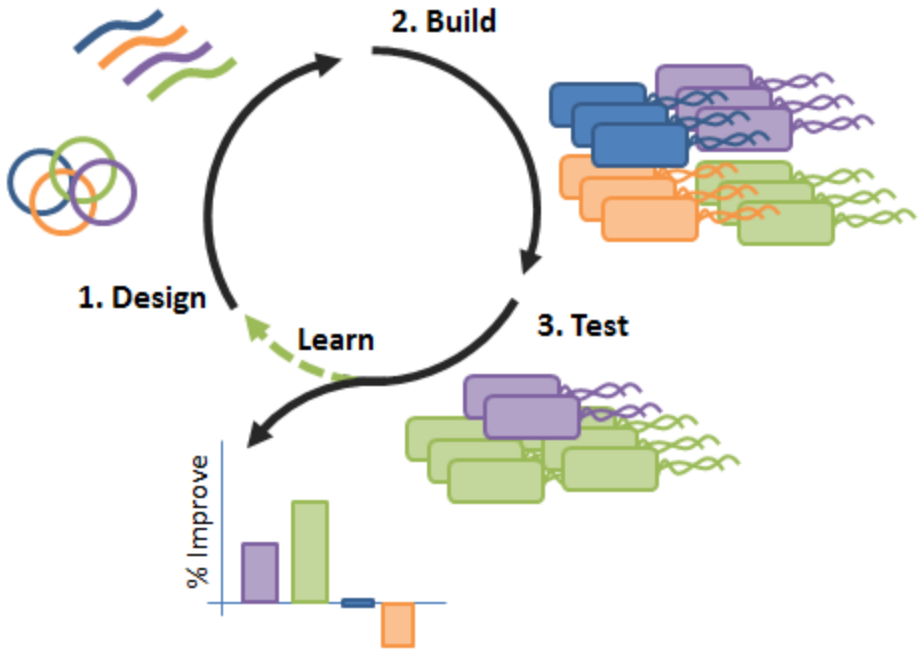


Figure 1.1 – Design-Build-Test cycle for genome engineering. Design entails choosing methods in which to augment wild-type populations. In one method of designing diversity, gene dosage can be increased through cloning onto multi-copy plasmids. Alternatively, synthetic DNA oligomers or constructed double-stranded cassettes can be used directly via recombineering to edit the genome for insertions, mutations, or deletions. The approach used during Build is prescribed by the design metrics of the library creation, but normally results in a mutant pool of $>10^3$ genotypically diverse clones. Mutants with altered phenotypes of interest (e.g., tolerance) can be enriched within the population through use of competitive growth-based selections, where fit mutants with improved growth characteristics increase in population frequency while less fit mutants are diluted within the population. High-throughput screens can also be performed on the diverse population, where the desired phenotype is not necessarily linked to improved growth (e.g., production of compound or increased enzymatic activity). Isolates from the screen or selection can then be confirmed for their improvements. Some reasons for mutants not being enriched during selection could be that the mutation provides no effect on growth (blue strain) or are deleterious (orange strain). By assessing trends of improved strains, an additional iteration of the cycle can be informed where new Design is informed by the success of previous round(s).

second phase, propagates a diverse library of mutants which vary based on the metrics outlined during Design. The final phase, Test, enriches strains with beneficial mutations through the use of growth-based selections. During the Test phase, we map genotype-to-phenotype relationships based on the trackability of the mutations we originally designed. Many tools have been created to allow targeted and rapid workflow through the Design-Build-Test phases, some of which are used in the following chapters. Ultimately though, implementing the Design-Build-Test approach is linear, unless we use the knowledge gained from a previous Test phase to inform future designs. By introducing a Learn phase, we enable iterative cycling through Design-Build-Test phases for engineering improved or new traits. Our objective here is that the progression through these studies achieves a complete Design-Build-Test cycling, where knowledge learned from initial studies directs a future design.

In Chapter 2, we motivate the phenotypes we are aiming to engineer, namely traits associated with overcoming some of the limitations in implementing cellulosic biofuel production. Lignocellulosic biomass is a renewable, sustainable, and importantly, non-food feedstock that contains large amounts of sugars in the form of cellulose and hemicellulose. Conversion of this saccharified, or "hydrolyzed," biomass is limited in part by microbial inhibition from chemical toxins within the hydrolysate. We focus initially on one such inhibitory compound, furfural, a degradation product of pentose sugars during feedstock pretreatment prior to fermentation. Additionally, as microbes ferment sugars within the hydrolysate to a biofuel product, they produce large quantities of a fuel that is, in turn, also toxic to the cells, reducing yields. We additionally discuss traditional and new approaches for genome engineering studies.

In Chapter 3, we perform our first studies towards applying the Design-Build-Test cycle by using the multiSCale Analysis of Library Enrichments (SCALEs) method for generating and tracking diversity. SCALEs employs multiple defined-size genomic libraries that test, in parallel, the effect of increased dosage for $>10^5$ clones under selective pressure. The frequency of each clone (library insert) is measured with microarrays. Using the SCALEs processing algorithms, a fitness score (i.e., frequency of a clone after selection compared to the frequency without selection) can be assigned to each gene, compiled from the contribution of each sized library insert containing a given gene. We hypothesized that use of the SCALEs method would identify novel genes that, when overexpressed, would confer tolerance to furfural. High-fitness genes were identified and verified for their beneficial effect on furfural tolerance. Two of these alleles are novel furfural tolerance targets: *lpcA*, which is involved in lipopolysaccharide (LPS) biosynthesis, and *groESL*, a protein folding chaperone. Assessment of the increased-fitness dataset suggests that cell membrane and wall formation is important under furfural challenge.

In Chapter 4, we apply the TRackable Multiplex Recombineering (TRMR) approach towards identifying additional furfural tolerance genes. The TRMR method uses two libraries of 'Up' and 'Down' mutants, engineered for increased or decreased expression, respectively, of virtually every gene in *E. coli* (~4,000 genes x 2 libraries = 8,000 distinct clones). 'Up' mutants contain a strong promoter and RBS, while 'Down' mutants disrupt the native promoter and RBS sequence. Each mutation is barcoded with a unique 20 bp sequence, which we used to track clone enrichments to calculate fitness scores. We performed various selection regimens with these libraries to gauge if application of selective pressure varied the enriched genes. Another comparison was also performed by overlapping enriched genes from the SCALEs selection to the 'Up' TRMR library enrichments to identify genes that potentially would confer tolerance

over a range of increased expression. Through this analysis, we identified four additional novel furfural tolerance genes: *ahpC*, *dicA*, *rna*, and *yhjH*. Growth improvements from these clones were only observed for one of either the SCALEs or TRMR constructs, but not both. Assessment of expression from these strains highlights that differences library design and build strategies can impact the results found during testing.

During Chapters 3 and 4, we used a linear approach through Design-Build-Test phases. While the previous chapters focused on furfural tolerance, we transitioned towards engineering biofuel tolerance (e.g., *n*-butanol and isobutanol) in Chapter 5. Furfural is a hydrophobic compound, as are many biofuel candidates. Through observations from the previous studies identifying genes important for hydrophobic tolerance (*lpcA* in Chapter 3 and enriched 'Up' mutants for LPS biosynthesis in Chapter 4), we hypothesized that we could target LPS biosynthesis to improve tolerance to hydrophobic biofuels. In this way, we introduced a learning step, completing the Design-Build-Test cycle (Fig. 1.1).

In this effort, we targeted increased expression of all LPS biosynthesis genes by inserting a promoter upstream of these targets. We used the Multiplex Automated Genome Engineering Approach (MAGE) approach to create these directed mutants. The MAGE method employs a lambda phage protein to insert user-defined mutations, encoded as single-stranded DNA oligomers, into the chromosome via homologous recombination (often referred to as recombineering). The resulting combinatorial library contained mutants of varying degrees of promoter insertion (i.e., a promoter inserted in front of one gene vs. promoter insertions at a number of sites). We confirmed up to 50% growth improvement in *n*-butanol of library isolates screened for the increased LPS phenotype. In addition, we pioneered use of a recently developed approach in the Gill Lab by Ramsey Zeitoun and Andrew Garst for multiplex mutation tracking,

with the aim of reducing sequencing costs from $110/clone by an order of magnitude. A selection was performed with these libraries in various alcohols (ethanol, *n*-butanol, isobutanol, and isopentenol), and characterization of clones from the enriched population was performed. Surprisingly, the greatest gains in tolerance were attributed to spontaneous (non-directed) mutations. Our findings highlight the spontaneous divergence from wild-type of recombineered libraries, an important metric needing to be accounted for during genotype-to-phenotype mapping efforts. Despite the propagation of random mutations within the population, our genotyping efforts of tolerant clones identified a specific locus with promoter insertion in front of LPS genes, supporting our hypothesis that targeting increased LPS biosynthesis gene expression can confer tolerance to hydrophobic biofuel compounds.

Chapter 2

Feedstock and Product Toxicity and Tolerance Mechanisms in *Escherichia coli*

Parts of this chapter have been reported in "Cellulosic hydrolysate toxicity and tolerance mechanisms in *Escherichia coli*." Biotechnology for Biofuels 2:26 (2009)

Authorship: Mills, T.Y.*
Sandoval, N. R.*
Gill, R.T.
*equal contribution for the manuscript as published

2.1 – Introduction

Governments worldwide have been issuing mandates for increased biofuel production over the past decade in order to meet growing energy demands. Focus has shifted from using food crops (e.g., corn) or land allocated for food production, which can lead to an increase in greenhouse gas emissions [1] to using lignocellulosic biomass. Advanced biofuels, like those derived from lignocellulosic biomass can offer greenhouse gas savings, particularly when drawn from feedstocks not requiring land use change [2]. The United States has set legislation to require cellulosic advanced biofuel production at 16 billion gallons in 2022 (44% of total biofuels mandated) [3], a far difference from the 20,000 gallons produced in 2010 [4].

Despite the large efficiency gains required to meet current and future mandates, advances in molecular, metabolic, synthetic, and genome engineering are enabling smart construction of cellular biocatalysts to meet the improvements made on the processing side ([5] and for a recent review see [6]). Conversion of sugars to fuels by microbes is a process exploited and refined by humans over millennia, but has primarily focused on idealized feedstocks and production of ethanol. Liquid transportation fuels are currently substituted with varying percentages of

ethanol, but the transition to advanced biofuels provides benefits like higher energy density and pipeline fungibility. Potential advanced biofuels include alcohols of varying chain length [7-9], diesel [10], and alkanes [11], and can be produced from feedstocks like agricultural biomass [10,12] and seaweed [13]. Desirable traits for engineered strains include high yield and titers while utilizing a range of substrates within the feedstocks (e.g., pentoses and hexoses), robustness to process conditions (e.g., pH, temperature, and salinity), as well as feedstock and product tolerance.

Escherichia coli, *Saccharomyces cerevisiae*, *Clostridium acetobutylicum, a*nd *Zymomonas mobilis* are some of the most promising industrial biocatalysts for cellulosic biofuel production [14]. However, each microbe has varying degrees of limitations, like non-ideal native substrate utilization (e.g., co-consumption of pentose, hexose, and cellobiose), lack of known metabolic pathways to produce the fuel compound, and low tolerance to both feedstocks and products. Unlike *S. cerevisiae* or *Z. mobilis*, *E. coli* natively ferments both hexose and pentose sugars. And although *C. acetobutylicum* has been in use for a century for its ability to produce acetone, ethanol, and *n*-butanol, it has been side-lined due to the prevalence of petroleum sources from which to derive acetone instead [15]. *C. acetobutylicum* is also an obligate anaerobe, a requirement that complicates process design, its metabolism of xylose and arabinose was only just characterized last year [16,17], and the toolbox for genetic manipulation is not as advanced as other biofuel-producing microbes—resulting in reduced momentum to engineer it for cellulosic biofuel production. Alternatively, ethanologenic *E. coli* has been shown to be similar or superior in performance to other fermentative microbes when comparing production levels and tolerance to hydrolysate inhibitors [18-21]. In addition, production of some of the most promising biofuel candidates has been originally engineered into *E. coli* [8-11].

These benefits, taken together with the vast knowledge of its genome and methods to alter it, make *E. coli* an attractive host for cellulosic biofuel production engineering. Here, we are especially concerned with engineering tolerance to furfural and biofuel products, which has shown promise for improved production of cellulosic biofuels [12].

2.2 – Furfural Toxicity and Tolerance

Furan derivatives, like furfural and 5-hydroxymethylfurfural (HMF), result from sugar dehydration during pretreatment of lignocellulosic biomass. Furfural and HMF are the primary furan derivatives appearing in lignocellulosic hydrolysate, with concentrations typically ranging between 0 to 5 g/l for each compound [22-25]. While dilute acid hydrolysis is a common method for pretreatment, acidic conditions are known to cause dehydration of a small fraction of the sugar monomers. While new methods are being developed to reduce the amount of furfural and HMF formed during pretreatment [26-28], industrial-scale technology and knowledge about process kinetics currently favors more traditional processes like dilute sulfuric acid treatment [29-31]. Therefore, it is important to improve understanding of the genetic and metabolic mechanisms underlying tolerance to furan derivatives. This section will focus primarily on furfural toxicity and tolerance, due to its higher abundance than HMF in a variety of hydrolysate preparations [32], but overlap between strategies benefitting furfural tolerance to HMF tolerance have been observed [33].

Aldehydes are, in general, known to have detrimental effects in microorganisms. Formaldehyde has been shown to denature and interact with polynucleotides [34] and can cause protein-protein cross-linking [35]. *In vitro* experiments with crude cell extracts identified a

glutathione-dependent formaldehyde dehydrogenase that is responsible for conferring aldehyde tolerance [36]. Two previously uncharacterized proteins, FrmB and YeiG, have also been identified for their role in conferring formaldehyde tolerance via a glutathione-dependent formaldehyde hydrolysis pathway [37]. Besides enzymatic detoxification, altered outer membrane protein composition has also been indicated as conferring increased tolerance to formaldehyde, acetylaldehyde, and glutaraldehyde [38]. Furthermore, methylglyoxal, a dicarbonyl compound, inhibits *E. coli* growth and protein synthesis at concentrations of 0.07 g/l [39,40].

Furfural has been identified as a key inhibitor in lignocellulosic hydrolysate because it is toxic by itself and also acts synergistically with other inhibitors [19]. It is a hydrophobic compound, which is a commonly regarded metric of an organic compound's toxicity [18,19,41,42]. Highly hydrophobic compounds have been shown to compromise membrane integrity [43]. Interestingly, perceptible membrane damage in *E. coli* resulting from furfural exposure has not been observed, suggesting that intracellular sites are more likely to be the primary inhibition targets of furfural and HMF [19]. In contrast, other furan derivatives like 2-furoic acid and furfuryl alcohol have been shown to cause significant membrane leakage [18,20]. Furfuryl alcohol also exhibits synergism when in binary combinations with other inhibitors, while 2-furoic acid results in additive toxicity [18,19].

Ethanol production is inhibited by furfural, suggesting a direct effect on glycotic and/or fermentative enzymes [19]. Glycotic dehydrogenases, like alcohol dehydrogenase (ADH), have been implicated as a potential site of inhibition via NAD(P)H-dependent aldehyde reduction into furfuryl alcohol [44]. A study performed *in vitro* has confirmed that acetaldehyde to ethanol conversion was inhibited by both furfural and HMF [45]. Subsequent *in vitro* enzymatic assays

demonstrated that furfural was a substrate for ADH (EC 1.1.1.1), albeit at a five-fold increase in K_m and five-fold decrease in V_{max}. In the same study, furfural inhibition on aldehyde dehydrogenase (AlDH; EC 1.2.1.5) and the pyruvate dehydrogenase (PDH) complex were investigated and determined to be more significant than ADH, as evidenced by more than 80% activity reduction in the presence of 0.12 g/l furfural, whereas ADH activity was only inhibited by 60%. These findings suggest that furfural may detrimentally affect multiple glycotic enzymes essential to central metabolism.

E. coli metabolizes furfural to furfuryl alcohol under aerobic conditions [46]. The bioconversion occurs via a NADPH-dependent furfural reductase (FFR), which was the first of its kind to be reported in the class of alcohol-aldehyde oxidoreductases [47]. In the same study, the FFR showed an increased rate of NADPH oxidation when acting on benzaldehyde compared to furfural, showing that the FFR can utilize a variety of aldehydes as substrates.

A recent long-course adaptation experiment found that furfural tolerance is conferred by silencing certain NADPH-dependent oxidoreductases [48]. Genes of special interest in this work were *yqhD* and *dkgA*, both of which encode enzymes with low K_m's for NADPH. Miller *et al.* proposed that the primary furfural toxicity mechanism derives from a competition for NADPH between furfural reduction and biosynthesis. The authors proposed that the lag phase cells initially undergo during furfural treatment was a result of decreased biosynthesis due to the NADPH pool being used for furfural reduction. The mutant with silenced *yqhD* and *dkgA* genes was able to concurrently reduce furfural and grow, providing further support for the proposed inhibition pathway. YqhD is also reported to play an important role in protecting *E. coli* from aldehydes derived from lipid-peroxidation via a glutathione-independent, NADPH-dependent reduction mechanism [49]. An example of furfural-induced lag phase in *E. coli* is provided in

Fig. 2.1. Here, cells were grown with either furfural or molar equivalents of furfuryl alcohol. A significant lag phase is observed with increasing furfural concentration, while additional furfuryl alcohol loading did not affect growth compared to a lower concentration. It has been shown that the minimum inhibitory concentration of furfural is much lower (< 4 g/l) than furfuryl alcohol (20 g/l), indicating that at the concentrations relevant to hydrolysate preparations (5 g/l or less of furfural), are within the range of complete furfural toxicity, but not for total inhibition by furfuryl alcohol [19,42].

Interesting to note is that the mutant obtained from the long-course adaptation study also overexpressed eight oxidoreductases that can use NADPH as an electron donor, but have substrate specificities. For example, one such enzyme, YajO, is highly specific for utilizing 2-carboxybenzaldehyde as a substrate in comparison to a variety of other aldehydes [50]. NADH- and NADPH-dependent reduction of furan derivatives has also proved paramount for hydrolysate inhibitor tolerance in *S. cerevisiae* and *P. stipitis* [51-55]. Recent studies in *E. coli* have found that NADPH-dependent furfural reduction inhibits sulfur assimilation [56]. Overexpression and increased activity of an NADH oxidoreductase [57,58], or of a predicted oxidoreductase [59], overexpression of the NADPH-restoring transhydrogenase PntAB [56], and combinations of some of these mutations have all been shown to confer furfural tolerance [12].

Furfural and HMF have shown cytotoxic characteristics towards both bacteria and yeast [19,60-62]. Furfural is a known dietary mutagen and has been under investigation for direct effects on DNA. A series of studies confirmed that furfural-DNA interactions occur [63], and that furfural-treated DNA leads to single-strand breaks after undergoing *in vitro* incubation with furfural, primarily at sequence sites with three or more adenine or thymine bases [63,64].

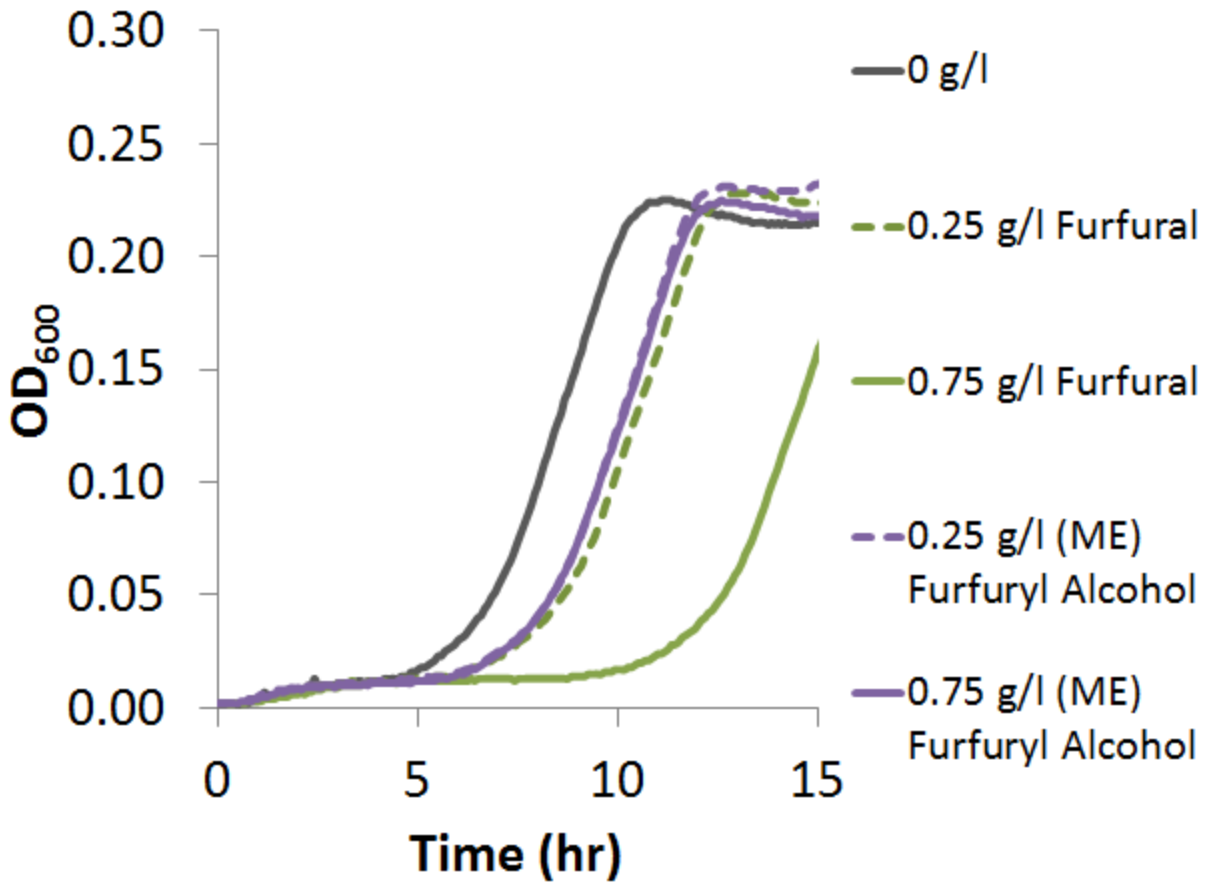


Figure 2.1 – Effect of increasing furfural and furfuryl alcohol concentrations. *E. coli* was inoculated at the same initial optical density from cultures grown to mid-exponential phase in MOPS minimal medium with varying concentrations of furfural or the molar equivalent (F.M.E. stands for furfural molar equivalent) of furfuryl alcohol. Cells were grown in a microtiter plate and optical density was monitored at 600 nm every 5 minutes for 15 hours. Materials and methods for this procedure are similar to those used in Chapter 3, with the exception of the use of a microtiter plate as a vessel.

Later, plasmids treated with furfural were observed to cause either an increase (in high furfural concentrations) or decrease (in low furfural concentrations) in plasmid size via insertions, duplications, or deletions [65]. Recently, overexpression of thymidylate synthase, encoded by *thyA*, has been shown to confer tolerance to furfural, presumably due to its role in pyrimidine biosynthesis, which might be required during furfural treatment [66]. *In vivo* experimentation suggests the importance of the *polA*-mediated DNA repair pathway for tolerating scissions caused by furfural [67]. *E. coli* has been observed to repair damaged DNA, reducing the frequency of furfural-induced mutagenic events to that of random mutation found in untreated cultures [68].

Additionally, furfural treatment increases reactive oxygen species (ROS) formation in yeast and *E. coli* [69,70] and global regulator engineering of an exogenous regulator provides increased tolerance to furfural [70]. Previous studies on hydrolysate have also identified genes involved with ROS detoxification [32,71].

2.3 – Toxicity and Tolerance of Other Inhibitors and Products

Although fermentation of glucose to ethanol is the current standard for biofuel production, higher chain alcohols present a variety of benefits over ethanol, namely higher energy density and lower hydroscopicity. Recent advances in production of higher chain alcohols has been exhibited in *E. coli* for *n*-butanol, isopropanol, isobutanol, and other branched chain alcohols [8,72-74]. Current titers of isobutanol production are ~20 g/l [8], but can produce up to ~50 g/l with *in situ* removal of isobutanol during production [75]. Since the production routes to these desirable fuels have only recently been metabolically engineered, the known

toxicity and tolerance mechanisms of these higher-chain alcohol fuels are still being elucidated. Targets currently known to play a role in hydrolysate and product tolerance are shown in Fig 2.2.

Solvent treatment elicits changes in membrane fatty acid composition, and the concentration at which these changes has been related to the alcohol's hydrophobicity [41]. Treatment with various alcohol and acid compounds that are also hydrolysate inhibitors (e.g., acetate and coniferyl alcohol) leads to leakage of intracellular contents [18,42]. Transcriptomic analyses of *E. coli* under ethanol, *n*-butanol, and isobutanol stress have identified unique differences based on the inhibitor used; distinction between ethanol and the C_4 alcohols is specifically highlighted [76-79], suggesting that while membrane maintenance is important, it is not the sole contributor towards tolerance.

The use of efflux pumps to confer tolerance to hydrophobic biofuel candidates has also received attention [32,80]. In a bioprospecting search, 43 efflux pumps from various microbes were expressed in *E. coli*, and pumps conferring tolerance to geranyl acetate, geraniol, α-pinene, limonene, and farnesyl hexanoate were discovered [81]. Meanwhile, no pumps were found to restore growth under *n*-butanol or isopentanol treatment, again highlighting the differences required while engineering tolerance to biofuel candidates.

Metabolic pathways are also inhibited under conditions relevant for hydrolysate tolerance. Studies on hydrolysate, acetate, and various other individual inhibitors have identified amino acid supplementation or pathway augmentation as methods for conferring tolerance [32,56,82-84]. Identifying bottlenecks in amino acid biosynthesis is especially important, not only for cell growth, but also for non-fermentative biofuel production pathways that utilize amino acids as substrates [8].

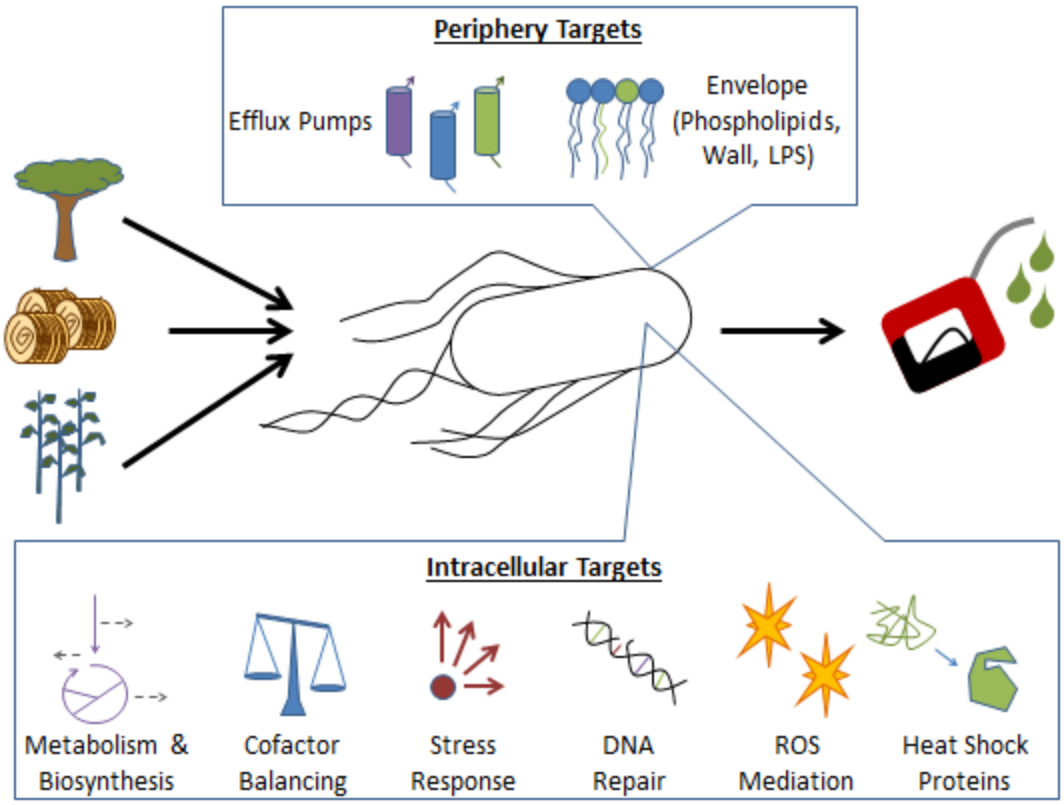


Figure 2.2 – Targets for engineering tolerance to inhibitory compounds that reduce production efficiency during cellulosic biofuel production. A cellulosic biorefinery uses feedstocks from non-food sources (e.g., switchgrass or corn stover) that are primarily made of cellulose, hemicellulose, and lignin. These feedstocks are pretreated to break apart the fibers and hydrolyze hemicellulose into C_5 and C_6 sugars. Pretreatment is typically performed under harsh conditions with high temperatures and low pH, resulting in the formation or release of inhibitory compounds, like furfural and acetate. Saccharification of cellulose is performed with a cocktail of enzymes to release glucose. The hydrolyzed mixture is then fed into fermentation, where production of biofuels like ethanol and butanol occurs. Microbial conversion of the sugars to the biofuels is inhibited by toxicity of compounds like furfural, as well as the formation of the products themselves. Mechanisms of toxicity and methods for engineering tolerance have been identified and fall into eight general categories: two directed at the cell's periphery and the remaining six involved with primarily cytosolic-based reactions.

The overexpression of heat shock proteins has also been implicated in solvent tolerance [85,86] and low pH conditions [87], which is integrated into the stress response of *E. coli*. The global transcription factor cAMP receptor protein (CRP), involved in conferring isobutanol tolerance [76], has recently been engineered to improve increased isobutanol tolerance while concurrently reducing ROS levels during isobutanol treatment [88]. Engineering transcription factors for global response improvements has been a fruitful undertaking with improvements made in solvent, hydrolysate inhibitor, and oxidative stress tolerance [7,70,89-92].

2.4 – Methods for Engineering Tolerance

Traditional engineering approaches tend to employ long-course adaptation studies. In this approach, a culture is inoculated into media with a chemical inhibitor and allowed to grow for many doublings until spontaneous mutants propagate. This approach, in essence, skips the Design phase while relying on random mutations to build diversity. While simple to perform, the process is typically time-consuming. Moreover, the process evolves strains with unknown mutations, which requires whole-genome sequencing to identify the mutations and lengthy subsequent analysis to determine which mutation(s) is beneficial. Despite these setbacks, strains adapted for ethanol (over 3 months) [9] and isobutanol (over 6 months or with 45 serial transfers in a different study) [77,93] tolerance have been isolated successfully and have contributed towards significant advances in engineering biofuel production. Use of chemical mutagens can increase the speed by which mutations are propagated and accumulated and have been used to engineer isobutanol [94], 3-methyl-1-butanol [95], and vanillin tolerance [96]. But chemical mutagenesis has also been reported to be one of the weakest approaches for creating beneficial mutations under acid stress [97,98].

2.4.1 – Genome-wide search tools

Because long-course adaptation and random mutagenesis studies can propagate both beneficial mutations as well as mutations not contributing towards the desired phenotype, attempting to map the relationship between a given genotype responsible for the desired phenotype can be hindered. To address this issue, a variety of tools have been developed to streamline the Design-Build-Test phases and allow for genotype-to-phenotype mapping of clones. Emphasis is focused here towards tools which provide whole-genome coverage since the mechanisms of toxicity, and thereby location of the genetic controls for conferring tolerance, might not be known.

The first application of such genome-wide genotype-to-phenotype mapping, enabled by the use of quantifying population frequencies with microarrays, was performed for transposon insertion libraries [99] and later, plasmid-based genomic libraries [100]. A further advance in genomic library genotype-to-phenotype mapping was performed with the multiSCale Analysis of Library Enrichments (SCALEs), where multiple defined-size libraries were developed to identify, at roughly 125 nucleotide resolution, allele frequency, and therefore enrichment, at over ten times 99.9% complete genome coverage. The SCALEs approach has been used for a variety of applications, including the identification of ethanol and acetate tolerance genes [78,83,101].

Integration of mutations onto the chromosome is both more industrially relevant and enables continued searching and strain construction without having to use many plasmids. Although designed for single mutation genotype-to-phenotype mapping efforts, TRackable Multiplex Recombineering (TRMR) integrates mutations directed at increasing or decreasing expression of virtually every gene in *E. coli* [71]. The TRMR approach has been used to identify mutations conferring tolerance to acetate, hydrolysate, and low pH [71,82]. The TRMR

approach expands the search to assess decreased gene expression, compared to SCALEs, which only assesses the effect of overexpression.

2.4.2 – Combinatorial approaches

Typically, improvements in tolerance rely on the effect of multiple mutations working together in concert. This has been true for both engineering furfural and isobutanol tolerance [12,77]. For this reason, the Coexisiting/Coexpressing Genomic Libraries (CoGeL) method has been developed to assess genome-wide epistatic interactions between genes through the use of plasmid and fosmid libraries [102], but is limited to assessing the epistasis of only two genomic library inserts.

The advent of Multiplex Automated Genome Engineering (MAGE) in 2009 has opened a door for a variety of engineering applications and can be used to target multiple mutations [103]. In this approach, lambda-mediated recombination integrates designer mutations, synthesized as single-stranded DNA oligomers, onto the chromosome. The recombineering typically occurs over recursive rounds, due to inherently low efficiency of recombination (typically ~5 %, but varies with the strain and type of oligomer and mutation [82,103-107]).

Although powerful in approach, MAGE is limited since it requires the user to define the location for mutation, by way of including homology arms on the oligomer to direct recombination. Last year, our lab introduced a strategy to combine genome-wide searches and multiplex engineering [82]. In this strategy, we used TRMR to guide a genome-wide search for alleles conferring tolerance to hydrolysate, acetate, and low pH. High-fitness genes identified with TRMR were then targeted with MAGE to alter expression over a wide range based upon the direction indicated by TRMR. Combinatorial mutants were created through recursive rounds of

recombineering, and then were run through a selection to enrich for tolerant mutants. This strategy was novel in the combination of genome-wide searching and multiplex genome engineering, but highlighted the importance of epistasis, when it was discovered that the fittest mutants after selection had only one mutation at a TRMR-directed allele, instead of at multiple sites. It was concluded that using TRMR-identified alleles that would have positive epistasis would increase the ability to select for multiplexed mutants, and that choosing just the top high-fitness genes would not ensure this outcome. Through our use of the Design-Build-Test cycle here, we use information gathered from our genome-wide searches to direct the targets chosen for multiplex engineering in order to support positive epistasis.

Chapter 3

Genome-wide mapping of furfural tolerance genes in *Escherichia coli*

In press at *PLOS ONE.*

Authorship: Mills, T.Y.
Sandoval, N.R.
Reeder, P.J.
Schilling, K.D.
Zhang, M.
Gill, R.T.

3.1 – Introduction

Genome engineering strategies are limited by the massive combinatorial search space created when multiple genetic units must be optimized in tandem [82,108]. While early efforts focusing on engineering a small number of genetic parts have resulted in several impressive results [109-111], efforts focused on the engineering of complex phenotypes have remained a key challenge for the field. This challenge is especially true when the genetic bases of the targeted phenotypes are poorly understood, as is the case for many tolerance phenotypes [9,77,78,82,83,101].

Advances in methods for mapping genotype-to-phenotype relationships have helped address this issue ([32,71,100,102,112-114] for a detailed review see [115]). Mapping approaches enable rapid identification of novel gene targets for strain design. These strategies generally employ well-defined libraries that allow for tracking of all members in parallel during a high-throughput screen or selection. Importantly, multiplex genome-modification strategies can be used to then develop combinatorial mutants of multiple alleles identified during genome

mapping [89,103,116,117]. Together, these strategies represent an approach for rationally searching genetic space during genome engineering efforts [82].

Here, we have applied one of these new methods for genome mapping, multiSCale Analysis of Library Enrichments (SCALEs) [113], to engineer furfural tolerance, an important phenotype for improving microbial biofuel production from lignocellulosic hydrolysate. Lignocellulosic biomass (e.g., switchgrass and corn stover) is a proposed feedstock for next-generation biofuel production [118], since it is a renewable and sustainable source of sugars (from hemicellulose and cellulose). Biomass pretreatment and saccharification release sugars into the liquid hydrolysate, which can be fermented into biofuels, but also release a variety of inhibitory compounds. Furfural, is a heterocyclic aldehyde formed from pentose degradation during pretreatment, and is one of the key inhibitory compounds in hydrolysate ([19] for a review on hydrolysate toxicity see [119]).

