OVERCOMING CLONAL INTERFERENCE IN ESCHERICHIA COLI USING GENDERLESS HIGH FREQUENCY RECOMBINATION STRAINS

A Dissertation

by

JAMES DANIEL WINKLER

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Katy C. Kao
Committee Members,	Zhilei Chen
	Arul Jayaraman
	Michael Manson
Head of Department,	M. Nazmul Karim

May 2014

Major Subject: Chemical Engineering

ABSTRACT

Adaptive laboratory evolution (ALE) is a powerful tool for strain improvement, and has been applied successfully to improve a range of desirable phenotypes in model organisms through continuous cultivation under a selective pressure of interest. Despite its demonstrable utility, one limiting factor for the effectiveness of ALE is competition between beneficial mutants that exist contemporaneously within an evolving population. This phenomenon of *clonal interference* arises from the fact that the majority of microbes are obligate asexual organisms that cannot exchange DNA between cells. Mutants that arise must therefore compete for resources until the fittest mutant drives the others to extinction. The resulting loss of genetic information from the population slows the overall rate of adaptation, and decreases the amount of information that can be extracted from a given ALE experiment.

To overcome these limitations, we have developed a novel *in situ* mating system based on the F plasmid to allow continuous DNA exchange between *E. coli* cells in liquid culture, allowing mutants to potentially combine their mutations into a single genetic background. The utility and limitations of an existing recombination method, genome shuffling, are also explored to demonstrate the advantages of this new method. The design and initial testing of the *in situ* mating system is first validated, and the system is used for a subsequent evolution experiment under osmotic stress to validate the industrial applicability of the mating system. Adaptive mutants generated in the course of these experiments are then used to test whether tolerant mutants can be formed via conjugation. Finally, additional side projects focusing on strain or population characterization tools are discussed, followed by recommendations for future work.

ACKNOWLEDGEMENTS

There are many people I would like to thank for their support during graduate school. First and foremost are my parents, Brad and Sarah, and siblings and sisters-in-law: Matthew, Jeff, Chris, Vanessa, Shawn, and Kim. I also appreciate the many scientific insights I have gained through conversations with my nieces, Sutton, Sloan, and soon Antonella, especially those regarding solutions to complex puzzles and modeling mass transportation systems. Guidance from the rest of the Winkler and Batt families has also been a great help over the years while I have been pursuing this degree.

In addition to my family, I must acknowledge the past efforts by my many teachers and professors to both educate me in the specifics of math, physics, and chemical engineering, and help me engage with the broader scientific world. I specifically thank Caroline Ferron, J. Kirsten Statt, and Pamela Liebman for laying the foundation for my later studies in engineering and social science (and their seemingly endless patience with teenagers). Several Rice professors also had a profound impact on my approach to science and the world in general: S. Lisa Biswal, Leonardo Dueñas-Osorio, Robert (Bob) Stein, Devika Subramanian, and Rick Wilson, and deserve recognition for their efforts. I would also like to thank Ratna Sharma-Shivappa for my initial introduction into the world of biotechnology and pointing out where free ice cream was handed out on NC State's campus back in 2008, and my Chinese teacher Meng Gang for his encouragement.

Although there are too many people to list, I would like to specifically thank some of my friends (in no particular order) for their stalwart support over the years before and during graduate school: Nine Lam, Josh M. Levin, Alec Walker, (Vanessa)

Johnson, Laura Timmerman, Ting Liu, Jun Axup, Sophie Bonifaz, Luvena Ong, and Xiaoxue Wang. I also appreciate the helpful conversations with Cory Klemashevich, Ana Maria Chamoun, Miguel Ramirez, Chi-Chou Lin, and Nandita Kohli over the years as well.

Finally, I thank my colleagues in the Kao laboratory for their assistance, advice, and allowing me to borrow endless reagents during our time together, as well as my advisor for her guidance and support. Several interns have also contributed significantly to the work presented here and deserve recognition for their tireless efforts: Carlos Garcia, Quynh Nguyen, and Emily Callaway. I also would like to extend my gratitude to my committee members for taking the time to advise me as well.

As scientists, we stand on the shoulders of giants who came before us, and so I must acknowledge my intellectual debt to those researchers past who have laid the foundation upon which I have built my own research. I specifically would like to thank C. Gross, A. Typas, and J. Cronan for providing bacterial strains, D. Siegele for providing P1vir lysate for transduction, and C. Potgieter and D. Hyduke for advice on modeling and statistical analysis.

TABLE OF CONTENTS

Page

ABSTRACT ii

ACKNOWLEDGEMENTS iii

TABLE OF CONTENTS v

LIST OF FIGURES ix

LIST OF TABLES xvi

1. INTRODUCTION 1

1.1 What is Adaptive Laboratory Evolution? 1
1.2 Microbial Sex and Evolution 6
1.3 Conjugation: a Natural Method for Cell-Cell DNA Transfer 9
1.4 Designing a Sexual Evolution System for *E. coli* 13
1.5 Outline 14

2. NOVEL *ESCHERICHIA COLI* HYBRIDS WITH ENHANCED BUTANOL TOLERANCE 16

2.1 Summary 16
2.2 Introduction 16
2.3 Methods and Materials 17
2.3.1 Bacterial strains and media 17
2.3.2 Protoplast fusion 18
2.3.3 Measurements of growth kinetics 18
2.3.4 16s rDNA determination 19
2.4 Results and Discussion 19
2.4.1 Generation of *E. coli*-*L. brevis* hybrids 19
2.4.2 Hybrid characterization 20
2.4.3 Butanol tolerance in the hybrids 22
2.5 Conclusions 24

3. HARNESSING RECOMBINATION TO SPEED ADAPTIVE EVOLUTION IN *ESCHERICHIA COLI* 25

3.1 Summary 25
3.2 Introduction 25
3.3 Methods and Materials 29
3.3.1 Strain construction 29

3.3.2 Strain characterization . . . 30
3.3.3 Conjugation efficiency assays . . . 32
3.3.4 Luria-Delbruck fluctuation tests . . . 32
3.3.5 Evolution experiments . . . 33
3.4 Results and Discussion . . . 35
3.4.1 Strain characterization . . . 35
3.5 Evolution Experiments . . . 39
3.5.1 Streptomycin . . . 41
3.5.2 Chloramphenicol . . . 44
3.5.3 Trimethoprim . . . 46
3.5.4 Implications for evolutionary engineering . . . 48
3.6 Conclusions . . . 49

4. UNRAVELING OSMOTOLOLERANCE IN *ESCHERICHIA COLI* USING A PSEUDOSEXUAL MODEL SYSTEM . . . 52

4.1 Summary . . . 52
4.2 Introduction . . . 52
4.3 Methods and Materials . . . 55
4.3.1 Bacterial strains and growth media . . . 55
4.3.2 Adaptive evolution . . . 55
4.3.3 Mutant isolation and screening . . . 57
4.3.4 Fitness distribution analysis . . . 58
4.3.5 Mutation rate under osmotic stress . . . 58
4.3.6 Hyperosmotic shock tests . . . 59
4.3.7 Genome sequencing and verification . . . 59
4.3.8 Transcriptional analysis . . . 60
4.3.9 Screening genes of interest . . . 61
4.3.10 Indole measurements . . . 62
4.4 Results and Discussion . . . 63
4.4.1 Adaptive evolution under NaCl challenge . . . 63
4.4.2 Mutant characterization . . . 65
4.4.3 Osmotic tolerance and other complex phenotypes . . . 67
4.4.4 Genetic patterns of adaptation . . . 70
4.4.5 Mutation-induced transcriptional perturbations . . . 78
4.5 Conclusions . . . 86

5. TRANSCRIPTIONAL RESPONSE OF *LACTOBACILLUS BREVIS* TO N-BUTANOL AND FERULIC ACID STRESSES . . . 89

5.1 Summary . . . 89
5.2 Introduction . . . 90
5.3 Materials and Methods . . . 93
5.3.1 Sigma factor identification, cloning, and expression . . . 93
5.3.2 Bacterial strains and growth conditions for sigma factor expression . . . 94

5.3.3 Bacterial strains and growth conditions for transcriptional analysis . . . 95
5.3.4 Extraction of total RNA . . . 96
5.3.5 Labeled cDNA generation and microarray hybridization . . . 96
5.3.6 Microarray data analysis . . . 97
5.3.7 Transcriptional target screening . . . 98
5.3.8 Fatty acid methyl ester analysis . . . 98
5.4 Results and Discussion . . . 98
5.4.1 Transcriptional response to ferulic acid stress . . . 99
5.4.2 Transcriptional response to n-butanol stress . . . 103
5.4.3 Developing a high-throughput screening tool for Lactobacillus genes . . . 111
5.5 Conclusions . . . 114
6. COMPUTATIONAL IDENTIFICATION OF ADAPTIVE MUTANTS USING THE VERT SYSTEM . . . 116
6.1 Summary . . . 116
6.2 Introduction . . . 116
6.3 Methods and Materials . . . 118
6.3.1 Experimental procedures . . . 118
6.3.2 Computational procedures . . . 119
6.4 Results and Discussion . . . 119
6.4.1 Statistical classification of population dynamics data . . . 120
6.4.2 Definition of the population state model . . . 123
6.4.3 Properties of the population state model . . . 127
6.4.4 Example application: analysis of a yeast chemostat . . . 128
6.4.5 Distribution of adaptive events . . . 131
6.4.6 Application to other evolution systems . . . 133
6.5 Conclusions . . . 133
7. HARNESSING FLUX BALANCE ANALYSIS FOR SIMULATING BIOLOGICAL EVOLUTION . . . 135
7.1 Summary . . . 135
7.2 Introduction . . . 136
7.3 Methods and Materials . . . 138
7.3.1 Software packages and data sources . . . 138
7.3.2 Genotype definition . . . 139
7.3.3 Evolution simulation . . . 140
7.4 Results and Discussion . . . 141
7.4.1 Simulator properties . . . 141
7.4.2 Evolution in non-optimal media . . . 143
7.4.3 Evolution under stress conditions . . . 145
7.4.4 Reducing the target sequence space . . . 151
7.4.5 Targets for further development . . . 153

7.5 Conclusions 154
8. CONCLUSIONS AND RECOMMENDATIONS 156
8.1 Summary of Results 156
8.2 Recommendations for Further Research 157
REFERENCES 159

LIST OF FIGURES

FIGURE Page

1.1 Two examples of artificial adaptive evolution. A) Domestication of silver foxes. The initially feral foxes are selected for breeding based on their interactions with human handlers. Photo credits to Zefram and Dan Child (BBC). After a sufficient number of generations, many of the foxes behave similarly to domesticated canines. B) Evolution of a microbe in the laboratory. The microbe is repeatedly propagated under selection (nutrient utilization or chemical inhibitors) until improved mutants are isolated. 2

1.2 Cartoon depicting the evolution of large A) asexual and B) sexual populations under an arbitrary selective pressure. The genome in this case consists of five alleles with two possible variants (*abcde* versus *ABCDE*) Clonal interference in the asexual populations eliminates adaptive mutations that are potentially informative but have lower fitness than other mutants in the population. Genotypes can also be combined only through sequential mutation, limiting the pace of adaptation. Sexual populations, in contrast, are able to use genetic recombination to generate new and varied genotypes competing lineages as they arise. Both populations suffer from stochastic loss of beneficial mutants (genotype *aBCde*). Adapted from [74]. 6

1.3 The process of conjugation between F^+ and F^- *E. coli* cells. A. The F^+ cell uses its sex pilus to "ratchet" the F^- cell until contact is made between their membranes. B. A mating bridge is formed, allowing for DNA transfer to begin between the cells. C. A single-stranded DNA copy of the F plasmid is transferred to the recipient, along with primase (required to synthesize the dsDNA plasmid) and several other proteins. D. The mating pair dissolves, leaving two F^+ cells. 11

2.1 Inhibition of growth by butanol in LB medium at a. 30 °C and b. 37 °C. Values are based on four replicate measurement for each strain at each condition. Error bars are 95% confidence intervals. Hybrids with growth inhibition levels significantly less ($P < 0.05$, Student's t test) than *E. coli* at each butanol concentration are marked with an asterisk. Note that the growth of KKOB6 at 37 °C in 1.5% butanol was only observed in three out of the four replicates. 23

3.1 A summary of the general F-mediated conjugation process in *E. coli*. **A.** The pilus of the male (Hfr) strain, which contains an F plasmid integrated into the genome, latches onto a F^- recipient. **B.** After cell-cell contact is made, a mating bridge is formed to allow for stable DNA transfer. Note that *traS* and *traT* are thought to inhibit this step by interfering with the mating bridge. **C.** Single-stranded DNA is transferred from the donor to recipient from the origin of transfer on the donor chromosome. **D.** Once transfer is complete or disrupted in some way (mechanically, chemically), the donor and recipient separate. Any transferred DNA may be integrated into the recipient chromosome via homologous recombination. 29

3.2 Basic design of the Hfr strains used in this study. The strain background is BW25113. Points A and C denote integrated oriT sites with transfer directionality (arrows) inferred from the orientation relative to the original F plasmid; Point B is an integrated F conjugated from a BW25113 strain obtained from CAG31031 [324]. 36

3.3 Growth of Hfr-2xSFX+, 2xSFX-, BW25113 2xOriT, and BW25113 strain in M9 minimal media supplemented with 0.5% glucose (w/v) and 25 μg/ml tryptophan. The Hfr-2xSFX- strain grows somewhat more slowly (6-17%) than Hfr-2xSFX+ and 2xOriT. 37

3.4 Number of recombinant progeny obtained when using Hfr-2xSFX+,- as donors to SFX+,- recipients (X892 and XΔ). Hfr-2xSFX- was significantly better at mating with SFX- recipients compared to Hfr-2xSFX+. Note that surface exclusion on the donor can also disrupt mating, as TraS and TraT interact with recipient proteins (TraG and OmpA [120]) to prevent Hfr-Hfr matings. When mating with the X892 Hfr strain, many fewer recombinant progeny are generated by both strains, though the Hfr-2xSFX- is still more efficient at conjugation in this case. 38

3.5 The average concentration of streptomycin during the serial batch evolution experiment. Six replicates are used per strain. Given that large differences in replicate antibiotic concentrations correspond to only one or two step-wise increases, error bars are omitted for clarity. The p-values for doubling time comparisons are: Hfr 2xSFX+ versus Hfr 2xSFX-: $P = 0.21$, Hfr 2xSFX+ versus 2xOriT: $P = 0.23$, and Hfr 2xSFX- versus 2xOriT: $P < 0.01$. 42

3.6 A. The average concentration of chloramphenicol during the serial batch evolution experiment. Six replicates are used per strain. Errors are omitted due to the large amount of variation caused by the exponential ramping scheme. B. Antibiotic concentrations in each replicate. Note that not all replicates are visible due to overlapping antibiotic concentration curves over the course of the experiment. P-values for each doubling time comparison are as follows: Hfr 2xSFX+ versus Hfr 2xSFX-: $P < 0.014$, Hfr 2xSFX+ versus 2xOriT: $P = 0.21$, and Hfr 2xSFX- versus 2xOriT: $P < 0.012$. 45

3.7 The average concentration of trimethoprim (TM) during the serial batch evolution experiment. Six replicates are used per strain. Error bars are also omitted in this case due to the exponential ramping scheme. 47

4.1 Average fitness improvements for the Hfr-2xSFX-, 2xOriT, and Hfr-2xSFX+ populations, relative to the 2xOriT parent strain during adaptive evolution in the presence of high sodium chloride concentrations. Fitness is defined as $S = \mu_{\text{pop}}/\mu_{\text{2xOriT}} - 1$ in this case. Error bars are 95% confidence intervals using the Student's t distribution. . 64

4.2 Histogram of relative fitness (to 2xOriT) in randomly screened isolates from six Hfr-2xSFX- and 2xOriT populations. All isolates were challenged with 0.65 M NaCl. Differences in the underlying fitness distributions in the Hfr-2xSFX- and 2xOriT populations are highly significant ($P = 1.81 \cdot 10^{-4}$, Kolmogorov-Smirnov test). 65

4.3 Improvement (ratios of inhibition) of Hfr-2xSFX-, 2xOriT, and Hfr-2xSFX+ isolates during growth under 0.65 M NaCl relative to the parental 2xOriT strain, after normalizing for difference in unchallenged growth rates. Error bars are 95% confidence intervals using the Student's t-distribution. 67

4.4 Relative fitness of Hfr-2xSFX-, 2xOriT, and Hfr-2xSFX+ isolates under several abiotic stressors, including A) 0.65 M NaCl, B) 0.3 M glucose, C) 0.8% n-butanol, D) growth at 42 °C, and E) pH 6. Error bars are 95% confidence intervals using the Student's t-distribution. . 69

4.5 Summary of n-acetylglucoasmine (NAG) catabolism in *E. coli*. Pathway information adapted from Peri et al. [249], Yadav et al. [354], and Ecocyc [166]. Metabolite abbreviations: NAG, n-acetylglucosamine, NAG-6P, NAG-6-phosphate, GlcN-6P, glucosamine-6-phosphate, GlcN-1P, glucosamine-1-phosphate, UDP-GlcNAc, uridine diphosphate NAG, Fru-6P, fructose-6-phosphate. Protein abbreviations: NagE, NAG PTS permease, NagA, NAG-6 phosphate deacetylase, NagB, glucosamine-6-phosphate deaminase, NagC, DNA-binding transcriptional dual regulator, GlmS, L-glutamine:D-fructose-6-phosphate aminotransferase, GlmM, phosphoglucosamine mutase, and GlmU, fused NAG-1-phosphate uridyltransferase glucosamine-1-phosphate acetyltransferase. Arrows indicate chemical or transport reactions while bars indicate repression of transcription or activity. 72

4.6 Batch growth screening of overexpression, knockout, and compensatory strains to identify their osmotolerance phenotypes in glucose minimal media supplemented with 0.55 M NaCl. A) Fitness of overexpression strains relative to the empty vector pCA24N control. Only *ydjK* expression results in a significant improvement of NaCl tolerance. B) Fitness of knockout strains relative to BW25113; strain genotypes are denoted as follows: $\Delta(nagC)$ = C, $\Delta(nagA)$ = A, $\Delta(fimA)$ = F, and $\Delta(proV)$ = P. C) Overexpression of NagC, NagA in mutants with possibly inactivating mutations in these genes; fitness is relative to the corresponding mutant with pCA24N. D) Overexpression of FimA, ProV in mutants with possibly inactivating mutations in these genes, with the empty vector as a reference. All error bars are 95% confidence intervals based on the Student's distribution. 79

4.7 Levels of indole, normalized by biomass density, for G2, G3, G5, G6, A2, and A4 mutants, along with the G parent Hfr-2xSFX- and A parent 2xOriT. Error bars are standard deviations. G2, G5, and G6 have statistically significant increases in indole accumulation versus Hfr-2xSFX- ($P < 0.05$, Student's t-test). 83

4.8 Summary of observed genetic and transcriptional responses to continual osmotic stress in the A) cytoplasm and B) the membrane. Overexpressed and repressed genes are denoted with red and green text, respectively, while genes with identified mutations are numbered. A detailed examination of how these responses relate to the observed osmotolerance phenotype is given in the text. Overall, significant changes in outer membrane protein expression coupled with altered amino-sugar and osmoprotectant metabolism, control of cell shape, and sulfonate transporter expression appear to be major adaptive mechanisms for the mutants. Abbreviations: NAG, n-acetyl-glucosamine, CCM, central carbon metabolism, IM, inner membrane, PP, periplasm, PG, peptidoglycan layer, and OM, outer membrane. 87

5.1 The essential *L. brevis* stress response to ferulic acid. A. Expression of many uncharacterized, multi-domained membrane proteins that may function as ion transporters, ferulic acid, or 4-vinylguaicol exporters, among several possibilities. B. Upregulation of fumarase (fum), malate dehydrogenase (mdh), and malate permases involved in the citric acid cycle. NADH dehydrogenase and glutathione reducatase may also play a role in NAD+/NADH recycling and superoxide generation. C. Expression of a small HSP to ensure proper folding of heat labile phenolic acid decarboxylase (PAD) which converts ferulic acid to 4-vinylguaicol [182]. **D.** Representation of membrane damage and ion leakage triggered by ferulic acid induced lipid disruption. . . 100

5.2 A visualization of the Type II fatty acid synthesis pathway found in most bacteria [340]. The AccABCD proteins carboxylate acetyl-CoA to form malonyl-CoA, followed by condensation of malonyl-CoA with acetyl-CoA to form acetoacetyl-CoA by FabH. Other proteins including FabG, FabZ, FabI and FabF elongate the acetoacetyl-CoA by two carbons every pass through the cycle. Transcription of each FAS gene is only slightly upregulated immediately following n-butanol addition but increases significantly after 75 and 135 minutes. 109

5.3 Growth of all 285c strains and the parent strain at 30 °C. Each strain was inoculated from overnights into LB with 25 μg/ml streptomycin (antibiotic omitted for P9OA5c) and cultured at 30 °C to determine the growth kinetics at the permissive temperature. As can be seen, the growth rates and the yields of the strains are similar. 112

5.4 Complementation challenge at 42 °C. Each strain was inoculated from overnights into LB with 25 μg/ml streptomycin (antibiotic omitted for P9OA5c) and cultured at 42 °C to determine if the heterologous sigma factors could complement a temperature sensitive defect in the native σ^{70} protein. Robust growth is observed for 285c/pCLE-Lb and 285c/pCLE-Lp while the 285c/pCLE strain is not viable at the elevated temperature. 113

6.1 Population dynamics from an yeast population (KK-Large1-2007) selected for growth in glucose limited media. 119

6.2 Decoding of the hidden Markov states for each labeled subpopulation occurs as follows. (1) The set of emission symbols O_k for a subpopulation is generated from the statistical classifier for all n measurements. (2) The forward Viterbi decoder generates the most likely set of hidden states by choosing the path of maximum likelihood through the system trellis (green lines) based upon the known Markov parameters and O_k. (3) The output set X_k is assembled from these predictions for all observations. 126

6.3 Using the experimental dynamics in 6.1 and the PSM, the timing of each adaptive event in the chemostat is calculated and displayed for the user as shaded time points. 130

6.4 Following the identification of adaptive events, estimates of optimal sampling points, as described in the text, are then computed to assist in mutant isolation. 131

6.5 The relative proportions of adaptive events in each subpopulation, calculated using the PSM, in the three chemostat systems considered here. The neutrality of expression of the fluorescent protein implies that there should be no consistent bias of adaptive events toward any particular color, and this assumption holds here for all chemostats. Statistically significant differences in the abundance of adaptive events between the labeled populations would imply the presence of jackpot mutants. 132

7.1 A) Turbidostat without recombination, B) Turbidostat with recombination, C) Serial batch without recombination, and D) Serial batch with recombination. All simulations were run for 2260 iterations (equivalent to 15 days of experimental evolution). 144

7.2 Representative inheritance patterns in turbidostats undergoing evolution A) with and B) without recombination. Directed edges connect parent to child genotypes. All genotypes are descended from a single initial genotype, and each simulation is seeded with a specified number (three, in this case) of labeled but otherwise isogenic strains. Note that clonal interference is attenuated when recombination is permitted between different mutant lineages. 146

7.3 Representative population dynamics during evolution of the *E. coli* models after being constrained with n-butanol and isobutanol gene expression data. A turbidostat without recombination was used for all simulations. A) Core model, isobutanol stress; and B) core model, n-butanol stress. 148

LIST OF TABLES

TABLE Page

2.1 Estimated efficiencies of protoplast formation, regeneration, and hybrid formation. The efficiency of hybrid formation is estimated based on randomly isolating 50 colonies from the fusion products after PEG treatment and testing for their abilities to grow on both MRS and LB. 20

2.2 The differential growth of *L. brevis* and *E. coli* on several media (+: growth; -: no growth). 21

2.3 Morphology of hybrid strains compared with *E. coli* and *L. brevis*. KKOB3-KKOB5 appear Gram positive, while KKOB2 and KKOB6 are Gram negative. 21

3.1 Bacterial strains and plasmids 31

3.2 Primers used in this study 34

3.3 Summary of evolution experiments 40

4.1 Partial list of strains, plasmids used in this study 56

4.2 Mutations identified with genome sequencing 73

4.3 Overexpressed genes of interest 81

4.4 Repressed genes of interest 85

5.1 Expression of select upregulated genes during ferulic acid stress . . . 102

5.2 Expression of select downregulated genes during ferulic acid stress . . 104

5.3 Expression of select upregulated genes during 2% n-butanol stress . . 106

5.4 Expression of select downregulated genes during 2% n-butanol stress . 108

6.1 Population state model parameters 125

6.2 Analysis of population dynamics 128

6.3 Population state model error analysis 129

1. INTRODUCTION

1.1 What is Adaptive Laboratory Evolution?

Competition for space and resources is an inherent feature of life. All organisms seek to maximize their ability to survive and pass down genetic material, either as individuals or as part of more complex collectives. Although in most cases offspring will have genomes that are faithful copies of their parent (in the case of asexual organisms) or a mix of two organisms (for sexually reproducing species), random errors during DNA replication occur that change the offspring genotype. In most cases, these errors will have a detrimental effect on the ability of the offspring to survive, or at best, no effect at all. However, on rare occasions, mutations that improve fitness will occur, generating an organism that can survive better and reproduce more frequently in a given environment. Adaptive evolution as a whole is essentially a cycle of random mutation and selection based on organism fitness that can be exploited to alter characteristics of complex organisms [321, 322] and rapidly reproducing microbial organisms [255, 96] as needed. Figure 1.1 describes the basic process of controlled evolution carried out using two distinct organisms, namely silver foxes and microbes. Once a desired behavior or property has been identified, such as docility for foxes or improved growth in the presence of ethanol for fermentative yeast, the experimenter can artificially apply selection by manually testing foxes for lack of aggression or continuously growing yeast in the presence of ethanol. After a sufficient number of generations, individuals within the selected populations will accumulate mutations that improve their fitness, resulting in an increase in their frequency. Given sufficient time, these mutants can displace the original background population and establish themselves as the dominant genotype [300, 110].

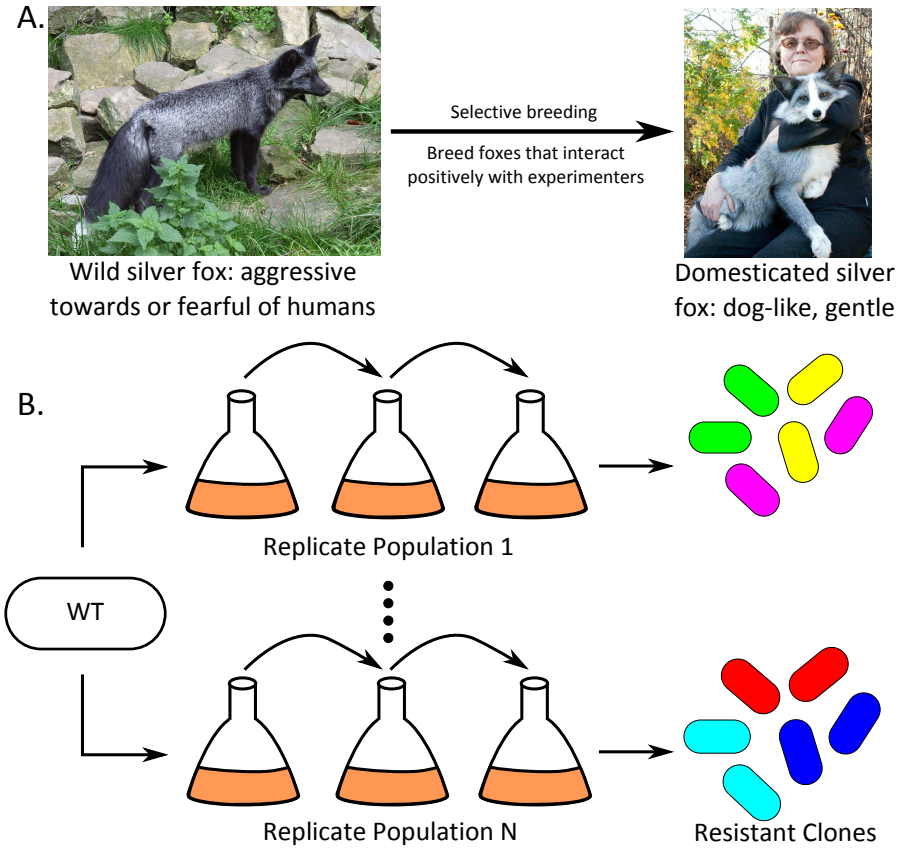

Figure 1.1: Two examples of artificial adaptive evolution. A) Domestication of silver foxes. The initially feral foxes are selected for breeding based on their interactions with human handlers. Photo credits to Zefram and Dan Child (BBC). After a sufficient number of generations, many of the foxes behave similarly to domesticated canines. B) Evolution of a microbe in the laboratory. The microbe is repeatedly propagated under selection (nutrient utilization or chemical inhibitors) until improved mutants are isolated.

Due to the difficulty of evolving large, slowly reproducing organisms, most evolution experiments focus on quickly reproducing microbes to study evolutionary theory, structure of evolving populations, and to improve microorganism characteristics for industrial use. Experimental design for adaptive evolution depends heavily on whether tolerance (substrate, product, or inhibitor) or enhanced metabolite production is the desired end goal. The former is usually accomplished by growing the

microbe of interest under the relevant selective pressure for several hundred generations [142, 24, 127, 217, 269, 108, 97]. Growth on non-preferred carbon sources can also be improved in a similar fashion [151, 179, 5]. The bulk of phenotype improvement appears to occur early during most adaptive evolution experiments, though fitness will continue to slowly improve afterward as demonstrated by the long term evolution experiment still being executed in the Lenski laboratory [349]. The rate of apparent adaptive improvement depends heavily on the number of loci involved with the phenotype of interest, as polygenic phenotypes will generally improve more smoothly than those with only a limited number of possible beneficial mutations [320]. However, since evolution is an inherently stochastic process, it is possible to have sudden jumps in fitness if beneficial mutations are simultaneously rare and have large phenotypic effects [320] even after a significant amount of time. Structure of the underlying fitness landscape [12, 16] therefore has a large impact on the outcome of evolution experiments by restricting how fitness improvements may accumulate for various genotypes in the evolving population.