Furfural is a known DNA mutagen in *Escherichia coli* [63,64,120]. In addition, growth inhibition induced by furfural has been linked to the reduction of furfural to furfuryl alcohol by NADPH-dependent oxidoreducatses [47]. This reduction elicits a variety of negative responses in the cell, causing starvation of available NADPH necessary for biosynthetic processes such as sulfur assimilation [56] and pyrimidine synthesis necessary for DNA repair [66]. Alleviation of NADPH-starvation can be obtained by silencing NADPH-dependent oxidoreductases [121], increasing NADH-dependent reductase expression [57] and activity [58], increasing expression of a predicted oxidoreductase [59], and overexpressing the NADPH-restoring transhydrogenase PntAB [56]. A recent study combined many of these mutations together to improve production of ethanol and succinate from hydrolysate [12]. Similar toxicity mechanisms and genetic manipulations have been beneficial for engineering *E. coli* for tolerance to 5-

hydroxymethylfurfural [33], a hexose degradation product in hydrolysate. In addition to directly redox related mechanisms, reactive oxygen species (ROS) accumulation has been observed in yeast cells [69] and *E. coli* [70] when treated with furfural, which is a common phenotype associated with DNA damage [122], as well as more generally with chemotoxicity [123].

We hypothesized that use of the SCALEs method would identify additional novel targets for engineering furfural tolerance. SCALEs employs four genomic libraries, each with distinct insert sizes (1, 2, 4 or 8 kb) to test, in parallel, the effect increased dosage of insert sequence (containing gene(s) and/or operon(s)) has under a selective pressure. Individual clone frequencies are calculated using microarray technology and the SCALEs signal processing algorithms, as described by Lynch et al. [113]. The multiscale analysis algorithm assigns the microarray signals according to the contribution from each library size. This method produces genome-wide fitness data at approximately 125 nucleotide resolution, thus allowing for precise mapping of the genetic basis of high fitness clones. The SCALEs method has previously been used to map genotype-to-phenotype relationships in a variety of applications, including: engineering tolerance to anti-metabolites [124], solvents [78,101,125], organic acids [83,126,127], antibiotics [128,129], as well as identifying genes restoring redox balance [130]. Here, we applied the SCALEs method to simultaneously map furfural related fitness effects resulting from overexpression of all *E. coli* genes (a total of $>10^5$ individual clones were evaluated). Follow-up studies confirmed novel furfural tolerance genes.

3.2 – Materials and Methods

3.2.1 – Bacteria, plasmids, and media

E. coli BW25113 Δ*recA*::Kan was obtained from the Keio Collection [131], and the kanamycin resistance cassette was removed according to the previously designed protocol [132] to yield BW25113 Δ*recA*::FRT, which was used as the host for all studies here, as similarly reported [78,83]. The pSMART-LCK (Lucigen) vector was used for library and clone construction. Ligated vector with no insert was used as the control. All cultures were grown at 37°C. Kanamycin was used where appropriate (30 μg/ml). Selections and growth tests were performed in MOPS minimal medium [133] with 0.2 w/v% glucose. Luria-Bertani (LB) medium was used for routine applications.

3.2.2 – Genomic libraries, selection, and microarray analysis

Genomic libraries were prepared previously by Warnecke et al. [127], by extracting genomic DNA from *E. coli* K-12 (ATCC #29425) to construct 1, 2, 4, and 8 kb SCALEs libraries in pSMART-LCK. Plasmid libraries were extracted from originally prepared cells with a Plasmid Midi Kit (Qiagen) and freshly transformed into the BW25113 Δ*recA*::FRT host. Samples of the transformants were diluted to confirm a minimum of 10x 99% library coverage ($>10^5$ cells) [113]. After a one hour recovery following transformation, the libraries were diluted into a single MOPS minimal medium culture and grown to early exponential phase. Aliquots of 50 μl were spread onto 20 MOPS minimal medium plates (control) or MOPS minimal medium plates with 0.75 g/l furfural ($>10^5$ cells total plated for each condition). Plates were incubated until growth appeared (one day for control plates and three days for furfural plates). Cells were harvested from the plates and plasmids were extracted with a Plasmid Midi Kit (Qiagen). Samples were digested and prepared for microarray analysis according to the method of Lynch et al. [113]. Analysis of the resulting data file was performed with the SCALEs software [113] as previously described [83], with plasmids from minimal medium plates without furfural serving

as the control sample. Fitness, W, is calculated for an individual clone, i, by $W = frequency_{i,furfural}/frequency_{i,control}$. Because overlapping clones may contain part of all of a particular gene, individual gene fitness scores were calculated as a summation of clones containing a given gene, weighted by the fraction of the gene contained in the clone. Analysis of Gene Ontology term enrichment was performed with the Batch Genes tool available on the GOEAST website [134] using default settings.

3.2.3 – Clone construction

Primers for gene amplification were designed to amplify the native promoter and open reading frame for each target and are listed in Table 3.1. Phosphorylated cassettes were ligated into pSMART-LCK according to manufacturer directions and then transformed into electrocompetent cells. Plasmid constructs were confirmed by gel electrophoresis and sequencing.

Table 3.1 - Primers used in creating plasmids for high-fitness SCALEs genes.

Clone_Direction	Sequence (5'-3')
thyA_for	CGTTGCAAAATTTCGGGAAGGCGT
thyA_rev	GCTGCTGCTGGAAGGTGTGGT
ybiY_rev	ATGCGGTCGCTGAGCGTGTC
ybiY_rev	CCTGGGCAAACAGACGCCCC
groESL_for[a]	GAGACCGGAATTCCGGTGACGGCGATGAAGAAATTGCGA
groESL_rev[a]	GAGACCGGAATTCCGGACATTTCTGCCCGGGGGTTTGT
lpcA[b]_for	AAGCCCCTTACTTGTAGGAGGTCTGA
lpcA[b]_rev	TCGCATCAGGCATCAGCGCACAAAT
ybaK_for	GCCGCTGGATGTGAGTGTTT
ybaK_rev	AAGCGACGGTGTAACTCGAT

[a] Due to the large size of this insert, it was constructed using cohesive-end cloning and contains the EcorI site.
[b] *lpcA* plasmid was constructed by Woodruff et al [78].

3.2.4 – Growth curves and plating assays

Cultures inoculated from freezer stocks were grown overnight in LB medium. Seed cultures were inoculated with 2 v/v% overnight cultures into MOPS minimal medium, grown into exponential phase, and diluted to OD_{600} 0.195-0.200 to be used as innocula for test cultures at 10 v/v%. Growth curve studies were performed in 15 ml conical tubes with 5 ml liquid volume. Furfural was added to a concentration of 0.75 g/l. Growth was monitored at 600 nm for 24 hours (n = 3).

For plating assays, normalized seed cultures were diluted by half, from which 1 µl (~10^4 cells) was streaked onto MOPS minimal medium plates with furfural (0-1.5 g/l). Plates were incubated at 37°C for up 72 hours.

3.2.5 – Furfural reduction measurements

Furfural was measured with a spectrophotometer at 284 nm [135]. A standard curve was prepared in MOPS minimal medium and fit by linear regression. Standards and samples were diluted 1:1000 in water. Samples were collected from growth curve cultures during cell density measurements and stored at 4°C for a maximum of 12 hours prior to analysis. Furfural measurements were normalized to cell density, and reduction rate was calculated from the regression line during the transition from lag phase to exponential phase, where reduction trends were linear. Samples were collected over 24 hours, at which point furfural was no longer observed in the cultures.

3.2.6 – Mutation frequency analysis

Mutation frequency was measured by proxy with frequency of rifampin resistance [136,137]. Cell cultures were grown overnight, harvested by centrifugation, diluted 10-fold into 25 ml of MOPS minimal medium, and incubated for 30 minutes to allow for growth to begin. Furfural was added to 0.75 g/l and cultures were incubated for 3 hours. Cells were then harvested and diluted accordingly for measuring total CFU count (LB agar) and spontaneous mutants (LB agar with 100 µg/ml rifampin). Mutation frequency was calculated by dividing the number of rifampin resistant mutants by total CFU (n = 4).

3.2.7 – qPCR expression analysis

Strains were prepared and grown according to the same procedure used for growth curve analysis with the following exception: strains were inoculated into MOPS minimal medium without furfural and grown for 6.5 hours (into exponential phase). Aliquots of 1 ml were harvested by centrifugation, decanted, and immediately frozen in a dry ice-ethanol bath, and stored at -80°C until further use. For RNA extraction, 400 µl of RNAProtect Cell Reagent (Qiagen) was added to pellets, mixed by pipetting, and then processed with an RNAEasy Mini Kit (Qiagen). RNA samples were analyzed with an iTaq Universal SYBR Green One-Step Kit (Bio-Rad). Expression of *cysG* was used as a housekeeping reference gene [138] for calculating relative fold-change (n = 2-3).

3.2.8 – Site-directed mutagenesis clone construction and testing

Mutants were constructed using a QuikChange Lightning Kit (Agilent Technologies) according to manufacturer's instructions with either the *lpcA* or *groESL* pSMART-LCK construct (Lucigen) as the template. Primers were designed to introduce point mutations as

follows: *lpcA*(E65Q) using TGCACTTTGCCGAACAGTTGACCGGTCGCTACCG and its complement sequence; *groES*(M1R) using CTCAAAGGAGAGTTATCACGGAATATTCGTCCATTGCATGATCG and its complement sequence; and *groEL*(M1R) with AAGGAATAAAGATACGGGCAGCTAAAGACG and its complement sequence. Growth studies were prepared as done for growth curve analysis, with the OD_{600} readings measured at 20 hours. Percentage improvement, compared to blank vector control, was used for comparison of the clones (n = 3).

3.2.9 – Statistical analyses

Sample averages were calculated for all phenotypic analyses and are plotted and reported with ± one standard error. Student's t-test was used to calculate one-tailed *p*-values. Values are reported within the text with ± one standard error.

3.3 – Results and Discussion

3.3.1 – Application of SCALEs method to identify furfural tolerance genes

SCALEs libraries containing $>10^5$ clones were selected on solid minimal medium with 0.75g/l furfural (Fig. 3.1A). Libraries cultured on minimal medium plates with no furfural served as the control in order to account for growth on minimal medium alone. The selection was performed on plates to provide a microenvironment where clones were spatially isolated, in an effort to remove population effects (e.g., decreased local furfural concentration due to increased reduction by certain clones) that might interfere with assessing individual clone fitness [66]. Colonies were harvested from plates after growth appeared (one day for control and three

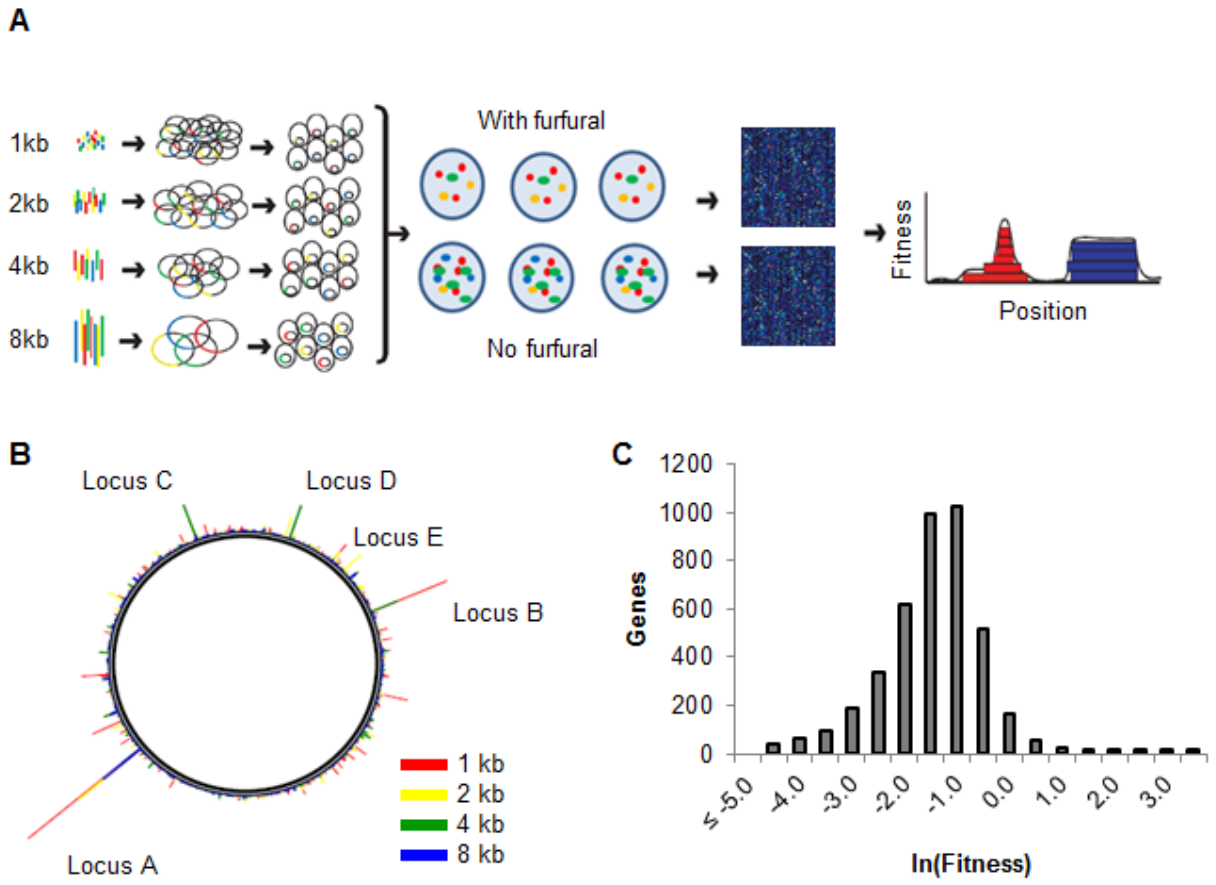


Figure 3.1 - Overview of furfural selection and SCALEs analysis. A) 1, 2, 4, and 8 kb fragments were prepared from *E. coli* genomic DNA and ligated into pSMART-LCK vector. Each sized genomic library was transformed into BW25113 Δ*recA*::FRT host cells, recovered, mixed together, and then grown on minimal medium plates (control) or solid minimal medium with 0.75 g/l furfural. Cells were harvested from the plates and microarrays (square boxes) were run with plasmid extracts from both the furfural and control plates in order to determine individual gene fitness scores (W). The fitness vs. position plot illustrates how different clones (stacked rectangles) can contribute to an individual gene's fitness. The red "triangle" has contribution from various sized clones, but is centered around a specific locus, whereas the blue "rectangle" represents a high fitness score from the presence of one single sized clone (e.g., requiring a large operon where smaller library sizes would not be found). B) Genome plot depicting clone fitnesses for the different library sizes. Loci corresponding to the top gene fitness scores are labeled. C) Histogram of log-transformed gene fitness scores, where increased fitness corresponds to $\ln(W) > 0$.

days for furfural treatment) and plasmids were extracted and analyzed with microarrays to determine clone concentration at approximately 125 bp resolution (Fig. 3.1B). A fitness score was calculated for each gene to determine those that were differentially enriched with furfural selection. High-fitness genes were identified across the entire genome and more than one size of library insert contributed to loci with the highest fitness scores (see Fig. 3.2 and Appendices 8.1-5).

A total of 268 genes, or ~6% of all *E. coli* genes, were enriched through selection (Fig. 3.1C), indicating that a strong selective pressure was applied (all genes with increased fitness during furfural selection are provided in Table S2). Using the Batch Genes program [134], we analyzed the increased fitness genes by Gene Ontology (GO) terms and found that significantly enriched terms were primarily associated with cell membrane (e.g., enterobacterial common antigen) and wall (e.g., peptidoglycan) biosynthetic processes, suggesting that membrane and wall formation are important for furfural tolerance (Fig. 3.3). No cellular component or molecular function GO terms were significantly enriched.

3.3.2 – Confirmation of furfural tolerance

Based on the gene-specific fitness scores (Appendix 8.6), we determined that the top 19 genes mapped to only five distinct loci (labeled A-E, Fig. 3.1B). Visual inspection of the clone fitness patterns associated with each loci suggested specific genes that were the primary (or sole) contributor towards fitness (Fig. 3.2 and Appendices 8.1-5). We then constructed individual clones for each of the hypothesized fitness-contributing gene(s) from the top five loci (Table 3.2): locus A (*thyA*), locus B (*ybiY*), locus C (*groESL*), locus D (*lpcA*), and locus E (*ybaK*).

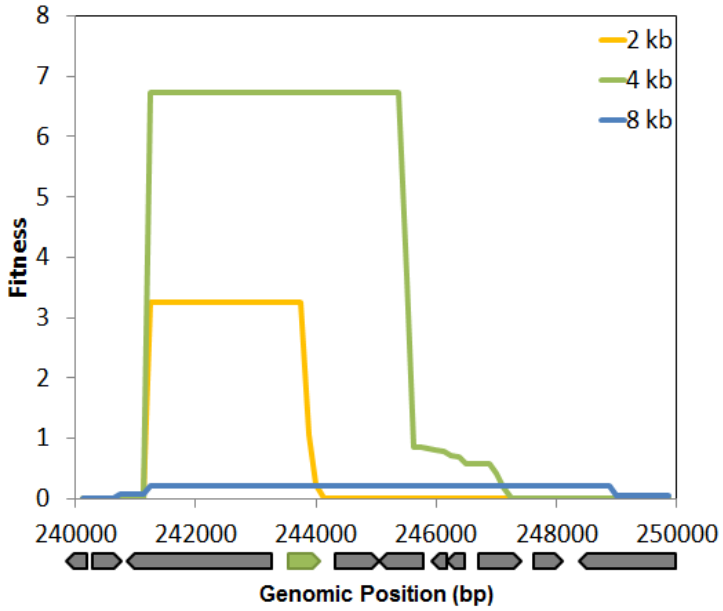


Figure 3.2 - Genomic position alignments for gene fitness assignments. SCALEs clone fitness scores from the 2, 4, and 8 kb libraries are based on clone frequency with and without selective pressure. The *lpcA* gene is shown in green, with neighboring genes shown in gray.

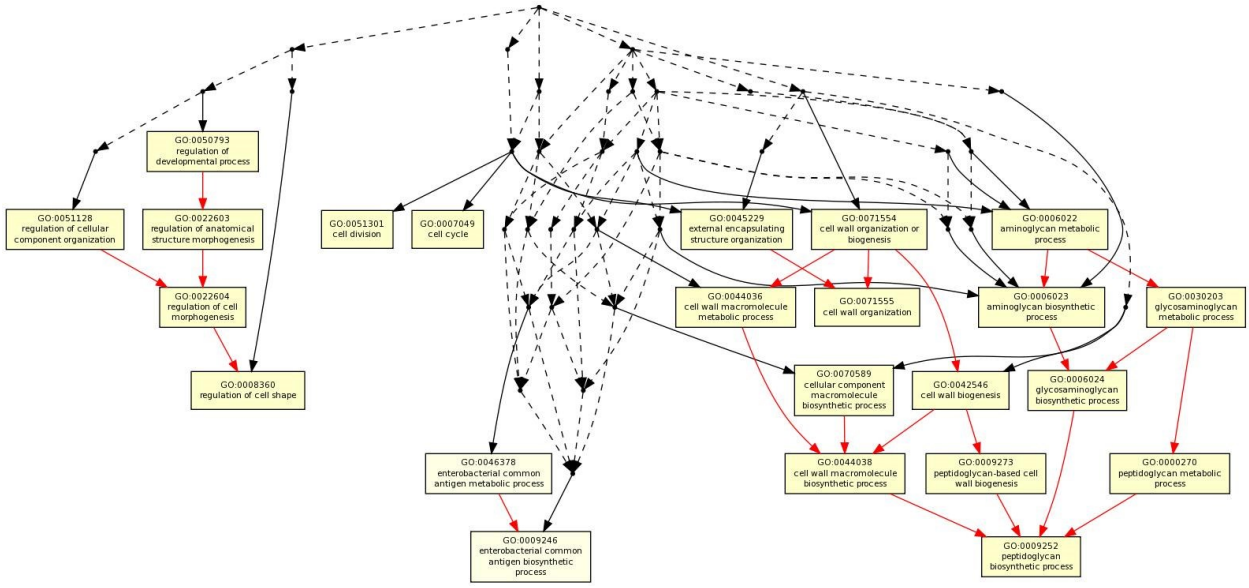


Figure 3.3 - Enriched biological processes GO terms in SCALEs selection. Yellow boxes represent significantly enriched GO terms and non-significant terms are condensed to nodes. Red arrows connect two significantly enriched GO terms, whereas black arrows connect a non-significantly enriched term (node) to a significantly enriched term (yellow box). Analysis was performed with the Batch Genes GOEAST online tool as described in the text.

Table 3.2 - Gene(s) cloned for confirmation studies.

Rank	Gene	Fitness	Locus	Function
2	*ybiY*	20.2	B	Predicted pyruvate formate lyase activating enzyme
6	*thyA*	14.8	A	Thymidylate synthase
8	*groEL*	13.5	C*	GroEL chaperone
12	*lpcA*	10.5	D	D-sedoheptulose 7-phosphate isomerase
17	*groES*	7.5	C*	GroES chaperone
19	*ybaK*	7.1	E	Cyc-tRNA^{Pro} and Cyc-tRNA^{Cys} deacylase

**groESL* operon was cloned into a single plasmid

We first attempted to confirm tolerance of the hypothesized fitness-contributing gene(s) under the same conditions used in our growth selections (i.e., improved growth on solid minimal medium with furfural). Cultures of each of the five clones were streaked onto solid medium supplemented with furfural at 0, 0.75 g/l, or 1.5 g/l (corresponding to 0, 1 and 2x selection concentrations). Growth was monitored for three days, consistent with the time of furfural selection. At both furfural treatment levels, growth appeared first from *thyA*, followed by *lpcA* and *groESL* clones (Fig. 3.4). Clones overexpressing *ybiY* or *ybaK* were not observed to confer improved tolerance compared to vector control and were thus removed from further study. Based on our previous experience with SCALEs [78,83,101,113,124-130], we expect that the lack of observed tolerance phenotypes from *ybiY* and *ybaK* is likely due to these genes requiring other genes in the enriched loci, although we cannot eliminate the possibility that they were false positives [127].

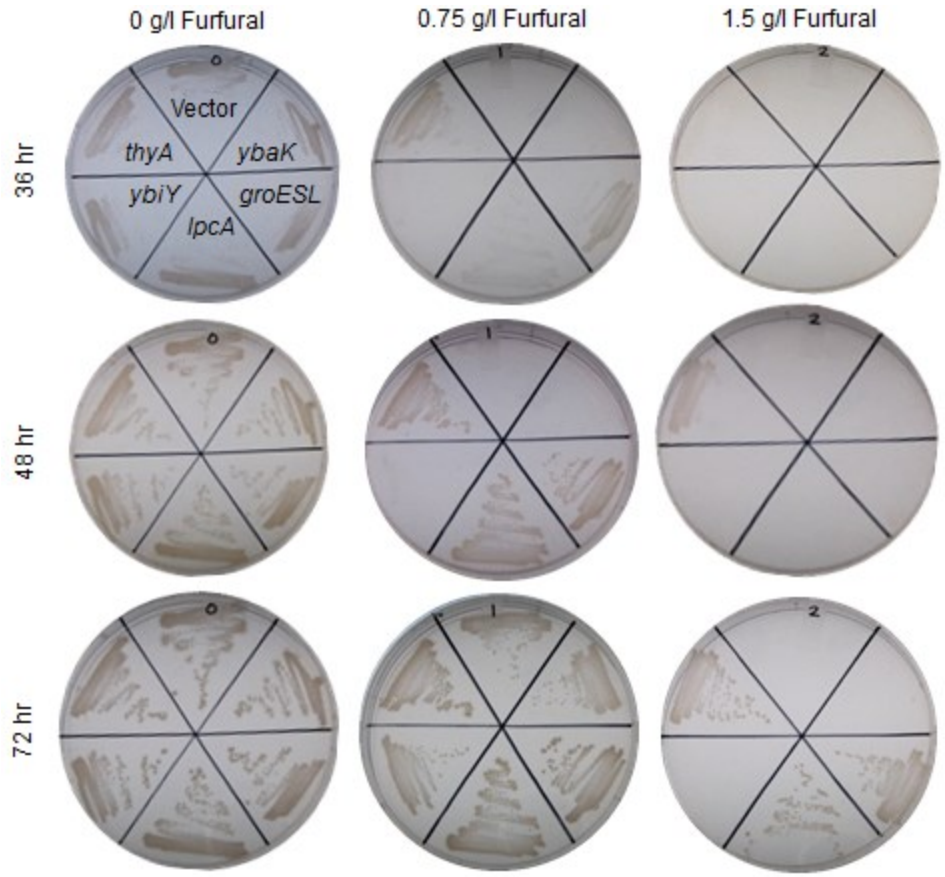


Figure 3.4 - Plating assay of hypothesized tolerance-conferring clones identified in SCALEs selection. Cells (10^4) were streaked onto solid minimal medium with 0, 0.75, or 1.5 g/l furfural (0, 1, or 2x selection pressure) and growth was observed for 72 hours.

We next tested each confirmed tolerance clone for improved growth in planktonic cultures. Growth curves of *thyA*, *lpcA*, and *groESL* overexpression clones were performed and we observed improved growth from all three strains tested (Fig. 3.5). Interestingly though, *thyA*, which was the first strain with visible growth on the solid medium with furfural (Fig. 3.4), had a longer lag phase than the *lpcA* clone, which was the first clone to leave lag phase in planktonic cultures. Additionally, both the *groESL* and *lpcA* clones had higher density at 24 hours than the *thyA* clone or the empty vector control, at which point we stopped sampling due to the complete disappearance of furfural.

ThyA, LpcA, and GroEL-ES are involved in relatively distinct cellular processes. Thymidylate synthase, encoded by *thyA*, catalyzes the conversion of dUMP to dTMP during *de novo* pyrimidine biosynthesis. ThyA overexpression has previously been observed to confer furfural tolerance [66], presumably by increasing dTMP availability for increased DNA repair suspected to occur during furfural treatment.

The isomerase encoded by *lpcA* catalyzes the first committed step in lipopolysaccharide (LPS) core biosynthesis by routing a pentose phosphate pathway (PPP) metabolite, D-sedoheptulose 7-phosphate, towards heptose formation and subsequent incorporation into the inner core region of LPS. Functional LPS formation is widely documented as important for tolerance to hydrophobic compounds [139-141]. Also, the PPP is a major source of NADPH in *E. coli*, and increased upper pathway flux through this pathway (to make up for losses due to increased LPS synthesis) could lead to increased NADPH formation, limitations of which are thought to play an important role in furfural toxicity [12,47,56-59,121]. Previous studies for furfural tolerance targets have not previously identified *lpcA* or LPS formation, but previous SCALEs studies from our laboratory have identified *lpcA* as a highly enriched locus in acetate

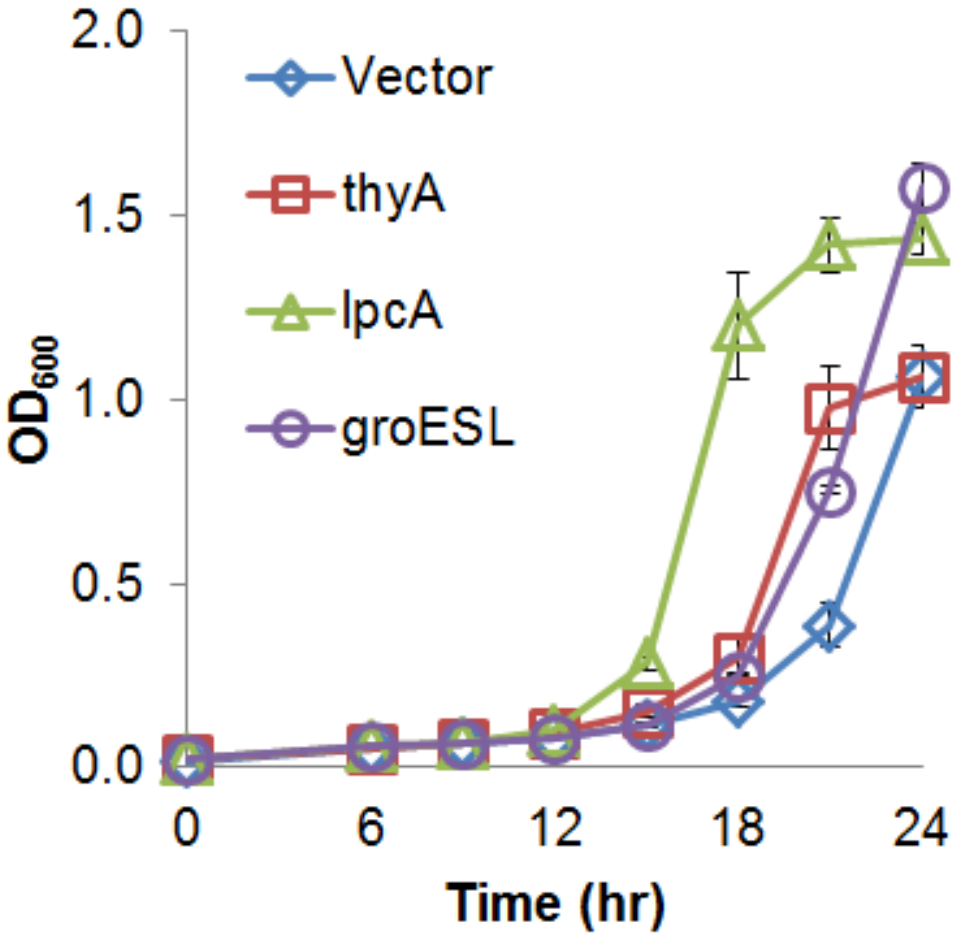


Figure 3.5 - Growth curve analysis of tolerant clones in 0.75 g/l furfural. Seed cultures were inoculated into furfural at the same initial density and grown for 24 hours. Optical density was recorded every 3-6 hours. Error bars represent one standard error (n=3). Double asterisks denote $p < 0.05$.

and ethanol selections, where *lpcA* overexpression was confirmed to improve ethanol tolerance several fold. [78,83].

The GroEL-ES chaperonin complex, encoded by *groESL*, is essential for cell growth under a range of temperatures [142], is required for proper folding of some essential proteins [143], and is a well-known stress associated protein [86,144,145]. Moreover, overexpression of *groESL* has been found to confer ethanol and butanol tolerance [85,86].

Given the varied functions encoded by these furfural tolerance genes, and the common reduced lag phase observation, we sought to better understand if these genes were conferring tolerance through previously implicated physiological mechanisms. Specifically, we assessed the effect of overexpression of each of these genes on furfural reduction and DNA mutation rates.

3.3.3 – Increased furfural reduction from lpcA overexpression

Furfural is known to be reduced to the less toxic furfuryl alcohol in *E. coli* [46,47]. This reduction has been primarily linked to the action of a low K_M NADPH-dependent oxidoreductase encoded by *yqhD* [121]. It is thought that the increased oxidation of NADPH required for furfural reduction limits the availability of NADPH reducing equivalents that are required for key biosynthetic reactions like sulfur assimilation [56] and nucleotide synthesis [66]. Indeed, for our fastest growing strain in liquid culture, *lpcA*, we measured 32 ± 10% increase in furfural reduction rate compared to control (Fig. 3.6A). This observation is consistent with our speculation that increased flux through the PPP could lead to elevated NADPH flux and thus increased reduction rates. Neither the *thyA* or *groESL* clones were observed to alter furfural reduction rates.

Furfural was found to induce a significant lag phase longer than cells grown without furfural treatment, which is traditionally linked to the aforementioned NADPH starvation concomitant with furfural treatment [46]. Despite a substantial lag phase, growth during furfural reduction was observed in our growth curve assessments (Fig. 3.5). Approximately 60-70% of the furfural still remained after 12 hours, roughly coinciding with the onset of exponential phase. All strains had reduced virtually all of the furfural within 20-24 hours (data not shown).

Assessing the redox state and furfural tolerance of cells overexpressing *lpcA*, and other enzymes related to PPP flux, could be a potential path for future research to complement the transhydrogenase overexpression approach recently used [12,56]. This approach could serve as an alternative to strategies directed at replacing NADPH-dependent oxidoreductase reduction with NADH-dependent oxidoreductases [12,57,58].

3.3.4 – Tolerance genes do not alter DNA mutation frequency

Since furfural is a known DNA mutagen [63,64,120], we hypothesized that our furfural tolerance genes might affect DNA mutation frequencies and thereby lead to tolerance. The mutation frequency was measured by treating cell cultures with furfural and then plating with rifampin to measure the number of spontaneous mutants, compared to total viable cells (Fig. 3.6B) [136,137]. Surprisingly, no clones exhibited significantly altered DNA mutation frequency from control ($p > 0.05$ for all). Although the *groESL* clone did appear to increase mutation frequency ~3-fold, statistical analysis indicated that this increase was not significant ($p > 0.08$).

We had hypothesized that we would observe altered DNA mutation frequency for the *thyA* clone based on its presumed role of increasing dTMP availability required for DNA repair

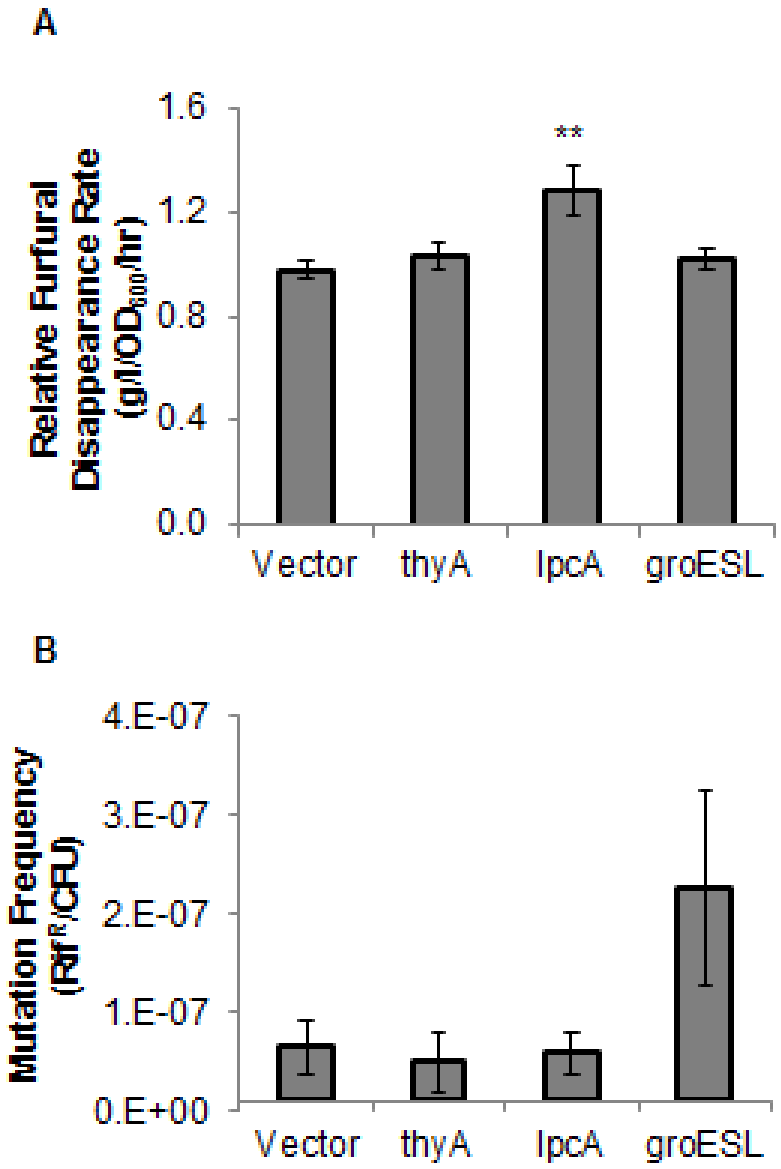


Figure 3.6 - Phenotypic analysis of tolerant clones for furfural reduction and DNA mutation frequency. A) Samples were collected for measuring furfural in growth curve cultures with 0.75 g/l furfural initial concentration. Furfural concentrations were normalized to cell number (optical density) for each value and disappearance rate was calculated during the transition from lag to exponential phase (n=3). B) Frequency of rifampin resistance of cells treated with furfural (n=4). Error bars represent one standard error. Double asterisks denote $p < 0.05$.

under furfural treatment [66] and for the *groESL* clone due to the chaperone's role in stress response and its ability to stabilize mutated proteins [146]. It is possible that the level of furfural treatment here did not deplete DNA repair pathways enough in order to elicit an observable difference, although previous studies have also indicated that furfural treatment does not always elevate mutation frequencies beyond what native repair mechanism can handle [68]. It is also worthwhile to note that ThyA is involved in formyl-tetrahydrofolate biosynthesis (converting THF to 5,10-methylene-THF during the dUMP to dTMP reaction), which is a pathway previously associated with tolerance to acetate [83] and 3-hydroxypropionic acid [126], and thus might suggest a more general role for ThyA in chemical tolerance beyond pyrimidine biosynthesis and DNA repair. In the case of the *groESL* clone, our data suggest that any role GroESL has in stabilizing mutations that might arise from furfural treatment is not significant, which suggests that GroESL may rather be acting to stabilize wild-type proteins whose function or formation is altered in the presence of furfural.

3.3.5 – Validation that lpcA and groESL overexpression confer tolerance

Because *lpcA* and *groESL* have not been previously identified to confer furfural tolerance, we aimed to verify that our plasmid constructs resulted in increased transcription for the targeted genes. Transcript levels were observed for *lpcA* to be 98 ± 24 fold-increase over the control strain. The *groESL* construct had increased expression of its *groES* and *groEL* genes of 150 ± 86 and 126 ± 48 fold, respectively.

While this data confirmed increased expression from the plasmid based constructs, we further wanted to verify that tolerance was conferred by functional expression of LpcA or GroESL. To do so, we introduced a missense mutation into the coding sequences of each of these genes. For *lpcA,* we targeted a residue in the active site with the E65Q mutation, which has previously been reported to confer undetectable enzymatic activity [147]. For *groESL,* we replaced the start codon (ATG with CGG for an M1R mutation) of *groES* or *groEL*. When tested for growth in 0.75 g/l furfural, the *lpcA* plasmid conferred 429 ± 7% improvement in growth over blank vector control, whereas *groESL* conferred 111 ± 4% improvement (Fig. 3.7). The missense mutation clone *lpcA*(E65Q) conferred a slight improvement in tolerance (68 ± 12%; $p < 0.05$), which could be a result of low enzymatic activity levels below the threshold of activity of the previous assay [147], but is markedly below the improvement conferred by the wild-type sequence. Additionally, the M1R missense mutation in *groES* conferred no difference in growth compared to blank vector ($p > 0.1$), and the M1R missense mutation in *groEL* conferred a decrease in growth (reduction of 30 ± 2%). Taken together, our data suggests that at the expression levels conferred by expression on the pSMART-LCK vector rely on functional expression of the enzyme LpcA enzyme or GroESL complex in order to confer tolerance to furfural.

3.4 – Conclusions

Much research has been performed over the past decade to uncover mechanisms of furfural toxicity and to engineer furfural tolerance in *E. coli* [12,33,46,47,56-59,66,70,121].

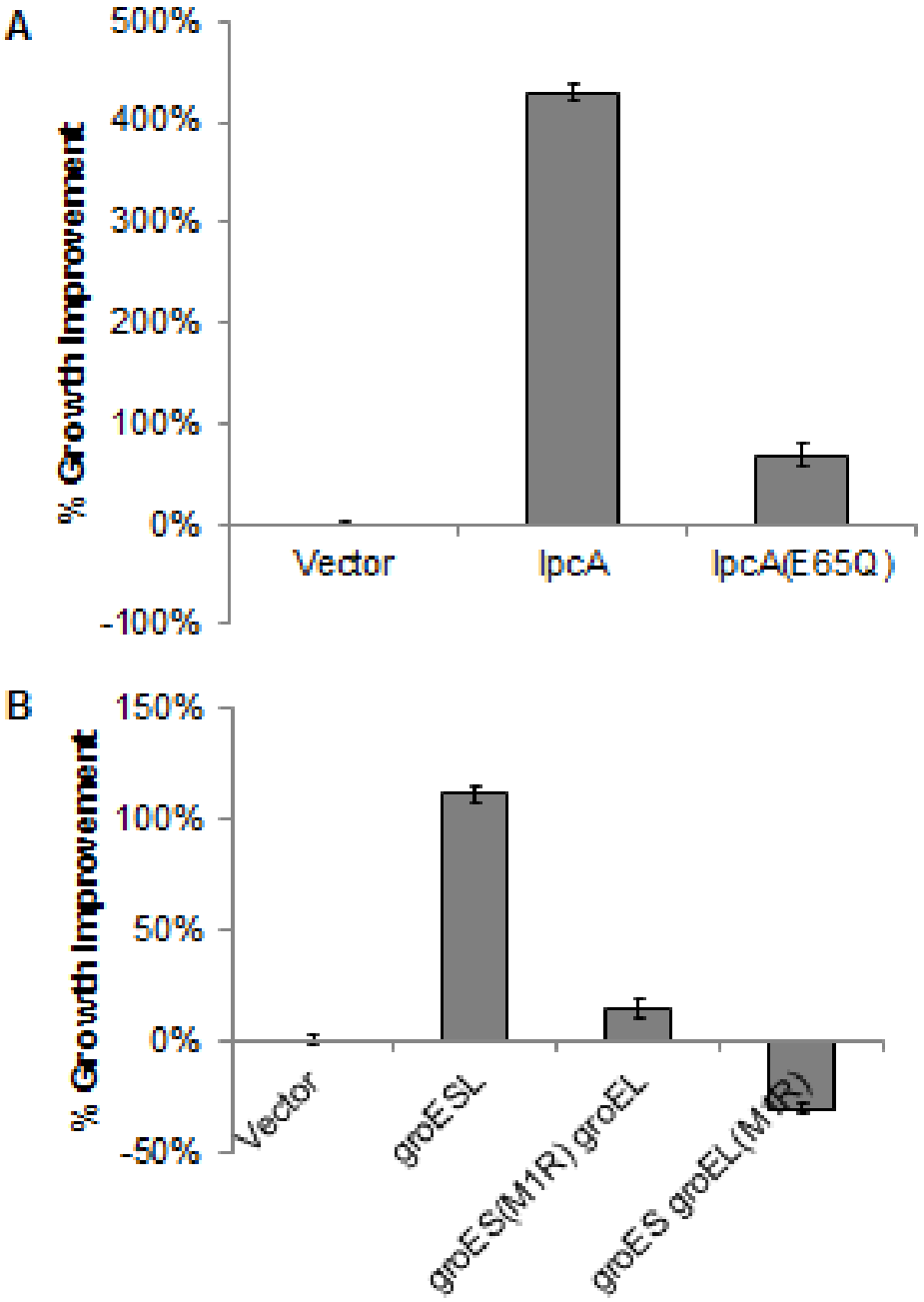


Figure 3.7 - Mutational studies on *lpcA* and *groESL* clones. Mutations were introduced onto the plasmids within the coding sequence targeted for (A) *lpcA* or (B) *groESL*. Cultures were grown in 0.75 g/l furfural for 20 hr. (n = 3; error bars represent standard error). Percentage improvement was calculated as the difference of the test strain subtracted from the control, divided by the control.

Here, we used the SCALEs method [113] to not only map fitness effects across the entire *E. coli* genome, but also to identify and confirm both novel (*lpcA* and *groESL*) and previously identified (*thyA* [66]) furfural tolerance genes. We determined that overexpression of *lpcA* increased observed furfural reduction. LPS core formation, for which LpcA plays a part, is vital for tolerance to chemical inhibitors [78,83,139-141]. To this end, analysis of GO term enrichment from our high-fitness genes suggests that membrane and wall biosynthesis is important for furfural tolerance. Alternatively, *lpcA* overexpression may increase flux through the PPP and thereby increase NADPH availability for furfural reduction. Overexpression of *groESL* also conferred increased growth, but did not alter the rate of furfural reduction or mutation frequency. It is possible that furfural elicits responses similar to those from solvent stress, where *groESL* overexpression has been shown to confer tolerance [86]. ThyA overexpression did not alter DNA mutation frequency even though it has previously been implicated in increasing DNA repair under furfural stress [66].

Robust microbes for lignocellulosic biofuel production must be engineered for multiple functions—production of a desired product, tolerance to feedstock and product, co-utilization of feedstock carbon sources—that all work in concert together. Our study here expands the understanding of furfural tolerance genes and thus provides additional targets for engineering furfural tolerance. Ultimately, finding genetic manipulations that are beneficial to multiple biocatalyst functions will enable rapid, reliable, and improved biofuel production in the future.

Chapter 4

Use of TRackable Multiplex Recombineering for Identification of Furfural Tolerance Genes

In preparation for submission to *Biotechnology and Bioengineering*

Authorship: Mills, T.Y.
Gillis, J.H.
Sandoval, N.R.
Gill, R.T.

4.1 – Introduction

Enabled by the decreasing price of DNA synthesis and sequencing, library generation is now reaching up to 4 billion designed mutants/day [103]. "Omics"-driven tool development and use is an important approach to select or screen for the best performing strains within these large populations. A variety of tools have been developed, including an *E. coli* knockout collection [131], genomic library strategies [102,113], global transcription machinery engineering [7,89], recombineering-enabled technologies multiplex automated genome engineering [103], as well as CRISPR-based strategies [148]. With the ever-increasing knowledge of contextual-dependence on expression, as is the case for plasmids by varying copy number or selection marker [149] or multiple-cloning site usage [150], efforts have been made to standardize expression through controlled, or context-free expression design [151,152].

In Chapter 3, we applied the SCALEs method to map genotype-to-phenotype relationships of increased gene dosage towards conferring furfural tolerance. Here, we apply another tool developed in the Gill Laboratory, TRackable Multiple Recombineering (TRMR)

[71]. The TRMR approach uses chromosomal mutations to confer increased ('Up') or decreased ('Down') expression by way of an integrated cassette containing the $P_{LtetO-1}$ promoter and RBS ('Up') or displacing these native sequences ('Down') in front of virtually every gene in *E. coli*. The mutations are barcoded, allowing for clone concentration to be tracked throughout selection with either microarray or next-generation sequencing technology. Allele fitness scores are calculated based upon clone frequency before and after selection; genes have two fitness scores from TRMR, one value for the 'Up' mutant and another score for the 'Down.'

Here, we highlight the differences in enrichment patterns from three furfural selection using TRMR, as well as compare increased fitness genes from a TRMR 'Up' selection to those identified during the SCALEs selection in Chapter 3. We identified a subset of genes that had increased fitness in both the SCALEs and TRMR plate-based selections, cloned each identified gene, and tested for improved furfural tolerance. We identified two novel clones with increased growth in furfural. Surprisingly though, when we compared plasmid-based clones to their TRMR 'Up' counterparts, we found that tolerance was not conserved (i.e., a TRMR clone was tolerant for two additional genes, but the plasmid overexpression strains for those genes were not). We measured expression from these clones and discuss the impact of expression on identifying furfural-tolerance genes, as well as the broader impact that differences in genome search strategies do not necessarily identify the same alleles.

4.2 – Materials and Methods

4.2.1 – Bacterial strains, plasmids, and media

E. coli MG1655 and derivatives thereof were used as hosts for studies. Primers were designed for cloning to include the native promoter and open reading frame. Genes were amplified from MG1655 genomic DNA extracted with a Genomic DNA Extraction Kit (Qiagen). Primers were designed to amplify the promoter and ORF of the gene because pSMART-LCKan is a promoterless vector. In the cases of *ninE*, *rimM*, *ybgF*, and *yfiB*, the genes fall within an operon, so the entire operon and native promoter sequence was cloned. Cleaned amplified PCR product was ligated into pSMART-LCKan using the CloneSmart LCKan Blunt Cloning Kit (Lucigen). Ligation products were transformed into NEB Turbo Competent *E. coli* (New England BioLabs). Plasmids were then harvested and transformed into MG1655 cells that were made electrocompetent by serial washing with ice-cold water. Confirmation of plasmids from individual colonies was performed by agarose gel electrophoresis and sequencing. TRMR 'Up' mutants are derivatives of *E. coli* MG1655 with a synthetic cassette containing a blasticidin resistance gene, the $P_{LtetO-1}$ promoter, and RBS sequence [71]. The *ahpC_Up* strain was created and confirmed previously by Warner et al [71]. The *yhjH_Up* strain was created in the Gill laboratory by R.I. Zeitoun and S.A. Lynch according to the methods prescribed by Warner et al.

Luria-Bertani (LB) broth was used for routine applications. All growth assessments were performed in MOPS minimal medium [133]. Kanamycin was used where appropriate (30 μg/ml). Cultures were grown at 37°C. Furfural was added to media at the concentrations indicated within the text.

4.2.2 – Furfural TRMR selection

TRMR overexpression ('Up') and decreased expression ('Down') libraries were prepared by Warner et al. [71]. Libraries were combined and spiked with JWKan, a neutral-mutation

barcoded mutant [71] at 1:400 ratio of a single barcode, and an initial sample (10^9 cells) was harvested by centrifugation and stored at -80°C. The mixed culture was applied at 10^4 cells per plate to 20 MOPS minimal medium plates with 0.75 g/l furfural. Cells were harvested after growth appeared, pelleted, and frozen until further processing. TRMR barcodes were amplified from genomic DNA, cleaned, prepared for microarray analysis with control barcode loading, and analyzed with microarrays as previously described [71]. Analysis of the resulting data files to calculate allele fitness was performed as previously described [71,82]. A standard curve was created, relating microarray signal to DNA concentration, as determined through the use of the control barcodes. From this curve, allele fitness (W) was calculated as the frequency of a given allele, i, after selection divided by the frequency of a given allele before selection, $W = frequency_{i,selection}/frequency_{i,initial}$. Gene ontology analysis was performed with the Batch Genes online program [134].

4.2.3 – Growth analyses

Overnight cultures were inoculated from freezerstocks and grown overnight in LB medium. Seed cultures in MOPS minimal medium [133] were inoculated with 2 v/v% overnight culture and grown approximately 2 hours. Seed cultures were normalized to OD_{600} = 0.195 - 0.200, and used to inoculate test cultures at 10 v/v %. Test cultures were grown in MOPS minimal medium and 1.5 g/l furfural. Optical density readings were monitored at 600 nm at the timepoints indicated.

4.2.4 – Gene expression analysis

Cultures were prepared similarly to growth analyses, with the exception of being inoculated into MOPS minimal medium without furfural. Mid-exponential cells were harvested

by centrifugation, flash frozen, and used for RNA extraction with an RNAeasy Plus Mini Kit Qiagen) after being treated with RNAprotect Cell Reagent (Qiagen). Expression analysis was performed using the iTaq Universal SYBR Green One-Step Kit (Bio-rad). The gene *cysG* was used as a housekeeping reference gene to account for differences in sample loading.

To calculate translation initiation rate of the different transcript sequences, we used V 1.1 of the RBS Calculator in reverse engineering mode. Reported translation initiation rates are reported for the gene's annotated start codon.

4.3 – Results and Discussion

4.3.1 – Identification of TRMR gene fitnesses

For our TRMR selection, we continued to use a plate-based selection with furfural added to solid medium to control an individual clone's microenvironment based upon the previous success we had in Chapter 3 for identifying furfural-tolerant alleles. In addition, we aimed to compare furfural enrichment patterns for TRMR clones with furfural treatment dosed according to regimens used by selections in hydrolysate [71] and acetate [82]. The hydrolysate and acetate selections employed two different regimens for liquid serial batch selections. The hydrolysate selection was performed by serial transfer into decreasing concentrations of hydrolysate (i.e., batch 1 at 20% hydrolysate, batch 2 at 19% hydrolysate, batch 3 at 18% hydrolysate) in order to increase selection specificity and sensitivity [127]. The acetate selection was performed in serial batches of constant concentration (all prepared at 16 g/l acetate). We performed three furfural selections in parallel to mimic these previous schemes: plate-based, decreasing furfural concentrations in each batch, and constant initial furfural concentration in each batch (Fig 4.1 A).

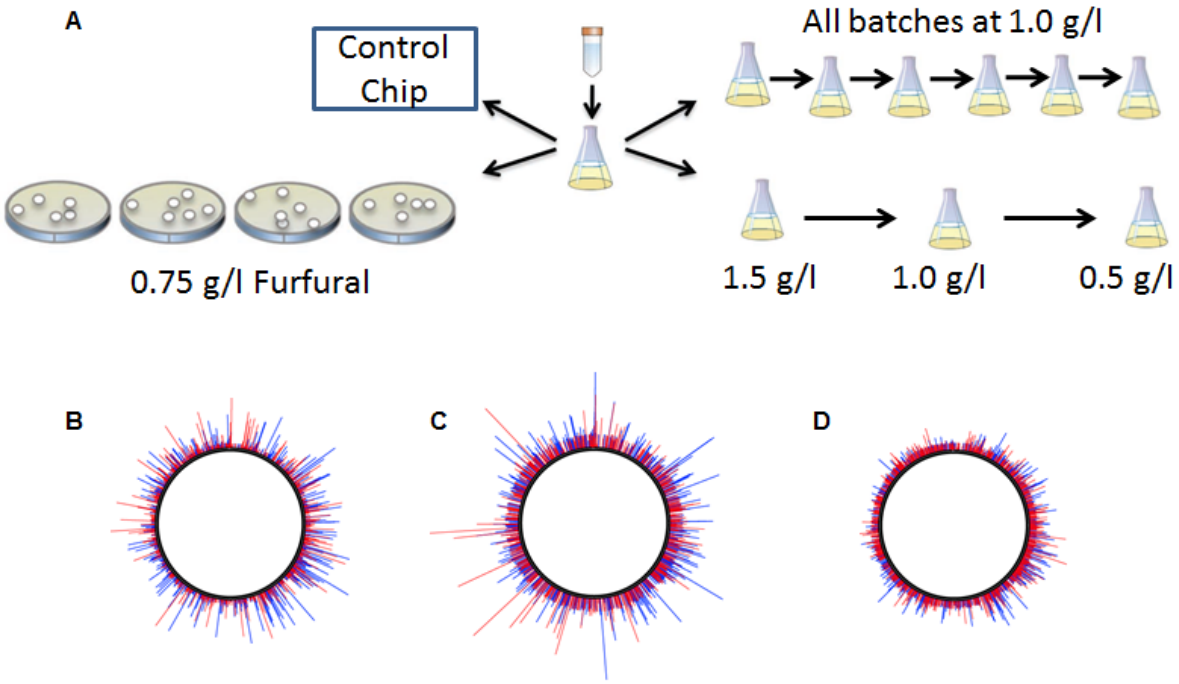


Figure 4.1 - TRMR furfural selection under various exposure regimens and resulting gene fitnesses. A) Three different schemes were performed for furfural selections: plate-based, constant initial concentration (1.0 g/l furfural) in liquid serial batches, and decreasing furfural concentration at each batch transfer (1.5 g/l to 1.0 g/l to 0.5 g/l). B-D) Log-transformed gene fitness scores for the three selections are plotted with reference to their position on the genome presented in a clockwise fashion. Red lines indicate 'Up' fitness values and blue indicate 'Down' fitnesses for B) plate-based, C) decreasing concentration, and D) constant concentration furfural selections. For reference, the maximum ln(fitness) score in (C) for the 'Up' library is 11.9 (more details on summary statistics are provided in Table 4.1).

Gene fitnesses were calculated according to clone frequency change through selection (Fig. 4.1 B-D; lists of the top 100 gene fitnesses for each selections are provided in Appendix 8.7-12). Based on the data visualized on the circle plots, we noticed that the plate-based and decreasing concentration selection fitness scores had larger amplitudes than the constant concentration selection. Summary statistics of the three selections are listed in Table 4.1. The maximum fitness values for the constant concentration were roughly half those of the other selections, yet the total number of genes with increased fitness was highest from this selection. Using the fitness calculated from the control barcode in JWKan, a TRMR clone with the cassette inserted in a neutral site that does not affect *E. coli* phenotypes, we found that the decreasing furfural selection was the strongest (i.e., lowest ln(fitness) value for JWKan), followed by the plate-based selection, and finally the constant concentration selection being the weakest (i.e., highest ln(fitness) value of the control tag). The fitness patterns between the plate-based and decreasing concentration selections were more comparable to each other than the plate-based selection and the constant concentration, as visualized by the trends of the ln(fitness) values plotted together (Fig. 4.2). Together, these findings together suggest that the decreasing concentration selection and plate-based selection are alternative but complementary methods to creating strong selective pressures.

Table 4.1 – Summary statistics of furfural TRMR selections.

	Plate-based		**Decreasing Conc.**		**Constant Conc.**	
	Up	**Down**	**Up**	**Down**	**Up**	**Down**
# Increased Fitness Genes	745	684	666	671	1059	873
Maximum Increased ln(Fitness)	6.7	8.4	11.9	10.9	5.9	5.2
JWKan ln(Fitness)	-3.3		-5.7		-1.1	

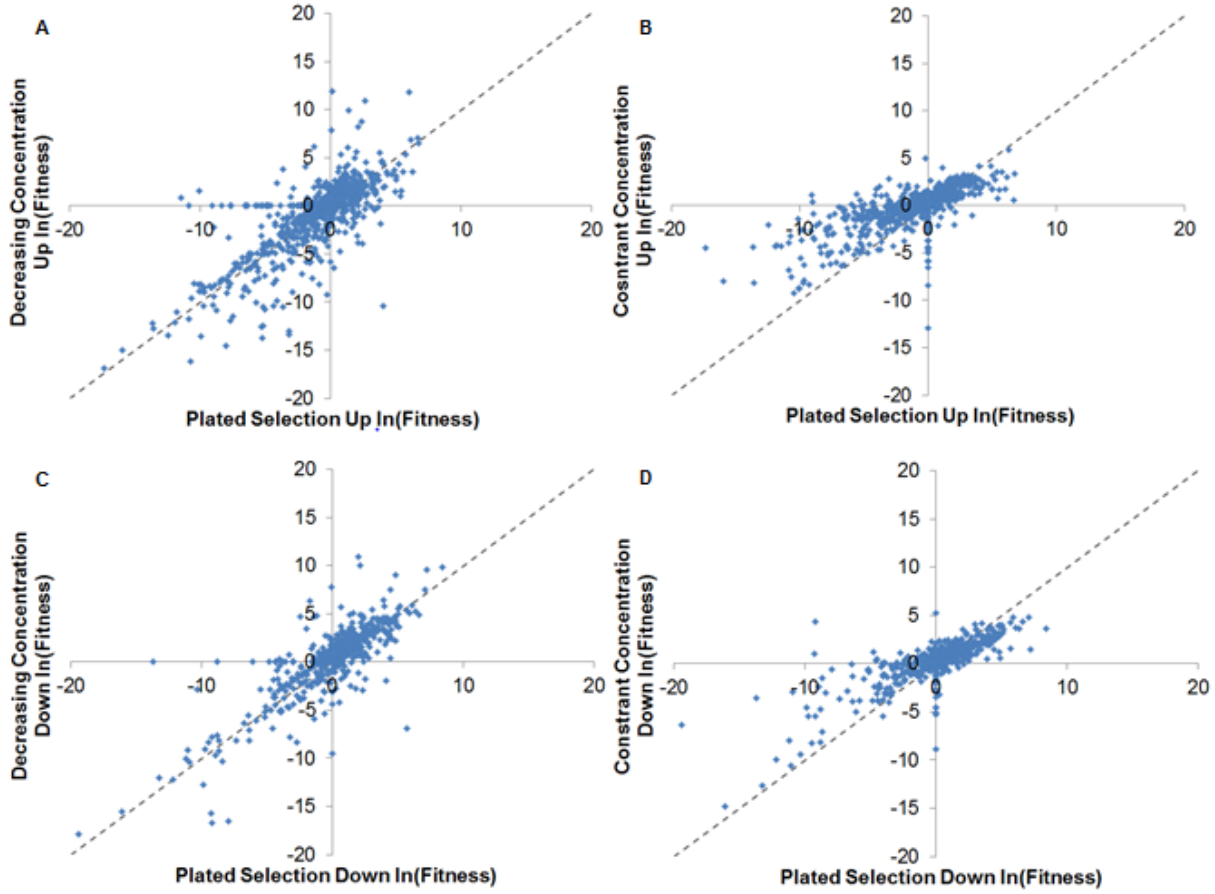


Figure 4.2 – Comparison of gene fitnesses from batch transfer furfural TRMR selections to the plate-based selection. A-B) 'Up' ln(fitness) values and C-D) 'Down' gene ln(fitnesses). The dotted line is y = x as a linearity reference. R^2 values for plots are A) 0.60, B) 0.51, C) 0.64, and D) 0.60. Quadrant 1 genes represent increased fitness genes (enriched) under both selections.

Interesting to note is that some common genes were identified through all three selections (Table 4.2, Appendices 8.7-12). Specifically, *csiD*, was the only 'Up' gene identified as a top 10 enrichment in all three furfural TRMR selections. CsiD is a predicted protein that is induced under carbon starvation [153], but is otherwise of unknown function. Genes appearing within the two strongest selections include 'Up' mutations for *ydaL, yeeN, talB, smg*, totaling 5 of the top 10 enrichments from these two selections. All of these genes, with the exception of *talB*, which is involved in the pentose phosphate pathway (PPP), are conserved or predicted proteins, limiting the hypotheses to be developed towards testing their role in furfural tolerance. TalB, with its function in the PPP, could potentially play a role in maintaining flux through the PPP in order to restore NADPH during constitutive furfural reduction, similar to a potential tolerance mechanism for *lpcA* overexpression we presented in Chapter 3 (LpcA coverts one of the same metabolites as TalB). Mutually shared 'Down' mutations are *nagA* and *ydiA*. NagA is an *N*-acetylglucosamine-6-phosphate deactylase, which produces glucosamine-6-phosphate [154], a key metabolite in the production of lipopolysaccharide. YdiA, also known as PpsR, is a phosphoenolpyruvate synthetase regulatory protein [155], which regulates genes involved in growth on three carbon substrates. None of these genes have previously been implicated in furfural tolerance, so our data here present potential new targets for engineering furfural tolerance.

We also performed gene ontology analysis [134] to look for significantly enriched terms as we had done in Chapter 3. The only significant terms found in all three domains (cellular component, biological process, and molecular function) from the plate-based selection were related to flagella/motility (Fig. 4.3-5). No GO terms were found to be significantly enriched for the 'Down' fitness scores.

Table 4.2 – Top 10 genes enriched in TRMR furfural selections.

	Plate-based		Decreasing Conc.		Constant Conc.	
Rank	**Up**	**Down**	**Up**	**Down**	**Up**	**Down**
1	*csiD*	*ddpF*	*yqcC*	*fruK*	*ydaL*	*aspS*
2	*talB*	*nagA*	*smg*	*yaaX*	*dhaK*	*ydiA*
3	*ydaL*	*ydiA*	*yihY*	*ddpF*	*ycjU*	*sdaA*
4	*yeeN*	*ycjN*	*yqhD*	*nagA*	*tonB*	*ycjN*
5	*smg*	*rseB*	*yhdE*	*nagC*	*ddpA*	*ygcP*
6	*yneG*	*yohN*	*zipA*	*ybjL*	*yneG*	*yecT*
7	*yjaH*	*sdaA*	*alaS*	*pldB*	*rplI*	*frmR*
8	*acpT*	*yecT*	*talB*	*ydiA*	*csiD*	*sra*
9	*rplI*	*yahF*	*yeeN*	*ilvM*	*ydiJ*	*paaI*
10	*yccF*	*potE*	*csiD*	*pdhR*	*yohK*	*fliO*

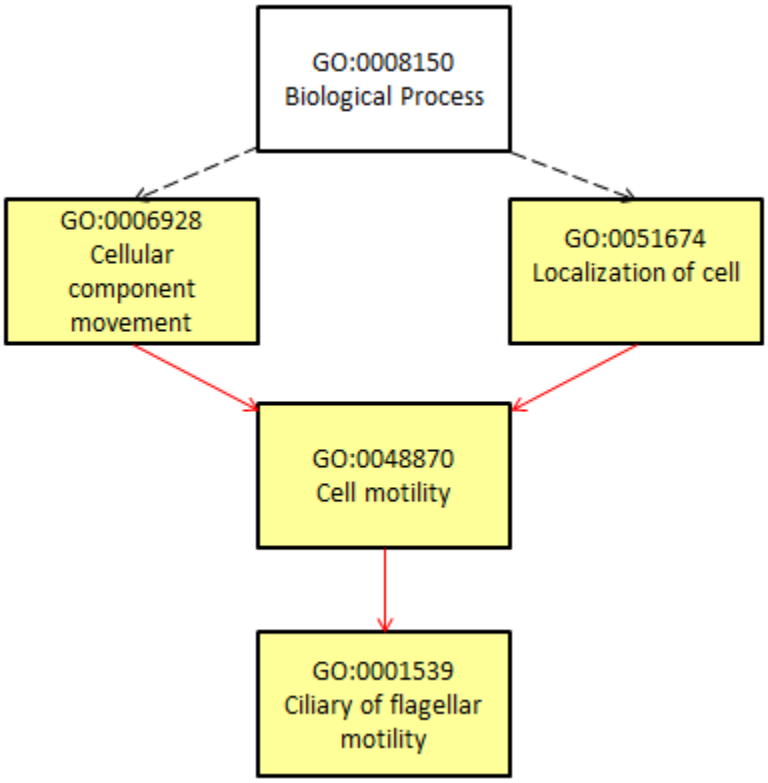


Figure 4.3 – Enriched gene ontology terms in the biological processes domain in 'Up' increased fitness genes in the plate-based TRMR selection. Yellow boxes represent significantly enriched terms as analyzed by the GOEAST online program [134]. Dotted black lines represent skipping of non-significant terms from the parent ontology. Red lines represent connections between nested significant ontologies.

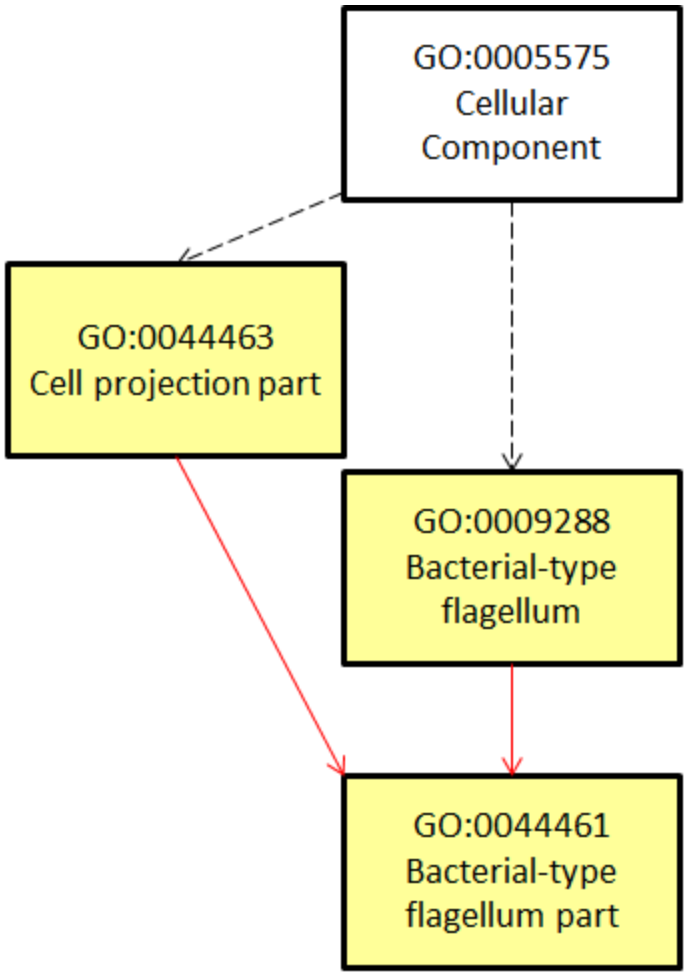


Figure 4.4 – Enriched gene ontology terms in the cellular component domain in 'Up' increased fitness genes in the plate-based TRMR selection. Yellow boxes represent significantly enriched terms as analyzed by the GOEAST online program [134]. Dotted black lines represent skipping of non-significant terms from the parent ontology. Red lines represent connections between nested significant ontologies.

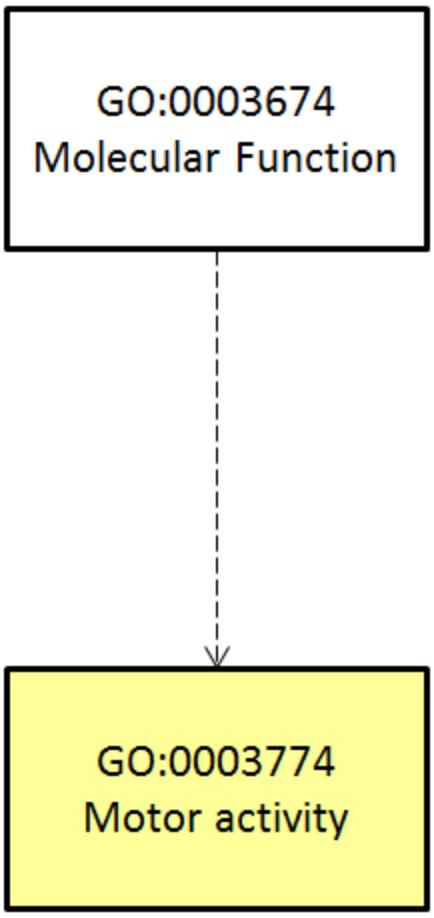


Figure 4.5 – Enriched gene ontology term in the molecular function domain in 'Up' increased fitness genes in the plate-based TRMR selection. Yellow box represents the significantly enriched term as analyzed by the GOEAST online program [134]. Dotted black line represents skipping of non-significant terms from the parent ontology.

4.3.2 – Comparison of TRMR and SCALEs furfural selections

We compared the SCALEs gene fitness scores to TRMR 'Up' gene fitness scores in order to search for enriched genes under both overexpression scenarios (Fig. 4.6). Surprisingly, only 35 genes had increased fitness (ln(fitness > 0)) in both the SCALEs and TRMR selections, representing less than 1% of all *E. coli* genes assessed. The functions of these mutually enriched genes were varied (Table 4.3), but had overlap to existing knowledge of furfural tolerance genes in three specific cases. Thymidylate synthase, encoded by *thyA*, has been previously confirmed to confer furfural tolerance through plasmid-based overexpression [156], purportedly due to its role in *de novo* pyrimidine biosynthesis which could repair furfural-related DNA damage [68]. Additionally, we had identified this gene and confirmed its growth improvements in furfural in Chapter 3. An alternative mechanism known for conferring tolerance is mediation of reactive species formation (ROS) induced by furfural treatment [69]. Genes related to ROS mediation have been implicated as providing resistance to hydrolysate [32], including the TRMR mutant *ahpC_Up* for hydrolysate tolerance [71]. Finally, *yhjH*, is involved in flagellar motility regulation [157], which aligns to the enriched GO terms for the TRMR 'Up' results (Fig. 4.3-5). The identification of chemotaxis/flagellar terms as being significantly enriched in our TRMR selection corroborates with the recent report that chemotaxis genes undergo repression during furfural treatment in *Clostridium beijerinckii* [156], but is a new link to furfural tolerance in *E. coli*.