In the context of the biotechnology industry, the execution of these experiments typically relies on batch or continuous (chemostat) cultivation systems [348] to maintain the evolving populations in a state of relatively constant growth under the selective pressure of interest. Ideally the experimental conditions should closely match those expected during actual production to ensure that any evolved mutants behave as expected when utilized downstream. Batch growth requires little investment in equipment compared to the pumps, reactors, and tubing required for most chemostats, and as a result, is the cultivation method of choice for most laboratories. One significant disadvantage of batch growth is that the fraction of culture transferred to fresh media every day exacerbates drift, the random loss of mutant genotypes from an evolving population [140, 102], slowly the accumulation of adap-

tive mutants significantly. Continuous systems have much lower levels of drift due to absence of this large, externally imposed population bottleneck and therefore retain more newly arisen mutants over a given length of time. Selection also differs between each system, as growth in batch systems never reaches a steady state similar to that observed in most chemostats, resulting in the selection of mutations affecting lag phase length in batch systems and the loss of signaling networks in chemostat systems [192, 180]. Clearly, the choice of culture system must be considered carefully to ensure mutants with the desired phenotypes are identified.

Beyond the direct selection for improved growth in unfavorable environments, adaptive evolution can also be applied to optimize for the production of certain biochemicals. Experimental design to enhance metabolite production is challenging due to the lack of a direct selection for the phenotype in question. Artificial essentiality can be engineered for certain metabolites by manipulating redox balancing [55, 160] or using the product of interest to protect against transient environmental stressors such as oxidative insult [270, 212]; the resulting strains can then be subjected to growth based selection to increase metabolite production. For other metabolites lacking pathways suitable for redox rescue or that do not confer tolerance phenotypes, no general method linking growth to production exists. Increasing the production of non-growth linked metabolites typically relies upon one of several types of random mutagenesis (for a review, see [242]). Mutants are then analyzed to determine if they produce a statistically significant increase in the product titer compared to the parent strain followed by additional rounds of mutagenesis and screening if desired [332, 187]. Since production of industrially desirable compounds represents a diversion of carbon and energy from cell growth, mutations that increase the cellular growth rate will often result in decreased or no synthesis of the desired product. In more characterized host organisms such as *E. coli* and *S. cerevisiae*, metabolic sim-

ulations are gaining traction as tools for rational engineering [292]. Efforts have also been made to create general linkages of metabolite production to mutation rate [65] to rapidly generate mutants with improved production levels. Given the difficulty in rational optimization of metabolic pathways, both computational and evolutionary approaches to improving metabolite production will find increasing use in the near future.

One of the main factors reducing the effectiveness of adaptive evolution experiments in producing enhanced phenotypes of interest is competition between clonal populations of microbes with different but beneficial mutations [121, 102, 86, 90]. This phenomenon, known as clonal interference, pits mutants with different beneficial mutations against one another in a competition for resource assimilation. The importance of clonal interference in shaping evolutionary outcomes is becoming clearer due to the advent of lower cost and high coverage genome sequencing, which has enabled deep sequencing of populations over the course of entire evolution experiments for the first time [40, 180, 184]. Competition between beneficial mutants and the dynamics of each genotype are easily observed using this approach, but other more visual approaches have also confirmed significant clonal interference during asexual evolution [269, 163]. An example of how population dynamics vary between clonal populations and those where DNA is exchanged between mutants is shown in Figure 1.2. Due to the lack of a mechanism for sexual recombination in most microbes, the competition inevitably ends with the fittest clone relative to current background population achieving a high frequency after several generations, in contrast to the recombinant genotypes obtained from sexual reproduction [74]. This outcome is perfectly optimal from the perspective of the surviving mutant, as it may now dominate the culture without competition for resources, but from a scientific viewpoint this loss of genotypic history is a severe impediment to understanding which mutations

enhance industrially relevant complex phenotypes [244].

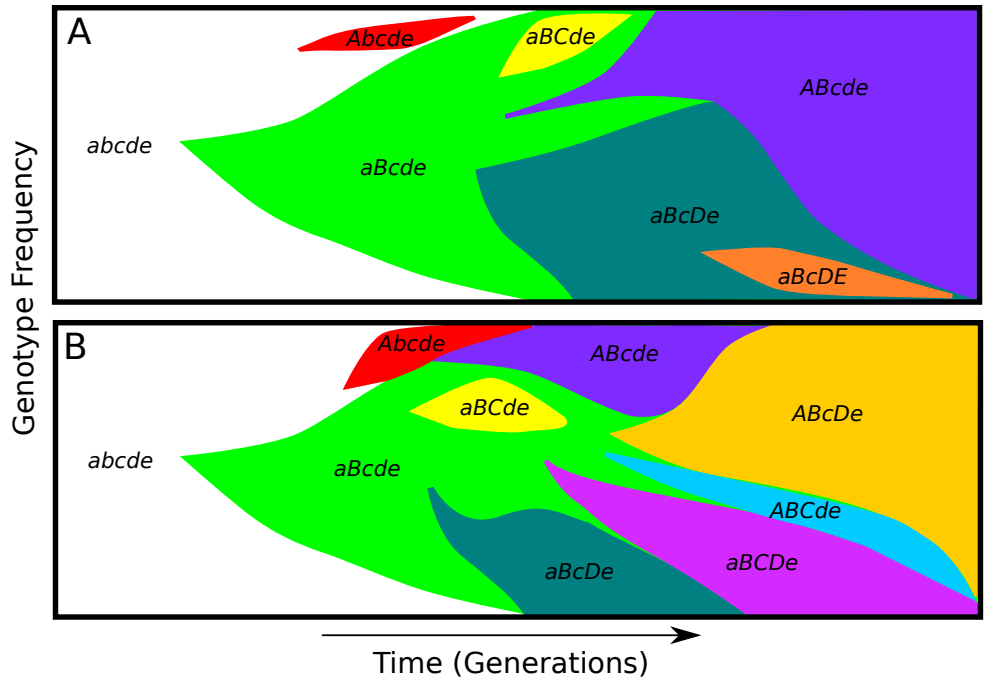

Figure 1.2: Cartoon depicting the evolution of large A) asexual and B) sexual populations under an arbitrary selective pressure. The genome in this case consists of five alleles with two possible variants (*abcde* versus *ABCDE*) Clonal interference in the asexual populations eliminates adaptive mutations that are potentially informative but have lower fitness than other mutants in the population. Genotypes can also be combined only through sequential mutation, limiting the pace of adaptation. Sexual populations, in contrast, are able to use genetic recombination to generate new and varied genotypes competing lineages as they arise. Both populations suffer from stochastic loss of beneficial mutants (genotype *aBCde*). Adapted from [74].

1.2 Microbial Sex and Evolution

Attenuating the effect of clonal interference in evolving populations requires a mechanism for exchanging DNA between individuals, allowing for beneficial mutations to be unlinked from their original genetic background and spread throughout

the population [70]. Genetic exchange of this sort would effectively abolish the concept of an unbroken lineage linking a given mutant to the unevolved parental strain, reducing the effect of competition on evolutionary outcomes. While most eukaryotic organisms have some form of sexual reproduction [30], microbes typically reproduce by binary fission: every new individual in the population is genetically identical to its the original parent cell, excepting cases where random mutations occur. However, a type of sexual genetic exchange occurs in the presence of parasitical episomes known as conjugative factors [230] which encode machinery for the transfer of DNA from cell to cell to enable their own propagation. On rare occasions, these episomes can integrate into the host chromosome by homologous or site-directed recombination, enabling transfer of host DNA between cells. The details of one such conjugation system in *Escherichia coli*, the F plasmid, are discussed below. The existence of these conjugative factors greatly simplifies the challenge of implementing a form of microbial sex to reduce the impact of clonal interference on population dynamics during evolution.

There is substantial theoretical evidence supporting its utility for adaptive laboratory evolution. A recent review of competing models examining the relative benefit of sexual versus asexual reproduction summarizes the work to date [33]. Some of the first theoretical studies examining the difference between sexual and asexual population dynamics were carried out independently by Fisher and Muller [81, 221, 169], resulting in the formulation of the Fisher-Muller model that suggested that sex had two main benefits: the removal of deleterious mutations from the genome and the efficient spread of beneficial mutations throughout a population. This hypothesis and its evolutionary implications have been extensively reviewed elsewhere [237]. Crow and Kimura [74] later showed explicitly that adaptive mutations would tend to approach 100% frequency (fixation) more quickly in sexual populations for mu-

tations with a given fitness benefit, at least under the conditions modeled. Other researchers also confirmed the essential correctness of the Fisher-Muller hypothesis for finite populations [107, 344, 2]. However, the predicted benefit of sexual recombination varies significantly depending on modeling assumptions, such as the number of mutable loci, the rate of recombination, and other factors [169, 298, 4, 214, 2]. Simulations with digital organisms have also indicated that sexual recombination is more effective at adapting to changing environments [218, 219].

In addition to mathematical and computational approaches for investigating sexual population genetics, predictions about the benefit of sex have been experimentally validated using a variety of experimental systems, including *E. coli*, *S. cerevisae*, and others. Due to current expense of deep sequencing for populations, comparisons between sexual and asexual adaptive rates are typically made phenotypically, either by tracking fitness relative to a standard over time or by challenging populations with progressively higher levels of stress until growth failure occurs. Organisms that are infected with conjugative plasmids or are naturally sexually competent adapt to certain types of selective pressures more quickly than comparable asexual organisms [72, 69, 357, 181, 129, 58, 28, 254]. A recent review by W.R. Rice summarizes older experimental tests of sexual recombination as well [271]. Sex itself was also associated with a temporary fitness cost [129], reflecting the energetic costs of maintaining the necessary molecular machinery for DNA exchange. The success of sexual systems in a variety of environmental conditions (minimal media, NaCl challenge, thermal stress) demonstrates the broad utility of a sexual evolution system for adaptive laboratory engineering. Given these numerous successes, it is clear that an effective mating system would be an useful evolutionary engineering tool for model organisms like *E. coli*, but the question of implementation naturally arises. Design of such a system requires extensive knowledge of the DNA transfer system, the ability to mate

under typical cultivation conditions (batch, liquid culture), and very high mating efficiency (i.e. frequent allele exchange). Only one tractable conjugation system in *E. coli* is known to fit these design criteria: the F plasmid.

1.3 Conjugation: a Natural Method for Cell-Cell DNA Transfer

One of the first discovered conjugation systems, the F plasmid is a large single copy plasmid found in F^+ *E. coli* [117]. Strains with chromosomal F integrations (known as Hfr for high frequency recombination) are capable of transferring chromosomal DNA due to the insertion of the F origin of transfer (*oriT*) sequence into genome. F-mediated conjugation is a four step process summarized in Figure 1.3 [343]. Males use their sex pilus to attach and ratchet F^- recipients until their cell membranes make contact, followed by stabilization of the mating pair. The donor F plasmid or chromosome (Hfr strains) is then transferred through a DNA secretion system [185] as single-stranded DNA in a rolling circle fashion along with several accessory proteins (e.g. primase to allow for dsDNA synthesis) to the recipient. Transfer is initiated at the origin of transfer (*oriT*) and proceeds along the entire plasmid or host chromosome. Based on known DNA transfer rates of 40,000 bp/min at 37 °C, the F plasmid requires approximately 3 minutes to transfer while 100 minutes are needed for the *E. coli* chromosome. This transfer process is sufficiently robust to permit mating in liquid culture [155], unlike other conjugation plasmids. The probability of transfer of a given locus decreases exponentially as distance from the origin of transfer (m) increases according to the empirical equation 1.1 [297]. The pre-exponential factor A is usually interpreted as representing the efficiency of mating pair formation ($A \in [0.2, 0.5]$). DNA transfer can be prematurely terminated by physical and chemical disruption of the mating bridge or by nalidixic acid treatment [37]. The main requirement for the recipient strains is that they be recombination

proficient to integrate DNA into their genome, if mating with a Hfr strain.

Other closely related gram negative bacteria (e.g. *Salmonella*) can also maintain F plasmids. F has also acquired numerous transposons over time that have affected *tra* regulation, specifically an insertion into the coding sequence of the fertility control protein FinO [59, 117]. This transposon deactivated the FinO RNA binding protein that acts to repress *tra* expression indirectly, effectively depressing the entire *tra* operon, thus enabling continuous expression of all transfer proteins. The FinO protein acts to stabilize a RNA-RNA helix of *finP* and *traJ* transcripts which is then subsequently degraded by the host [189]. This mutation was a key factor in the original detection of the plasmid as without FinO the sex pilus and other genes required for successful mating are naturally derepressed, allowing for constant F transfer outside stationary phase [118]. While continuous mating likely represents a significant burden on cellular resources, it should also make the generation of recombinant genotypes containing DNA from two or more cells more frequent than transient expression of transfer proteins.

$$f \approx A\exp\left(-0.065m\right) \tag{1.1}$$

While detailed information about each *tra* gene product is available, the "gender" proteins TraS and TraT warrant special consideration, given their central role in this proposal. The functional roles of these proteins were originally identified by analyzing the properties of Hfr mutants that lacked a property known as surface exclusion (*sfx*) [296, 1, 120], defined as the apparent inability of F^+ (Hfr) cells to mate with other F^+ individuals. Mutants that were sfx^- were competent recipients for other Hfr or F^+ cells in any growth phase but tended to suffer reduced growth rates due to excessive membrane damage [238]. As TraS and and TraT were

characterized further, their actual biochemical roles became more clear: TraS is a highly hydrophobic membrane protein that interacts with the transfer protein TraG to destabilize mating aggregates; TraT on the other hand appears to interact with membrane porins (OmpA) to prevent mating, though the precise inhibition mechanism is unknown [120]. Despite the possible utility of sexual Hfr strains, no attempt appears to have been made to utilize surface exclusion deficient mutants for combining different *E. coli* lineages though Hfr strains had been used to test the advantage of sex previously [72]. Disrupting TraS and TraT may therefore increase cell to cell DNA transfer, resulting in more efficient conjugation and ultimately better transfer of adaptive mutations between individuals in an evolving population.