We hypothesized that the subset of high fitness genes (ln(fitness) > 1 in either SCALEs or TRMR selection; genes labeled in Figure 4.6) would confer increased furfural tolerance. In order to test this hypothesis, we first cloned the 13 hypothesized tolerance genes into pSMART-LCKan. The functions of these genes are provided in Table 4.3 and primers used for clone

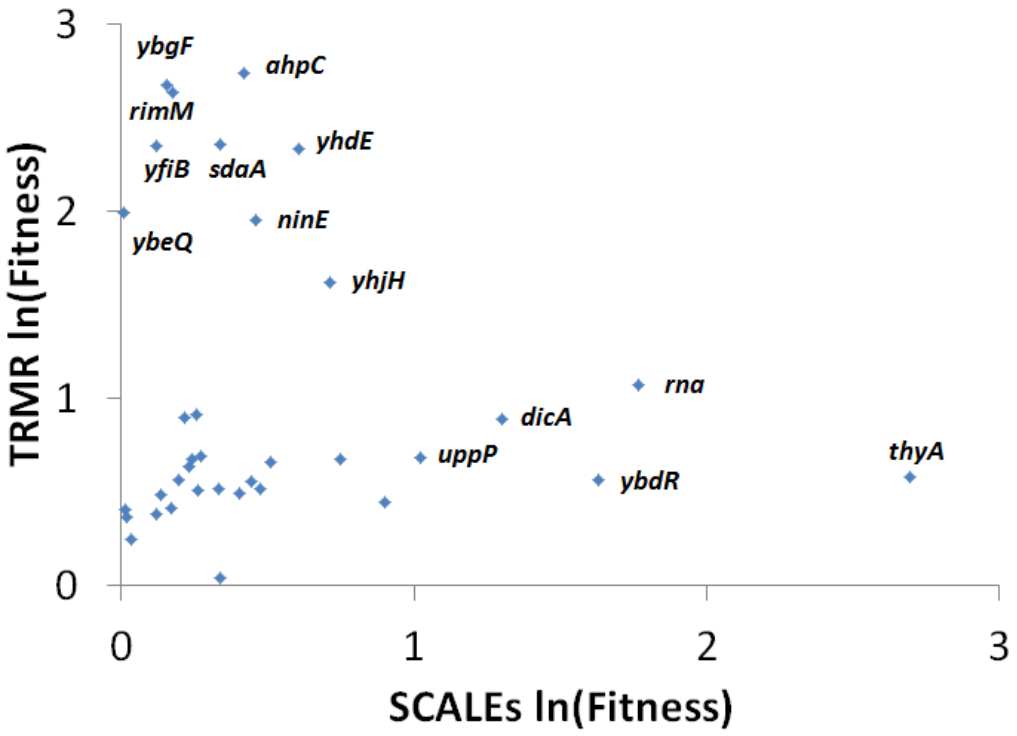


Figure 4.6 – Comparison of increased-fitness genes in TRMR and SCALEs furfural selections. Selections were performed by plating TRMR or SCALEs libraries on MOPS minimal medium plates with 0.75 g/l furfural. Fitness scores for genes were calculated based on the gene frequency change in selection. Genes with ln(fitness) > 1 in either selection are labeled.

construction are listed in Table 4.4. Nomenclature in this chapter follows the following method to provide clarity when referring to different mutations: plasmid constructs are labeled *gene-pSM* while chromosomally mutated strains are labeled as the targeted TRMR mutation and then the pSMART-LCKan vector they contain (i.e., *gene_Up pSM*).

These clones where then tested for improved growth in 1.5 g/l furfural (Fig. 4.7). Two of these clones, *rna-pSM* and *dicA-pSM*, showed improved growth. The gene *rna* codes for ribonuclease I, which has not previously been associated with furfural tolerance, but is important for degradation of RNA, especially during recovery from starvation [158]. Carbon-induced starvation was also associated with expression of the *csiD_Up* high fitness we saw enriched in all three TRMR selections. DicA inhibits DicB expression [159], which is part of the DicB-MinC complex that inhibits FtsZ at the septal ring [160]. With increased DicA expression in *dicA-pSM*, FtsZ inhibition might be reduced, resulting in differences in cell division. Identification of both *rna* and *dicA* as furfural tolerance genes highlights the ability of genome-wide searches like TRMR to identify non-obvious tolerance alleles.

Surprisingly though, four clones, *yhdE-pSM*, *ahpC-pSM*, *sdaA-pSM*, and *ybgF-pSM*, showed significant susceptibility under furfural treatment. In the case of*ybgF-pSM*, the measured optical density was less than the initial loading density, implicating increased cell susceptibility in this clone. These genes were identified with higher TRMR fitness than in SCALEs, suggesting that the level of expression might dictate the ability of the gene to confer tolerance under the conditions tested. We note here that these strains were tested in liquid culture (Fig. 4.7), but they had been identified from plate-based selections at half the concentration; they might confer tolerance under selection conditions, but not those tested here. Originally, we had hypothesized that *ahpC-pSM* would confer tolerance due to its previous

Table 4.3 – Genes with high fitness in either SCALEs or TRMR furfural selections.

Gene	SCALEs ln(fitness)	TRMR 'Up' ln(fitness)	Protein/Function
thyA	2.69	0.59	Thymidylate synthase
rna	1.77	1.07	RNase I
ybdR	1.63	0.56	Predicted oxidoreductase
dicA	1.30	0.89	DNA-binding transcriptional dual regulator for cell division process genes
uppP	1.02	0.68	Undecaprenyl pyrophosphate phosphatase
yhjH	0.71	1.62	c-di-GMP phosphodiesterase
yhdE	0.60	2.33	Conserved protein
ninE	0.46	1.95	DLP12 prophage
ahpC	0.42	2.74	Alkylhydroperoxide reductase
sdaA	0.34	2.35	L-serine deaminase I
rimM	0.17	2.63	Ribosome maturation protein
ybgF	0.15	2.68	Predicted periplasmic protein
yfiB	0.12	2.35	Predicted outer membrane protein
yebQ	0.01	0.41	Conserved protein

Table 4.4 – Primers used to amplify genomic regions plasmid clones.

Gene target	Forward Primer (5' → 3')	Reverse Primer (5' → 3')
rna	GGGAATTCCTCAATGCAGCG	ATAAATCATCACGCCCGCCA
ybdR	ACATCAACTGGCATAATGATTGTCT	CGGCGATTATCGTCATGGCT
dicA	CCTGCTGCTTGTGCAAGTTT	TCGGTCAAGTGATTTTGTATGCT
uppP	GTCGCATCAGGCGTTGATTG	GGCGTAATCATTGAGCGTGG
yhjH	ATTCTTCCTGTGCCAGTCCT	TCCGTTGTGGAGTGAGGAAA
yhdE	CGCGCCTCACGTTCAATATG	GTGATTAACGTCTCTTTCAGACCG
ninE	AGCTTCCAGAGAGTAAAAGTGTT	AAAATCAGACCAGAACGCCA
ahpC	TGCAAAAGTCGAGTAAAAGGCA	GAATCCCCGGGAGCTTACAC
sdaA	GCGCTGCAAATTGGTGTGA	CCTGACGCAACAGTGGAAGT
rimM	CGCAGTGGTGATTACTACCC	CCGACGGCCTTTTACAGCA
ybgF	CTGGCACCCGATGGTATGTT	TCGATGTAGATCACTCAACTGCT
yfiB	TGGCAAAAAGGGGCTGATGA	GAGCGTCTTAACTAAGATTTCGCT
yebQ	TCCAGATCTGCGTTAGCCAT	GCTCTGCGTACGGGTGAATA
ahpCF	TGCAAAAGTCGAGTAAAAGGCA	CCGGCGGGGCTTTTTAATG

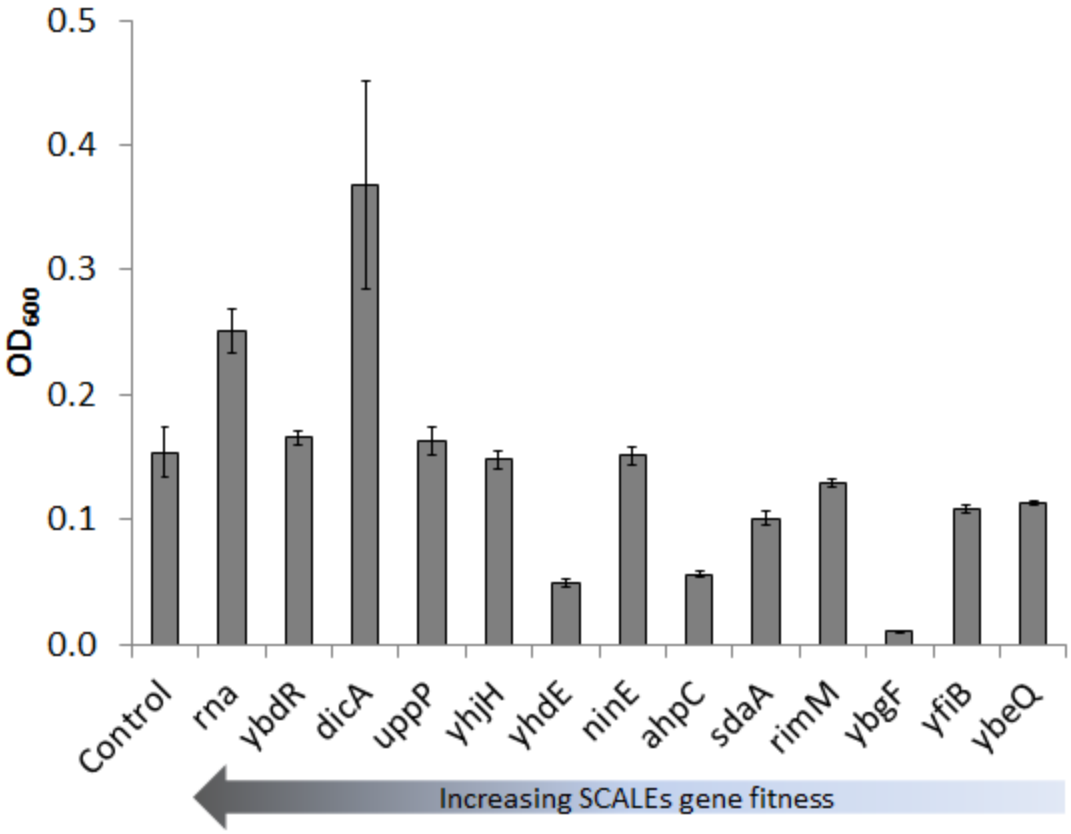


Figure 4.7 - Growth of increased fitness genes on plasmids in 1.5 g/l furfural. Genes listed were cloned to include their native promoter and ligated into the promoterless pSMART-LCK vector. Cultures were inoculated at normalized initial cell density (OD_{600} = 0.02) and grown at 37°C in MOPS minimal medium with furfural. Cell density was measured after blank vector control entered exponential phase (OD600 > 0.1, 26 hours). Average values (n = 3) are plotted; bars represent standard error.

identification as a hydrolysate tolerant allele [71], but found the opposite in this case for the gene cloned onto a plasmid in our tests. This surprising result motivated us to further test the effect of SCALEs or TRMR-based expression of this gene.

4.3.3. – Assessing tolerance differences

The *ahpC* gene is the leading gene in the *ahpCF* operon, meaning that these genes can be transcribed onto a single RNA molecule (there are three proposed promoters transcribing this region: two for *ahpCF* and one for *ahpF*). These two peptides work together as the AhpCF complex, an alkylhydroperoxide reductase, responsible for reducing peroxides to alcohols as a means of ROS detoxification [161]. Because *ahpC* expression is induced under sulfate-starvation, which maps to known inhibition caused by furfural treatment [56], we hypothesized that overexpression of *ahpC* without concurrent overexpression of *ahpF* (which would be the case in our *ahpC-pSM* clone) would not provide tolerance. In comparison, the *ahpC_Up* TRMR construct might confer tolerance since it could potentially also result in increased *ahpF* expression since the transcriptional terminator lies downstream from *ahpF*. This hypothesis was also supported that *ahpF* has also been increased during the SCALEs selection (Appendix 8.6). We constructed the *ahpCF-pSM* plasmid to test overexpression of the operon compared to the *ahpC-pSM* and *ahpC_Up pSM* clones in order to verify this hypothesis.

When grown in 1.5 g/l furfural, we did not find growth improvements in *ahpCF-pSM* as hypothesized (Fig. 4.8). The TRMR mutant *ahpC_Up pSM* conferred 82 ± 14% increased growth relative to control, which corroborates with the original TRMR study that identified *ahpC_Up* as a hydrolysate, and presumably furfural tolerant, mutant [71]. Both plasmid-based constructs, *ahpC-pSM* and *ahpCF-pSM* conferred decreased cell growth relative to control, -54 ±

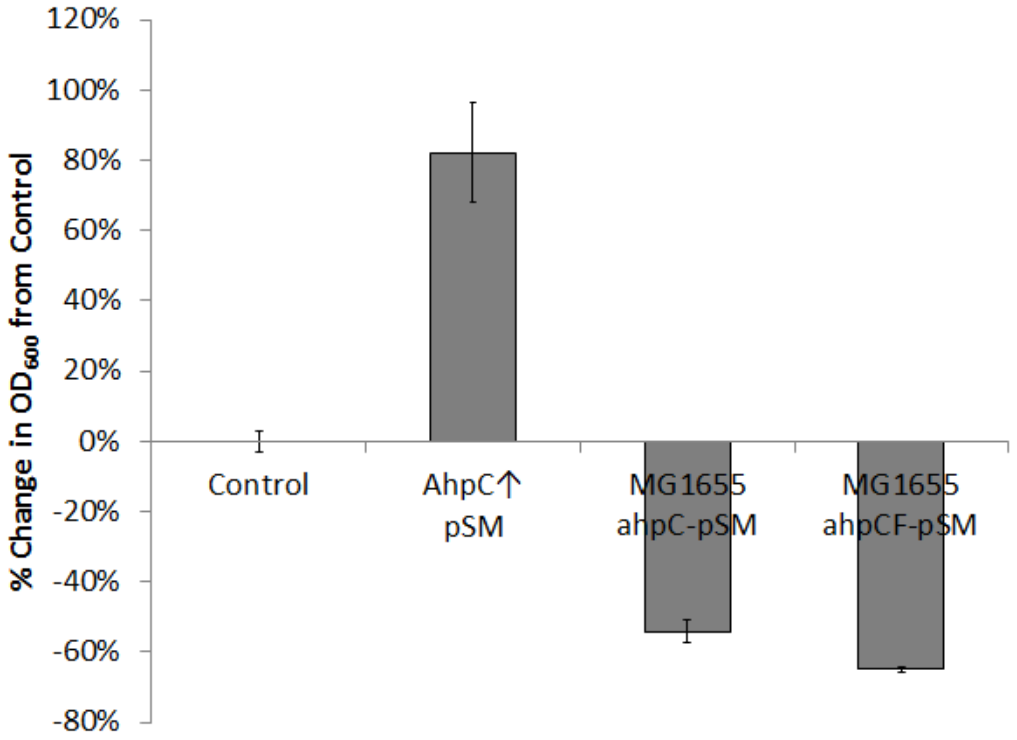


Figure 4.8 – Growth of *ahpC* constructs in 1.5 g/l furfural. Cultures were inoculated at normalized initial cell density (OD_{600} = 0.02) and grown at 37°C in MOPS minimal medium with furfural. Cell density was measured after 34 hr. Average values (n = 3) are plotted; bars represent standard error.

3% and -65 ± 1%, respectively. Additionally, the *ahpCF-pSM* construct was more inhibited in the furfural than the *ahpC-pSM* construct, indicating that the reduced growth of *ahpC-pSM* was not due to the lack of increased dosage of *ahpF*. Instead, we turned to look at the level of *ahpC* overexpression in the *ahpC-pSM* and *ahpC_Up pSM* strains.

We tested the expression levels of *ahpC* and one additional gene from the overlapping SCALEs and TRMR data, *yhjH*. We chose to look at *yhjH* due to its role in regulation of flagellar motility, which was related to the only enriched GO terms from the TRMR selection (Fig. 4.3-5). The *yhjH-pSM* construct also did not confer increased growth (Fig. 4.7), but as anticipated by the enriched flagella and motility related terms from the TRMR selection, *yhjH_Up pSM* did confer increased growth in 1.5 g/l furfural (290 ± 24 % increase over control; Fig. 4.9). The difference in growth improvements between the TRMR and plasmid constructs additionally motivated us to assess the significance of expression from the TRMR and plasmid constructs for both of these genes.

When we tested the expression of *ahpC* and *yhjH* from these clone constructs we observed that the plasmid constructs *ahpC-pSM* and *yhjH-pSM* resulted in increased expression (approximately 1-1.5 $\log_{10}$ increase compared to control; Fig 4.10). Conversely, the chromosomally integrated mutations in *ahpC_Up pSM* and *yhjH_Up pSM* did not alter expression. This result was in direct contrast to the intention of the 'Up' mutation cassette in the TRMR design, where the $P_{LtetO-1}$ promoter is reported to be strong (i.e., increase expression) [162]. Our study here is the first transcriptional analysis of TRMR constructs, and suggests that, at least for some TRMR clones, increased transcription might not occur as anticipated from a TRMR 'Up' mutation.

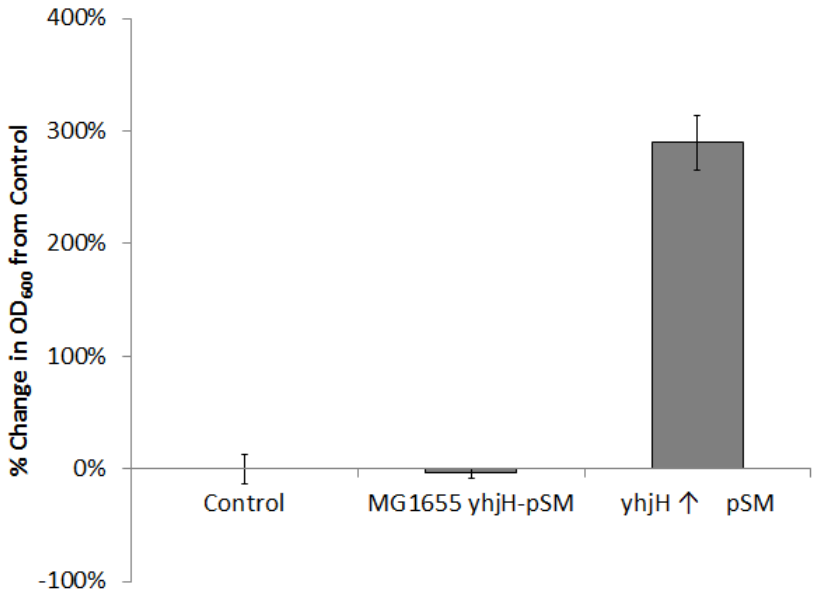


Figure 4.9 – Growth of *yhjH* constructs in 1.5 g/l furfural. Cultures were inoculated at normalized initial cell density (OD_{600} = 0.02) and grown at 37°C in MOPS minimal medium with furfural. Cell density was measured after 26 hr. Average values (n = 3); plotted and bars represent standard error.

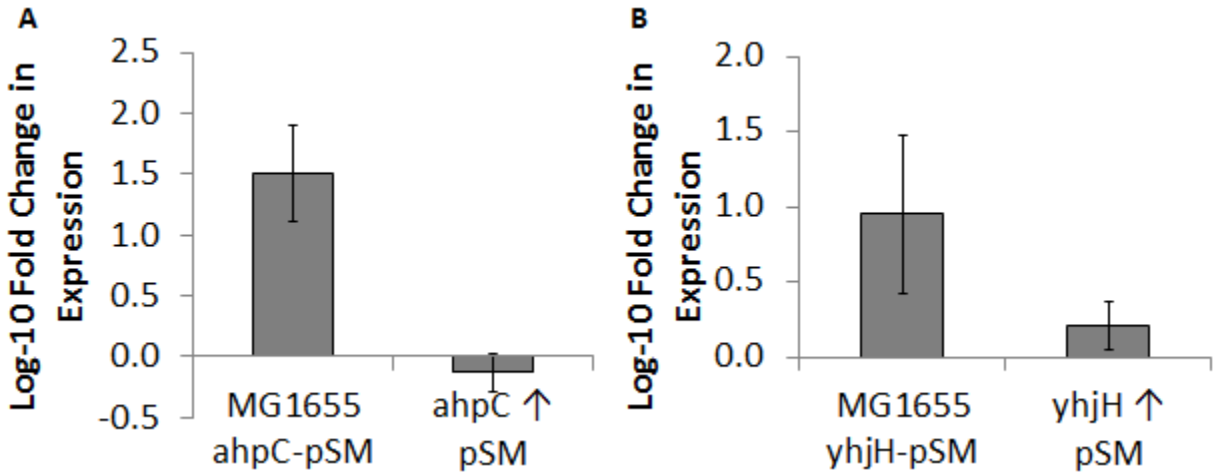


Figure 4.10 – Expression comparison from chromosomally integrated and plasmid-based constructs. Samples were harvested for RNA extraction from cultures growing in MOPS minimal medium. Transcript levels were measured with qPCR, normalizing expression to *cysG* expression. Fold-improvements were calculated for A) *ahpC* and B) *yhjH* with the $-\Delta\Delta C_T$ method compared to control. Average values (n = 3) are plotted; bars represent standard error.

We recognized though that although we did not observe a difference in expression from the 'Up' mutations, protein expression could still be altered as a result of translational changes from the TRMR designed RBS compared to the native RBS strength and/or mRNA stability.

To test differences in translation, we used the RBS calculator reverse engineering function [163,164]. The RBS calculator is modeled to predict the translation initiation rate (a.u.) based upon the thermodynamic properties of mRNA transcripts and has been used to accurately predict expression levels of fluorescent proteins from various RBS sequences [164]. When we performed analyses on the transcript sequences, the translation initiation rates for the *ahpC-pSM* (native RBS and mRNA sequence), translation initiation rate was 5.1×10^4 (a.u.) and 3.1×10^4 (a.u.) (recalling that there are two annotated promoters for *ahpC*). Comparatively, the translation initiation rate was 9.5×10^4 (a.u.) for *ahpC_Up pSM*. These values would suggest that the translation initiation, and thereby protein expression, is increased for *ahpC_Up pSM*, by up to half or two-thirds, depending on which transcript is made and read. Translation initiation rate from *yhjH* transcripts exhibited the same trend, with *yhjH_Up pSM* mRNA sequence returning 2.0×10^5 (a.u.) compared to *yhjH-pSM*, which was calculated at 2.2×10^3 (a.u.), showing an approximately two orders of magnitude increase for the TRMR clone. These calculations suggest that although we did not observe increased expression on a transcriptional level, translation is likely increased in the TRMR mutants, which would lead to increased expression of these genes.

4.4 – Conclusions

We used the TRMR method to track enrichments of 'Up' and 'Down' mutations in front of virtually every gene in *E. coli* during furfural selection. We performed three different selection schemes: plate-based growth with furfural added to the solid medium, a serial batch transfer with decreasing concentrations of furfural at each transfer, and a serial batch transfer with constant concentrations of furfural. Gene fitnesses from the selections suggested that the plate-based and decreasing concentration serial batch transfer schemes most closely map to each other and provided stronger selective pressure than the constant concentration selection. By analyzing the plate-based selection for enriched gene ontologies, we identified terms for motor activity, bacterial-flagellum part, and ciliary or flagellar motility as significantly enriched. Supporting the role that flagella function might play in furfural tolerance, we confirmed improved growth for a TRMR 'Up' mutation in front of *yhjH*, a gene involved in flagellar motility regulation. Through our assessments, we also observed growth improvements for *rna, dicA,* and *ahpC* overexpression. By comparing some plasmid-based constructs to their chromosomally mutated counterparts, we observed that expression levels at the transcriptional and translational level significantly impact the benefit of gene overexpression.

Our data here relate to three key findings. The first follows the adage "you get what you select for." We saw marked differences in fitness scores (total number of genes enriched, maximum enrichment values, and loss of the control tag in the population) based upon the type of selection we used. By these metrics, we determined that the constant concentration selection, which is a method often employed when performing directed evolution studies, was the "weakest." The "strongest" selection, and perhaps most effective at identifying furfural tolerance genes, was a serial batch transfer with decreasing inhibitor loadings at each transfer.

This method was originally described as a favored selection scheme that increased both selection specificity and sensitivity [127], traits that are desirable for genotype-to-phenotype mapping studies. Our data here parallel those findings and support this non-traditional approach for future selection designs.

Secondly, by using the TRMR method, we were able to rapidly identify non-obvious furfural tolerance genes. These genes and their functions can be added to the ever-growing collection of genetic manipulations for conferring furfural tolerance, with potential applications to hydrolysate and its inhibitors, other hydrophobic compounds, or ROS generators.

Lastly, we have shown that the capability of a gene to confer tolerance is highly dependent on its expression, both at the transcriptional and translational levels. In fact, plasmid-based overexpression of *ahpC* proved to decrease growth, whereas use of a synthetic promoter and RBS sequence integrated on the chromosome conferred improved growth. These findings strongly support the development of genome searching tools that are designed to provide controlled and context-free expression. Gains are being made in this area by designing synthetic sequences to control expression independent of surrounding sequence, quantifying the effect of trackable marker (e.g., antibiotics) and multiple-cloning site usage, as well as the rapidly occurring reports discussing the effect of codon bias in *E. coli* [149,150,152,165]. Moving forward, tools will need to designed to account for contextual dependence, or to control expression, in order to achieve efficient genotype-to-phenotype relationships.

Chapter 5

Targeted Multiplex Genome Engineering for Improved Biofuel Tolerance

In preparation for submission.

Authorship: Mills, T.Y.
Gillis, J.H.
Zeitoun, R.I.
Garst, A.D.
Reynolds, T.S.
Gill, R.T.

5.1 – Introduction

In Chapters 3 and 4, we have performed Design-Build-Test steps for identifying furfural tolerance genes with the SCALEs and TRMR methods. We have confirmed tolerance conferred by a number of genes involved in various cellular functions. Moving forward, we aimed to use these identifications to motivate a Learn phase, to re-enter the Design-Build-Test cycle, while targeting a phenotype hypothesized to be related to furfural tolerance. Our objective was to transition towards engineering tolerance to hydrophobic compounds, specifically those that are potential biofuel products, like isobutanol. Many long-course adaptation studies identify multiple mutations which support a specific phenotype or pathway [9,77,93,121]. Recognizing that in order to achieve large improvements in tolerance, we would likely need to target multiple mutations within a single strain, we assessed our data from the previous chapters for genes that might share common tolerance mechanisms between furfural (a hydrophobic compound) and hydrophobic alcohols. This analysis led us to consider one of our most tolerant genes, *lpcA*, which was identified in Chapter 3, and the potential benefit overexpression of genes involved in the same biosynthetic pathway might have towards engineering hydrophobic molecule tolerance.

We focused on targeting a specific pathway instead of the top-fitness mutants from our furfural selections for two reasons. The first, we recognized that not all furfural tolerance genes would also confer tolerance to hydrophobic compounds. We needed to direct the choice towards pathways that supported a hypothesis in which its function could dually confer tolerance to other hydrophobic compounds. Secondly, previous work in our lab has emphasized the importance of epistatic interactions towards engineering mutations in multiplex (i.e., multiple mutations within a single cell) [82]. In this previous study, the top-fitness genes from hydrolysate, acetate, and pH TRMR selections were targeted for altered expression by way of ribosome binding site (RBS) mutations according to the Multiplex Automated Genome Engineering (MAGE) method [103]. MAGE utilizes a lambda phage recombineering protein to insert synthetically designed mutations into specific regions of the chromosome in a recursive fashion to build up a population of multiplexed mutants. Surprisingly though, in our study with TRMR-directed targets, the most tolerant mutants contained only single mutations, rather than multiple mutations, suggesting that the mutations targeted were not positively synergistic with each other. A recent study combining singly identified furfural tolerance genes into one strain also highlighted the unpredictable epistasis of these interactions towards conferring tolerance [12]. In order to avoid potentially antagonistic epistasis between 'competing' mutations, we chose to focus on one specific pathway.

LpcA, or sedoheptulose-7-phosphate isomerase, catalyzes the first committed step towards heptose formation for lipopolysaccharide (LPS) biosynthesis [166]. LPS, a vital component of the *E. coli* outer membrane, is well known to act as a barrier towards hydrophobic compounds in microbes. Strains which lack inner core heptose, have been shown to be more susceptible to hydrophobic compounds in *E. coli* and *Salmonella typhimurium* [140,141,167],

and the weakened immunity of these mutants is attributed to a lack of structural integrity maintenance normally obtained through cross-linkage of neighboring phosphorylated heptose units with divalent cations [139,168].

There are 29 genes involved in the LPS superpathway. Each of these genes catalyzes steps for the formation of either lipid A, keto-deoxyoctulosonate (KDO), or heptose, or the assembly thereof (Fig. 5.1). Lipid A is the amphipathic component, serving as the anchor for LPS in the outer leaflet of the outer membrane. KDO is integral for *E. coli* viability, and connects lipid A to the core region, and two of its biosynthesis genes, *kdsA* and *kdsB*, are essential [131]. In non-pathogenic *E. coli*, the core is first composed of heptose molecules and then a small number of other sugar molecules like galactose and glucose [169]. We searched for enrichments of each of these genes in our own furfural SCALEs and TRMR 'Up' (Appendices 8.6-7, 8.9, and 8.11), and discovered that many of the genes had increased fitness in one of more of our selections (Fig. 5.1 A).

Two recent reports also supported our hypothesis that overexpressing LPS biosynthesis genes might confer tolerance to candidate biofuel compounds. First, during a long-course adaptation to evolve isobutanol tolerance, glucosamine-6-phosphate, the substrate for which GlmM catalyzes in lipid A biosynthesis, was identified as an important metabolite [77]. Also in this study, deletion of *yhbJ*, which results in increased *glmS* expression (the upstream step from GlmM, Fig. 5.1A) [170], was also important for conferring isobutanol tolerance. The second report, from our own lab, identified *lpcA* overexpression through a SCALEs ethanol selection [78]. Although transcriptional analysis of ethanol and isobutanol-treated cells differs [76,79],

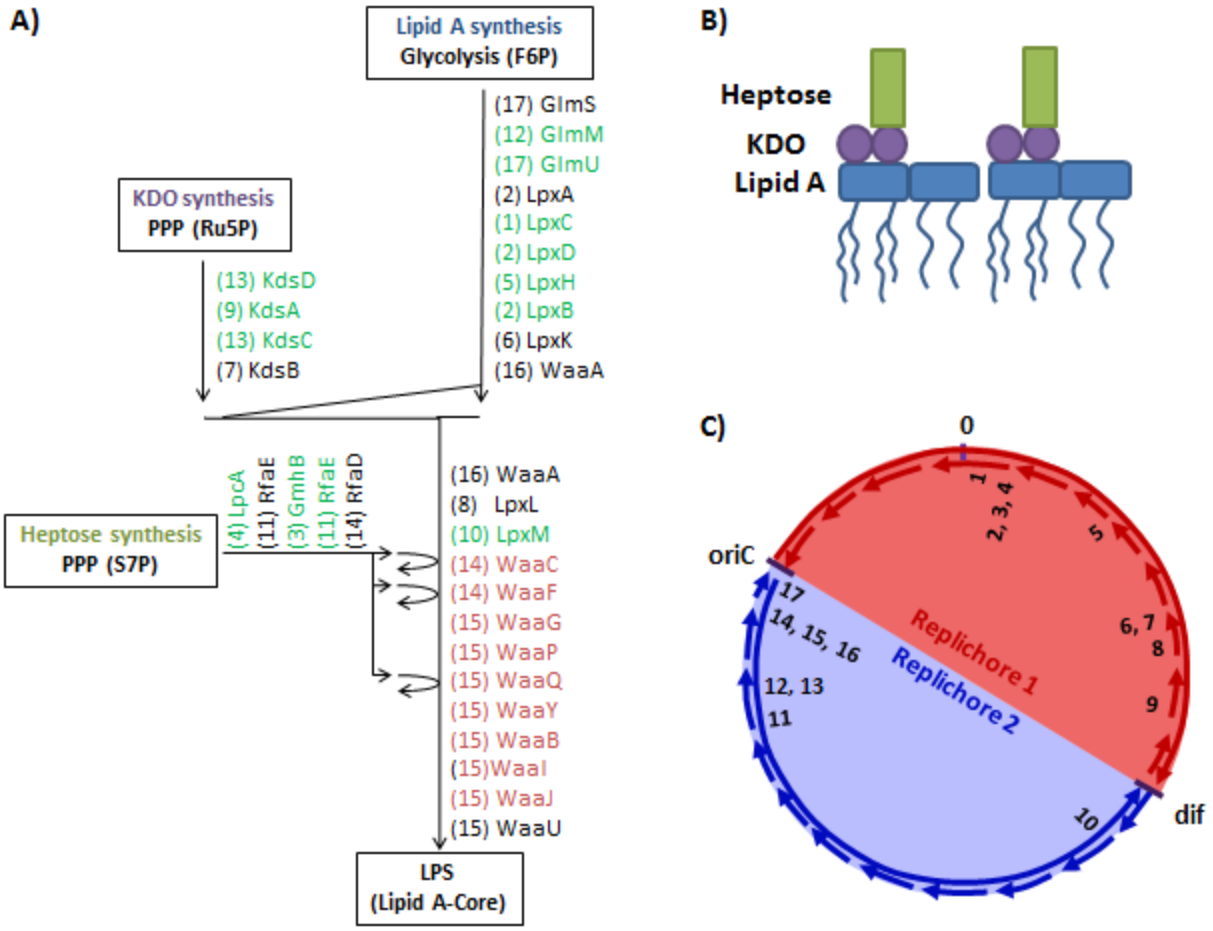


Figure 5.1 – LPS biosynthetic pathway, structure, and biosynthesis gene locations on the chromosome. A) Superpathway of LPS biosynthesis containing lipid A, KDO, and heptose formation and assembly. Numbers next to enzymes denote the operon (1-17) in which the gene is found. Gene names in green text were identified with increased fitness through SCALEs and TRMR furfural selections in chapters 3 and 4. Gene names in red text were not designed in the original TRMR library due, in part, to their overlapping architecture on the chromosome, and were thus not candidates to be identified in TRMR gene fitness searches. B) Schematic of LPS structure. LPS sits in the outer leaflet of the outer membrane, with lipid A facing phospholipids of the inner leaflet. KDO attaches the sugar chain to lipid A, and together, represent the LPS inner core. C) Genomic position of the 17 operons that contain the 29 enzymes involved in LPS biosynthesis.

alcohols or varying length are known to associate and disrupt *E. coli* membranes [41]. These reports, supported our hypothesis that overexpression of LPS biosynthesis genes might confer tolerance to biofuel alcohol candidates like isobutanol, and potentially ethanol, *n*-butanol, and isopentenol.

In order to target these 29 genes, which are organized into 17 distinct operons across the genome (Fig. 5.1C and Fig. 5.2), we decided to use a MAGE-like approach, relying on lambda-mediated single stranded DNA (ssDNA) oligomer recombineering [105]. In this system, ssDNA oligomers are transformed into cells, where the lambda protein *beta* serves as an ssDNA-binding protein, protecting the oligomer from exonuclease activity. It is proposed that these oligomers can then replace Okasaki fragments during DNA replication on the lagging strand, introducing the mutation onto the chromosome [171]. Deleting *mutS*, which is part of the methyl-directed mismatch repair system [172], has been shown to increase recombineering efficiency for point mutations and small insertions [104]. Recombineering efficiencies are typically low enough that only a small portion of the population incorporates the mutation. For this reason, we mimic the recursive recombineering approach presented in MAGE to insert an orthogonal RNA polymerase (RNAP) derived from the T7 RNAP [173]. This approach and design is similar to a recent study directed towards altering amino acid biosynthesis with multiple T7 promoter insertions [107]. In this scheme, recombineered populations are recovered, again and then subjected to subsequent rounds of recombineering to create increasingly multiplexed members of the population. The population created after ten recursive rounds of recombineering, which contained members of increasing diversity, was then subjected to selections under ethanol, *n*-butanol, isobutanol, and isopentenol (3-methyl-3-buten-1-ol).