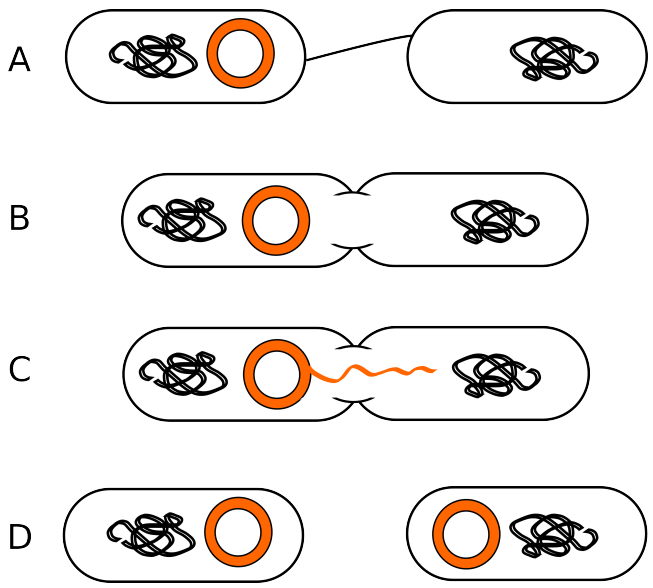

Figure 1.3: The process of conjugation between F^+ and F^- *E. coli* cells. A. The F^+ cell uses its sex pilus to "ratchet" the F^- cell until contact is made between their membranes. B. A mating bridge is formed, allowing for DNA transfer to begin between the cells. C. A single-stranded DNA copy of the F plasmid is transferred to the recipient, along with primase (required to synthesize the dsDNA plasmid) and several other proteins. D. The mating pair dissolves, leaving two F^+ cells.

Hfr strains are of particular interest as they are capable of transferring chromosomal DNA from the donor to a F^- recipient. The process of Hfr mating at its core involves the formation of a mating bridge between the Hfr donor and F^- recipient, followed by transfer of ssDNA derived from the host chromosome. The recipient has a high probability of integrating the transferred DNA via homologous recombination due to the genetic similarity of most *E. coli* strains [18]. Hfr strains can theoretically transfer their entire chromosome, but in practice this rarely occurs due to DNA breakage or disruption of the mating pair. Mating specificity (i.e. gender) controlled by TraS and TraT efficiently acts to prevent Hfr-Hfr DNA transfer [342, 1, 120]; genetic transfer in a mixed population of Hfr and F^- cells is therefore unidirectional. Creating true bidirectional (i.e. Hfr $\rightleftharpoons$ Hfr mating) conjugation therefore requires manipulation of *tra* expression and the gender proteins to ensure that Hfr-Hfr matings occur under typical growth conditions. Several studies have shown that F^+-F^+ and Hfr-Hfr matings are possible if TraS or TraT mutations are present [342, 1]. An Hfr strain with inactive TraST proteins is therefore capable of transmitting and receiving chromosomal DNA from other individuals in the population, representing a form of sexual exchange in an otherwise asexual organism.

Though most Hfr strains arise naturally due to chance homologous recombination events between shared insertion sequence on the *E. coli* chromosome and F plasmid, they can also be artificially engineered as well. François et al. [116] cloned the HindIII fragment of F containing the entire transfer operon of the plasmid into a randomly integrating temperature-sensitive vector, and were able to generate Hfr strains with integrations uniformly around the genome. The resulting pseudo Hfr strains were similar to natural Hfrs in terms of DNA transfer efficiency and rate, though some integrants did exhibit cryosensitivity that may have resulted from the use of a temperature-sensitive replicon. These artificial Hfr strains later found use as

the basis of a novel method to rapidly generate double-knockout strains by conjugation [324] in one demonstration of the usefulness of conjugation as a tool for strain engineering. The genetic tractability of the F conjugation system and its ability to efficiently catalyze DNA transfer under common cultivation conditions, coupled with the wealth of existing Hfr strains, genetic tools, and experience with evolutionary engineering of *E. coli*, make this conjugation system the ideal basis for a sexual evolution system.

1.4 Designing a Sexual Evolution System for *E. coli*

Currently, the only broadly applicable method available for surmounting bacterial clonal interference is genome shuffling [308], where two cells physically merge together and produce recombinant progeny via interchromosomal homologous recombination [250]. Even though this technique has been successfully applied to improve several desirable complex phenotypes [246, 359, 78, 236, 347], genome shuffling has relatively low yields of stable recombinants (0.05-0.7%) [79, 347], and many of these individuals revert to parent phenotypes [148]. A subtler disadvantage of genome shuffling is that fusion cannot take place under selective conditions, necessitating additional investigation to identify isolates with the desired phenotypes. Allowing for continuous genetic exchange *in situ* sidesteps this issue by allowing strains with competing beneficial mutations to exchange genetic information while remaining under the desired selective pressure. Recombinant progeny will ideally accumulate combinations of beneficial mutations continuously throughout the evolutionary experiment, thereby leading to largely automatic identification of multiple mutants that exhibit positive gene epistasis while allowing for the repair of any deleterious alleles.

Based on the hypothesis that increasing cell-to-cell DNA transfer by eliminating surface exclusion from *E. coli* Hfr strains will result in less clonal interference, and

hence faster adaptation during evolution, we proposed to develop an effectively "genderless" Hfr strain that is competent as a DNA donor and recipient in conjugation with other genderless cells for use as a platform for evolutionary engineering. A transfer operon from the F plasmid was integrated into the strain and the TraS and TraT surface exclusion proteins disrupted, allowing for efficient bidirectional DNA transfer between individuals in an evolving population. This approach exploits the ability of Hfr strains to efficiently transfer chromosomal DNA for intrapopulation recombination between mutants during evolution. Mating occurs continuously under selective conditions without the need for any external intervention. Genetically heterogeneous populations of genderless cells constantly exchange genetic material, increasing the rate of adaptation within a population while improving evolutionary outcomes in terms of increased growth rates. Unlike other methods for recombining *E. coli* genomes, genderless mating can occur under a constant selective pressure without external intervention, reducing the need to screen potential recombinants generated by other procedures while directly selecting for positive gene interactions among existing mutations.

1.5 Outline

The dissertation begins with an application of the main existing method for achieving cell-cell recombination, genome shuffling, by generating interspecific hybrids between *E. coli* and *Lactobacillus brevis* with improved solvent tolerance. This study was hampered by known issues with genome shuffling, specifically the low rate of recombinant formation, that proved difficult to overcome even after method optimization. The next section details how this obstacle was overcome through the initial development and validation of a Hfr-based sexual evolution using evolution on well-defined fitness landscapes, followed by the use of the sexual evolution system

to improve the industrially relevant osmotic tolerance phenotype to demonstrate the broader applicability of the system. An attempt to exploit the conjugation proficiency of these evolved donors is then explored in the following section. Shifting away from the sexual mating system, two other aspects of strain improvement are analyzed: transcriptional responses to stress and the expression of heterologous genes in *E. coli*, and computational interrogation and generation of evolution data. Conclusions from these studies and recommendation for future research in this area are then discussed.

2. NOVEL *ESCHERICHIA COLI* HYBRIDS WITH ENHANCED BUTANOL TOLERANCE*

2.1 Summary

We successfully generated hybrid strains between *Escherichia coli* and *Lactobacillus brevis* via protoplast fusion. Growth kinetics of five hybrid strains and *E. coli* were used to evaluate the butanol tolerance of the novel strains under different temperatures and nutritional conditions. The hybrid strains were able to tolerate up to 2% butanol (v/v) compared to the 1% maximum for *E. coli*. The growth inhibitory effects of butanol were also significantly less in several of the hybrids compared to *E. coli*. These results demonstrate the potential use of protoplast fusion to generate butanol tolerant strains.

2.2 Introduction

Butanol toxicity is one of the challenges currently facing the microbial production of biobutanol. Even the native producer, *Clostridium acetobutylicum*, only tolerates up to 1-2% of this organic solvent. Butanol tolerance is a complex phenotype, involving multiple loci [241, 172], rendering the engineering of strains for enhanced resistance to this solvent difficult. Thus far, efforts for enhancing butanol tolerance have mainly focused on the enrichment of mutants via serial transfers [197, 303]. More recently, mutagenesis of the RNA polymerase in *E. coli* resulted in a mutant with enhanced resistance to butanol toxicity from 80% growth inhibition to 50% growth inhibition in 0.9% (v/v) butanol [172]. These efforts have relied on modifications of existing machinery in the native organisms, which may limit success. A

*Reprinted with permission from "Novel *Escherichia coli* hybrids with enhanced butanol tolerance"' by J. Winkler, M. Rehmann, and K.C. Kao, 2010. Biotechnology Letters, Volume 32, pp. 915-920, Copyright 2010 by Springer.

recent report surveyed a wide range of microorganisms for their inherent tolerance to butanol; most organisms analyzed exhibited significant growth inhibition in 1-2% of butanol with the exception of *Lactobacillus brevis* (*L. brevis*), which was able to sustain growth in 3% butanol [175].

Lactobacillus brevis, a gram-positive bacterium, may be a promising candidate to be engineered as a biobutanol producer due to its high inherent tolerance to the product. However, *L. brevis* lacks many biosynthetic genes and has high nutritional requirements [209]. In addition, it is not a widely studied organism, and the molecular mechanisms involved in the higher butanol tolerance in *L. brevis* are unknown. On the other hand, *Escherichia coli* (*E. coli*), one of the most well-studied organisms, is non-fastidious and has relatively low nutritional requirements. Recent efforts have successfully metabolically engineered *E. coli*, a non-native producer, to produce butanol [15, 154]. However, *E. coli* has low tolerance to butanol (up to 1%). The ability to generate strains that contain the benefits of both organisms can potentially be useful for engineering biobutanol producers. Protoplast fusion and genome shuffling have been demonstrated as a useful tool for generating strains with novel properties [246, 162]. We have successfully fused the genomes of *E. coli* and *L. brevis* to isolate hybrids with enhanced resistance to butanol compared to *E. coli* and lower nutritional requirements than *L. brevis*.

2.3 Methods and Materials

2.3.1 Bacterial strains and media

The bacterial strains used for the present study are *Escherichia coli* BW25113 and *Lactobacillus brevis* ATCC 367. *E. coli* was grown on LB media at 37 °C and *L. brevis* was grown on MRS (BD) media at 30 °C. The hybrids were grown on LB media unless noted otherwise.

2.3.2 Protoplast fusion

Cultures of both *E. coli* and *L. brevis* were grown to an OD_{600} between 0.7 and 1.0 from 2% inoculum of overnight cultures. Protoplasts of *E. coli* were generated using the protocol previously described on LB soft agar with 0.5 M of sucrose [79]. The protoplasts of *L. brevis* were generated using a modified protocol from Cocconcelli et al. [68]. Briefly, the *L. brevis* cells were collected by centrifugation, washed once with 0.01 M Tris-HCl pH 8.0, washed twice with SMM buffer (0.5 M sucrose, 20 mM sodium maleate monohydrate, 20 mM $MgCl_2$, pH 7.1), and the cell wall was digested with 2 mg/ml of lysozyme (Sigma) and 30 μg/ml of mutanolysin (Sigma) at 30 °C for 20 minutes. A small portion of the digested cells were serially diluted in sterile water and plated on MRS plates to estimate protoplast formation efficiencies or in SMM buffer and plated on regeneration medium (RM; MRS, 1.5% agar, 0.5 M sucrose, 20 mM $MgCl_2$, 2.5% gelatin, 0.5% heat inactivated bovine serum albumin) plates for estimation of protoplast regeneration efficiencies. The resulting protoplasts of *E. coli* and *L. brevis* were mixed together and DNA from lysed cells was digested with DNase I (Promega) at room temperature for 10 minutes. The protoplasts were incubated at room temperature for 10 minutes and centrifuged at 2000g and 4 °C for 20 minutes. After centrifugation, the protoplasts were resuspended in PEG buffer (SMM + 0.4 g/ml PEG 6000, 5% v/v DMSO, 10 mM $CaCl_2$) and incubated at room temperature for 6 minutes. The protoplasts were centrifuged again at 2000g and 4 °C for 20 minutes, and then resuspended in SMM buffer, plated on RM plate at various dilutions, and incubated at 30 °C for 3 days.

2.3.3 Measurements of growth kinetics

Cultures were grown overnight in either LB or M9 minimum media supplemented with 0.5% glucose at either 30 °C or 37 °C. Growth kinetic measurements for each

strain in varying concentrations of 1-butanol were taken in 96 well plates with the Infinite M200 plate reader (Tecan) with shaking and incubating capabilities. The total volume of the culture in each well was 100 μl. Replicate measurements for each strain and condition were measured in different wells within the 96 well plate to ensure that potential differences in oxygen transfer rates did not bias the differences in the growth kinetic measurements for the different strains. The growth kinetic parameters of the specific growth rate (μ) and maximum cell density achieved were calculated for each strain in each condition.

2.3.4 16s rDNA determination

The genomic DNA of all hybrids were purified using the DNeasy kit (Qiagen). The presence of 16S rDNA sequences for *E. coli* and *L. brevis* were determined via PCR with the purified genomic DNA using the following primer pairs: for the *E. coli* 16S rDNA (BW25113_16S_for: 5'-GCTTGCTTCTTTGCTGACGAGTG-3' and BW25113_16S_rev: 5'-TACGCATTTCACCGCTACACC-3') and for the *L. brevis* 16S rDNA (ATCC367_16S_for: 5'-TGAAAGGTGGCTTCGGCTATC-3' and ATCC367_16S_rev: 5'-GCGGAAACCCTCCAACACTTAG-3').

2.4 Results and Discussion

2.4.1 Generation of E. coli-L. brevis hybrids

Protoplast fusions between two different species were first reported in 1984 by Gokhale et al. [124]. Since then, the generation of interspecific protoplast fusions has been successfully used to combine desirable properties from different organisms [60, 338, 162]. The resulting interspecific hybrids are reported to be stably maintained [60, 338], and thus are likely to be haploid recombinants. No interspecific hybrids between *E. coli* and *Lactobacillus* have been reported. Here we report the generation of hybrids between *E. coli* and *L. brevis* using protoplast fusion tech-

niques. The resulting efficiencies for protoplast formation and regeneration for both organisms are shown in Table 2.1. For the identification of true fusion products, we first characterized differential permissive growth conditions for *E. coli* and *L. brevis* by growing each strain on different solid media (LB, MRS, RM, and M9 minimum medium). The results are shown in Table 2.2. As expected, *L. brevis* has higher nutritional requirements compared to *E. coli* and is not able to grow on LB or M9 minimum medium. On the other hand, *E. coli* was not able to grow on either MRS or RM, most likely due to the 5 g/L of sodium acetate present in the medium, as acetate disrupts intracellular pH and inhibits growth [203]. Based on the results, the protoplast fusion products selected on regeneration medium were most likely a combination of pure *L. brevis* and true hybrids. Thus, true hybrids were identified as fusants that are able to grow on both MRS and LB media, since neither *E. coli* nor *L. brevis* are able to grow on both. We isolated a total of 70 hybrids and chose 5 (named KKOB2-6) to further characterize.

Table 2.1: Estimated efficiencies of protoplast formation, regeneration, and hybrid formation. The efficiency of hybrid formation is estimated based on randomly isolating 50 colonies from the fusion products after PEG treatment and testing for their abilities to grow on both MRS and LB.

Step	*E. coli*	*L. brevis*	Hybrids
Protoplast formation	>99%	>98.5%	
Protoplast regeneration	$\tilde{1}$0%	30%	
Recovery			$\tilde{0}$.3%

2.4.2 Hybrid characterization

The primary mechanisms for generating hybrid genotypes through genome shuffling are homologous recombination (if the parental strains are genetically similar)

Table 2.2: The differential growth of *L. brevis* and *E. coli* on several media (+: growth; -: no growth).

Strain	LB	M9	RM	MRS
E. coli	+	+	-	-
L. brevis	-	-	+	+

or non-homologous end joining (NHEJ) [312]. Given the significant genotypic differences between *E. coli* and *L. brevis*, the latter mechanism most likely plays the dominant role in generating viable hybrids. The genetic heterogeneity of the hybrids is apparent from the readily apparent morphological differences in the five hybrids we examined in detail (Table 2.3). The KKOB2 and KKOB6 strains are most similar to *E. coli*, while the KKOB3-5 strains are more similar to *L. brevis*. 16S rDNA inheritance of the hybrids was also estimated using *E. coli* and *L. brevis* specific primers. It appears that all 5 hybrids had inherited the *E. coli* 16S rDNA. No products were observed with the *L. brevis* specific primers in any of the hybrids, possibly suggesting that the hybrid genome is composed of more *E. coli* than *L. brevis* sequences.

Table 2.3: Morphology of hybrid strains compared with *E. coli* and *L. brevis*. KKOB3-KKOB5 appear Gram positive, while KKOB2 and KKOB6 are Gram negative.

Strain	Morphology
L. brevis	Long rods
E. coli	Short rods
KKOB2	Short rods
KKOB3	Long rods
KKOB4	Long rods
KKOB5	Long rods
KKOB6	Short rods, circular

2.4.3 Butanol tolerance in the hybrids

To establish the butanol tolerance of the hybrid strains generated in this study, growth kinetics were evaluated in LB medium with the addition of 0-3% (v/v) of butanol for comparison with the tolerance level of native *E. coli*. These results are shown in Figure 2.1. All hybrids grew robustly in LB only media (without butanol), achieving specific growth rates between 69% to 100% and final optical densities similar to that of *E. coli*. Two hybrids (KKOB2 and KKOB6) exhibited the same levels of butanol tolerance as *E. coli*, while the other three (KKOB3, KKOB4, and KKOB5) exhibited increased resistance to butanol. Though *E. coli* growth was dramatically inhibited at 1% v/v butanol, the growths of the three hybrids were not inhibited at 1% butanol at 30 °C. At the lower temperature, the hybrids were able to grow in up to 2% butanol and achieve a final optical density of approximately 66% of that in 0% butanol (data not shown). After growth in LB with 1.5% (v/v) butanol at 30 °C, the three most butanol tolerant strains (KKOB3-5) maintained their cell morphology (long rods) and their ability to grow on MRS (data not shown). Increase in temperature has been shown to enhance the toxicity of butanol [175]. This effect was observed for all hybrids tested. At 37C, the maximum concentration of butanol permitting growth in all strains was 1%. However, the hybrids (KKOB3-5) still exhibited higher tolerance at 37C compared to *E. coli* at 37 °C.

Among the five characterized hybrids, only two (KKOB2 and KKOB6) were able to grow robustly in M9 minimum medium. Similar to growth on LB, these hybrids exhibited butanol tolerance very close to that of *E. coli*. The temperature dependent tolerance to the organic solvent was also observed in M9 minimum medium. At 30C, the three strains were able to grow in up to 1% v/v butanol, whereas at 37 °C, the strains were only able to grow in up to 0.7% butanol (data not shown). The growth

characteristics of the KKOB2 and KKOB6 strains may suggest that they lack the *L. brevis* genes that impart the enhanced butanol tolerance seen in the other hybrids.

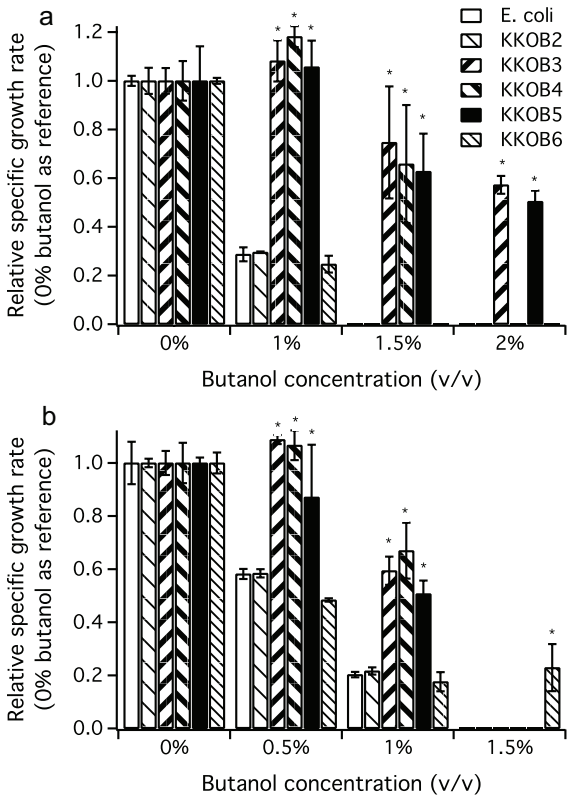

Figure 2.1: Inhibition of growth by butanol in LB medium at a. 30 °C and b. 37 °C. Values are based on four replicate measurement for each strain at each condition. Error bars are 95% confidence intervals. Hybrids with growth inhibition levels significantly less ($P < 0.05$, Student's t test) than *E. coli* at each butanol concentration are marked with an asterisk. Note that the growth of KKOB6 at 37 °C in 1.5% butanol was only observed in three out of the four replicates.

Interestingly, some samples exhibited growth at a higher concentration of bu-

tanol (but not at a lower concentration of butanol) in one out of four replicates. These outlier growths are always accompanied by a long growth lag, suggesting that growth at these higher concentrations originates from adaptive mutations conferring improved butanol tolerance. It is unlikely that these instances were the results of contamination, as no such growth were observed in the negative controls (no cell controls) included for all butanol concentrations in every run. In addition, two replicates in each 96 well plate were from the same overnight inoculum. Contamination of the overnight inoculum should therefore result in observed growth in both replicates within the plate, but in this case growth was only observed in a single replicate. The reasons for these outlier growths are under investigation.

2.5 Conclusions

This study establishes a system for generating hybrids between *E. coli* and *L. brevis* towards the goal of engineering strains with higher butanol tolerance. Further improvements in reducing the nutritional requirements and growth in higher butanol concentrations may be achieved by genome shuffling between the different hybrids and their parents. Since the genomes of both *E. coli* and *L. brevis* are both fully sequenced, whole genome sequencing of the hybrids can potentially reveal the underlying genetic bases for hybrid solvent tolerance and their other observed phenotypes.

3. HARNESSING RECOMBINATION TO SPEED ADAPTIVE EVOLUTION IN *ESCHERICHIA COLI**

3.1 Summary

Evolutionary engineering typically involves asexual propagation of a strain to improve a desired phenotype. However, asexual populations suffer from extensive clonal interference, a phenomenon in which distinct lineages of beneficial clones compete and are often lost from the population given sufficient time. Improved adaptive mutants can likely be generated by genetic exchange between lineages, thereby reducing clonal interference. We present a system that allows continuous *in situ* recombination by using an *Esherichia coli* F-based conjugation system lacking surface exclusion. Evolution experiments revealed Hfr-mediated recombination significantly speeds adaptation in certain circumstances. These results show that our system is stable, effective, and suitable for use in evolutionary engineering applications.

3.2 Introduction

Evolutionary engineering encompasses a range of powerful methodologies that have been harnessed to improve a range of microbial phenotypes of industrial interest [299, 26, 47, 350]. One of the principal challenges that reduces the effectiveness of adaptive evolution experiments in producing enhanced phenotypes of interest is competition between clonal populations of microbes with different but beneficial mutations [121, 102, 86, 300]. This phenomenon, known as clonal interference, pits

*Reprinted with permission from "Harnessing Recombination to Speed Adaptive Evolution in *Escherichia coli*" by J. Winkler and K.C. Kao, 2012. Metabolic Engineering, Volume 14, pp. 487-495, Copyright 2012 by Elsevier.

mutants with different beneficial mutations against one another in a contest for survival. Due to the lack of a mechanism for sexual recombination in most microbes, this competition generally results in the loss of adaptive genotypes from the population over time. Recombination between these lineages, in contrast, would reduce clonal interference by allowing multiple lineages to combine beneficial mutations into a single background, preventing the extinction of these alleles [74]. The loss of genotypic history due to clonal interference impedes understanding of industrially relevant complex phenotypes [244] by limiting our knowledge of the underlying molecular mechanisms governing the phenotype in question. Beneficial mutations must be acquired sequentially in an asexual population as well, theoretically limiting the overall pace of adaptation. Sexual populations avoid this problem by continuously exchanging beneficial mutations through mating, resulting in increased adaptation rates [169], reduced mutational load [272], along with many other benefits [87].

Currently, the only broadly applicable method available for inducing recombination in asexual organisms is protoplast fusion [308], where two cells physically merge together and produce recombinant progeny via interchromosomal recombination [250]. Even though this technique has been successfully applied to improve several desirable complex phenotypes [245, 359, 78, 236, 347, 269], protoplast fusion generates low yields of stable recombinants (0.05-0.7%) [79, 347] and many of these individuals revert to parental phenotypes over time [148]. Additionally, genome shuffling cannot be performed under selective conditions; recombinants generated using this procedure must therefore be recovered and enriched later, which increases experimental complexity significantly. Allowing for continuous genetic exchange *in situ* sidesteps this issue entirely by allowing strains with competing beneficial mutations to exchange genetic information while remaining under the desired selective pressure.

DNA transfer between *Escherichia coli* cells is mediated by fertility plasmids such

as F, originally discovered by Esther Lederberg and colleagues in the early 1950s as a "sexual factor" that could be physically transferred between different *E. coli* strains [53]. F-conjugation is unique in that it supports mating in liquid media, rather than on filters or agar as required for other conjugative systems [155], making it well-suited for use in evolutionary engineering. Decades of subsequent work by generations of scientists uncovered various aspects of F biology, including its occasional integration into the host chromosome to create high frequency recombination (Hfr) strains, gene function within the transfer (*tra*) operon [117], and eventually biotechnological applications of F such as exploration of genetic interactions or bacterial artificial chromosomes [324]. Hfr strains are capable of transferring chromosomal DNA from the integrated F origin of transfer, making these strains useful for strain construction. Overall, the F plasmid is well characterized and amenable to engineering for the creation of a sexual recombination system in *E. coli*. Hfr strains and yeast evolution incorporating mating have also been previously used without modification to demonstrate experimentally that recombination, in certain circumstances , leads to improved evolutionary outcomes [72, 128].

High frequency recombination strains are of particular interest as they are capable of transferring chromosomal DNA from the donor to a F^- recipient. The process of Hfr mating at its core involves the formation of a mating bridge between the Hfr donor and F^- recipient, followed by transfer of single stranded DNA derived from the host chromosome, as demonstrated in Figure 3.1. The recipient has a high probability of integrating the transferred DNA via homologous recombination due to the genetic similarity of most *E. coli* strains [18]. Hfr strains can theoretically transfer their entire chromosome, but in practice this rarely occurs due to DNA breakage or disruption of the mating pair. Mating specificity is controlled by two proteins within the *tra* operon (TraS, TraT) which efficiently act to reduce Hfr-Hfr DNA transfer by

100-300 fold via surface exclusion [342, 1, 120]; genetic transfer in a mixed population of Hfr and F^- cells is therefore unidirectional. Efficient bidirectional (*i.e.* Hfr $\rightleftharpoons$ Hfr mating) conjugation therefore requires manipulation of *tra* expression and the gender proteins to ensure that Hfr-Hfr matings can occur under typical growth conditions. Several studies have shown that F^+-F^+ and Hfr-Hfr matings can occur if TraS or TraT are mutated [342, 1]. An Hfr strain with inactive TraST proteins is therefore capable of efficiently transmitting and receiving chromosomal DNA from other individuals in the population, representing a form of sexual exchange in an otherwise asexual organism.

We propose to harness this phenomenon by engineering an effectively "genderless" Hfr strain that is competent as a DNA donor and recipient in conjugation with other Hfr individuals for use in adaptive evolution. This approach exploits the well-known ability of Hfr strains to efficiently transfer chromosomal DNA to permit recombination between evolved mutants without external intervention. Mating occurs continuously under selective conditions without the need for any external intervention. Genetically heterogeneous populations of genderless cells will constantly exchange genetic material, potentially increasing the rate of adaptation within a population and improving evolutionary outcomes for improved tolerance or growth. Unlike other methods for recombining *E. coli* genomes, mating can occur under a constant, arbitrary selective pressure, reducing the need to screen potential recombinants generated by other procedures while selecting directly for improved genotypes composed of multiple, independent adaptive mutations.

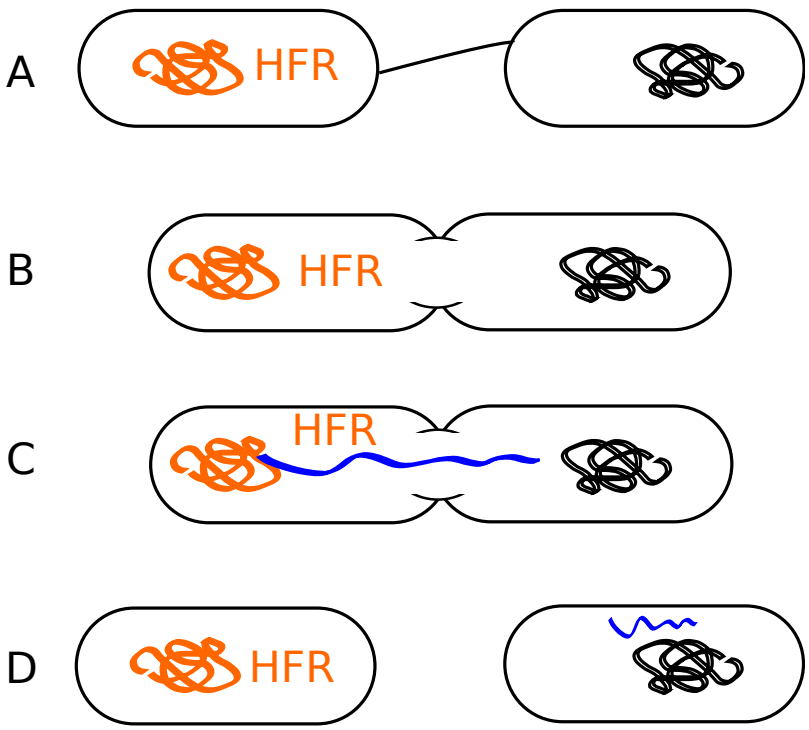

Figure 3.1: A summary of the general F-mediated conjugation process in *E. coli*. **A.** The pilus of the male (Hfr) strain, which contains an F plasmid integrated into the genome, latches onto a F^- recipient. **B.** After cell-cell contact is made, a mating bridge is formed to allow for stable DNA transfer. Note that *traS* and *traT* are thought to inhibit this step by interfering with the mating bridge. **C.** Single-stranded DNA is transferred from the donor to recipient from the origin of transfer on the donor chromosome. **D.** Once transfer is complete or disrupted in some way (mechanically, chemically), the donor and recipient separate. Any transferred DNA may be integrated into the recipient chromosome via homologous recombination.