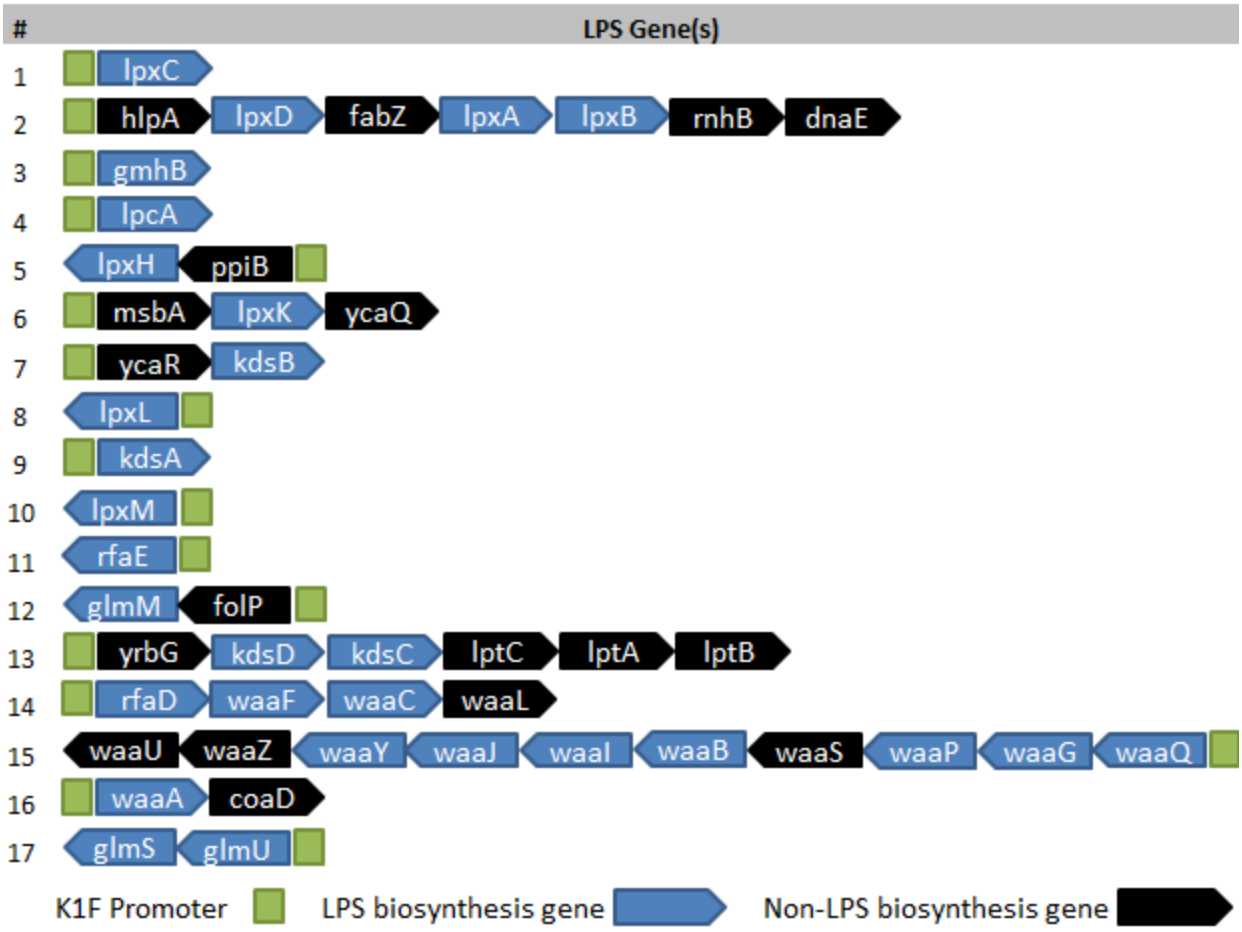


Figure 5.2 – Architecture of LPS biosynthesis operons. The position for P_{K1F} insertion is designated in front the gene in the operon. Black arrows indicate genes that are not involved in LPS biosynthesis, but have the potential to be transcribed from the promoter insertion mutation. Direction of the operon as it is orientated in the genome is indicated by the genes facing clockwise or counterclockwise. Orientation and replichore (Fig. 5.1C) were accounted for in designing oligomers to target the lagging strand.

5.2 – Materials and Methods

5.2.1 – Bacteria, plasmids, and media

E. coli EcNR2 (MG1655 *bioA/bioB::λ-bet mutS::cm*^{R}), HEM6 (W3110 *galK*$_{tyr145UAG}$ Δ*lacU169* [*λ c*I857 Δ(*cro-bioA*)]), and HEM63 (HEM6 *mutS::amp*^{R}*)* were used for recombineering studies. The K1F T7 variant plasmid (pK1F) was provided from the Voigt Lab at MIT and contains spectinomycin resistance and the K1F RNA polymerase (RNAP) under an IPTG-inducible promoter [173]. Low-salt Luria Bertani broth (LS LB) was prepared to contain 10 g tryptone, 5 g yeast extract, and 5 g sodium chloride per liter. All cultures were grown at 30°C. The following concentrations were used as appropriate: IPTG, 250 μM; spectinomycin dihydrochloride pentahydrate, 100 μg/ml; X-gal, 40 μg/ml. MOPS minimal medium [133] was used for selections and growth analyses. MacConkey agar plates were used for *galK* screening with 1 v/v% galactose [104]. *E. coli* SimD70, a recombineering strain previously used in our lab [82], was transformed with pK1F and used for MacConkey agar assays.

5.2.2 – Promoter oligomer design

Oligomers were design according to the method of Wang et al. [107]. Briefly, the insertion site was designed for insertion 35 bases upstream of the gene start codon, with 35 bases of MG1655 chromosome homology flanking the insertion site on either side. The 20 bp K1F promoter sequence, TAATAACTATAACTATAGGG, was designed between the two homology arms. Sequences were designed to target the lagging strand during recombination. Sequences are given in Table 5.1, where asterisks designate phosphorothioated linkages. Oligomers were used for measuring recombination efficiency at a 1:20 ratio with the promoter oligomers, and targeted *lacZ* (EcNR2) or *galK* (HEM6 and HEM63) The sequences for the *lacZ* inactivation

oligomer (lacZ_mut_1) and the *galK* restoration oligomer (galK_144) were designed as previously reported [103,106], with the four 5' bases phosphorothioated.

Table 5.1 – Oligomer sequences synthesized for K1F promoter insertions.

K1F_1	C*G*T*A*TTATCTCGCCAAATTACCTATCCAACCGAAGCCCTATAGTTATA GTTATTATGTACTATACATTCGGCGGGCCAGTTTAGCACAAA
K1F_2	A*A*T*A*AACTCCTTACCATCCCATTTGCACCGGAGGGCCCTATAGTTAT AGTTATTATGCAGTTCTTTGCGTGGCCCGGCGATCTTATATTG
K1F_3	C*T*T*T*TATAGCTCCTTAATAAGGCATGTGACGCTAGCCCTATAGTTATA GTTATTATATCGCATGTTTCGACCTGCAAGAAAGTGCTCTTC
K1F_4	G*A*G*G*ATATCCTTCAGCATAAATGTAATAGACAAAACCCTATAGTTAT AGTTATTATGCAGTGTACCGGATACCGCCAAAAGCGAGAAGTA
K1F_5	T*G*C*T*ATAACACCACCCTATATATGACCCGAACTGGTAATAACTATAA CTATAGGGGTTGAAGCACCAATCAAACGGAACAGGATGCAAAA
K1F_6	T*C*A*A*AAAACCAGCATTTGTTGAAATAGCCGCATATCCCTATAGTTAT AGTTATTATCTACCCGTTATCCACTGGCACGCCAAACCACTGA
K1F_7	A*G*T*T*CCTCCAGATGGATCGGGTTATGAATGCATAACCCTATAGTTAT AGTTATTAATCTTATCATAATCATTTATGCGACGGGGTCTATT
K1F_8	A*T*T*T*TGGTTGCGGGCGAAAAAATGCGACAATACATTAATAACTATAA CTATAGGGACAATTGCCCGAATAGGTTGAAAAACAGGATTGAT
K1F_9	A*G*G*A*TCGCCTTAATCTTGAGTAAAATGTCGATTAACCCTATAGTTAT AGTTATTATTAATGCAGCACAATATGTTTATGCGCGATGTTAT
K1F_10	G*C*T*T*TTCCAGTTTCGGATAAGGCAAAAATCAATCTCCCTATAGTTATA GTTATTAGGTGATAGTGTAGCGGCGCAACTTGCCCCGCACCA
K1F_11	T*C*C*T*GTCTCCTGAGAGATTCAAAATTTGCGCGCGACCCTATAGTTATA GTTATTATAATACCATACTTCATTCTTCCACCAGCCACTTCT
K1F_12	G*A*T*G*TTATCCCTGGTATGAATTGATAAGAAAAAACCCCTATAGTTAT AGTTATTACCCGGAGCACGCCCCGGGGTTTTCGGTACAAATAC
K1F_13	T*T*A*A*GCATCCGTTACGGCTTTCTGAAAATCTTCAGTAATAACTATAAC TATAGGGCGGACCGGCGAGTATACCTGAAGAAAGGACGTTAG
K1F_14	A*T*G*T*GATGGTATGATTACAGACATTCGTGTCTGAGTAATAACTATAA CTATAGGGATTGTCTCTGACTCCATAATTCGAAGGTTACAGTT
K1F_15	C*T*T*T*ATGACCAGGATTTTTCGAAATGGCTTTTCCACCCTATAGTTATA GTTATTACTAGCGACTCTTTTGTGTGATTGTCTGGTTAAGTT
K1F_16	G*G*A*A*AAGTAATGGTAAAGCCACAGCTAAATACATATAATAACTATA ACTATAGGGGAATCCCCAGCACATCCATAAGTCAGCTATTTACT
K1F_17	A*C*G*C*GTCCTGACTGTAATTTGAGAACGAATTTAAACCCTATAGTTAT AGTTATTACCGCTTCACCTTGAAAAAACTACATTTTTTTCATC

5.2.3 – Recursive recombineering for library creation

Recombineering strains EcNR2, HEM6, and or HEM63 with the pK1F plasmid were grown in LS LB with IPTG and spectinomycin overnight. A 2 v/v% inoculation was made into 25 ml fresh LS LB medium and grown until mid-late exponential phase ($0.6 < OD_{600} < 1.0$). Cultures were then induced at 42°C for 15 minutes to induce recombineering protein expression. Cells were then chilled on ice, and 1 ml aliquots (~10^{7-8} cells) were serially washed twice with ice-cold ultra-purified water. Pellets were resuspended in 50 µl ice-cold ultra-pure water, and electroporated at 1800 kV with 1 µl oligomer mix (equimolar amounts of all 17 targeting oligos and a 1:20 dilution of either the *lacZ* or *galK* oligos). Cells were recovered in 3 ml LS LB, IPTG, and spectinomycin until entering exponential phase again. This process was repeated for a total of 10 rounds, with cultures being stored at 4°C or -80°C, as needed. Dilutions of cultures were plated on LS LB agar with IPTG and X-gal or MacConkey agar plates with galactose and IPTG for estimation of recombination efficiency. Population dynamics (percentage of population with varying number of mutations) were calculated using equations previously derived [174]. After the fifth and tenth round of recombineering, cultures were plated in dilutions to isolate individual colonies while the remaining population was stored at -80°C until further use. Control recombineering reactions (1 µl of water transformed instead of oligomer mix) were performed in parallel for all strains. Cultures grown from this water-transformed population is referred to as 'Adapted' within the text.

5.2.4 – Linking PCR

Primers were designed with the Linking PCR algorithms developed by Zeitoun et al. in the Gill laboratory. In short, 60 bp sequences of the chromosome upstream and downstream of

the targeted insertion site are input into the MATLAB code which designs compatible 'linking' sequences (i.e., optimized GC content, limited secondary structure, and minimization of homology to chromosome sequences) for each internal primer. Two sets of linking PCR primers were designed to yield PCR products of approximately 1000 bp and 800 bp constructs (Tables 5.2-3).

Two reactions were performed with Phusion polymerase (New England Biolabs). The first was an assembly PCR where 1 µl 10^{-3} dilution of an equimolar primer solution of all primers in the set was initially loaded into the reaction mixture and cycled for 12 rounds of PCR. Then an additional 1 µl of 10^{-1} dilution of primers was introduced to the reaction mixture and cycled for another 40 rounds of PCR. The PCR products were either used directly or cleaned to remove remaining primers with a Qiaquick PCR Purification Kit (Qiagen). This eluted mixture was then used as the template for an amplification reaction, which only used the external flanking primers of each set to amplify the entire linked together construct. The products were then separated with gel electrophoresis, extracted from the gels by excision, and purified from the agarose with a Qiaquick Gel Extraction Kit (Qiagen). Cleaned products of the expected size were sequenced to determine sites with promoter insertion.

5.2.5 – Growth analyses

Overnight cultures were inoculated from freezerstocks into LS LB with IPTG and spectinomycin. Overnight cultures grown to stationary phase were then diluted to OD_{600} = 1.0, and 1 ml aliquots were centrifuged at 14,000 rpm for 2 minutes. The supernatant was decanted and the pellet was resuspended in 500 µl LS LB to concentrate the cells by a factor of two. An aliquot of 100 µl was then inoculated into MOPS minimal medium with IPTG and

Table 5.2 – Linking PCR primers used to amplify Set 1 single sequencing construct.

Gene Target	Forward (5'→ 3')	Reverse (5'→ 3')
lpxC	TAAACTGGCCCGCCGAATGTATAG	TGGGTGAGTCCTGAGCTGCCGACCGTTATTCCTTTGTTTGATcatCGTATTATCTCGC
hlpA	ATAACGGTCGGCAGCTCAGGACTCACCCAGTGTAGCGATGACTTTAGGCGATCA	ACCTTGAGACCATTGCCATCCTCCGCCAGCGAGACCTGCAGCTAATAACCACTT
gmhB	CTGGCGGAGGATGGCAATGGTCTCAAGGTATCAGGTTTATGCGAAGAGCACTTTCT	AGTCGTTGCGGGTTGGGTTATGTCACGGCCTCTTCGCCACCTTTTATAGCTCCT
lpcA	GCCGTGACATAACCCAACCCGCAACGACTACCTGCCCGTACTTCTCGCTT	TCGGCTAAGAACACTGGTGACGCATCGGGTCGTTACGAATAAGATCCTGGTACATGAG
msbA	CCCGATGCGTCACCAGTGTTCTTAGCCGATTGGCGTGCCAGTGGATAACG	TGCTCGGTCGCTGGTATGTAGCCCGTTGACGTAGAGAGATCTTTGTCGTTATGCATTC
ycaR	TCAACGGGCTACATACCAGCGACCGAGCATCTTTACTGATTGCCGCACCGG	TCCACAGTCCAGCACAGTTCTTCCCGCCTACGATGATCCATAGTTCCTCCAGATG
kdsA	AGGCGGGAAGAACTGTGCTGGACTGTGGACGCGCATAAACATATTGTGCTGCA	TTGTGCCATGCTTCGGGATCGGAGAGCTGTTGATGTCGCCAATGCTAACCACTT
yrbG	CAGCTCTCCGATCCCGAAGCATGGCACAACCGCATAAAGTCAAAATTAAGCATCCG	ACCATCCGCCCATCAACTTCAACCGCACGGCATCTAACGTCCTTTCTTCAGGTATACT
rfaD	CGTGCGGTTGAAGTTGATGGGCGGATGGTGATTCGGATGTGATGGTATGATTACAGAC	GGTAACGATGATCATAACTGTAACCTTCG

Table 5.3 – Linking PCR primers used to amplify Set 2 single sequencing construct.

Gene Target	Forward (5'→ 3')	Reverse (5'→ 3')
waaA	CGGGAAAAGTAATGGTAAAGCCACAG	AGCCACAGTTTCCCTCAGCACAGGCAGGAGCGGTGTAAAGCAATTCGAGCATAG
ppiB	TCCTGCCTGTGCTGAGGGAAACTGTGGCTCGCCGTGATTGGTGTGGAAAGT	TAGACGGGCTACCACCCTTTGCCAGCGAAGATCCGATCGTGCGGTTATGCT
lpxL	TTCGCTGGCAAAGGGTGGTAGCCCGTCTAAAGCAGTGCGGTGGAGAACTTG	AGCACCAAACGCAGCATCCTCACATCGCCCGCCCAGTCTTCAGCCACAAA
lpxM	GGCGATGTGAGGATGCTGCGTTTGGTGCTATGCTTTTCCAGTTTCGGATAAGGCA	AACGCCACACAGAGGGAGCAAGTGCCTTGATTAACATCCATTCGCAGCCGGTAC
rfaE	CAAGGCACTTGCTCCCTCTGTGTGGCGTTCGTTACTTTCATTCCTGTCTCCTGAGA	ATAACGCTTGTGTGGTCACGGCAGGGCGAGGCTGGTGGAAGAATGAAGTATGGT
folP	TCGCCCTGCCGTGACCACACAAGCGTTATGCAAAGAGTTTCATGATGTTATCCCTGG	GAGCTGGATTCGGAAGGAGGCGAGTTCGACGCATCAGATGACTGTATTTGTACCG
waaQ	TCGAACTCGCCTCCTTCCGAATCCAGCTCCCATGATATCGCATCTTTATGACCAGG	GCGAAGCGAATGAAAGCTGCCTTGGGCTCCTTAACCAGACAATCACACAAAAGAGTCG
glmU	GAGCCCAAGGCAGCTTTCATTCGCTTCGCTTCAACATACGCGTCCTGACTGTAATT	CAATTTATCCTCTGTCCATTTCACGATGA

spectinomycin and 0.5 v/v % *n*-butanol (initial OD_{600} ~ 0.04). Cultures were grown and measured for optical density readings after 6 hours.

5.2.6 – Alcohol selections

Freezerstock aliquots of the libraries after 10 rounds of recombineering (HEM6 and HEM63 P_{K1F} libraries and their respective water-transformed 'adapted' populations) were thawed on ice. A 500 µl aliquot was inoculated into 20 ml of LS LB with IPTG and spectinomycin and grown until mid-exponential phase (~8 hours). Cells were harvested by centrifugation and then resuspended in LS LB and inoculated 2 v/v% into 15 ml of MOPS minimal medium (OD_{600} ~0.04) with IPTG, spectinomycin, and an alcohol (ethanol: 3 v/v %; *n*-butanol: 0.5 v/v %; isobutanol: 0.5 v/v %; and isopentenol: 0.5 v/v %). Cultures were grown until they reached early or mid-exponential phase and then inoculated into a second round at the same inhibitor concentrations at OD_{600} ~0.04. The ethanol, *n*-butanol, isobutanol selections were inoculated into a third and final round. Cultures were dilution plated to isolate colonies for individual clone genotyping and phenotyping.

5.3 – Results and Discussion

We used recursive rounds of lambda-mediated recombineering to create libraries of mutants with an orthogonal RNAP promoter sequence in front of the 17 LPS biosynthesis operons. We hypothesized that altering LPS biosynthesis gene expression would allow us to increase tolerance to hydrophobic biofuel candidates. Secondary to this goal, we also wanted to test the effect that *mutS* deletion mutants would have on propagating unintentional mutations

over the course of recursive recombineering rounds, since the deletion of *mutS*, and thus the methyl-directed mismatch repair system, has recently been suggested to contribute significantly to increased mutation frequency [175].

We also aimed to apply a recently developed method in the Gill Lab by Zeitoun et al. for tracking targeted mutations across the genome. In this 'Linking PCR' method, primers are designed to amplify sequences flanking the intended mutation site, and have linking sequences so that the first amplification product shares a complimentary linker sequence to another mutational target site. This process is reminiscent of overlap-extension PCR [176], but is being optimized for mutation tracking by development of a computation program to design ideal linker sequences to minimize unintended PCR products. Our application of this method is the first use of this method for tracking the diversity of a designed mutational library. Linked PCR constructs can be sequenced in a single Sanger sequencing reaction. By linking together sequences from various sites, the cost of genotyping of our mutations can be reduced by approximately an order of magnitude.

The current standard for assessing presence of a mutation in recombineered populations is using Multiplex Allele-SpecifiC (MASC) PCR [174]. This method, in comparison, relies on two forward primers, one with homology to the wild-type sequence and one with homology on the 3' end to the intended mutation, and one reverse primer with homology to the un-mutated template. PCR products are only amplified from the allele-specific primer if the 3' end has homology (i.e., anneals to the mutated sequence). An amplified product from this reaction is inferred as a mutated sequence while lack of product is inferred as a wild-type sequence. Problems arise with MASC PCR due to the lack of positive controls (mutated templates to ensure the allele-specific forward primer works), or control for non-specific binding yielding a

product of the expected length but wrong sequence. The issue of multiple chromosomes within a single cell [32,104,106,177], where one chromosome is mutated and another is not, can also introduce a number of false-positives in MASC PCR. In addition, the method requires individual PCR reactions for each target. Primer sets can be designed *de novo* so that they will amplify different sized products to be screened in a single lane by gel electrophoresis, but this requires considerable time in designing them and quality assurance testing that they work as intended. Additionally, sequences must ultimately be verified, and MASC PCR only designs constructs for single sites, resulting in high costs for sequencing individual constructs. For these reasons our linking PCR method provides advantages to us in this study.

5.3.1 – Creation of LPS biosynthesis libraries with EcNR2

We performed 10 rounds of recursive recombineering with oligomers designed to insert an orthogonal RNAP promoter sequence in front of all genes/operons of the lipopolysaccharide biosynthesis pathway. Initially, we used an *E. coli* MG1655 derivative (EcNR2) designed for high efficiency ssDNA recombineering with a *mutS* knockout and chromosomal integration of lambda *beta* [103]. In this case, we also spiked in an oligomer that introduces a premature stop codon in *lacZ*, thereby inactivating its function and enabling mutants to be screened with blue/white screening (mutants that were recombineered and incorporated the oligomer would be white when grown in the presence of IPTG and X-gal compared to wild-type strain that would be blue from wild-type metabolism of X-gal). We performed 10 rounds of recombineering in parallel, using both our LPS/*lacZ*-targeting oligomers and a water control transformation (to create a control 'adapted' population). We plated the 'adapted' population after the first and tenth round and observed the propagation of a white colony sub-population (Fig. 5.3). All of the colonies from this 'adapted' population should have been blue, since they were not exposed to

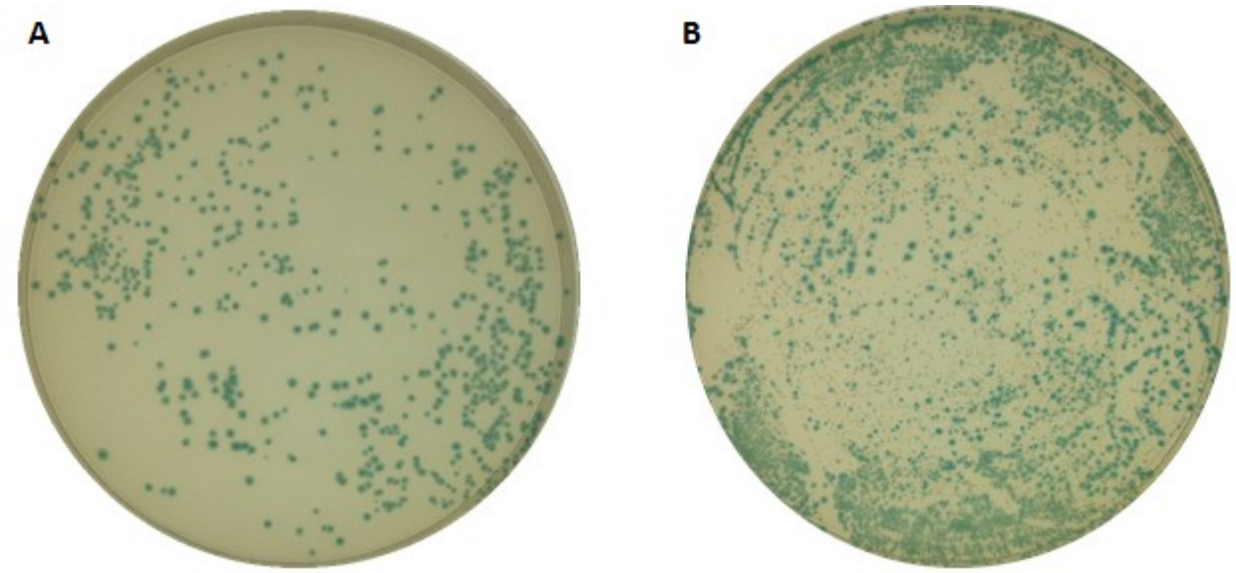


Figure 5.3 – Blue/white screening of EcNR2 'adapted' populations. Cultures are plated on LS LB, IPTG, X-gal plates. Colonies with functional lacZ (wild-type) can metabolize X-gal, producing a blue appearing colony. White colonies are unable to metabolize X-gal, and indicate a mutation deactivating *lacZ* function or one resulting in a *lacZ*$^-$ phenotype. A) EcNR2 pK1F 'adapted' population plated after the first round of recombineering. B) EcNR2 pK1F 'adapted' population plated after ten rounds of recombineering.

the *lacZ*-inactivating oligomer. This finding suggested to us that deletion of *mutS* might be responsible for allowing random mutations to propagate within our recombineered population that might result in a *lacZ* mutant phenotype, as has been implied by recent reports of increased mutation frequency in $mutS^-$ strains[116,174].

5.3.2 – Creation of LPS biosynthesis libraries with HEM6 and HEM63

Due to the propagation of white colonies during the course ten recombineering rounds in the EcNR2 host, we decided to re-create the libraries in a new host system, one that would allow us to monitor the effect *mutS* deletion had on the population in comparison to a $mutS^+$ strain. For this purpose, we used HEM6 ($mutS^+$) and HEM63 ($mutS^-$) developed by the Court Lab [105]. These strains have mutations deleting the *lac* operon and cannot be screened with blue/white screening, but have the $galK_{tyr145UAG}$ mutation, which allows for pink/white screening based upon the metabolism, and resulting acidification (creating a pink color based upon a pH shift of the dye) of colonies with restored *galK* function.

We performed ten rounds of recombineering with both the HEM6 pK1F and HEM63 pK1F populations, yielding four populations total: HEM6 pK1F transformed with LPS-targeting oligomers, HEM6 pK1F 'adapted,' HEM63 pK1F transformed with LPS-targeting oligomers, and HEM63 pK1F 'adapted.' We plated dilution cultures after recombineering onto MacConkey agar plates with galactose, IPTG, and spectinomycin to estimate recombineering efficiency (the number of pink colonies/total colonies). We noticed though, that colonies were unable to grow on MacConkey agar supplemented with the normal working concentration of spectinomycin, leading to no growth on pink/white screening plates. The lack of growth is potentially due to a slightly more acidic environment than LS LB, as this has been shown to significantly alter the

toxicity of spectinomycin [178]. An example of reduced *E. coli* growth with the pK1F plasmid on MacConkey agar with spectinomycin compared to growth on LS LB agar with spectinomycin can be found in Appendix 8.13. Blue/white and/or pink/white screening are used to estimate recombineering efficiency by measuring the frequency of loss or gain of a particular phenotype conferred by the replacement of a single point mutation. The efficiencies related to point mutation incorporation and insertion are known to be different, with the point mutation incorporation being favored [174], which is a limitation (i.e., would provide an over-estimation) of using this type of screening for our library design. The unforeseen ability to measure a potentially inaccurate estimation of recombineering efficiency led us to identify another way to estimate efficiency by phenotypic analysis.

After five rounds of recombineering, we had plated diluted cultures to isolate colonies for sequencing to confirm promoter insertion in the targeted populations (Fig. 5.4). We observed a "large" colony phenotype that, when sequenced, confirmed insertion of the pK1F promoter sequence upstream of *lpcA*. This observation was consistent with large colonies we had observed when creating the *lpcA* plasmid in Chapter 3, and enabled us to estimate recombineering efficiency based upon the large colony phenotype as an estimator of promoter insertion in front of LPS targets in the population. This phenotype was observed in ~20 % of the LPS-targeting oligomer library, while no difference in colony size was observed in the 'adapted' population (Fig. 5.4). A similar percentage was observed in the HEM63 pK1F population.

Using the 20% recombineering efficiency estimation from Fig. 5.4, we calculated the predicted degree of mutations in the population. Fitting this information to equations previously derived for population diversity analysis [174], we were able to estimate a recombineering

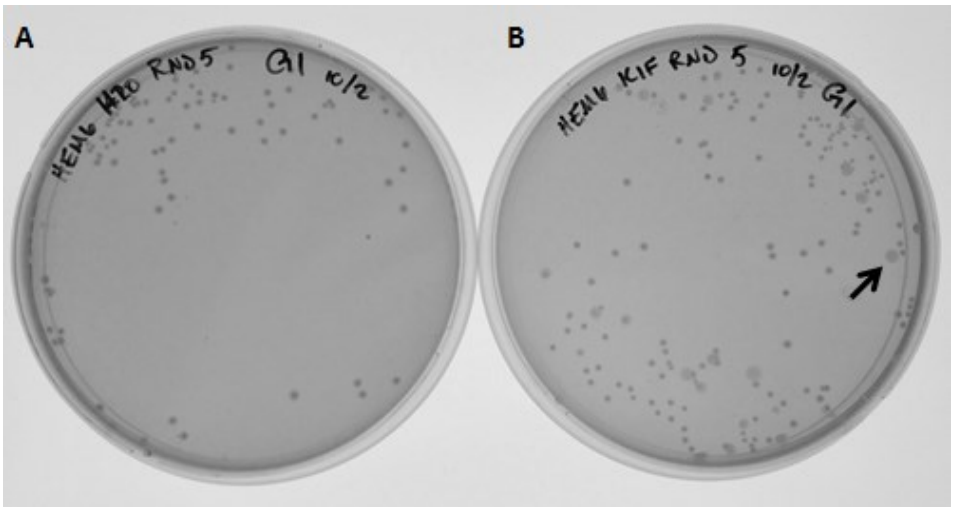


Figure 5.4 – Dilution plates of HEM6 pK1F populations after five rounds of recombineering. A) 'Adapted' culture transformed with only water, and B) Population transformed with LPS-targeting oligos. "Large" colony confirmed for P_{K1F} promoter insertion in front of *lpcA* indicated with arrow.

efficiency of ~5% per round. Our estimates are in good agreement with the predicted recombineering efficiency of a 20 bp insertion based on the same equation derivations (predicted at ~4%). We recognize that this is an estimate that could potentially be conservative, based on the reports that oligomers targeting chromosomal regions within 500,000 bp of each other are more likely to go in during the same recombination event. As seen in Fig. 5.1C, a number of these mutations are targeted in close proximity to each other, and thus might help increase the actual number of recombinants. Based upon 5% recombineering efficiency, Fig. 5.5 shows the estimated population diversity over the course of 10 rounds of recombineering. At this efficiency, it is estimated that 1% of the population has three promoter insertions, 0.1 % of the population has four promoter insertions, while 0.01 % has five mutations.

Although we were unable to perform a pink/white screening on this population, similar to how we had performed blue/white screening with EcNR2, dilution plates from the end of recombineering showed homogeneity of size within the 'adapted' water-control population (similar in appearance to colonies in Fig. 5.4A) in both the HEM6 pK1F and HEM63 pK1F populations, a trait that was not observed in the previous library with the EcNR2 host (Fig. 5.3), suggesting that we maintained a lower level of random mutations with them HEM6 derivatives than with EcNR2, even though HEM63 also is a *mutS* knockout.

5.3.3 – Confirming hydrophobic tolerance of LPS library isolates

Before performing selection on the LPS-targeted populations, we wanted to confirm improved growth characteristics from library isolates. We picked five colonies from each of the

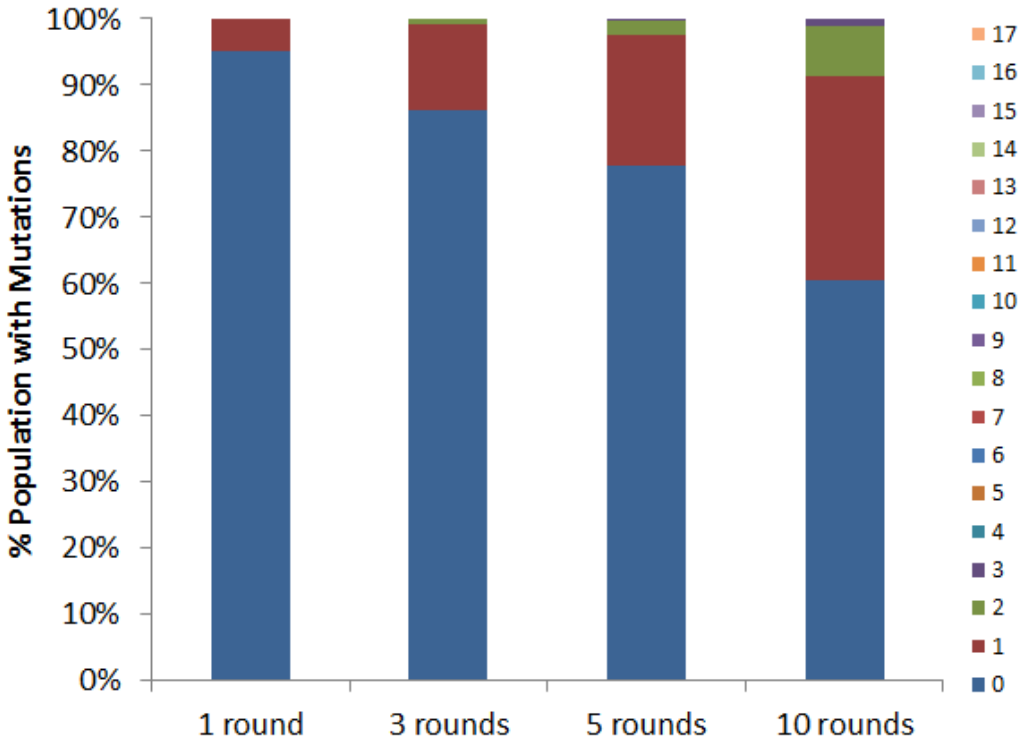


Figure 5.5 – Estimated population diversity of LPS-targeted strains through recursive rounds of recombineering. Recombineering efficiency was estimated at 5% based on phenotypic and sequencing results from isolates after five rounds. The number of mutations within a single cell is indicated by the color of the bar and can range from 0 (wild-type) to 17.

targeted populations and measured their growth with 0.5 v/v % *n*-butanol (Fig. 5.6-7). These isolates were tested against both the parent strain (HEM6 pK1F or HEM63 pK1F) and the 'adapted' population from ten rounds of recombineering with water. This double comparison was performed to assess if mutations propagated within the population ('adapted') during library creation that might contribute to growth in *n*-butanol. All of the five colonies isolated showed improved growth for the HEM6 pK1F LPS-targeted variants (6-10A-E). Isolate 6-10D was 50 ± 7% improved from parent, while the adapted strain also showed improvements of 20 ± 5% improvement, suggesting that spontaneous growth-promoting mutations were introduced into this population.

In comparison, the HEM63 pKIF 'adapted' population was not significantly improved compared to parent ($p > 0.05$), while two of the isolated clones, 10-63B and 10-63C were improved, 24 ± 6% and 33 ± 3 %, respectively (Fig. 5.7). Three of the isolated clones exhibited reduced growth in *n*-butanol, suggesting that not all of the mutation combinations we targeted confer improved growth. Linking PCR has been performed for these strains, and constructs of the expected size have been sent for sequencing to determine the location of the beneficial and detrimental mutations. Figure 5.8 shows an example of the PCR products obtained from the linking PCR reactions for these clones. Individual site Sanger sequencing reactions were also performed, and HEM6-10B and HEM6-10D were found to have K1F promoter insertions upstream of the *waaQ* gene (which is the leading gene in the largest LPS biosynthesis operon, Fig. 5.1-2).

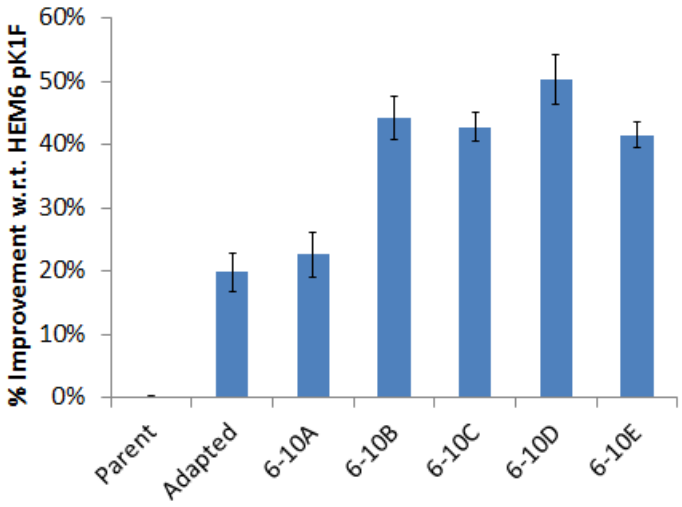


Figure 5.6 – Growth of HEM6 pK1F library isolates in 0.5 v/v % *n*-butanol. HEM6 pK1F and derivatives were inoculated at OD_{600} = 0.04 and grown for six hours at which point optical density was recorded (Parent OD_{600} ~ 0.20 at this time). 'Adapted' refers HEM6 pK1F population that underwent ten rounds of recombineering but was transformed with water as a control for spontaneous mutations. 6-10A-E are isolates from the LPS-targeted library after ten rounds of recombineering. Average improvements (n = 3) are plotted with one standard error.