3.3 Methods and Materials

3.3.1 Strain construction

All strains used for evolution experiments in this study are derivatives of BW25113 (Coli Genetic Stock Center, CGSC). A list of all strains used in this study is given in Table 3.1. The origin of transfer sequence from X892 (CGSC) was PCR amplified (Phusion, NEB) and cloned into the Pst1 site of the pKD13 (*kan*) and pKD32

(*cat*) disruption plasmids developed by Datsenko and Wanner [82]. Clones containing origin of transfer (oriT) fragments in the same orientation were used in all subsequent experiments. Two pseudogenes identified on EcoCyc [166], *mbhA* (5.39 min) and *hyfC* (56.08 min), were chosen to be replaced by the oriT cassettes. To create strains BW25113-A and B containing oriT sites at these locations, the cassettes from pKD13 and pKD32 were amplified using previously described procedures [17] and were then subsequently transformed into BW25113/pKD46. Transformants were verified with PCR and sequenced (MCLAB, CA) to ensure no mutations occurred in the oriT region. Following this verification step, P1 transduction was used to transfer the Δ(*hyfC*)::[oriT cat] marker from strain B to A, creating BW25113 2xOriT. CAG3101 [324] and CAG31031 Δ(*traST*))were separately used to introduce F into the genome of BW25113 2xOriT via conjugation to construct Hfr-2xSFX+ (active surface exclusion) and Hfr-2xSFX- (Δ(*traST*), surface exclusion deficient). Both strains were screened to ensure that an additional *oriT* linked to *tetAR* was not transferred as well. The 2xOriT, Hfr-2xSFX+, and Hfr-2xSFX- strains were then transformed with pCP20 and incubated at 43 °C overnight to eliminate the antibiotic markers in the cassettes. The final strains (hereafter referred to as the evolution strains) have either no antibiotic resistance cassettes (2xOriT) or are gentamicin resistant (Hfr-2xSFX+,-). Presence and orientation of the cassettes and the F plasmid in these strains was verified by PCR.

3.3.2 Strain characterization

Microplate assays (TECAN) were used to assess the growth phenotypes of the evolution strains and to test their sensitivity to several inhibitors (chloramphenicol, trimethoprim, butanol, acetate). Growth assays were performed in M9 minimal media supplemented with 50 μg/ml tryptophan and 0.5% (w/v) glucose. Antibiotic

Table 3.1: Bacterial strains and plasmids

Strain	Description	Ref.
BW25113	Δ(*araD-araB*)567, ΔlacZ4787 λ^-, *rph-1*, Δ(*rhaD-rhaB*)568, *hsdR514*	CGSC
BW25113-A	BW25113 Δ*mbhA*::[*oriT kan*]	This work
BW25113-B	BW25113 Δ*hyfC*::[*oriT cat*]	This work
BW25113 2xOriT-NF	BW25113 Δ*mbhA*::[oriT kan] Δ*hyfC*::[*oriT cat*]	This work
BW25113 2xOriT Hfr	2xOriT-NF *trp*::F (*gen*^{R})	This work
BW25113 2xOriT HfrΔ	2xOriT-NF trp::F[ΔtraST::*kan*] (*gen*^{R})	This work
BW25113 2xOriT	BW25113 Δ*mbhA*::*oriT* Δ*hyfC*::oriT	
Hfr-2xSFX+	2xOriT *trp*::F (*gen*^{R})	This work
Hfr-2xSFX-	2xOriT trp::F[ΔtraST] (*gen*^{R})	This work
BW25141/pKD13::oriT	Template for *oriT*, *kan* cassette	[82]
BW25141/pKD32::oriT	Template for *oriT*, *cat* cassette	[82]
CAG31031‡	BW25113 *trp*::F *ycdN*::(*oriT tetAR*)	[324]
CAG31031Δ(*traST*)‡	BW25113 *trp*::F[Δ(*traST*)::*kan*] *ycdN*::(*oriT tetAR*)	[324]
BW25113/pTB8†	BW25113 *amp*^{R}	This work
X892	K-12 prototroph, HFR	[278]
X892/pTB10†	X892 *amp*^{R}	This work
XΔ/pTB10†	X892 Δ(traST)::*kan amp*^{R}	This work

*: Strains used for antibiotic resistance evolution experiments.
†: Strains used to measure the relative conjugation efficiency of the evolution strains (except the asexual control) using F$^-$, Hfr (*sfx*$^+$), and Δ(*traST*) (*sfx*$^-$) strains.
‡: Strains used to conjugate the integrated F into the appropriate evolution strains.

tolerance was evaluated in M9 media supplemented with 0-8 μg/ml of drug (chloramphenicol (CM), streptomycin (STR), trimethoprim (TM)). Hfr-2xSFX- was slightly more sensitive to trimethoprim than the other evolution strains and did not initially tolerate TM concentrations greater than 0.8 μg/ml. Hfr-2xSFX+ was also found to be approximately 50% more resistant to streptomycin prior to the evolution experiment. The CM tolerances of the strains were similar. These differences could be the result of jackpot mutations or metabolic changes triggered by the presence of the

F plasmid in the strains. The chloramphenicol concentration that reduced the final biomass yield (optical density) by approximately 50%, 3 μg/ml, was selected as the initial concentration for the evolution experiment. A similar procedure was utilized to determine trimethoprim and streptomycin evolution starting points (0.8-1.6 and 2-3 μg/ml respectively).

3.3.3 Conjugation efficiency assays

Assessment of conjugation efficiency for the Hfr-2xSFX+, Hfr-2xSFX- strains (evolution strain donors) was accomplished using standard mating assays [10]. X892 / pTB10 (Hfr), XΔ/pTB10 (Hfr Δ(*traST*)), and BW25113/pTB8 (F^-) served as recipients for the evolution strain donors. The plasmids pTB8 and pTB10 only served to provide the ampicillin resistance for the dual selection scheme used for measuring mating efficiencies. Mating efficiency was assessed by mixing equal amounts of donors and recipients (both in exponential phase) in LB for 30 min without shaking, and then plating appropriate dilutions on selective plates (15 μg/ml gentamicin + 100 μg/ml ampicillin). Based on the expected structure of the integrated F [116], conjugation from one of the added oriT sites would be necessary for successful transfer of the gen^R marker. At least three biological replicates were performed in each case. Conjugation with an F^- recipient (BW25113/pTB8) was also performed to test for possible lethal zygosis [1] and to assess general transfer efficiency.

3.3.4 Luria-Delbruck fluctuation tests

The mutation rates of the Hfr-2xSFX- and BW25113 2xOriT strains were computed using standard Luria-Delbruck fluctuation tests [356] to determine if the disruption of *traST* was mutagenic. Single colonies of both strains were inoculated into LB. After growth to stationary phase, each strain was diluted 1000-fold into fresh media (10 replicates per strain) and allowed to grow to stationary phase The entire

cultures were then concentrated and plate onto LB agar supplemented with nalidixic acid (20 μg/ml) and incubated overnight. The mutation rate of each strain was then estimated using a standard maximum-likelihood fitting approach [134].

3.3.5 Evolution experiments

A feedback control scheme was utilized for the evolution experiments. First, six isogenic replicates of each evolution strain (BW25113 2xOriT, Hfr-2xSFX+, Hfr-2xSFX-) were inoculated into M9 media (0.5% (w/v) glucose and 50 $\mu g/ml$ tryptophan) containing the initial concentration of antibiotic (CM, STR, or TM). After a 24 hour period, the final biomass yield (OD_{600}) was measured. The density of unchallenged cultures (OD_{unchal}, no antibiotics) was used to determine the level of antibiotic inhibition in each replicate. Propagation into new media, assuming the current antibiotic concentration is C_{ab}, was decided using the following decision tree:

1. If ($OD_{\text{final}}/OD_{\text{unchal}} > 0.7$), increase the antibiotic concentration by $1.5C_{ab}$ in fresh media.

2. If ($0.15 \geq OD_{\text{final}}/OD_{\text{unchal}} < 0.7$), allow for 24 more hours of growth.

3. If ($OD_{\text{final}}/OD_{\text{unchal}} < 0.15$), propagate the previous culture into new media with 0.75C_{ab}

Replicates that fall under categories (2)-(3) are propagated once more at C_{ab} to ensure that the observed level of antibiotic resistance is genuine, rather than a result of antibiotic decomposition in the media. Once the entire dataset is gathered, exponential regressions can be performed on antibiotic concentration data to determine the parameter α ($exp(\alpha t)$) so that the antibiotic doubling time may be computed as follows:

$$t_{d,ab} = \frac{ln(2)}{\alpha} \tag{3.1}$$

These rules are designed to ensure that the antibiotic concentration increases in a consistent manner for each strain, depending on the acquisition of mutations conferring resistance. The rate of evolution for each population (or replicate) can then be quantified simply by observing the rate of improvement in antibiotic resistance, avoiding the need to individually screen clones and determine their antibiotic tolerance level. All replicates were screened by strain specific PCR (Primers 1+3 for BW25113, 11+12 for Hfr-2xSFX-, and 13+14 for Hfr-2xSFX+ (Table 3.2)) on a regular basis to detect contamination and verify the presence of F (if applicable).

Table 3.2: Primers used in this study

#	Name	Sequence
1	ForOriT_Pst1	AATCTGCAGATTTAACCCACTCCACAAAA
2	RevOriT_Pst1	AATCTGCAGATTCATAATGCAAACAGGGA
3	ForMbhA	CAGAAACCTCGGAAATACGC
4	RevMbhA	GCATTGCTCACCTCTCAACA
5	ForHyfC	GGCTGGCCAAAGAAATACAG
6	RevHyfC	AGATCAGCGACAACATGCAC
7	ForKW	TGTAGGCTGGAGCTGCTTCG
8	RevKW	ATTCCGGGGATCCGTCGACC
9	ForTraST_KO	GCTAATGTCATTTCTAATAACACAAAAGACAAG TGAAACACTCACTGCTGATTCCGGGGATCCGTCGACC
10	RevTraST_KO	GGCCGGTCAGACCAGCCTCCGGAAGATAATCAGA GAATATTTGCGATTGATGTAGGCTGGAGCTGCTTCG
11	ForTraST_Ver	TCATGGGGTGGATGATTTTT
12	RevTraST_Ver	TCCCGCAGATCCATCTTATC
13	ForTraST_F+	TTACCAGCATAAAGAATAATC
14	RevTraST_F+	GCCTCCGGAAGATAATC

3.4 Results and Discussion

3.4.1 Strain characterization

Hfr strains used in this work are derivatives of BW25113, with two added *oriT* sites to enhance gene transfer from around the chromosome by reducing the average distance from each gene to an *oriT* site (see [297] for transfer probabilities), and a F plasmid integrated into the *trp* locus to provide the conjugation machinery. The motivation behind inserting additional *oriT* sites stems from the exponential decline in transfer frequency as the distance between a given site and the *oriT* increase according to the empirical formula $f = \mathrm{Aexp}(-0.065\mathrm{m})$, where m is the distance in minutes separating the *oriT* from the target site, and the constant A represents the efficiency of mating pair formation. The inserted *oriT* sites act to decrease m for all sites on the genome, facilitating their enhanced transfer. If the user would like to transfer particular loci at very high frequency (for instance, genes that are likely to produce an enhanced phenotype once mutated), then an *oriT* can be inserted upstream to theoretically improve their transfer.

The Hfr strain design and the expected direction of DNA transfer from each origin is shown in Figure 3.2. Following their construction, we sought to determine how the heterologous *oriT* sites or the integrated F in Hfr-2xSFX+,- affected host physiology. Growth in minimal media supplemented with glucose and tryptophan, required by the Hfr strains due to the F integration in the *trp* locus [324], is shown in Figure 3.3. Growth of the BW25113 2xOriT strain appears normal compared to BW25113 in the same media, indicating that disrupting the *mbhA* and *hyfC* pseudogenes is not deleterious for growth in these conditions. Hfr-2xSFX- does have an extended lag phase and somewhat lowered growth rate (20%, $P = 0.004$) compared to the other strains, and a slightly lower biomass yield. It is likely that disrupting the surface

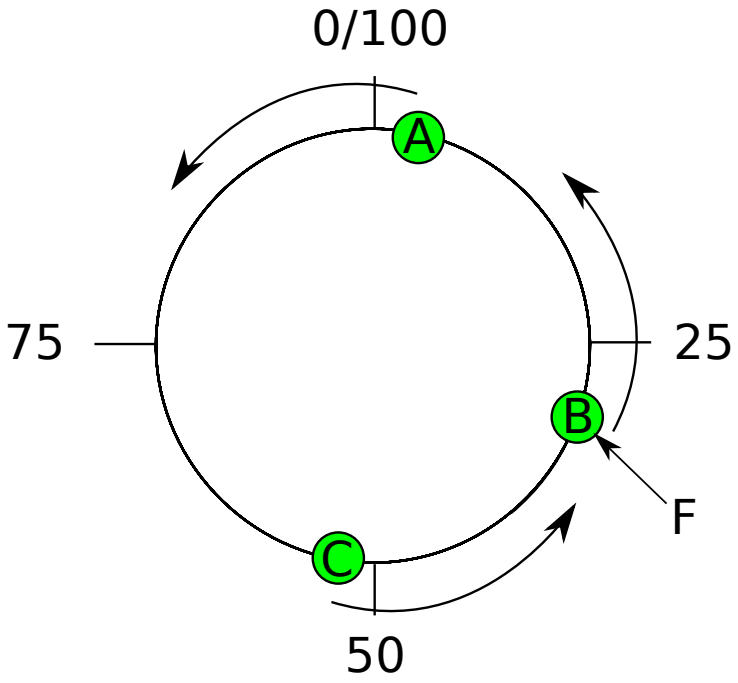

Figure 3.2: Basic design of the Hfr strains used in this study. The strain background is BW25113. Points A and C denote integrated oriT sites with transfer directionality (arrows) inferred from the orientation relative to the original F plasmid; Point B is an integrated F conjugated from a BW25113 strain obtained from CAG31031 [324].

exclusion genes in Hfr-2xSFX- leads to excessive membrane damage due to constant mating within a population, which is the mechanism of F-mediated cell killing (lethal zygosis) [238]. This effect would likely reduce stationary phase survival, leading to a lower proportion of viable cells in the inocolum, contributing to the extended lag phase and observed sensitivity to certain membrane irritants (n-butanol and acetate, data not shown) as well. The reduced growth rate of Hfr-2xSFX- may also be affected by the energetic requirements for DNA synthesis and transfer.