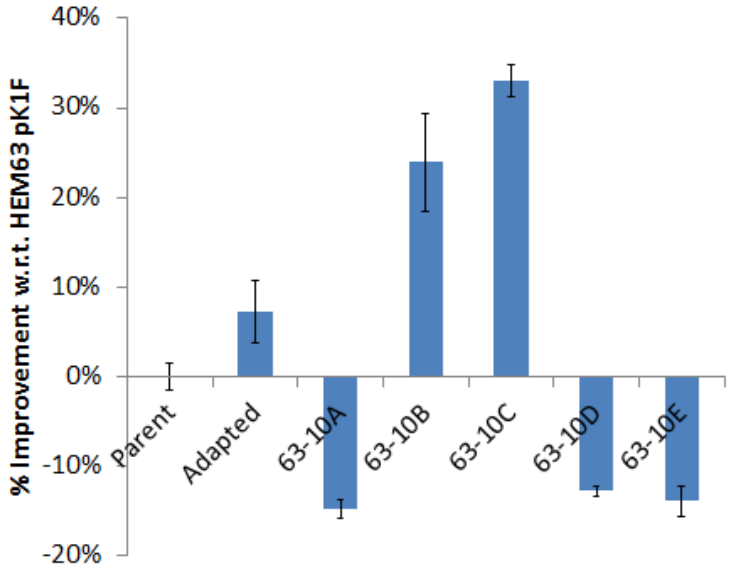


Figure 5.7 – Growth of HEM63 pK1F library isolates in 0.5 v/v % *n*-butanol. HEM63 pK1F and derivatives were inoculated at OD_{600} = 0.04 and grown for six hours at which point optical density was recorded (Parent OD_{600} ~ 0.20 at this time). 'Adapted' refers HEM63 pK1F population that underwent ten rounds of recombineering but was transformed with water as a control for spontaneous mutations. 63-10A-E are isolates from the LPS-targeted library after ten rounds of recombineering. Average improvements (n = 3) are plotted with one standard error.

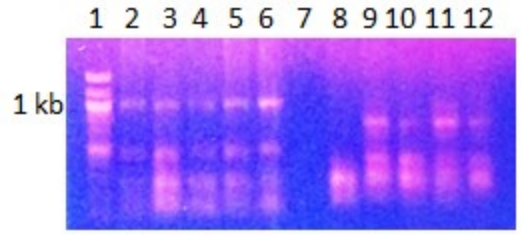


Figure 5.8 – Linking PCR products from 63-10 A-E isolates. Lane 1) ladder with 1 kb marker indicated, 2-6) set 1 targets for 63-10 A-E, respectively. Lane 7 is blank. Lanes 8-12) set 2 targets for 63-10 A-E. Set 1 targets amplify to ~1 kb construct while set 2 targets amplify to ~800 bp. Gel extractions were performed prior to sending samples for sequencing, or in the case of lowly amplified sets (e.g., lane 8), a new amplification reaction was performed.

5.3.4 – Alcohol selections on LPS libraries

Motivated by the promising results for LPS-targeted mutants conferring increased growth under *n*-butanol treatment, we performed multiple selections in parallel on our four populations (both the LPS-targeted populations and the 'adapted' populations for both host strains). Cultures were inoculated into ethanol (3 v/v %), *n*-butanol (0.5 v/v %), isobutanol (0.5 v/v %), or isopentenol (0.5 v/v %). As growth occurred within these cultures, we serially transferred them into subsequent rounds of selection. Figures 5.9-10 shows the optical density measurements of these cultures over time for the selections. The initial round of the selection saw the fastest growth, most likely due to the transfer of some LS LB medium into the MOPS minimal medium used during the selection.

Through the selections, we saw the LPS-targeted populations of HEM6 pK1F and HEM63 pKIF grow ~50 % and ~20% compared to the 'adapted' population in the isobutanol selection after the third round (Fig. 5.9C). We know from the testing performed on the post-recombineering isolates (Fig. 5.7) that not all cells within the LPS-targeted population confer beneficial phenotypes, which might partially explain our findings of slower growth from the LPS-targeted library under the other conditions tested. The isopentenol 'adapted' populations in both cases appear to have greatly improved growth characteristics compared to our targeted library, potentially due to our directed mutations not conferring a tolerant phenotype to this compound, or the success of a spontaneous mutant within the 'adapted' populations. The behavior of the 'adapted' cultures did not distinguish a growth vs. no growth distinction between the HEM6 pK1F and HEM63 pK1F alcohol selections, suggesting that the deletion of *mutS* in the HEM63 host does not confer tolerance to alcohols across the board, but might allow for increased spontaneous mutant propagation.

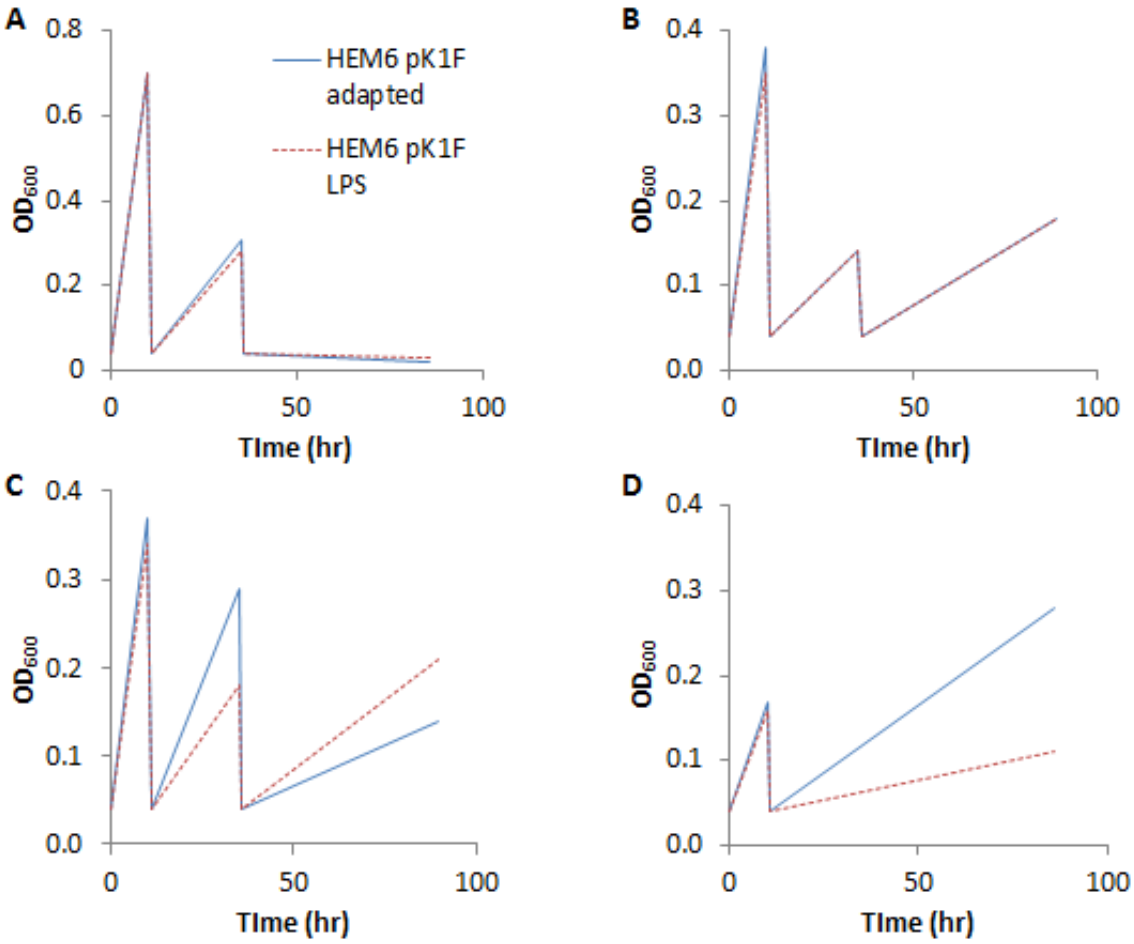


Figure 5.9 – Selection dynamics of HEM6 pK1F libraries in various alcohols. A) Selections in 3 v/v % ethanol, B) 0.5 v/v % *n*-butanol, C) 0.5 v/v % isobutanol, and D) 0.5 v/v % isopentenol. Cultures were grown until entering exponential phase ($OD_{600} > 0.1$) and then transferred into a subsequent batch of inhibitor.

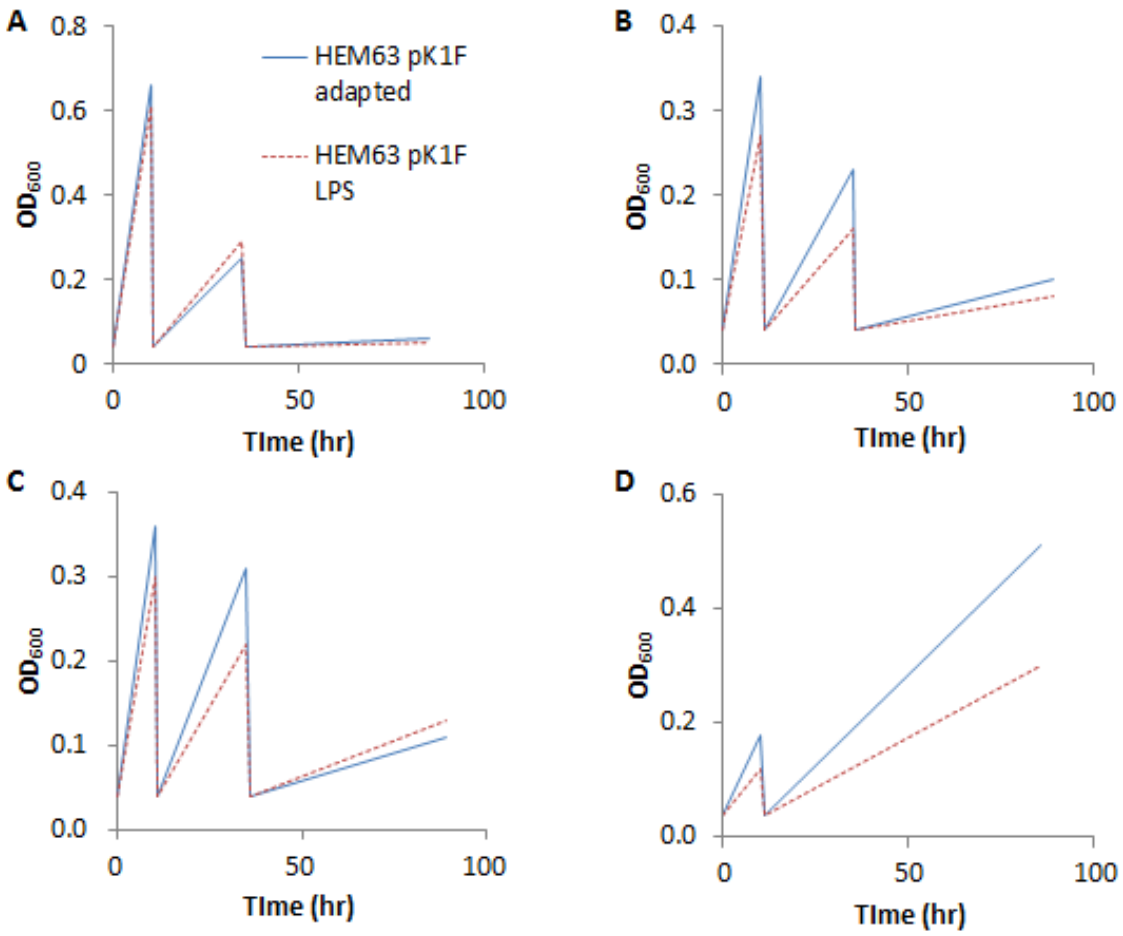


Figure 5.10 – Selection dynamics of HEM63 pK1F libraries in various alcohols. A) Selections in 3 v/v % ethanol, B) 0.5 v/v % *n*-butanol, C) 0.5 v/v % isobutanol, D) 0.5 v/v % isopentenol. Cultures were grown until entering exponential phase ($OD_{600} > 0.1$) and then transferred into a subsequent batch of inhibitor.

LPS-directed mutant isolates were collected from the isobutanol selection. Their performance in MOPS minimal medium (i.e., uninhibited condition) and various alcohols was tested (Fig. 5.11). It was observed that the largest differences in tolerance were between the parent and the 'Adapt/Sel' cultures ('adapted' population after isobutanol selection), rather than from LPS-directed isolates. Similar to our findings before selection (Fig. 5.6-7), it appears that growth-benefitting spontaneous mutations exist within both the control recombineering ('adapted') and selected populations ('Adapt/Sel'), as evidenced by the improvement of growth in MOPS minimal medium. The 'Adapt/Sel' population grew faster than the HEM6 pK1F parent strain under all conditions tested. Slight improvements are observed for some LPS-directed library isolates under some alcohols, but do not surpass the improvements gained by the 'Adapt/Sel' population.

In the future, these libraries will be used to source individual LPS-directed mutants for biofuel tolerance screening. One of the limitations in using recombineering to create multiplexed libraries is the large percentage of the population that remains non-recombineered (i.e., wild-type; as seen after 10 rounds of recombineering in Fig. 5.5). We hypothesized that using a selection would enrich for tolerance-conferring mutations, but enriching for LPS-directed mutations was confounded here by spontaneous mutations obtained through recombineering steps, as well as during selection. Moving forward, we will instead screen the library population for clones exhibiting the "large" colony phenotype (Fig. 5.4) indicative of LPS-directed mutations. This subset of library clones will then be screened for tolerance improvements, and ultimately genotyped to determine locations of the beneficial promoter insertions.

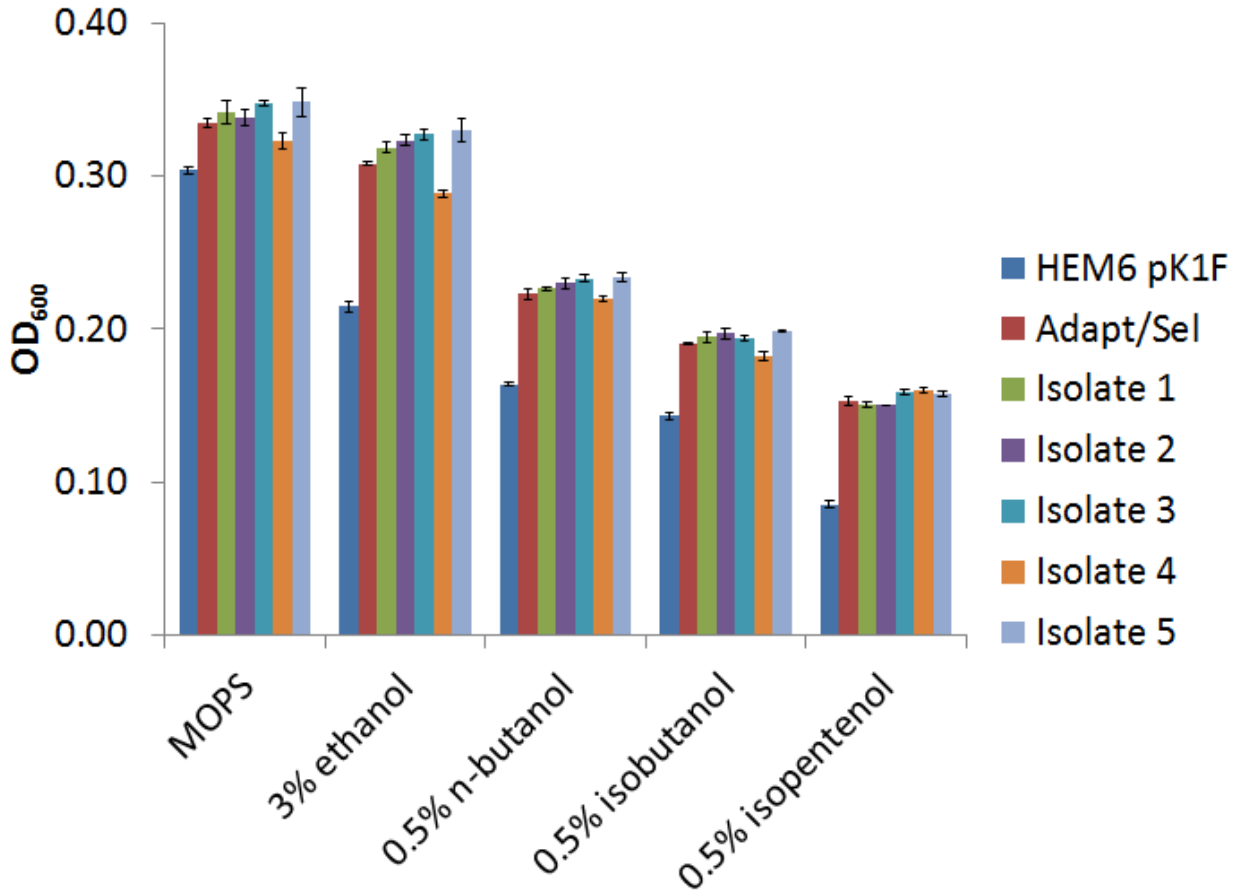


Fig. 5.11 – Growth of HEM6 pK1F selection isolates under various conditions. Cultures were inoculated at normalized cell density (OD_{600} ~0.04) and grown under the listed conditions for 6 hours. Growth was monitored by measuring absorbance at 600 nm (n =3; error bars represent standard error).

5.4 – Conclusions

Engineering tolerance to both hydrolysate inhibitory compounds and proposed biofuel products is important for increasing the viability of cellulosic biofuel production. Here, we analyzed data from the previous chapters to search for enrichments that were applicable to furfural tolerance but might also be beneficial for product tolerance. Previously, in Chapter 3, we had observed that overexpression of *lpcA* conferred improved growth in the presence of furfural, leading us to hypothesize that targeting LPS biosynthesis for altered expression would confer increased tolerance to a wide range of hydrophobic compounds. We created a library of clones with varying degrees of mutations through recursive rounds of recombineering. From this population, we isolated strains that show up to 50% growth improvements in *n*-butanol.

A surprising result was the propagation of spontaneous mutants when we used one of the most common recombineering strains, EcNR2. In order to assess if the deletion of *mutS* would impede our ability to map a genotype-to-phenotype relationship (i.e., promoter insertions at specific LPS biosynthesis operons and their ability to confer tolerance), we employed two recombineering strains that differed only by a deletion of *mutS*. When we analyzed library isolates against internal control 'adapted' populations, we observed that in the case of the *mutS*$^{+}$ strain, a beneficial phenotypic divergence had occurred, suggesting that multiple generations and/or stress incurred by growth, induction, chilling, and electroporation is capable of propagating random mutations. Some of these mutations are considered to confer growth improvements, which confounds our genotype-to-phenotype mapping efforts. It is important to note that improvements attributed to spontaneous mutations occurred even in a *mutS*$^{+}$ strain.

To limit the potential for additional mutations to be introduced within the genomes of these strains while performing selections to enrich for the most tolerant clones, we could look towards employing selection strategies similar to the plate-based selection used in Chapters 3 and 4, thereby reducing the number of generations populations are grown while still identifying beneficial mutations. Alternatively, we are pursuing an approach for screening the LPS-directed library for "large" colony phenotype clones and individually testing them for biofuel tolerance. Ultimately, beneficial mutations identified from our LPS-directed library isolates should be reconstructed (ideally through reduced recombineering steps) or otherwise confirmed to ensure that the benefit is conferred by the LPS-directed mutation and not an uncharacterized spontaneous mutation within the clone's genome.

This study is pioneering the use of a new method developed in the Gill Lab that is designed to track library creation and selection of non-barcoded, chromosomal mutations. Current technologies for genotyping our population, or isolates thereof, fall short in both cost and time compared to this linking PCR approach. Compared to traditional amplification of the region around the mutation for each of our 17 targeted sites, subsequent sample preparation, and individual sequencing reactions, we can reduce the cost by an order of magnitude or more by condensing 17 sites into four PCR reactions (two sets of targets, each with an assembly and an amplification reaction) and a traditional 17 sequencing reactions into four (bidirectional sequencing of each set). Our lab has had success linking larger constructs together, but we are not pursuing that method here since it does not provide increased cost reduction (the normal read length of Sanger sequencing is ~500-700 bases, where a 17-target linked construct would be roughly 2000 bp, thus requiring the same number of sequencing reactions as is in our current design).

Ultimately, our goal in this chapter was to use knowledge gained from the initial Design-Build-Test steps in Chapters 3 and 4 to inform a future design. From our growth assessments in this chapter, we have successfully isolated tolerant clones that were targeted for increased LPS biosynthesis, providing initial confirmation that our efforts to learn from genome-wide searches can direct the engineering of new, but related, phenotypes. Testing for improved cellulosic biofuel production from strains engineered for increased LPS biosynthesis will ultimately discern if this is a viable approach for strain design. We are encouraged by recent reports for ethanol and isobutanol production that strains containing mutations likely to increase LPS production within the cell did not hinder production, and in the case of isobutanol, helped cell viability in stationary phase, where the majority of product is produced [77,179].

Chapter 6

Concluding Remarks

The field of genome engineering is undergoing a renaissance. Enabled by reduced costs in DNA synthesis and sequencing technologies, we are able to ask increasingly complex questions that were unimaginable in previous years. As of 2000, a ground-breaking method for the creation of in-frame individual knockout mutants via homologous recombination in *E. coli* was developed [131]. In just over a decade, the *de novo* synthetic construction of a genome and subsequent cell control by the synthesized genome has occurred [180,181]. These examples highlight the advances of the field obtained in just a few years. Where we used to consider individual mutations by knockouts or overexpression on plasmids, we are now on the road to constructing designer genomes from scratch. We have already reached the capability of making billions of designer mutants in a matter of days [103]. Thus, we are entering a stage where the limitations are no longer the ability to read or write the genome, but instead limited by the capability to test all of this easily created diversity to identify winning strategies. Without effective means for making these determinations, we lose potential winners within the vast mix of diversity we are able to create.

Here, we have applied various tools to search the genome for genotypes of relevance to cellulosic biofuel production. Although many such approaches have been developed, we aimed to use genome-wide searches that would not bias our results towards known mechanisms. Through the first application, using the SCALEs method, we confirmed benefit from genes previously not associated with furfural tolerance. Both of these clones, *lpcA* and *groESL*,

perform different tasks within the cell and reinforce the strength of using genome-wide searches to return non-obvious results.

In our second application, we used the TRMR method to continue searching for alleles that with altered expression might confer tolerance to furfural. By performing analysis between different selection schemes, resulting gene ontology enrichments, and comparing these data with our SCALEs study, we were able to confirm increased furfural tolerance for an additional four genes.

We used information from these two initial studies to direct the design of multiplexed mutants for altered lipopolysaccharide biosynthesis, which we hypothesized would confer improved growth to strains in the presence of hydrophobic biofuel compounds. We isolated LPS-directed mutants with up to 50% growth improvements in *n*-butanol, suggesting a new approach for engineering tolerance to both feedstock inhibitors and biofuel products.

At the outset of this project, the goal was to assess large numbers of possible mutations to find winners for a given phenotype and then use what we learned about the winners to inform future designs, in the hopes that it would allow us to more rapidly engineer a trait that we had previously not analyzed. In order to do this, we used a Design-Build-Test cycle, where initial studies provided a number of results that could be fed into a subsequent Design phase. Our results appear promising, but ultimately relied on years of work here understanding the importance of lipopolysaccharide for tolerance to a variety of compounds. We therefore consider the term ‘rapid’ in a relative sense.

Moving forward, it would be beneficial to create real-time algorithms that could process results from different Test phases (e.g., promoter libraries, open reading frame libraries, and

combinations thereof selected for under a variety of conditions) to perform the learning required to iterate through the Design-Build-Test cycle on a truly rapid scale. If the field is entering an era where we are not limited in the Design and Build, then great efforts should be directed towards learning how to Learn.

Let us not skip the importance of the Test phase though. Without effective, rapid, and trackable testing procedures, we stand to lose all the potential benefits gained by technologies advancing the Design and Build phases (i.e., reduced DNA synthesis and methods for assembly). In the studies here, we observed how the selection scheme affected enrichment patterns. Some selections are more effective than others, and this knowledge can be applied in the future to aid in more rapid and effective Tests.

While the approach used here for our genome engineering applications might be perceived to focus on directing the evolution of our populations, it is also a case study on the evolution of genome engineering strategies. In our initial study, we used the SCALEs approach, which, while providing fairly high mapping resolution and being relatively easy to use, only allowed us to assess the benefit of increased dosage of a particular allele on our desired phenotype. As an alternative, but partially complementary approach, we employed the TRMR method to assess both over and under expression of a gene's effect on the same phenotype. We found overlapping data between the SCALEs and TRMR studies, but when tested head to head, found results highlighting differences between the approaches, and the effect expression level has on a gene's degree of conferred tolerance. Overexpressed or increased expression are terms commonly used in this field, without regard to the actual level—on a case by case basis (since the sequence context matters according to a growing number of reports)—of expression being changed. Our studies reinforce this point and serve as a warning (or perhaps, a partial

explanation) that even though a mutation was identified in one study, it might not perform the expected way in another study when slightly adapted (e.g., chromosomal mutation vs. plasmid-based expression).

Efforts are being made to address this sort of issue with the development of defined biological parts through the use Biobricks, a nationally funded and public-benefit facility (BIOFAB), and a recent call by our own lab for creating a metabolic strains commons of workhouse production strains [179]. The field's willingness to standardize the language in which we perform and share genome engineering efforts stands to have a large impact on its ability to keep pace with the technologies being developed to support it.

It is with profound awe that I close these studies. *E. coli*, and its 4.6 million DNA bases, has provided a fascinating study over the past years spent on this project. Despite the breadth of knowledge we supposedly have about this model organism, it still holds surprises, and finds ways to grow in the wide array of toxic chemicals we throw it in. It is my hope that this writing has presented an overview of factors both limiting and enabling current engineering efforts; with the ultimate hope, that through future studies performed by dedicated members of the field, these limitations will soon be memories, observed only in hindsight.

References

1. Searchinger T, Heimlich R, Houghton RA, Dong F, Elobeid A, et al. (2008) Use of U.S. croplands for biofuels increases greenhouse gases through emissions from land-use change. Science 319: 1238-1240.
2. Slade R, Bauen A, Shah N (2009) The greenhouse gas emissions performance of cellulosic ethanol supply chains in Europe. Biotechnol Biofuels 2: 15.
3. Randy Schnepf BY (2013) Renewable Fuel Standard (RSF): Overview and Issues.
4. Administration USEI (2012) Biofuels Issues and Trends.
5. Kumar D, Murthy GS (2011) Impact of pretreatment and downstream processing technologies on economics and energy in cellulosic ethanol production. Biotechnol Biofuels 4: 27.
6. Blanch HW (2012) Bioprocessing for biofuels. Current Opinion in Biotechnology 23: 390-395.
7. Alper H, Moxley J, Nevoigt E, Fink GR, Stephanopoulos G (2006) Engineering yeast transcription machinery for improved ethanol tolerance and production. Science 314: 1565-1568.
8. Atsumi S, Hanai T, Liao JC (2008) Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. Nature 451: 86-U13.
9. Yomano LP, York SW, Ingram LO (1998) Isolation and characterization of ethanol-tolerant mutants of Escherichia coli KO11 for fuel ethanol production. J Ind Microbiol Biotechnol 20: 132-138.
10. Steen EJ, Kang YS, Bokinsky G, Hu ZH, Schirmer A, et al. (2010) Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. Nature 463: 559-U182.
11. Schirmer A, Rude MA, Li XZ, Popova E, del Cardayre SB (2010) Microbial Biosynthesis of Alkanes. Science 329: 559-562.
12. Wang X, Yomano LP, Lee JY, York SW, Zheng H, et al. (2013) Engineering furfural tolerance in Escherichia coli improves the fermentation of lignocellulosic sugars into renewable chemicals. Proc Natl Acad Sci U S A 110: 4021-4026.
13. Wargacki AJ, Leonard E, Win MN, Regitsky DD, Santos CNS, et al. (2012) An Engineered Microbial Platform for Direct Biofuel Production from Brown Macroalgae. Science 335: 308-313.
14. Dien BS, Cotta MA, Jeffries TW (2003) Bacteria engineered for fuel ethanol production: current status. Appl Microbiol Biotechnol 63: 258-266.
15. Jones DT, Woods DR (1986) Acetone-butanol fermentation revisited. Microbiol Rev 50: 484-524.
16. Servinsky MD, Germane KL, Liu S, Kiel JT, Clark AM, et al. (2012) Arabinose is metabolized via a phosphoketolase pathway in Clostridium acetobutylicum ATCC 824. J Ind Microbiol Biotechnol 39: 1859-1867.
17. Liu L, Zhang L, Tang W, Gu Y, Hua Q, et al. (2012) Phosphoketolase pathway for xylose catabolism in Clostridium acetobutylicum revealed by 13C metabolic flux analysis. J Bacteriol 194: 5413-5422.

18. Zaldivar J, Ingram LO (1999) Effect of organic acids on the growth and fermentation of ethanologenic Escherichia coli LY01. Biotechnol Bioeng 66: 203-210.
19. Zaldivar J, Martinez A, Ingram LO (1999) Effect of selected aldehydes on the growth and fermentation of ethanologenic Escherichia coli. Biotechnol Bioeng 65: 24-33.
20. Zaldivar J, Martinez A, Ingram LO (2000) Effect of alcohol compounds found in hemicellulose hydrolysate on the growth and fermentation of ethanologenic Escherichia coli. Biotechnology and Bioengineering 68: 524-530.
21. Jarboe LR, Grabar TB, Yomano LP, Shanmugan KT, Ingram LO (2007) Development of ethanologenic bacteria. Biofuels 108: 237-261.
22. Larsson S, Palmqvist E, Hahn-Hagerdal B, Tengborg C, Stenberg K, et al. (1999) The generation of fermentation inhibitors during dilute acid hydrolysis of softwood. Enzyme and Microbial Technology 24: 151-159.
23. Larsson S, Reimann A, Nilvebrant NO, Jonsson LJ (1999) Comparison of different methods for the detoxification of lignocellulose hydrolyzates of spruce. Applied Biochemistry and Biotechnology 77-9: 91-103.
24. Ranatunga TD, Jervis J, Helm RF, McMillan JD, Wooley RJ (2000) The effect of overliming on the toxicity of dilute acid pretreated lignocellulosics: the role of inorganics, uronic acids and ether-soluble organics. Enzyme and Microbial Technology 27: 240-247.
25. Martinez A, Rodriguez ME, Wells ML, York SW, Preston JF, et al. (2001) Detoxification of dilute acid hydrolysates of lignocellulose with lime. Biotechnology Progress 17: 287-293.
26. Lopez MJ, Nichols NN, Dien BS, Moreno J, Bothast RJ (2004) Isolation of microorganisms for biological detoxification of lignocellulosic hydrolysates. Applied Microbiology and Biotechnology 64: 125-131.
27. Saha BC, Iten LB, Cotta MA, Wu YV (2005) Dilute acid pretreatment, enzymatic saccharification, and fermentation of rice hulls to ethanol. Biotechnology Progress 21: 816-822.
28. Saha BC, Iten LB, Cotta MA, Wu YV (2005) Dilute acid pretreatment, enzymatic saccharification and fermentation of wheat straw to ethanol. Process Biochemistry 40: 3693-3700.
29. Aden A. RM, Ibsen K., Jechura J., Neeves K., Sheehan J., Wallace B., Montague L., Slayton A., and Lukas J. (2002) Lignocellulosic Biomass to Ethanol Process Design and Economics Utilizing Co-Current Dilute Acid Prehydrolysis and Enzymatic Hydrolysis for Corn Stover. Golden, Colorado: National Renewable Energy Laboratory.
30. Sierra R, A. Smith, C. Granda, M.T. Holtzapple (2008) Producing fuels and chemicals from lignocellulosic biomass. Chemical Engineering Progress 104: S10-S17.
31. Yat SC, Berger A, Shonnard DR (2008) Kinetic characterization for dilute sulfuric acid hydrolysis of timber varieties and switchgrass. Bioresource Technology 99: 3855-3863.

32. Skerker JM, Leon D, Price MN, Mar JS, Tarjan DR, et al. (2013) Dissecting a complex chemical stress: chemogenomic profiling of plant hydrolysates. Mol Syst Biol 9: 674.
33. Miller EN, Turner PC, Jarboe LR, Ingram LO (2010) Genetic changes that increase 5-hydroxymethyl furfural resistance in ethanol-producing Escherichia coli LY180. Biotechnol Lett 32: 661-667.
34. Haselkorn R, Doty P (1961) The reaction of formaldehyde with polynucleotides. J Biol Chem 236: 2738-2745.
35. Sambrook J, and D. W. Russell (2001) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
36. Kummerle N, Feucht HH, Kaulfers PM (1996) Plasmid-mediated formaldehyde resistance in Escherichia coli: characterization of resistance gene. Antimicrob Agents Chemother 40: 2276-2279.
37. Gonzalez CF, Proudfoot M, Brown G, Korniyenko Y, Mori H, et al. (2006) Molecular basis of formaldehyde detoxification. Characterization of two S-formylglutathione hydrolases from Escherichia coli, FrmB and YeiG. J Biol Chem 281: 14514-14522.
38. Azachi M, Henis Y, Shapira R, Oren A (1996) The role of the outer membrane in formaldehyde tolerance in Escherichia coli VU3695 and Halomonas sp. MAC. Microbiology 142 (Pt 5): 1249-1254.
39. Egyud LG, Szent-Gyorgyi A (1966) On the regulation of cell division. Proc Natl Acad Sci U S A 56: 203-207.
40. Egyud LG, Szent-Gyorgyi A (1966) Cell division, SH, ketoaldehydes, and cancer. Proc Natl Acad Sci U S A 55: 388-393.
41. Ingram LO (1976) Adaptation of membrane lipids to alcohols. J Bacteriol 125: 670-678.
42. Zaldivar J, Martinez A, Ingram LO (2000) Effect of alcohol compounds found in hemicellulose hydrolysate on the growth and fermentation of ethanologenic Escherichia coli. Biotechnol Bioeng 68: 524-530.
43. Ingram LO (1976) Adaptation of Membrane Lipids to Alcohols. Journal of Bacteriology 125: 670-678.
44. Banerjee N, Bhatnagar R, Viswanathan L (1981) Inhibition of Glycolysis by Furfural in Saccharomyces cerevisiae. European Journal of Applied Microbiology and Biotechnology 11: 226-228.
45. Modig T, Liden G, Taherzadeh MJ (2002) Inhibition effects of furfural on alcohol dehydrogenase, aldehyde dehydrogenase and pyruvate dehydrogenase. Biochem J 363: 769-776.
46. Gutierrez T, Buszko ML, Ingram LO, Preston JF (2002) Reduction of furfural to furfuryl alcohol by ethanologenic strains of bacteria and its effect on ethanol production from xylose. Appl Biochem Biotechnol 98-100: 327-340.
47. Gutierrez T, Ingram LO, Preston JF (2006) Purification and characterization of a furfural reductase (FFR) from Escherichia coli strain LYO1--an enzyme important in the detoxification of furfural during ethanol production. J Biotechnol 121: 154-164.

48. Miller EN, Jarboe LR, Yomano LP, York SW, Shanmugam KT, et al. (2009) Silencing of NADPH-dependent Oxidoreductases (yqhD and dkgA) in Furfural-Resistant Ethanologenic Escherichia coli. Appl Environ Microbiol.
49. Perez JM, Arenas FA, Pradenas GA, Sandoval JM, Vasquez CC (2008) Escherichia coli YqhD exhibits aldehyde reductase activity and protects from the harmful effect of lipid peroxidation-derived aldehydes. J Biol Chem 283: 7346-7353.
50. Ko J, Kim I, Yoo S, Min B, Kim K, et al. (2005) Conversion of methylglyoxal to acetol by Escherichia coli aldo-keto reductases. J Bacteriol 187: 5782-5789.
51. Modig T, Almeida JR, Gorwa-Grauslund MF, Liden G (2008) Variability of the response of Saccharomyces cerevisiae strains to lignocellulose hydrolysate. Biotechnol Bioeng 100: 423-429.
52. Liu ZL, Moon J, Andersh BJ, Slininger PJ, Weber S (2008) Multiple gene-mediated NAD(P)H-dependent aldehyde reduction is a mechanism of in situ detoxification of furfural and 5-hydroxymethylfurfural by Saccharomyces cerevisiae. Appl Microbiol Biotechnol 81: 743-753.
53. Laadan B, Almeida JR, Radstrom P, Hahn-Hagerdal B, Gorwa-Grauslund M (2008) Identification of an NADH-dependent 5-hydroxymethylfurfural-reducing alcohol dehydrogenase in Saccharomyces cerevisiae. Yeast 25: 191-198.
54. Almeida JR, Roder A, Modig T, Laadan B, Liden G, et al. (2008) NADH- vs NADPH-coupled reduction of 5-hydroxymethyl furfural (HMF) and its implications on product distribution in Saccharomyces cerevisiae. Appl Microbiol Biotechnol 78: 939-945.
55. Almeida JR, Modig T, Roder A, Liden G, Gorwa-Grauslund MF (2008) Pichia stipitis xylose reductase helps detoxifying lignocellulosic hydrolysate by reducing 5-hydroxymethyl-furfural (HMF). Biotechnol Biofuels 1: 12.
56. Miller EN, Jarboe LR, Turner PC, Pharkya P, Yomano LP, et al. (2009) Furfural inhibits growth by limiting sulfur assimilation in ethanologenic Escherichia coli strain LY180. Appl Environ Microbiol 75: 6132-6141.
57. Wang X, Miller EN, Yomano LP, Zhang X, Shanmugam KT, et al. (2011) Increased furfural tolerance due to overexpression of NADH-dependent oxidoreductase FucO in Escherichia coli strains engineered for the production of ethanol and lactate. Appl Environ Microbiol 77: 5132-5140.
58. Zheng H, Wang X, Yomano LP, Geddes RD, Shanmugam KT, et al. (2013) Improving Escherichia coli FucO for Furfural Tolerance by Saturation Mutagenesis of Individual Amino Acid Positions. Appl Environ Microbiol 79: 3202-3208.
59. Wang X, Miller EN, Yomano LP, Shanmugam KT, Ingram LO (2012) Increased furan tolerance in Escherichia coli due to a cryptic ucpA gene. Appl Environ Microbiol 78: 2452-2455.
60. Ranatunga TD, Jervis J, Helm RF, McMillan JD, Hatzis C (1997) Identification of inhibitory components toxic toward Zymomonas mobilis CP4(pZB5) xylose fermentation. Applied Biochemistry and Biotechnology 67: 185-198.

61. Taherzadeh MJ, Gustafsson L, Niklasson C, Liden G (2000) Physiological effects of 5-hydroxymethylfurfural on Saccharomyces cerevisiae. Applied Microbiology and Biotechnology 53: 701-708.
62. Taherzadeh MJ, Gustafsson L, Niklasson C, Liden G (2000) Inhibition effects of furfural on aerobic batch cultivation of Saccharomyces cerevisiae growing on ethanol and/or acetic acid. Journal of Bioscience and Bioengineering 90: 374-380.
63. Hadi SM, Shahabuddin, Rehman A (1989) Specificity of the interaction of furfural with DNA. Mutat Res 225: 101-106.
64. Shahabuddin, Rahman A, Hadi SM (1991) Reaction of furfural and methylfurfural with DNA: use of single-strand-specific nucleases. Food Chem Toxicol 29: 719-721.
65. Khan QA, and S.M. Hadi (1993) Effect of furfural on plasmid DNA. Biochemistry and Molecular Biology International 29: 1153-1160.
66. Zheng H, Wang X, Yomano LP, Shanmugam KT, Ingram LO (2012) Increase in furfural tolerance in ethanologenic Escherichia coli LY180 by plasmid-based expression of thyA. Appl Environ Microbiol 78: 4346-4352.
67. Khan QA, Hadi SM (1994) Inactivation and repair of bacteriophage lambda by furfural. Biochem Mol Biol Int 32: 379-385.
68. Khan QA, Shamsi FA, Hadi SM (1995) Mutagenicity of furfural in plasmid DNA. Cancer Lett 89: 95-99.
69. Allen SA, Clark W, McCaffery JM, Cai Z, Lanctot A, et al. (2010) Furfural induces reactive oxygen species accumulation and cellular damage in Saccharomyces cerevisiae. Biotechnol Biofuels 3: 2.
70. Wang J, Zhang Y, Chen Y, Lin M, Lin Z (2012) Global regulator engineering significantly improved Escherichia coli tolerances toward inhibitors of lignocellulosic hydrolysates. Biotechnol Bioeng 109: 3133-3142.
71. Warner JR, Reeder PJ, Karimpour-Fard A, Woodruff LB, Gill RT (2010) Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. Nat Biotechnol 28: 856-862.
72. Cann AF, Liao JC (2008) Production of 2-methyl-1-butanol in engineered Escherichia coli. Appl Microbiol Biotechnol 81: 89-98.
73. Atsumi S, Cann AF, Connor MR, Shen CR, Smith KM, et al. (2008) Metabolic engineering of Escherichia coli for 1-butanol production. Metab Eng 10: 305-311.
74. Hanai T, Atsumi S, Liao JC (2007) Engineered synthetic pathway for isopropanol production in Escherichia coli. Appl Environ Microbiol 73: 7814-7818.
75. Baez A, Cho KM, Liao JC (2011) High-flux isobutanol production using engineered Escherichia coli: a bioreactor study with in situ product removal. Appl Microbiol Biotechnol 90: 1681-1690.
76. Brynildsen MP, Liao JC (2009) An integrated network approach identifies the isobutanol response network of Escherichia coli. Mol Syst Biol 5: 277.
77. Atsumi S, Wu TY, Machado IM, Huang WC, Chen PY, et al. (2010) Evolution, genomic analysis, and reconstruction of isobutanol tolerance in Escherichia coli. Mol Syst Biol 6: 449.

78. Woodruff LB, Pandhal J, Ow SY, Karimpour-Fard A, Weiss SJ, et al. (2013) Genome-scale identification and characterization of ethanol tolerance genes in Escherichia coli. Metab Eng 15: 124-133.
79. Gonzalez R, Tao H, Purvis JE, York SW, Shanmugam KT, et al. (2003) Gene array-based identification of changes that contribute to ethanol tolerance in ethanologenic Escherichia coli: comparison of KO11 (parent) to LY01 (resistant mutant). Biotechnol Prog 19: 612-623.
80. Dunlop MJ (2011) Engineering microbes for tolerance to next-generation biofuels. Biotechnol Biofuels 4: 32.
81. Dunlop MJ, Dossani ZY, Szmidt HL, Chu HC, Lee TS, et al. (2011) Engineering microbial biofuel tolerance and export using efflux pumps. Mol Syst Biol 7: 487.
82. Sandoval NR, Kim JY, Glebes TY, Reeder PJ, Aucoin HR, et al. (2012) Strategy for directing combinatorial genome engineering in Escherichia coli. Proc Natl Acad Sci U S A 109: 10540-10545.
83. Sandoval NR, Mills TY, Zhang M, Gill RT (2011) Elucidating acetate tolerance in E. coli using a genome-wide approach. Metab Eng 13: 214-224.
84. Roe AJ, O'Byrne C, McLaggan D, Booth IR (2002) Inhibition of Escherichia coli growth by acetic acid: a problem with methionine biosynthesis and homocysteine toxicity. Microbiology 148: 2215-2222.
85. Zingaro KA, Papoutsakis ET (2012) Toward a semisynthetic stress response system to engineer microbial solvent tolerance. MBio 3.
86. Zingaro KA, Terry Papoutsakis E (2013) GroESL overexpression imparts Escherichia coli tolerance to i-, n-, and 2-butanol, 1,2,4-butanetriol and ethanol with complex and unpredictable patterns. Metab Eng 15: 196-205.
87. Maurer LM, Yohannes E, Bondurant SS, Radmacher M, Slonczewski JL (2005) pH regulates genes for flagellar motility, catabolism, and oxidative stress in Escherichia coli K-12. J Bacteriol 187: 304-319.
88. Chong H, Geng H, Zhang H, Song H, Huang L, et al. (2013) Enhancing E. coli isobutanol tolerance through engineering its global transcription factor cAMP receptor protein (CRP). Biotechnol Bioeng.
89. Alper H, Stephanopoulos G (2007) Global transcription machinery engineering: A new approach for improving cellular phenotype. Metabolic Engineering 9: 258-267.
90. Chong H, Yeow J, Wang I, Song H, Jiang R (2013) Improving Acetate Tolerance of Escherichia coli by Rewiring Its Global Regulator cAMP Receptor Protein (CRP). PLoS One 8: e77422.
91. Chong H, Huang L, Yeow J, Wang I, Zhang H, et al. (2013) Improving ethanol tolerance of Escherichia coli by rewiring its global regulator cAMP receptor protein (CRP). PLoS One 8: e57628.
92. Basak S, Jiang R (2012) Enhancing E. coli tolerance towards oxidative stress via engineering its global regulator cAMP receptor protein (CRP). PLoS One 7: e51179.
93. Minty JJ, Lesnefsky AA, Lin F, Chen Y, Zaroff TA, et al. (2011) Evolution combined with genomic study elucidates genetic bases of isobutanol tolerance in Escherichia coli. Microb Cell Fact 10: 18.

94. Smith KM, Liao JC (2011) An evolutionary strategy for isobutanol production strain development in Escherichia coli. Metab Eng 13: 674-681.
95. Connor MR, Cann AF, Liao JC (2010) 3-Methyl-1-butanol production in Escherichia coli: random mutagenesis and two-phase fermentation. Appl Microbiol Biotechnol 86: 1155-1164.
96. Yoon SH, Lee EG, Das A, Lee SH, Li C, et al. (2007) Enhanced vanillin production from recombinant E. coli using NTG mutagenesis and adsorbent resin. Biotechnol Prog 23: 1143-1148.
97. Klein-Marcuschamer D, Stephanopoulos G (2008) Assessing the potential of mutational strategies to elicit new phenotypes in industrial strains. Proc Natl Acad Sci U S A 105: 2319-2324.
98. Bonomo J, Warnecke T, Hume P, Marizcurrena A, Gill RT (2006) A comparative study of metabolic engineering anti-metabolite tolerance in Escherichia coli. Metab Eng 8: 227-239.
99. Badarinarayana V, Estep PW, 3rd, Shendure J, Edwards J, Tavazoie S, et al. (2001) Selection analyses of insertional mutants using subgenic-resolution arrays. Nat Biotechnol 19: 1060-1065.
100. Gill RT, Wildt S, Yang YT, Ziesman S, Stephanopoulos G (2002) Genome-wide screening for trait conferring genes using DNA microarrays. Proc Natl Acad Sci U S A 99: 7033-7038.
101. Woodruff LB, Boyle NR, Gill RT (2013) Engineering improved ethanol production in Escherichia coli with a genome-wide approach. Metab Eng.
102. Nicolaou SA, Gaida SM, Papoutsakis ET (2011) Coexisting/Coexpressing Genomic Libraries (CoGeL) identify interactions among distantly located genetic loci for developing complex microbial phenotypes. Nucleic Acids Res 39: e152.
103. Wang HH, Isaacs FJ, Carr PA, Sun ZZ, Xu G, et al. (2009) Programming cells by multiplex genome engineering and accelerated evolution. Nature 460: 894-898.
104. Costantino N, Court DL (2003) Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. Proc Natl Acad Sci U S A 100: 15748-15753.
105. Ellis HM, Yu D, DiTizio T, Court DL (2001) High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. Proc Natl Acad Sci U S A 98: 6742-6746.
106. Sawitzke JA, Costantino N, Li XT, Thomason LC, Bubunenko M, et al. (2011) Probing cellular processes with oligo-mediated recombination and using the knowledge gained to optimize recombineering. J Mol Biol 407: 45-59.
107. Wang HH, Kim H, Cong L, Jeong J, Bang D, et al. (2012) Genome-scale promoter engineering by coselection MAGE. Nat Methods 9: 591-593.
108. Woodruff LB, Gill RT (2011) Engineering genomes in multiplex. Curr Opin Biotechnol 22: 576-583.
109. Alper H, Miyaoku K, Stephanopoulos G (2005) Construction of lycopene-overproducing E. coli strains by combining systematic and combinatorial gene knockout targets. Nat Biotechnol 23: 612-616.

110. Tyo KE, Ajikumar PK, Stephanopoulos G (2009) Stabilized gene duplication enables long-term selection-free heterologous pathway expression. Nat Biotechnol 27: 760-765.
111. Yim H, Haselbeck R, Niu W, Pujol-Baxley C, Burgard A, et al. (2011) Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol. Nat Chem Biol 7: 445-452.
112. Cho RJ, Fromont-Racine M, Wodicka L, Feierbach B, Stearns T, et al. (1998) Parallel analysis of genetic selections using whole genome oligonucleotide arrays. Proc Natl Acad Sci U S A 95: 3752-3757.
113. Lynch MD, Warnecke T, Gill RT (2007) SCALEs: multiscale analysis of library enrichment. Nat Methods 4: 87-93.
114. Shoemaker DD, Lashkari DA, Morris D, Mittmann M, Davis RW (1996) Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. Nat Genet 14: 450-456.
115. Boyle NR, Gill RT (2012) Tools for genome-wide strain design and construction. Curr Opin Biotechnol 23: 666-671.
116. Isaacs FJ, Carr PA, Wang HH, Lajoie MJ, Sterling B, et al. (2011) Precise manipulation of chromosomes in vivo enables genome-wide codon replacement. Science 333: 348-353.
117. Sandoval NR, Kim JY, Glebes TY, Reeder PJ, Aucoin HR, et al. (2012) Strategy for directing combinatorial genome engineering in Escherichia coli. Proc Natl Acad Sci U S A.
118. Rubin EM (2008) Genomics of cellulosic biofuels. Nature 454: 841-845.
119. Mills TY, Sandoval NR, Gill RT (2009) Cellulosic hydrolysate toxicity and tolerance mechanisms in Escherichia coli. Biotechnol Biofuels 2: 26.
120. Khan QA, Hadi SM (1993) Effect of furfural on plasmid DNA. Biochem Mol Biol Int 29: 1153-1160.
121. Miller EN, Jarboe LR, Yomano LP, York SW, Shanmugam KT, et al. (2009) Silencing of NADPH-dependent oxidoreductase genes (yqhD and dkgA) in furfural-resistant ethanologenic Escherichia coli. Appl Environ Microbiol 75: 4315-4323.
122. Cooke MS, Evans MD, Dizdaroglu M, Lunec J (2003) Oxidative DNA damage: mechanisms, mutation, and disease. FASEB J 17: 1195-1214.
123. Brynildsen MP, Winkler JA, Spina CS, MacDonald IC, Collins JJ (2013) Potentiating antibacterial activity by predictably enhancing endogenous microbial ROS production. Nat Biotechnol 31: 160-165.
124. Bonomo J, Lynch MD, Warnecke T, Price JV, Gill RT (2008) Genome-scale analysis of anti-metabolite directed strain engineering. Metab Eng 10: 109-120.
125. Gall S, Lynch MD, Sandoval NR, Gill RT (2008) Parallel mapping of genotypes to phenotypes contributing to overall biological fitness. Metab Eng 10: 382-393.
126. Warnecke TE, Lynch MD, Karimpour-Fard A, Lipscomb ML, Handke P, et al. (2010) Rapid dissection of a complex phenotype through genomic-scale mapping of fitness altering genes. Metab Eng 12: 241-250.

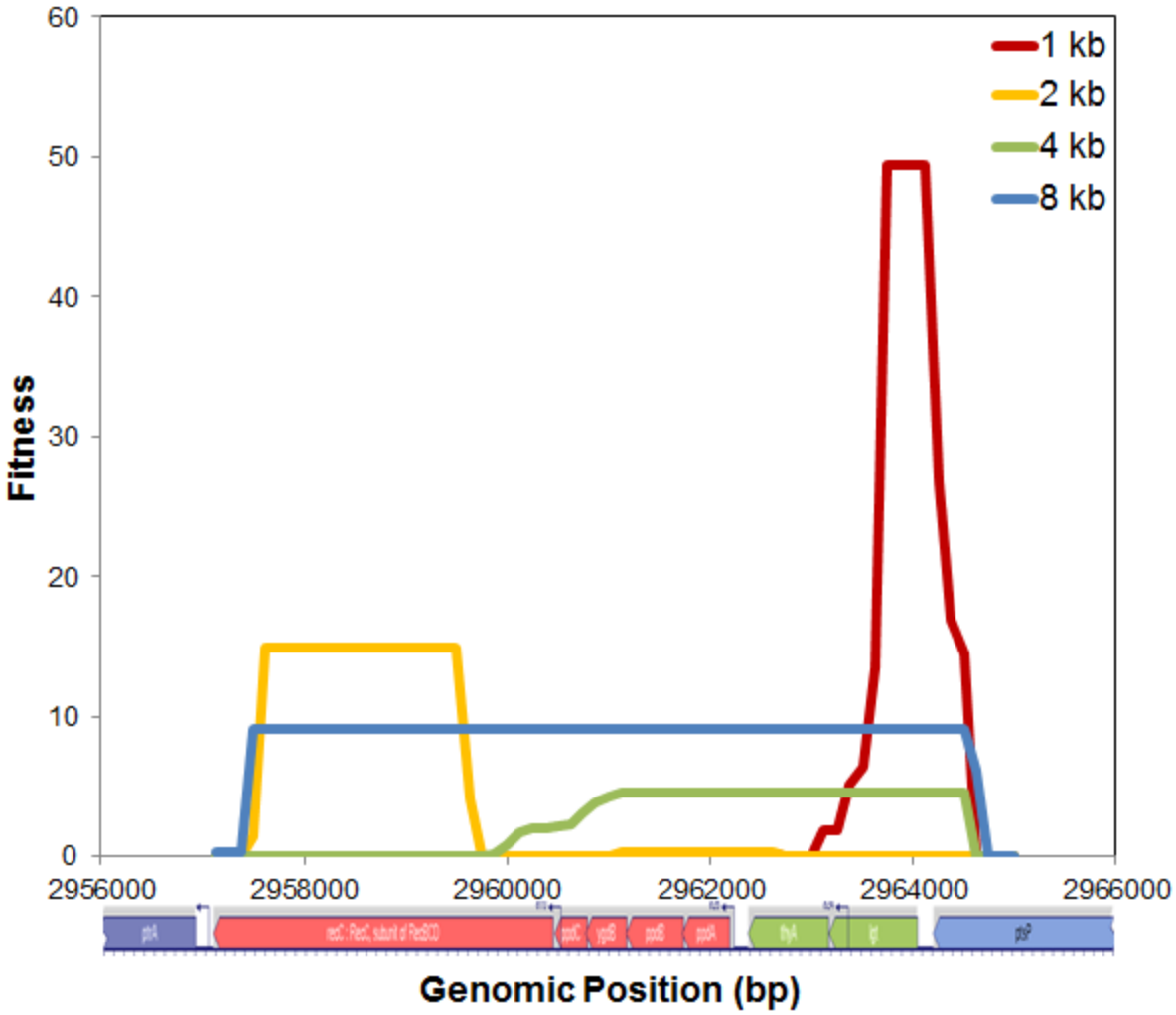
127. Warnecke TE, Lynch MD, Karimpour-Fard A, Sandoval N, Gill RT (2008) A genomics approach to improve the analysis and design of strain selections. Metab Eng 10: 154-165.
128. Struble JM, Gill RT (2009) Genome-scale identification method applied to find cryptic aminoglycoside resistance genes in Pseudomonas aeruginosa. PLoS One 4: e6576.
129. Spindler EC, Boyle NR, Hancock RE, Gill RT (2013) Genome-wide identification of genes conferring energy related resistance to a synthetic antimicrobial Peptide (bac8c). PLoS One 8: e55052.
130. Singh A, Lynch MD, Gill RT (2009) Genes restoring redox balance in fermentation-deficient E. coli NZN111. Metab Eng 11: 347-354.
131. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, et al. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol 2: 2006 0008.
132. Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci U S A 97: 6640-6645.
133. Neidhardt FC, Bloch PL, Smith DF (1974) Culture medium for enterobacteria. J Bacteriol 119: 736-747.
134. Zheng Q, Wang XJ (2008) GOEAST: a web-based software toolkit for Gene Ontology enrichment analysis. Nucleic Acids Res 36: W358-363.
135. Martinez A, Rodriguez ME, York SW, Preston JF, Ingram LO (2000) Use of UV absorbance To monitor furans in dilute acid hydrolysates of biomass. Biotechnol Prog 16: 637-641.
136. Garibyan L, Huang T, Kim M, Wolff E, Nguyen A, et al. (2003) Use of the rpoB gene to determine the specificity of base substitution mutations on the Escherichia coli chromosome. DNA Repair (Amst) 2: 593-608.
137. Singh A, Karimpour-Fard A, Gill RT (2010) Increased mutation frequency in redox-impaired Escherichia coli due to RelA- and RpoS-mediated repression of DNA repair. Appl Environ Microbiol 76: 5463-5470.
138. Zhou K, Zhou L, Lim Q, Zou R, Stephanopoulos G, et al. (2011) Novel reference genes for quantifying transcriptional responses of Escherichia coli to protein overexpression by quantitative PCR. BMC Mol Biol 12: 18.
139. Schnaitman CA, Klena JD (1993) Genetics of lipopolysaccharide biosynthesis in enteric bacteria. Microbiol Rev 57: 655-682.
140. Fralick JA, Burns-Keliher LL (1994) Additive effect of tolC and rfa mutations on the hydrophobic barrier of the outer membrane of Escherichia coli K-12. J Bacteriol 176: 6404-6406.
141. Nikaido H (1976) Outer membrane of Salmonella typhimurium. Transmembrane diffusion of some hydrophobic substances. Biochim Biophys Acta 433: 118-132.
142. Fayet O, Ziegelhoffer T, Georgopoulos C (1989) The groES and groEL heat shock gene products of Escherichia coli are essential for bacterial growth at all temperatures. J Bacteriol 171: 1379-1385.

143. Kerner MJ, Naylor DJ, Ishihama Y, Maier T, Chang HC, et al. (2005) Proteome-wide analysis of chaperonin-dependent protein folding in Escherichia coli. Cell 122: 209-220.
144. Gill RT, Cha HJ, Jain A, Rao G, Bentley WE (1998) Generating controlled reducing environments in aerobic recombinant Escherichia coli fermentations: effects on cell growth, oxygen uptake, heat shock protein expression, and in vivo CAT activity. Biotechnol Bioeng 59: 248-259.
145. Yamamori T, Yura T (1980) Temperature-induced synthesis of specific proteins in Escherichia coli: evidence for transcriptional control. J Bacteriol 142: 843-851.
146. Tokuriki N, Tawfik DS (2009) Chaperonin overexpression promotes genetic variation and enzyme evolution. Nature 459: 668-673.
147. Taylor PL, Blakely KM, de Leon GP, Walker JR, McArthur F, et al. (2008) Structure and function of sedoheptulose-7-phosphate isomerase, a critical enzyme for lipopolysaccharide biosynthesis and a target for antibiotic adjuvants. J Biol Chem 283: 2835-2845.
148. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol 31: 233-239.
149. Karim AS, Curran KA, Alper HS (2013) Characterization of plasmid burden and copy number in Saccharomyces cerevisiae for optimization of metabolic engineering applications. FEMS Yeast Res 13: 107-116.
150. Crook NC, Freeman ES, Alper HS (2011) Re-engineering multicloning sites for function and convenience. Nucleic Acids Res 39: e92.
151. Blazeck J, Garg R, Reed B, Alper HS (2012) Controlling promoter strength and regulation in Saccharomyces cerevisiae using synthetic hybrid promoters. Biotechnol Bioeng 109: 2884-2895.
152. Mutalik VK, Guimaraes JC, Cambray G, Lam C, Christoffersen MJ, et al. (2013) Precise and reliable gene expression via standard transcription and translation initiation elements. Nat Methods 10: 354-360.
153. Marschall C, Labrousse V, Kreimer M, Weichart D, Kolb A, et al. (1998) Molecular analysis of the regulation of csiD, a carbon starvation-inducible gene in Escherichia coli that is exclusively dependent on sigma s and requires activation by cAMP-CRP. J Mol Biol 276: 339-353.
154. White RJ, Pasternak CA (1967) The purification and properties of N-acetylglucosamine 6-phosphate deacetylase from Escherichia coli. Biochem J 105: 121-125.
155. Burnell JN (2010) Cloning and characterization of Escherichia coli DUF299: a bifunctional ADP-dependent kinase--Pi-dependent pyrophosphorylase from bacteria. BMC Biochem 11: 1.
156. Zhang Y, Ezeji TC (2013) Transcriptional analysis of Clostridium beijerinckii NCIMB 8052 to elucidate role of furfural stress during acetone butanol ethanol fermentation. Biotechnol Biofuels 6: 66.
157. Pesavento C, Becker G, Sommerfeldt N, Possling A, Tschowri N, et al. (2008) Inverse regulatory coordination of motility and curli-mediated adhesion in Escherichia coli. Genes Dev 22: 2434-2446.

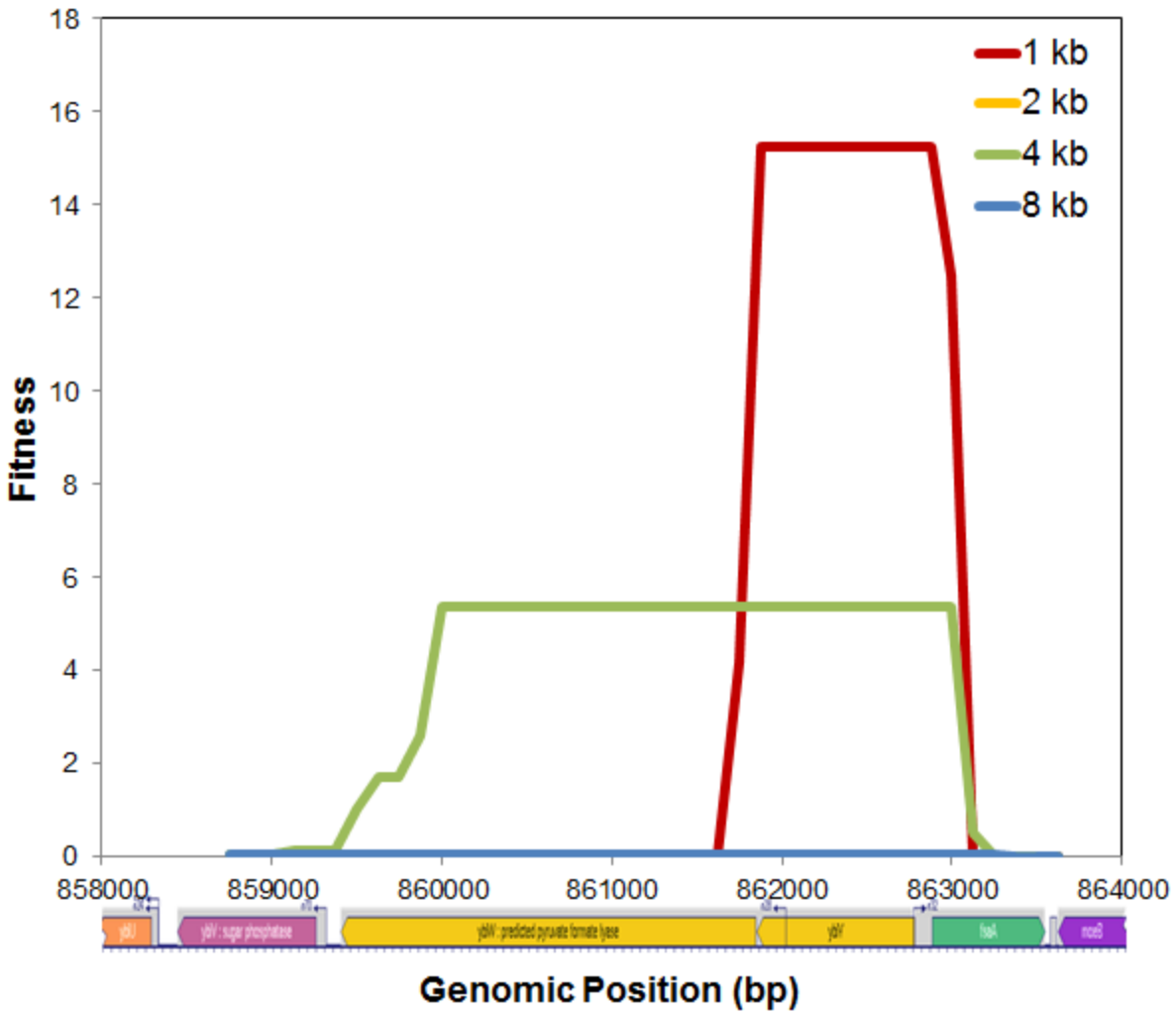
158. Kaplan R, Apirion D (1975) The fate of ribosomes in Escherichia coli cells starved for a carbon source. J Biol Chem 250: 1854-1863.
159. Bejar S, Bouche JP (1985) A new dispensable genetic locus of the terminus region involved in control of cell division in Escherichia coli. Mol Gen Genet 201: 146-150.
160. Johnson JE, Lackner LL, de Boer PA (2002) Targeting of (D)MinC/MinD and (D)MinC/DicB complexes to septal rings in Escherichia coli suggests a multistep mechanism for MinC-mediated destruction of nascent FtsZ rings. J Bacteriol 184: 2951-2962.
161. Seaver LC, Imlay JA (2001) Alkyl hydroperoxide reductase is the primary scavenger of endogenous hydrogen peroxide in Escherichia coli. J Bacteriol 183: 7173-7181.
162. Lutz R, Bujard H (1997) Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. Nucleic Acids Res 25: 1203-1210.
163. Salis HM (2011) The ribosome binding site calculator. Methods Enzymol 498: 19-42.
164. Salis HM, Mirsky EA, Voigt CA (2009) Automated design of synthetic ribosome binding sites to control protein expression. Nat Biotechnol 27: 946-950.
165. Lajoie MJ, Kosuri S, Mosberg JA, Gregg CJ, Zhang D, et al. (2013) Probing the limits of genetic recoding in essential genes. Science 342: 361-363.
166. Brooke JS, Valvano MA (1996) Biosynthesis of inner core lipopolysaccharide in enteric bacteria identification and characterization of a conserved phosphoheptose isomerase. J Biol Chem 271: 3608-3614.
167. Lee JH, Lee KL, Yeo WS, Park SJ, Roe JH (2009) SoxRS-mediated lipopolysaccharide modification enhances resistance against multiple drugs in Escherichia coli. J Bacteriol 191: 4441-4450.
168. Tamaki S, Sato T, Matsuhashi M (1971) Role of lipopolysaccharides in antibiotic resistance and bacteriophage adsorption of Escherichia coli K-12. J Bacteriol 105: 968-975.
169. Ramstedt M, Nakao R, Wai SN, Uhlin BE, Boily JF (2011) Monitoring surface chemical changes in the bacterial cell wall: multivariate analysis of cryo-x-ray photoelectron spectroscopy data. J Biol Chem 286: 12389-12396.
170. Kalamorz F, Reichenbach B, Marz W, Rak B, Gorke B (2007) Feedback control of glucosamine-6-phosphate synthase GlmS expression depends on the small RNA GlmZ and involves the novel protein YhbJ in Escherichia coli. Mol Microbiol 65: 1518-1533.
171. Mosberg JA, Lajoie MJ, Church GM (2010) Lambda red recombineering in Escherichia coli occurs through a fully single-stranded intermediate. Genetics 186: 791-799.
172. Au KG, Welsh K, Modrich P (1992) Initiation of methyl-directed mismatch repair. J Biol Chem 267: 12142-12148.
173. Temme K, Hill R, Segall-Shapiro TH, Moser F, Voigt CA (2012) Modular control of multiple pathways using engineered orthogonal T7 polymerases. Nucleic Acids Res 40: 8773-8781.

174. Wang HH, Church GM (2011) Multiplexed genome engineering and genotyping methods applications for synthetic biology and metabolic engineering. Methods Enzymol 498: 409-426.
175. Lajoie MJ, Rovner AJ, Goodman DB, Aerni HR, Haimovich AD, et al. (2013) Genomically recoded organisms expand biological functions. Science 342: 357-360.
176. Higuchi R, Krummel B, Saiki RK (1988) A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. Nucleic Acids Res 16: 7351-7367.
177. Boyle NR, Reynolds TS, Evans R, Lynch M, Gill RT (2013) Recombineering to homogeneity: extension of multiplex recombineering to large-scale genome editing. Biotechnol J 8: 515-522.
178. Rosenblatt JE, Gerdts AM (1977) Activity of spectinomycin against anaerobes. Antimicrob Agents Chemother 12: 37-39.
179. Woodruff LB, May BL, Warner JR, Gill RT (2013) Towards a metabolic engineering strain "commons": an Escherichia coli platform strain for ethanol production. Biotechnol Bioeng 110: 1520-1526.
180. Gibson DG, Benders GA, Andrews-Pfannkoch C, Denisova EA, Baden-Tillson H, et al. (2008) Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome. Science 319: 1215-1220.
181. Gibson DG, Glass JI, Lartigue C, Noskov VN, Chuang RY, et al. (2010) Creation of a bacterial cell controlled by a chemically synthesized genome. Science 329: 52-56.