The mutation rates of BW25113 2xOriT and Hfr-2xSFX- were also calculated using a standard fluctuation test [24] and compared to determine if the $\Delta(traST)$ disruption was mutagenic, as a raised mutation rate would generally impact evolu-

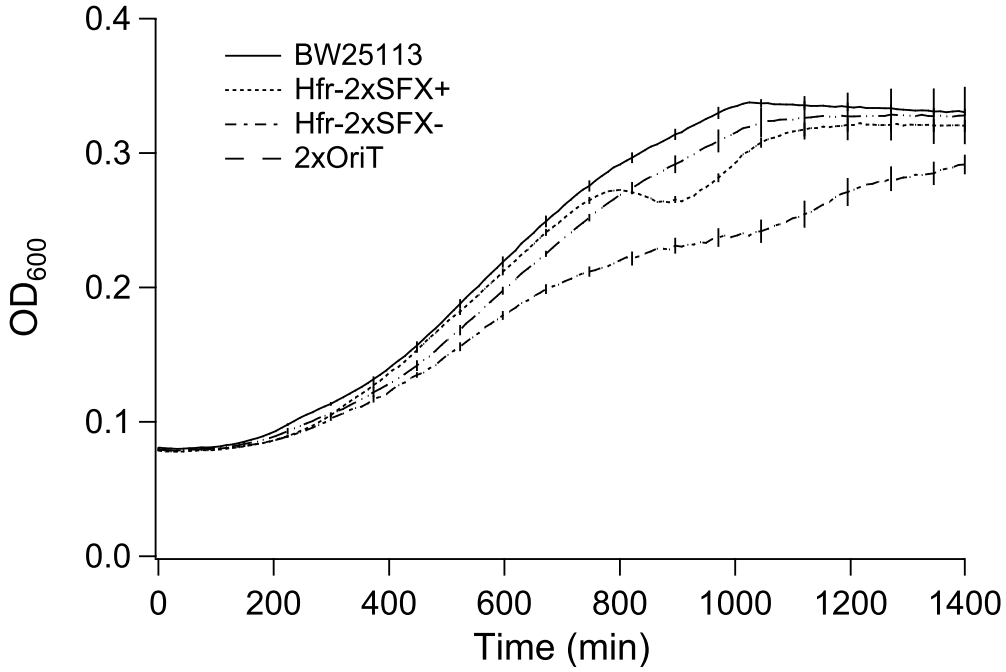

Figure 3.3: Growth of Hfr-2xSFX+, 2xSFX-, BW25113 2xOriT, and BW25113 strain in M9 minimal media supplemented with 0.5% glucose (w/v) and 25 μg/ml tryptophan. The Hfr-2xSFX- strain grows somewhat more slowly (6-17%) than Hfr-2xSFX+ and 2xOriT.

tionary trajectories. The estimated mutation rates for both strains, however, are not significantly different from one another using a 95% confidence interval (2xOriT: $[4.16 \cdot 10^{-9}$- $1.92 \cdot 10^{-8}]$, and Hfr-2xSFX-: $[4.91 \cdot 10^{-9}$,-$2.13 \cdot 10^{-8}]$), indicating that there are no untoward genetic effects caused by eliminating surface exclusion or the integrated F plasmid in the Hfr strains.

In addition to growth physiology of the strains, the ability of the Hfr strains to transfer DNA to recipients was assessed to determine how much recombination could be expected. Both Hfr strains served as donors (sources of DNA) in this assay, and SFX+,- recipients were utilized (X892 and X892 $\Delta(traST)$) to determine how much removing surface exclusion impacted overall transfer efficiency. Figure 3.4 shows that

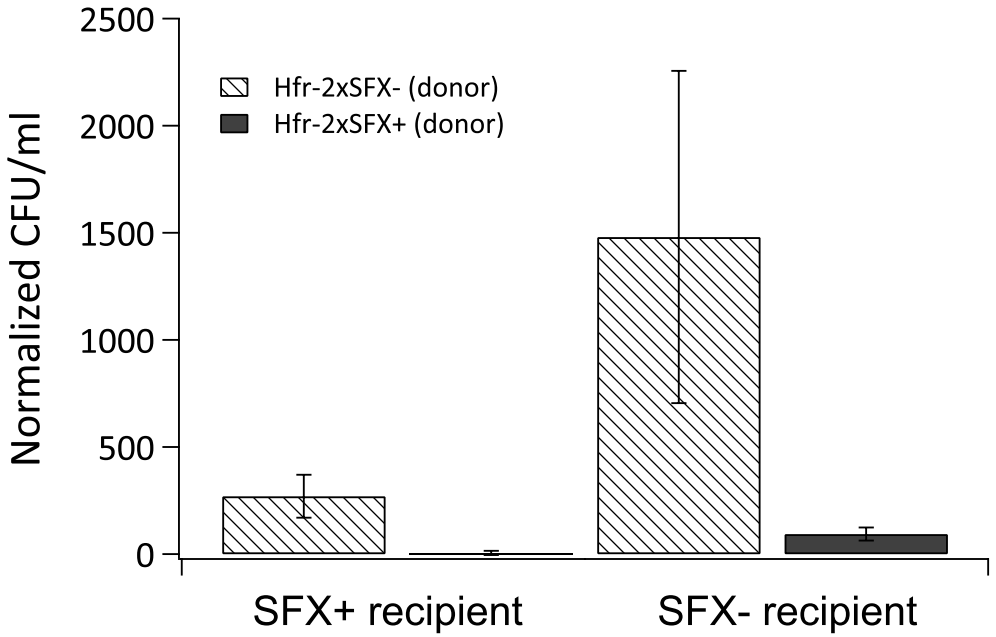

Figure 3.4: Number of recombinant progeny obtained when using Hfr-2xSFX+,- as donors to SFX+,- recipients (X892 and XΔ). Hfr-2xSFX- was significantly better at mating with SFX- recipients compared to Hfr-2xSFX+. Note that surface exclusion on the donor can also disrupt mating, as TraS and TraT interact with recipient proteins (TraG and OmpA [120]) to prevent Hfr-Hfr matings. When mating with the X892 Hfr strain, many fewer recombinant progeny are generated by both strains, though the Hfr-2xSFX- is still more efficient at conjugation in this case.

when mating with either type of recipient, the Hfr-2xSFX- strain is more efficient than the Hfr-2xSFX+ strain. This result is expected given that surface exclusion typically reduces Hfr-Hfr (F^+-F^+) mating by a factor of 100-300 fold [1]. One reason that Hfr-2xSFX+ is not able to efficiently mate with the surface exclusion deficient recipient is that TraS and TraT are thought to interact with recipient proteins (TraG and OmpA) that are always expressed [120]. These results indicate that recombination should be much more frequent within the Hfr-2xSFX- population compared to Hfr-2xSFX+. Mating each Hfr strain with F^- recipient (BW25113/pTB8), however, revealed that Hfr-2xSFX+ is more efficient (approximately 25-fold) at producing re-

combinant progeny than Hfr-2xSFX- (data not shown). This efficiency discrepancy stems from the fact that surface exclusion deficient strains are deficient at forming mating aggregates [1], a necessary step in the conjugation process. Despite this disadvantage, the size of this effect is small (25 fold versus 100-300 fold) compared to surface exclusion, leading to an enhancement in the overall frequency of recombination for $\Delta traST$-$\Delta traST$ matings. Given that evolving populations are initially isogenic and not composed of Hfr and F^- cells, the poor performance of Hfr-2xSFX- strain when mating with F^- recipients will not impact the utility of the strain.

3.5 Evolution Experiments

The ultimate goal of this work is to develop an evolutionary engineering system that incorporates continuous recombination to produce improved industrial strains. Given that most relevant complex phenotypes are poorly understood and involve unknown genetic mechanisms, we expect our system to expedite the rate of improvement in evolutionary experiments. As a proof of principle, we chose to compare the ability of the BW25113 2xOriT and Hfr strains to adapt to progressively increasing levels of antibiotic stress. Three antibiotics were chosen based on the characterized complexity of their respective adaptive landscapes: chloramphenicol [320], streptomycin [122], and trimethoprim [320, 122]. Chloramphenicol and streptomycin target different aspects of protein synthesis [103, 304]; a wide range of mutations are available to improve drug tolerance, ranging from increased expression of drug exporters to mutation of the appropriate ribosomal protein target site. The adaptive landscapes for both drugs likely contain many disparate beneficial mutations that can be used to improve tolerance. Recombination should be effective at increasing tolerance on this type of landscape by combining multiple mutations into a single strain [169, 271]. Trimethoprim resistance, in contrast, is principally due to a small number

of highly beneficial mutations in the promoter and coding sequence of dihydrofolate reductase (DHFR) [320]. Adaptation is accordingly constrained to a single, reproducible path characterized by repeated, rapid fixation of newly generated mutants. Given the minimal co-existence time of different beneficial mutations, recombination is not expected to improve the acquisition of trimethoprim resistance meaningfully.

Table 3.3: Summary of evolution experiments

Antibiotic	Maximum C_{ab}	Fold Inc.	Doubling Time (days)
CM			
Hfr 2xSFX+	103.55 (63.06)	34.52	2.64 (0.33)
Hfr 2xSFX-	182.40 (0.0)	60.8	2.15 (0.18)
2xOriT	77.65 (56.85)	25.88	3 (0.56)
TRI			
Hfr 2xSFX+	67.33 (66.55)	42.08	NC†
Hfr 2xSFX-	132.27 (73.16)	82.67	NC†
2xOriT	115.20 (85.29)	72	NC†
STR			
Hfr 2xSFX+	131.43 (47.75)	65.72	2.39 (0.35)
Hfr 2xSFX-	190.91 (81.64)	95.45	1.97 (0.20)
2xOriT	91.90 (27.39)	45.95	2.37 (0.21)

Summary of the evolution experiments performed for this work. Standard deviations (if provided) are given in parentheses. †: not calculated.

We tested these hypotheses by performing serial batch evolution experiments with BW25113 2xOriT (asexual), Hfr-2xSFX+ (Hfr, with surface exclusion), and Hfr-2xSFX- (Hfr, without surface exclusion). Broadly speaking, the results confirm that recombination does enhance evolutionary outcomes on complex fitness landscapes, while evolution on mutation-limited landscapes such as trimethoprim resistance see little benefit when recombination is introduced. No loss of mating efficiency or the integrated F plasmid was observed (with one exception detailed below) in any of the

Hfr strains. The specific results for each evolution experiment are detailed below. Statistical analysis was performed using the Student's T-test with a 95% confidence threshold.

3.5.1 Streptomycin

Replicate populations were evolved in minimal media supplemented with increasing amounts of streptomycin for 15 days according to the ramping rules laid out in the Materials and Methods section. Initially, the Hfr-2xSFX+ strain was less inhibited by streptomycin and required a higher initial STR concentration (3 μg/ml) to achieve the desired inhibitory effect. The average level of streptomycin over time for each set of replicate populations is shown in Figure 3.5. Hfr-2xSFX- populations generally achieved a higher level tolerance than either Hfr-2xSFX+ or BW25113 2xOriT replicates. Summary statistics for each population are provided in Table 3.3. The differences observed in the final antibiotic concentration between the populations are primarily due to instances where, following inoculation into media with an increased antibiotic concentration, a population failed to reach the OD necessary for subsequent transfer within 24 hours, leading to a temporary plateau in antibiotic resistance. Growth failures occurred more frequently and at lower STR concentrations for the BW25113 2xOriT replicates, but also occasionally occurred at higher antibiotic concentrations for both Hfr strains (Figure 3.5). After the conclusion of the experiment, the mating ability of the Hfr replicates was qualitatively assayed via spotting on plates following liquid conjugation with a F^- recipient. No significant loss of mating efficiency was observed compared to the parental controls (data not shown).

A direct method to determine whether these differences in antibiotic tolerance are significant is to compare antibiotic concentration doubling times using Equation

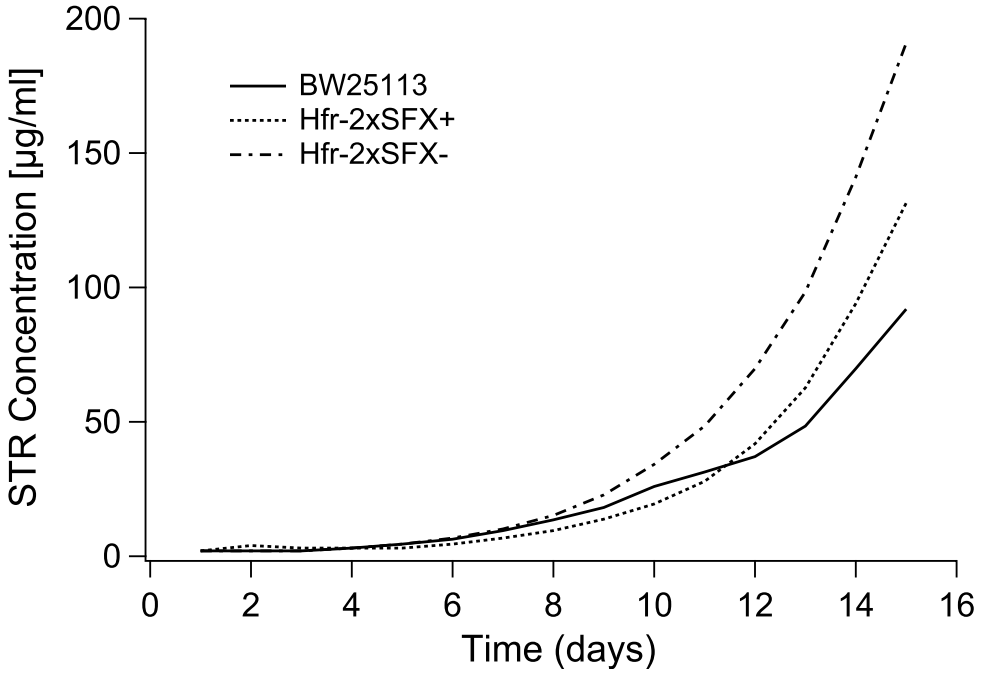

Figure 3.5: The average concentration of streptomycin during the serial batch evolution experiment. Six replicates are used per strain. Given that large differences in replicate antibiotic concentrations correspond to only one or two step-wise increases, error bars are omitted for clarity. The p-values for doubling time comparisons are: Hfr 2xSFX+ versus Hfr 2xSFX-: $P = 0.21$, Hfr 2xSFX+ versus 2xOriT: $P = 0.23$, and Hfr 2xSFX- versus 2xOriT: $P < 0.01$.

3.1; slower adaptation or lower tolerance will increase the antibiotic doubling time, while a smaller doubling time can be equated to more rapid tolerance acquisition, up to the limit imposed by the selective pressure increases. The doubling times given in Table 3.3 reveal that Hfr-2xSFX- has a significantly faster antibiotic doubling time ($P < 0.01$) compared to the asexual strain. The larger variance in doubling time for the Hfr-2xSFX+ strain makes it difficult to distinguish from BW25113 2xOriT or Hfr-2xSFX- ($P \approx 0.23$ for both cases). The magnitude of the Hfr-2xSFX- improvement compared to 2xOriT is relatively small (17%) which may suggest that streptomycin resistance loci are distant from the origin of transfer or suffer from

antagonistic interactions. It is also possible that some beneficial mutations conferring streptomycin resistance have large fitness effects that enable rapid fixation and reduce time available for recombination between mutant lineages.

Following the conclusion of the evolution experiments, we also attempted to determine if the ramping scheme correlated with actual tolerance by quantifying drug resistance. In this case, the inhibitory effect ((OD_{drug}/OD_{unchal})) of a range of streptomycin concentrations (generally 1X-4X the concentration used at the time of isolation) was assessed for population samples revived from frozen stocks after 5, 12, and 15 days of evolution. Cultures that could only tolerate lower antibiotic concentrations during the evolution experiment were consistently identified as being more sensitive to streptomycin in this assay. This result shows that the slow improvement for certain replicates in this experiment (and chloramphenicol tolerance; see below) was not an artifact of the ramping scheme, but actually indicative of insufficient antibiotic resistance. Cultures that met the criterion for antibiotic concentration increases without interruption over the course of the experiment were less inhibited by increased concentrations of streptomycin, as expected. Based on these data, the ramping scheme provides a strong selective pressure over the course of the experiment and correlates well with the actual antibiotic resistance of each population.