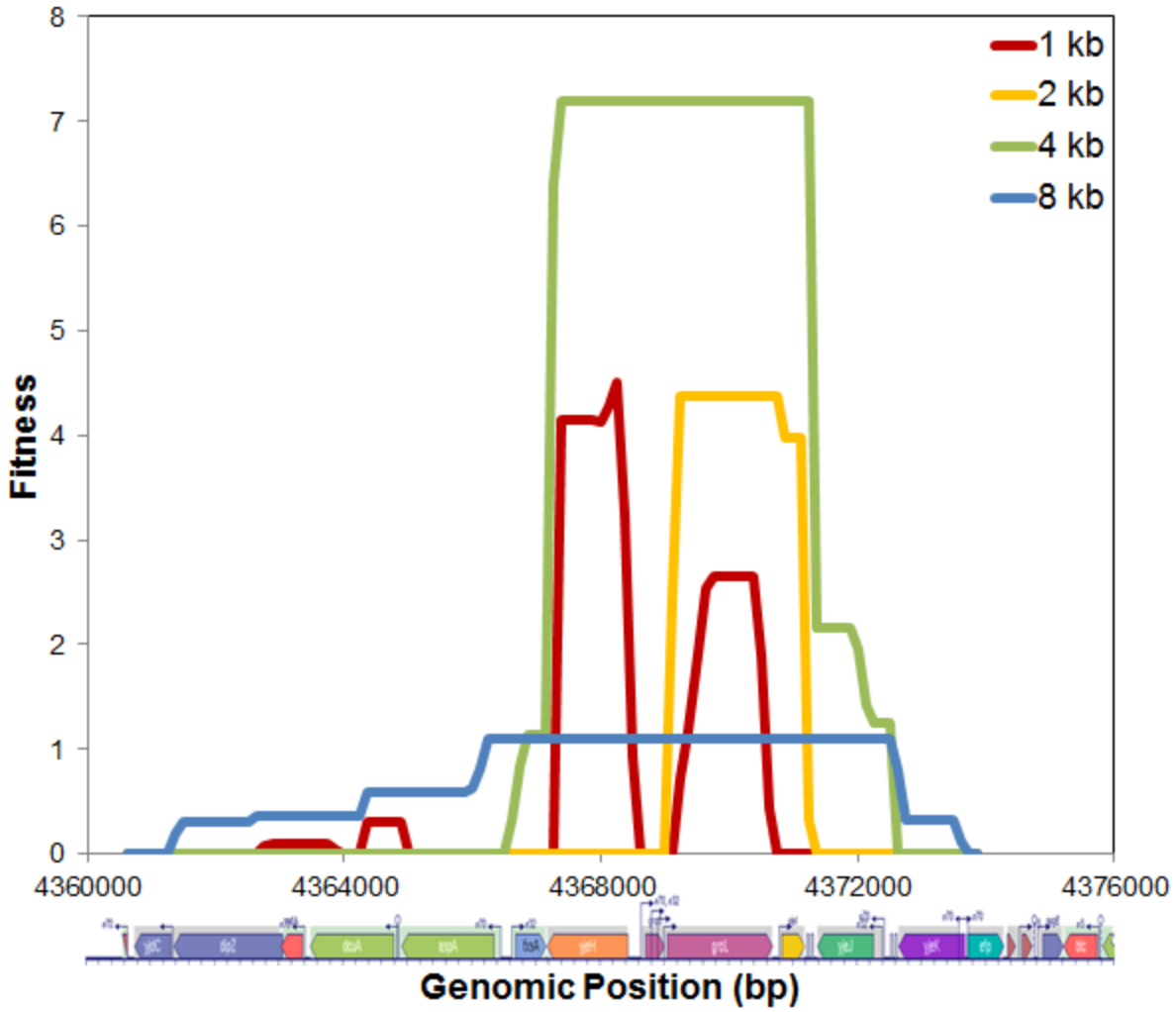
Appendix



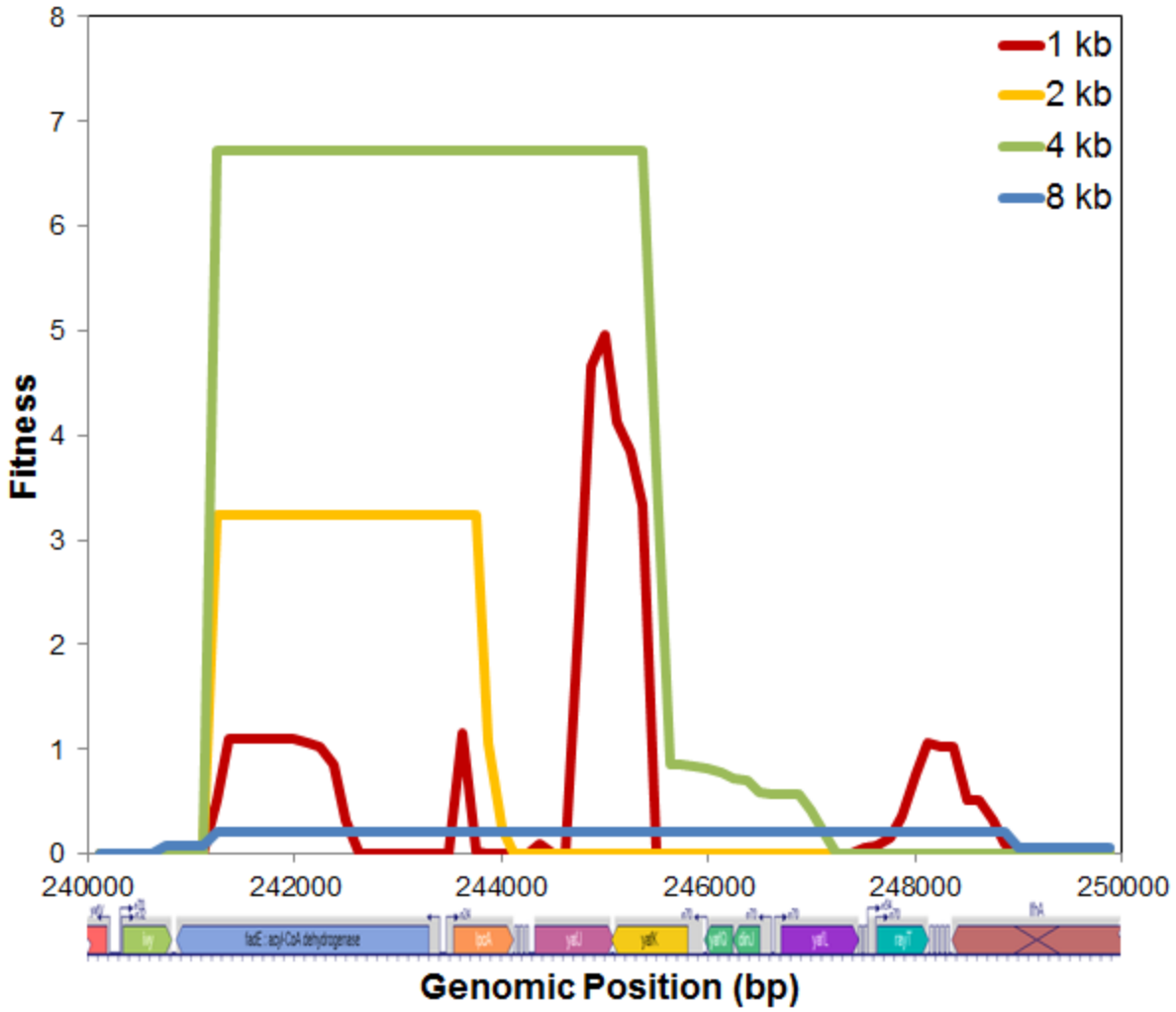
Appendix 8.1 – Blowup region of Locus A from SCALEs furfural selection. Clones fitness scores are shown according to assigned library size (1, 2, 4, or 8 kb). Fitness alignments to the genome are used in the SCALEs processing algorithms to turn clone fitness into gene fitness, based on the amount of gene coverage in a given clone. Genomic position graphic was obtained from ecocyc.org.



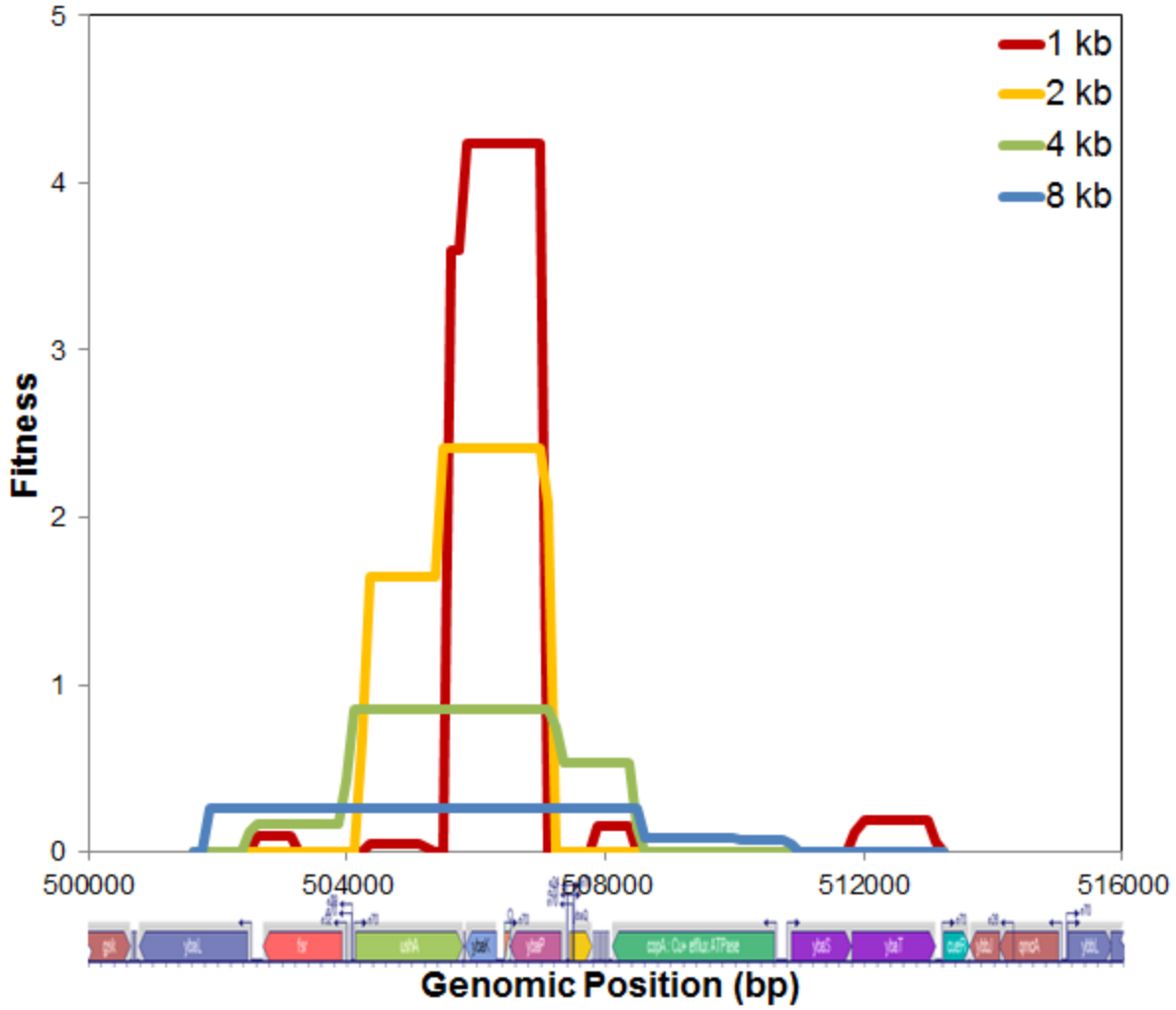
Appendix 8.2 – Blowup region of Locus B from SCALEs furfural selection. Clones fitness scores are shown according to assigned library size (1, 2, 4, or 8 kb). Fitness alignments to the genome are used in the SCALEs processing algorithms to turn clone fitness into gene fitness, based on the amount of gene coverage in a given clone. Genomic position graphic was obtained from ecocyc.org.



Appendix 8.3 – Blowup region of Locus C from SCALEs furfural selection. Clones fitness scores are shown according to assigned library size (1, 2, 4, or 8 kb). Fitness alignments to the genome are used in the SCALEs processing algorithms to turn clone fitness into gene fitness, based on the amount of gene coverage in a given clone. Genomic position graphic was obtained from ecocyc.org.



Appendix 8.4 – Blowup region of Locus D from SCALEs furfural selection. Clones fitness scores are shown according to assigned library size (1, 2, 4, or 8 kb). Fitness alignments to the genome are used in the SCALEs processing algorithms to turn clone fitness into gene fitness, based on the amount of gene coverage in a given clone. Genomic position graphic was obtained from ecocyc.org.



Appendix 8.5 – Blowup region of Locus E from SCALEs furfural selection. Clones fitness scores are shown according to assigned library size (1, 2, 4, or 8 kb). Fitness alignments to the genome are used in the SCALEs processing algorithms to turn clone fitness into gene fitness, based on the amount of gene coverage in a given clone. Genomic position graphic was obtained from ecocyc.org.

Appendix 8.6 – Increased fitness genes from furfural SCALEs selection.

Rank	Gene	ln(Fitness)	Rank	Gene	ln(Fitness)	Rank	Gene	ln(Fitness)
1	lgt	3.586	45	ahpF	1.099	89	mreD	0.591
2	ybiY	3.006	46	ytfF	1.086	90	katE	0.586
3	recC	2.900	47	ybbD	1.079	91	htrG	0.580
4	ppdA	2.704	48	yjeJ	1.070	92	pspA	0.564
5	ppdB	2.694	49	yhdW	1.032	93	yeaB	0.555
6	thyA	2.694	50	insH-9	1.025	94	ftsW	0.547
7	ygdB	2.632	51	uppP	1.020	95	yaiA	0.546
8	groEL	2.600	52	yhdP	1.011	96	ydeP	0.537
9	ppdC	2.521	53	ylbH	0.987	97	yacA	0.534
10	yjeI	2.455	54	ycbR	0.971	98	secB	0.514
11	yjeH	2.352	55	kdgK	0.961	99	fimF	0.510
12	lpcA	2.349	56	ynbC	0.941	100	ygeR	0.509
13	fadE	2.213	57	dmsD	0.939	101	gshB	0.508
14	fsaA	2.211	58	yddW	0.937	102	srmB	0.508
15	yafJ	2.141	59	mutS	0.931	103	fdnH	0.507
16	ptsP	2.140	60	citT	0.924	104	yjgD	0.505
17	groES	2.012	61	pspB	0.899	105	frvX	0.503
18	yafK	2.007	62	ilvB	0.880	106	rffG	0.501
19	ybaK	1.963	63	yjcP	0.867	107	frwD	0.500
20	ygiP	1.782	64	ushA	0.859	108	ade	0.487
21	rna	1.767	65	yqfA	0.835	109	murD	0.487
22	ybaP	1.756	66	ygbI	0.831	110	yncB	0.486
23	rnk	1.754	67	yoaD	0.822	111	ydiO	0.481
24	ygiH	1.656	68	mukF	0.815	112	citG	0.481
25	ybdR	1.632	69	yaeB	0.799	113	yijP	0.473
26	yqhA	1.610	70	uxuR	0.797	114	rffD	0.473
27	yciK	1.599	71	ydgJ	0.786	115	yccW	0.470
28	ybiW	1.556	72	uspG	0.784	116	tolB	0.466
29	yghA	1.475	73	yddV	0.776	117	tpx	0.464
30	yddM	1.326	74	yhjJ	0.747	118	ylbA	0.461
31	dicA	1.299	75	ddl	0.744	119	ygeQ	0.461
32	yicO	1.290	76	pphB	0.743	120	ninE	0.461
33	folB	1.289	77	yhjH	0.712	121	ybcO	0.461
34	adhP	1.283	78	feoA	0.709	122	rffH	0.459
35	yhgF	1.279	79	yghO	0.686	123	ycbQ	0.457
36	fdnI	1.258	80	ydfA	0.669	124	ynbD	0.446
37	murC	1.242	81	ycdS	0.668	125	yedK	0.443
38	yedL	1.215	82	yedP	0.667	126	yiaW	0.440
39	ycdR	1.197	83	ynfH	0.633	127	ylcG	0.432
40	dsrB	1.160	84	yjeO	0.623	128	grxC	0.432
41	proS	1.152	85	ydcW	0.623	129	ahpC	0.420
42	murG	1.139	86	rng	0.608	130	yhjR	0.408
43	dicC	1.124	87	maf	0.604	131	ybiR	0.401
44	nfrA	1.099	88	aroM	0.593	132	ybcQ	0.401

Appendix 8.6, continued.

Rank	Gene	ln(Fitness)	Rank	Gene	ln(Fitness)	Rank	Gene	ln(Fitness)
133	rusA	0.400	177	gpsA	0.251	221	mraY	0.123
134	sixA	0.393	178	gltS	0.240	222	yafY	0.121
135	trkD	0.390	179	cca	0.233	223	alr	0.121
136	yiaV	0.388	180	yijO	0.223	224	eno	0.120
137	yoaE	0.387	181	nei	0.222	225	proW	0.120
138	rffA	0.378	182	murF	0.219	226	yfiB	0.119
139	yfcV	0.375	183	ykfG	0.216	227	ydfX	0.113
140	pal	0.367	184	argH	0.215	228	yafZ	0.110
141	yieN	0.366	185	ptr	0.214	229	ykgF	0.096
142	rffE	0.357	186	ykfH	0.209	230	yibN	0.093
143	ydgD	0.356	187	ynfL	0.209	231	yicE	0.092
144	yncA	0.353	188	yfcU	0.208	232	recG	0.090
145	dacC	0.347	189	dgsA	0.206	233	ydfC	0.090
146	sdaA	0.339	190	hcaF	0.204	234	cbpA	0.088
147	dusB	0.339	191	ygfI	0.201	235	ymfB	0.087
148	ymgF	0.336	192	pbpG	0.198	236	artJ	0.085
149	yjhX	0.332	193	trmD	0.195	237	ytjA	0.080
150	yjiC	0.332	194	lpdA	0.194	238	ilvN	0.076
151	ybgL	0.331	195	ykfF	0.194	239	ycjG	0.076
152	rffC	0.330	196	yedE	0.193	240	yaiE	0.075
153	rcsC	0.319	197	folP	0.192	241	hcaC	0.071
154	ydcY	0.311	198	ynfK	0.186	242	yncC	0.066
155	yafX	0.311	199	ybfH	0.182	243	moaE	0.066
156	yfiD	0.305	200	artP	0.182	244	flhC	0.065
157	yfiK	0.305	201	ygbJ	0.175	245	greB	0.062
158	yggE	0.303	202	rimM	0.174	246	thiD	0.058
159	ydcZ	0.303	203	rpsP	0.170	247	yiiM	0.057
160	ymfC	0.296	204	yhjQ	0.169	248	potH	0.051
161	focB	0.289	205	ycbB	0.168	249	ygbM	0.048
162	ribA	0.286	206	uhpB	0.166	250	ybjF	0.041
163	mviN	0.285	207	wcaC	0.162	251	yafQ	0.037
164	hcaB	0.285	208	ftsQ	0.161	252	btuR	0.037
165	yedF	0.278	209	uhpA	0.158	253	ybiS	0.035
166	yihG	0.276	210	ygfO	0.154	254	yohC	0.033
167	frvR	0.272	211	ybgF	0.153	255	pppA	0.032
168	ttdA	0.270	212	acpD	0.146	256	ychP	0.031
169	potI	0.270	213	yjgJ	0.135	257	yiiL	0.026
170	ydfB	0.263	214	yggA	0.133	258	wzxE	0.026
171	fxsA	0.262	215	mdoH	0.131	259	yieM	0.024
172	kdpB	0.260	216	slt	0.131	260	rhaB	0.022
173	murE	0.260	217	yliD	0.129	261	speD	0.020
174	ydcX	0.255	218	pflC	0.129	262	fimH	0.015
175	yjjU	0.253	219	dos	0.129	263	ymjA	0.013
176	flhE	0.253	220	yccV	0.129	264	allC	0.012

Appendix 8.6, continued.

Rank	Gene	ln(Fitness)
265	yfiC	0.009
266	yghG	0.008
267	yebQ	0.008
268	yliC	0.002

Appendix 8.7 – Top 100 increased fitness 'Up' genes from plate-based TRMR selection.

Rank	Gene	ln(Fitness)	Rank	Gene	ln(Fitness)	Rank	Gene	ln(Fitness)
1	csiD	6.748	45	ychE	3.089	89	yebQ	1.993
2	talB	6.649	46	sulA	2.963	90	ydeM	1.984
3	ydaL	6.276	47	cyoC	2.799	91	ymiA	1.968
4	yeeN	6.153	48	ydiJ	2.748	92	tufA	1.963
5	smg	6.023	49	yedX	2.740	93	amyA	1.958
6	yneG	5.755	50	ahpC	2.739	94	ninE	1.954
7	yjaH	5.567	51	ybgF	2.676	95	flgC	1.932
8	acpT	5.439	52	metG	2.668	96	ydeQ	1.930
9	rplI	5.367	53	yihY	2.658	97	rpmI	1.927
10	yccF	5.198	54	ybiA	2.651	98	umuC	1.920
11	yohK	5.008	55	rimM	2.633	99	mioC	1.897
12	atpF	4.992	56	zntA	2.583	100	secG	1.846
13	pdhR	4.969	57	glpC	2.552			
14	ycjU	4.896	58	rstB	2.546			
15	yhfS	4.657	59	trxC	2.540			
16	ybgA	4.370	60	paaE	2.512			
17	fhuC	4.301	61	ybgQ	2.499			
18	yadG	4.143	62	shiA	2.437			
19	yciZ	4.023	63	ynjF	2.435			
20	yfdV	4.014	64	yihR	2.407			
21	amiB	4.001	65	sdaA	2.352			
22	tonB	3.969	66	yfiB	2.348			
23	xisE	3.891	67	adiC	2.343			
24	dnaG	3.792	68	yhdE	2.332			
25	aceF	3.745	69	nfsB	2.259			
26	talA	3.725	70	ybcV	2.252			
27	ynfO	3.698	71	ydiH	2.239			
28	lpxB	3.641	72	rimN	2.229			
29	envR	3.637	73	glpK	2.172			
30	dctA	3.620	74	rep	2.160			
31	ycdU	3.447	75	pepB	2.160			
32	ybdL	3.417	76	rsuA	2.150			
33	lolB	3.409	77	thiL	2.130			
34	ydeK	3.407	78	zipA	2.117			
35	creD	3.393	79	gfcA	2.104			
36	sodA	3.374	80	ymcE	2.099			
37	appA	3.281	81	yejG	2.096			
38	yajQ	3.279	82	ilvC	2.086			
39	ydcN	3.157	83	yjaA	2.083			
40	dsbA	3.152	84	yceO	2.074			
41	ybhL	3.150	85	rhlB	2.064			
42	yadL	3.145	86	tdk	2.064			
43	gatC	3.133	87	ompF	2.061			
44	ybiU	3.123	88	yhbY	2.013			

Appendix 8.8 – Top 100 increased fitness 'Down' genes from plate-based TRMR selection.

Rank	Gene	ln(Fitness)	Rank	Gene	ln(Fitness)	Rank	Gene	ln(Fitness)
1	ddpF	8.356	45	ychQ	3.889	89	hcaT	2.452
2	nagA	7.210	46	gabT	3.860	90	ycgN	2.427
3	ydiA	7.049	47	ydfK	3.838	91	sad	2.386
4	ycjN	6.613	48	yicR	3.766	92	yhhS	2.383
5	rseB	6.361	49	dmsD	3.651	93	yaiY	2.371
6	yohN	6.075	50	ycbS	3.565	94	ddpC	2.362
7	sdaA	5.898	51	bioC	3.551	95	yihO	2.352
8	yecT	5.699	52	frmR	3.501	96	phnD	2.350
9	yahF	5.687	53	yqcE	3.501	97	ilvB	2.349
10	potE	5.176	54	yebO	3.497	98	yejL	2.312
11	yfbU	5.109	55	ybjX	3.489	99	yqaA	2.289
12	gfcB	5.011	56	ghrA	3.437	100	htpX	2.283
13	paaI	4.999	57	yqcC	3.412			
14	hisB	4.868	58	yfgD	3.358			
15	ybhR	4.838	59	yjbE	3.217			
16	nagC	4.822	60	hslJ	3.203			
17	fliO	4.815	61	idi	3.142			
18	gadB	4.782	62	garL	3.136			
19	moaA	4.777	63	pflA	3.089			
20	ydcQ	4.715	64	yoaI	3.079			
21	yceF	4.579	65	lsrD	3.068			
22	ydaV	4.514	66	yidK	3.032			
23	eutJ	4.502	67	recC	3.022			
24	deoA	4.444	68	ycjF	3.012			
25	pldB	4.442	69	sra	2.973			
26	ybeT	4.435	70	pspD	2.932			
27	lpxM	4.414	71	ulaE	2.920			
28	ldcA	4.319	72	flgM	2.825			
29	feaR	4.315	73	sgcA	2.784			
30	dinJ	4.287	74	rplI	2.782			
31	ydeA	4.287	75	hdhA	2.769			
32	ymgF	4.269	76	ybfQ	2.757			
33	dinB	4.182	77	mdlB	2.731			
34	yedP	4.176	78	yaeP	2.669			
35	ydjA	4.157	79	yiaV	2.665			
36	ydeO	4.155	80	gcvA	2.658			
37	yraQ	4.093	81	argE	2.574			
38	fliE	4.029	82	yhjV	2.551			
39	ahpC	4.014	83	yijO	2.534			
40	fabH	3.971	84	pspB	2.500			
41	yjgM	3.925	85	yqiJ	2.493			
42	cmoA	3.907	86	rlmL	2.481			
43	ilvM	3.899	87	fliA	2.467			
44	dacB	3.895	88	ydaU	2.456			

Appendix 8.9– Top 100 increased fitness 'Up' genes from decreasing TRMR selection.

Rank	Gene	ln(Fitness)	Rank	Gene	ln(Fitness)	Rank	Gene	ln(Fitness)
1	yqcC	11.901	45	lpxB	3.174	89	plsB	2.178
2	smg	11.842	46	ydfH	3.094	90	umuC	2.168
3	yihY	10.949	47	yohK	3.063	91	tufB	2.165
4	yqhD	9.910	48	nfsB	3.049	92	uidA	2.154
5	yhdE	8.743	49	rdoA	3.037	93	mreC	2.145
6	zipA	8.203	50	ybiU	2.998	94	rstB	2.136
7	alaS	7.848	51	metR	2.992	95	ydiJ	2.133
8	talB	7.004	52	ilvE	2.983	96	fecC	2.114
9	yeeN	6.858	53	yjjX	2.887	97	zntA	2.092
10	csiD	6.531	54	ynbE	2.845	98	yciE	2.084
11	yfeW	6.112	55	yhhM	2.827	99	yceO	2.058
12	aceE	6.076	56	lolB	2.756	100	pgi	2.050
13	tufA	5.578	57	dctA	2.742			
14	talA	5.545	58	atpF	2.731			
15	yneG	5.371	59	ydcN	2.716			
16	xseA	5.320	60	soxS	2.670			
17	mdtM	4.988	61	fliN	2.665			
18	xisE	4.649	62	ynfO	2.652			
19	tonB	4.640	63	ygaZ	2.634			
20	metG	4.542	64	ilvM	2.612			
21	yccF	4.340	65	yfiB	2.598			
22	yciZ	4.338	66	ynjF	2.585			
23	yihX	4.240	67	nagC	2.563			
24	yfbV	4.074	68	yraI	2.562			
25	surA	3.971	69	rplM	2.535			
26	ddpA	3.958	70	shiA	2.456			
27	guaB	3.832	71	mdaB	2.446			
28	ycjU	3.689	72	gfcA	2.440			
29	uof	3.605	73	yceD	2.412			
30	nhaA	3.567	74	ydiH	2.388			
31	yjaH	3.542	75	sspA	2.342			
32	ydaL	3.523	76	glpC	2.334			
33	fhuC	3.517	77	hns	2.327			
34	tesA	3.495	78	dsbA	2.323			
35	yedX	3.489	79	yfaL	2.306			
36	yafC	3.447	80	appA	2.304			
37	sulA	3.394	81	rof	2.298			
38	gatC	3.382	82	amyA	2.286			
39	secG	3.348	83	araG	2.283			
40	trpD	3.253	84	yegS	2.277			
41	rep	3.239	85	fdhE	2.274			
42	yhfS	3.233	86	ycfJ	2.239			
43	ybgA	3.232	87	yceJ	2.200			
44	yajQ	3.231	88	mltC	2.179			

Appendix 8.10 – Top 100 increased fitness 'Down' genes from decreasing TRMR selection.

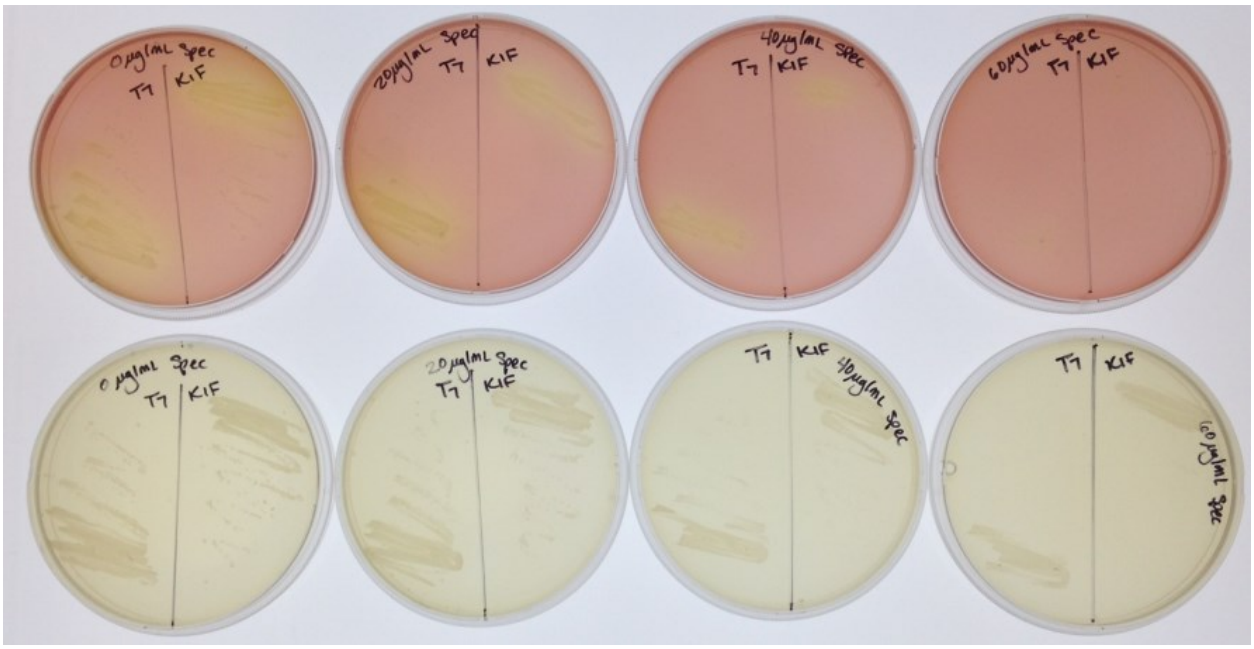
Rank	Gene	ln(Fitness)	Rank	Gene	ln(Fitness)	Rank	Gene	ln(Fitness)
1	fruK	10.885	45	yfeZ	4.116	89	cspA	3.034
2	yaaX	10.031	46	ulaE	4.053	90	yidK	3.030
3	ddpF	9.833	47	ychQ	4.053	91	ddpC	3.009
4	nagA	9.522	48	trpL	4.050	92	yaiE	2.997
5	nagC	9.034	49	dacB	4.047	93	narV	2.968
6	ybjL	7.737	50	ybhR	4.005	94	rimL	2.965
7	pldB	7.464	51	sra	3.909	95	yehE	2.960
8	ydiA	7.449	52	feaR	3.907	96	murB	2.948
9	ilvM	6.412	53	ymgF	3.901	97	rsmE	2.920
10	pdhR	6.311	54	dinJ	3.851	98	prpE	2.896
11	yohN	5.856	55	gcvA	3.768	99	ygbJ	2.871
12	potE	5.758	56	yqiJ	3.716	100	cspG	2.852
13	fabH	5.754	57	hisB	3.682			
14	yhdE	5.711	58	puuE	3.673			
15	gfcB	5.475	59	frmR	3.670			
16	ycjF	5.426	60	yfbU	3.661			
17	yecT	5.300	61	yciH	3.654			
18	rseB	5.227	62	cpxR	3.639			
19	mscS	5.075	63	hslJ	3.627			
20	sdaA	5.061	64	recC	3.562			
21	paaI	4.971	65	yjbE	3.558			
22	tufB	4.902	66	yciV	3.541			
23	plsX	4.888	67	serS	3.472			
24	ycjN	4.841	68	gabT	3.467			
25	ydbL	4.821	69	yqiA	3.461			
26	yjdM	4.816	70	tufA	3.410			
27	thrB	4.732	71	ydaV	3.382			
28	pyrL	4.704	72	ybjX	3.365			
29	moaA	4.700	73	ydfK	3.347			
30	cmoA	4.637	74	yqjB	3.269			
31	yedP	4.614	75	yqcE	3.245			
32	yihO	4.553	76	yhjV	3.244			
33	yoaI	4.528	77	ybfQ	3.231			
34	ydcQ	4.526	78	dusB	3.149			
35	yfgD	4.488	79	pspB	3.146			
36	yjgM	4.443	80	yqcC	3.136			
37	ycbS	4.437	81	pspD	3.127			
38	ybeT	4.334	82	lsrD	3.123			
39	bioC	4.280	83	ydeO	3.113			
40	gadB	4.257	84	idi	3.112			
41	yicR	4.190	85	ydcP	3.097			
42	eutJ	4.189	86	yhhS	3.089			
43	yebO	4.187	87	yeaX	3.077			
44	dinB	4.145	88	fliE	3.049			

Appendix 8.11 – Top 100 increased fitness 'Up' genes from constant TRMR selection.

Rank	Gene	ln(Fitness)	Rank	Gene	ln(Fitness)	Rank	Gene	ln(Fitness)
1	ydaL	5.870	45	amiA	2.397	89	lpxB	1.868
2	dhaK	5.005	46	yceJ	2.387	90	ycgZ	1.845
3	ycjU	4.168	47	ybdL	2.361	91	ynfG	1.843
4	tonB	4.131	48	fhuC	2.357	92	ychJ	1.834
5	ddpA	3.950	49	ybgA	2.334	93	ypfJ	1.800
6	yneG	3.514	50	umuC	2.329	94	smg	1.777
7	rplI	3.489	51	nhoA	2.324	95	pphA	1.776
8	csiD	3.389	52	frsA	2.301	96	pdhR	1.774
9	ydiJ	3.227	53	yadL	2.301	97	atpF	1.773
10	yohK	3.218	54	amyA	2.300	98	lpp	1.772
11	ybiU	3.206	55	ydeQ	2.285	99	secG	1.770
12	ydeK	3.072	56	appA	2.281	100	acpP	1.768
13	nfsB	3.049	57	yhdE	2.279			
14	yedX	3.038	58	ybcV	2.259			
15	yeeN	3.026	59	yhfS	2.249			
16	yajQ	2.964	60	mdtN	2.159			
17	ynfO	2.958	61	yjbF	2.147			
18	ydcN	2.929	62	yejG	2.120			
19	fliC	2.928	63	eutL	2.114			
20	ychE	2.920	64	talA	2.114			
21	yebQ	2.919	65	yihY	2.109			
22	rstB	2.900	66	sapF	2.096			
23	cpxP	2.847	67	cyoC	2.080			
24	aceF	2.756	68	ompF	2.068			
25	yccF	2.747	69	malY	2.062			
26	gatC	2.730	70	ybhL	2.045			
27	amiB	2.719	71	araG	2.028			
28	sieB	2.671	72	gfcA	2.020			
29	sdaA	2.663	73	yhbY	2.020			
30	sulA	2.661	74	znuC	2.019			
31	rimM	2.654	75	glpC	2.012			
32	priC	2.628	76	ydcX	2.005			
33	envR	2.624	77	yiiR	2.004			
34	ydiH	2.623	78	thiL	1.984			
35	ynbE	2.532	79	lolC	1.982			
36	yfiB	2.511	80	xisE	1.970			
37	shiA	2.501	81	soxS	1.970			
38	ycdU	2.457	82	rpmI	1.955			
39	ymcE	2.453	83	dnaG	1.948			
40	lolB	2.444	84	flgC	1.942			
41	trxC	2.422	85	yadG	1.934			
42	ybgF	2.417	86	ykfM	1.925			
43	ymiA	2.409	87	yceO	1.921			
44	fliN	2.401	88	clpB	1.899			

Appendix 8.12 – Top 100 increased fitness 'Down' genes from constant TRMR selection.

Rank	Gene	ln(Fitness)	Rank	Gene	ln(Fitness)	Rank	Gene	ln(Fitness)
1	aspS	5.218	45	dacB	2.631	89	yciZ	1.959
2	ydiA	4.808	46	yceF	2.629	90	queA	1.954
3	sdaA	4.765	47	ydfK	2.619	91	yjbE	1.950
4	ycjN	4.385	48	trpL	2.585	92	ydfZ	1.943
5	ygcP	4.352	49	nuoC	2.553	93	lsrF	1.939
6	yecT	4.245	50	rrmA	2.498	94	sgcA	1.935
7	frmR	4.130	51	ydaV	2.487	95	yqaA	1.935
8	sra	4.086	52	ydgU	2.480	96	yjgM	1.924
9	paaI	3.978	53	ppx	2.407	97	idi	1.915
10	fliO	3.954	54	ymgF	2.398	98	yqiJ	1.911
11	moaA	3.936	55	ybfQ	2.389	99	ybfM	1.911
12	yedP	3.824	56	dmsD	2.370	100	sbcB	1.903
13	gfcB	3.775	57	ydeO	2.353			
14	yohN	3.679	58	yqcE	2.346			
15	rseB	3.650	59	entD	2.346			
16	ddpF	3.617	60	fliF	2.344			
17	yncE	3.544	61	dinB	2.321			
18	ycjF	3.539	62	narV	2.300			
19	yoaI	3.506	63	puuE	2.300			
20	potE	3.493	64	bioC	2.300			
21	ychQ	3.491	65	mgrB	2.268			
22	feaR	3.410	66	recC	2.246			
23	lpxM	3.356	67	ybjX	2.229			
24	pldB	3.337	68	yeaX	2.223			
25	ybhR	3.213	69	pspD	2.151			
26	ydbL	3.211	70	ycjG	2.146			
27	yfbU	3.163	71	lacZ	2.118			
28	hslJ	3.093	72	sppA	2.105			
29	ahpC	3.030	73	yihO	2.103			
30	ybeT	3.021	74	tufB	2.099			
31	ydcQ	3.005	75	gabT	2.087			
32	potI	2.975	76	uidA	2.076			
33	eutJ	2.966	77	ymgD	2.076			
34	lsrD	2.947	78	yaiY	2.074			
35	ycbS	2.933	79	yqcC	2.062			
36	yfgD	2.896	80	hyfF	2.039			
37	cmoA	2.872	81	rzoR	2.019			
38	gadB	2.847	82	flgD	2.013			
39	fliE	2.842	83	ydjA	2.011			
40	ldcA	2.839	84	yebO	2.004			
41	fabH	2.759	85	yicR	1.994			
42	yaeP	2.715	86	htpX	1.993			
43	dinJ	2.639	87	pspB	1.978			
44	ydeA	2.635	88	yraQ	1.976			



Appendix 8.13 – Effect on growth of increasing spectinomycin concentrations with different culture medium. Top row of plates are filled with MacConkey agar, used for pink/white screening of galactose metabolism. Bottom row are plates filled with LS LB agar. Cultures of *E. coli* SIMD70 pT7 (left side of all plates) or SIMD70 pK1F (right side of all plates) were streaked onto solid medium with varying levels of spectinomycin concentrations: plates on the left have 0 µg/ml, the second column has 20 µg/ml, the third column has 40 µg/ml, and the right column has 60 µg/ml. Normal working concentration of spectinomycin is 100 µg/ml.