In this case, the benefit of recombination appears to be small despite the reported polygenic fitness landscape. In order to better confirm that our system effectively harnesses recombination to speed adaptation, we performed an additional evolution experiment evolving for chloramphenicol tolerance to determine if this enhancement was consistent for another phenotype with a qualitatively complex fitness landscape.

3.5.2 Chloramphenicol

Asexual populations exposed to chloramphenicol have been demonstrated to increase their antibiotic tolerance in an approximately continuous fashion as they acquired new mutations, suggesting that there are many alleles that have a small, beneficial effect on CM tolerance [320]. The use of recombination in this case would be ideal, as lineages carrying weakly beneficial mutations will coexist for extended periods of time, allowing for extensive recombination. Assuming that the majority of mutations do not have antagonistic effects, then free recombination should have a significant effect on the observed evolutionary outcome. The average antibiotic concentrations for the strains over 14 days, shown in Figure 3.6A, confirm this expectation; the Hfr-2xSFX- strain improves significantly faster and to a greater degree than the other strains (Table 3.3). After 8 days, the 50% increases in CM concentration per step appeared to be insufficient selective pressure; henceforth we double the antibiotic concentration in the replicates at each step. One Hfr-2xSFX+ replicate population lost the integrated F during the course of the evolution experiment, but in all other cases the integrated F appeared to be stable. The doubling times for each strain (Table 3.3) confirm that the Hfr-2xSFX- strain improves in tolerance significantly more quickly than either the Hfr-2xSFX+ ($P = 0.014$, 18.4% faster adaptation) or BW25113 2xOriT strains ($P = 0.012$, 28.1% faster adaptation). The Hfr-2xSFX- strain also improved more rapidly at lower concentrations of CM as well. This result represents a substantial reduction in the time required to develop the desired level of tolerance. Note that the difference in doubling between Hfr-2xSFX+ and BW25113 2xOriT is not significant ($P = 0.214$).

Doubling time variance appears to be inversely related to recombination proficiency: the Hfr-2xSFX- replicates tended to have broadly similar antibiotic tolerances

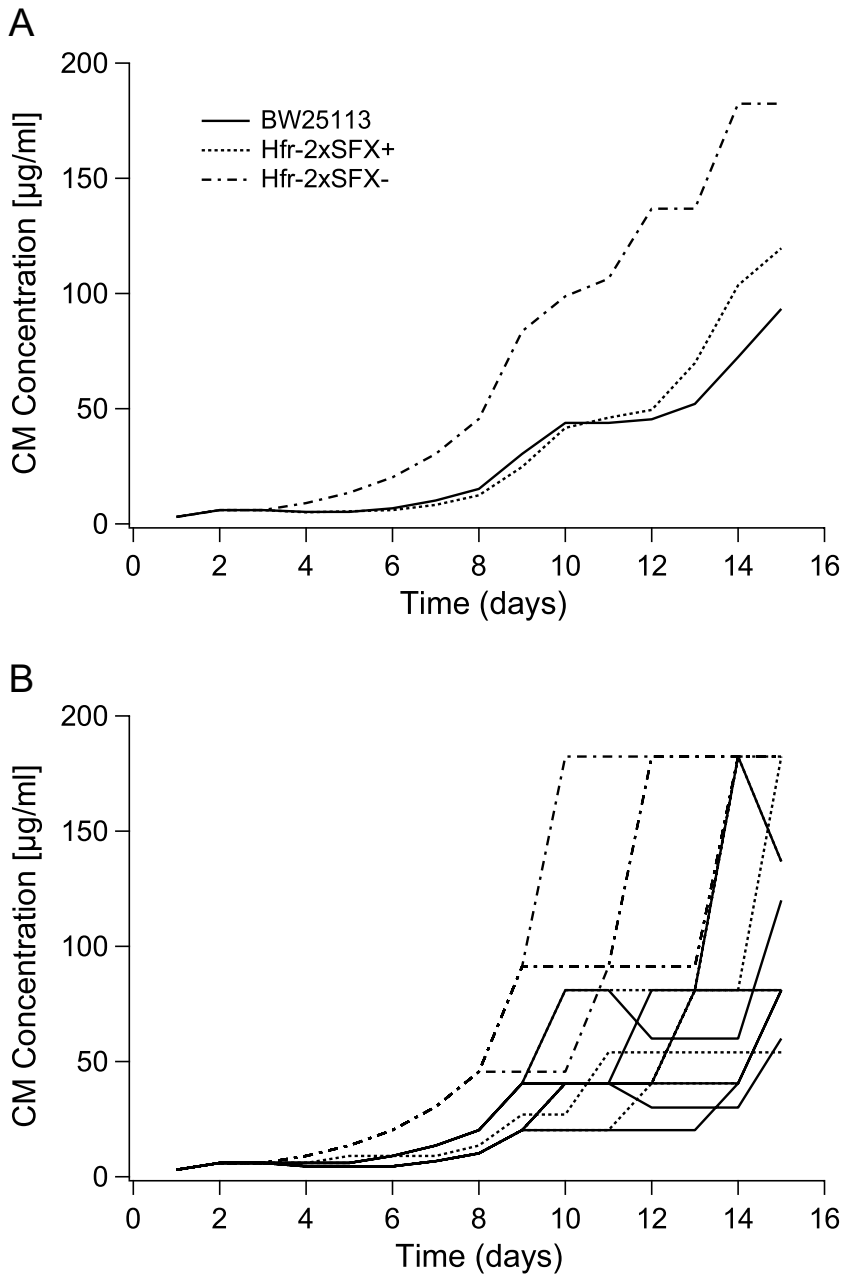

Figure 3.6: A. The average concentration of chloramphenicol during the serial batch evolution experiment. Six replicates are used per strain. Errors are omitted due to the large amount of variation caused by the exponential ramping scheme. B. Antibiotic concentrations in each replicate. Note that not all replicates are visible due to overlapping antibiotic concentration curves over the course of the experiment. P-values for each doubling time comparison are as follows: Hfr 2xSFX+ versus Hfr 2xSFX-: $P < 0.014$, Hfr 2xSFX+ versus 2xOriT: $P = 0.21$, and Hfr 2xSFX- versus 2xOriT: $P < 0.012$.

and rarely failed to miss growth targets, except at higher antibiotic concentrations, while BW25113 2xOriT replicates had widely varying antibiotic resistances and frequently required extended growth (*i.e.* more plateaus (Figure 3.6B)) as the concentration of CM increased. Clonal interference is likely the explanation for this result, as different beneficial mutations would be generated and lost in each replicate over the evolutionary time course, leading to disparate antibiotic tolerance levels. Mutants that are fittest at low CM concentrations likely arose in many of the BW25113 2xOriT replicates, but as the antibiotic concentration increased, a significant amount of time was spent waiting for the next adaptive mutant to arise on these existing genetic backgrounds based on the number of "pauses" observed before CM concentrations could be increased. The Hfr-2xSFX- replicates could harness positive epistasis between distinct mutations through recombination, leading to the observed increased adaptation rate.

3.5.3 Trimethoprim

Adaptation to chloramphenicol and (to a lesser extent) streptomycin represent ideal cases where recombination may be harnessed to combine multiple, independent lineages into a single mutant line. As stated previously, adaptation to trimethoprim seems antithetical to this ideal: effectively, only a few genes and regulatory regions appear to be involved [227, 320], and mutations are generally acquired in a particular order during the course of adaptation [320]. We expect that there will be little to no advantage to recombination on this type of landscape due to the rapid fixation of highly beneficial mutations conferring TM resistance, leading all of the strains to adapt to trimethoprim at a similar rate.

Overall, the data confirmed this hypothesis. Very little difference in average tolerance was indeed observed among the different strains (Table 3.3). Antibiotic

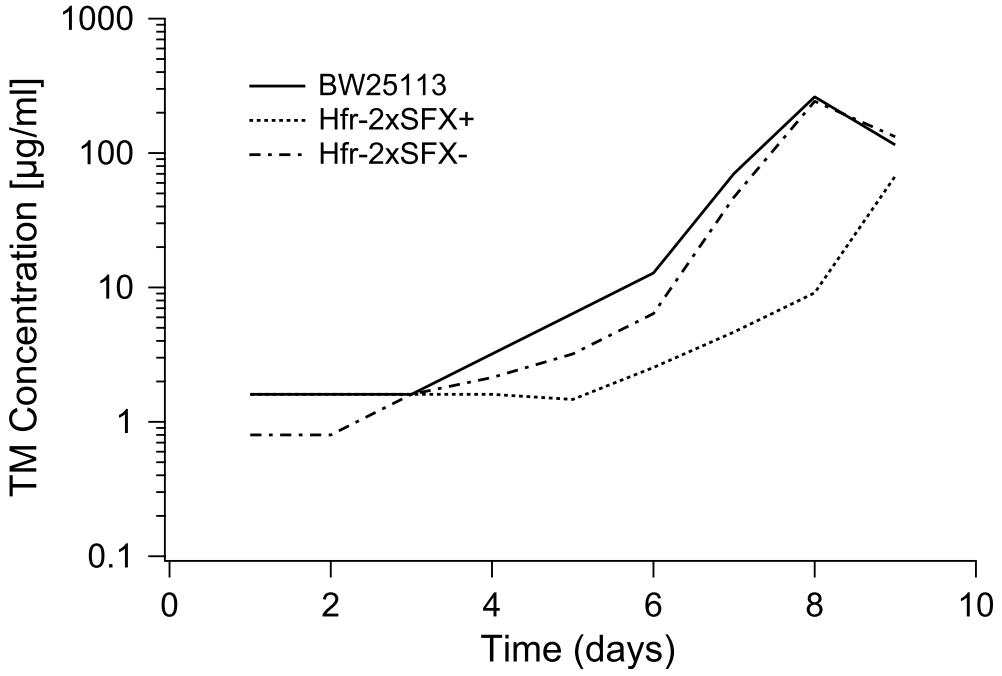

Figure 3.7: The average concentration of trimethoprim (TM) during the serial batch evolution experiment. Six replicates are used per strain. Error bars are also omitted in this case due to the exponential ramping scheme.

concentration were increased more rapidly than in the previous evolution experiments, as identified mutations appear to increase TM tolerance 10-fold individually. The average TM concentration for each strain is shown in Figure 3.7 on a log scale to account for this change. There is little difference among the strains, though Hfr-2xSFX- was initially slightly more sensitive to trimethoprim and required a lower initial TM concentration (0.8 μg/ml versus 1.6 μg/ml) than the other evolution strains. Despite this increased sensitivity, all strains approached the same level of tolerance on average, showing that fitness defects of this sort are rapidly abolished and should not limit the applicability of this system. Variation between replicates for each strain was quite large, indicating that clonal dynamics played a strong role in shaping the observed evolutionary outcomes by favoring the existence of a sin-

gle dominant clone in each replicate population. It is unlikely that each population would contain the same exact adaptive mutant, leading to a wide variation in antibiotic tolerance between replicates. Highly fit clones would also tend to displace other less fit adaptive mutants through competition and random drift, keeping the population heterogeneity to a minimum in each replicate. These conditions disfavor recombination as there is little time to facilitate genetic exchange prior to the extinction of all but a single, fittest genotype. The similarity in adaptation between the Hfr and asexual strains dovetails with the expectation that recombination is most useful when various adaptive mutants can co-exist, as expected.

3.5.4 Implications for evolutionary engineering

These findings support the assertion that if a compatible phenotype is selected for study, such as one where many small-effect mutations exhibiting positive epistasis contribute to overall strain fitness, the use of constant recombination under selective pressure during evolution results in faster improvement than standard serial batch evolution experiments with asexual *E. coli*. The success of genome shuffling, despite the low overall recombination efficiency of the method [79], to generate improved strains [359, 245, 308, 347, 269] is additional evidence for the utility of recombination as part of the evolutionary engineering toolbox. In order to further enhance evolutionary outcomes, whole cell mutagenesis could also be incorporated into this protocol to increase the genetic diversity of the initial population, providing additional grist for the evolutionary mill.

There are also situations where this method would likely yield little benefit over evolutionary engineering with purely asexual strains. For phenotypes with fitness landscapes conceptually similar to trimethoprim, given the predominance of individual mutations with large fitness effects in those cases, recombination would probably

be minimal and limited only to brief periods during population sweeps where genetic heterogeneity exists. Other methods for engineering complex phenotypes, such as high throughput promoter engineering [335, 6] or genomic libraries [205, 225, 268], to improve fitness may be more appropriate in these cases, as they do not depend on interactions between mutant strains to identify alleles of interest. However, it is very unlikely that any one method would be objectively superior to all other strain engineering techniques, no matter the target phenotype, given their fundamentally different underlying methodologies. All of these techniques have also been successfully applied to characterize the molecular mechanisms that govern various phenotypes of interest or to generate improved strains. The *in situ* recombination method developed here best represents a complementary approach to improve adaptive laboratory evolution by attenuating clonal interference, rather than a direct replacement of these alternative methodologies for complex phenotype engineering.

3.6 Conclusions

These results, taken altogether, support the conclusion that recombination-proficient *E. coli* are suitable for evolutionary engineering experiments from both the standpoints of stability and effectiveness. In the case of the former, only a single loss of the integrated F was observed for the 36 Hfr replicates used in the evolution experiments. Even if mutants with impaired or absent mating competency arise during the course of the experiment, recombination with the remaining mating competent individuals in the population should reduce the likelihood of defective F fixation. It is likely that the fitness benefit of mutations conferring tolerance to the phenotype of interest will fix more rapidly than potential F-disrupting mutations as well. Our attempts to use a F plasmid (rather than an integrated F), however, did show that the F[$\Delta(traST)$] plasmid is unstable in a host with multiple origins of transfer (data not shown). This

failure, along with the impact on host physiology detailed previously, does imply that excision or inactivation of the integrated F is under positive selection in the absence of surface exclusion. As a result, when performing a long-term evolution experiment, the Hfr strains developed in this work should be tested periodically to assess mating efficiency and ensure that the integrated F is still present in the population.

The Hfr-2xSFX- strain developed in this work was able to improve tolerance 17% to 28% more quickly than either control strain in response to streptomycin or chloramphenicol challenges, representing a significant reduction in time and cost over conventional evolutionary techniques. It may be the case that these landscapes are strong mutation/weak selection regimes [300], where various genotypes coexist for significant periods of time, but deep sequencing of the population is required to confirm the validity of this hypothesis. Performance of Hfr-2xSFX- on simpler landscapes, such as the one defined by trimethoprim resistance, was similar to that obtained from the asexual control. Previous literature suggests that evolution for trimethoprim resistance best resembles a strong mutation/strong selection regime, matching theoretical expectations. Taken together, we believe these results demonstrate that this system is well suited for use in evolutionary engineering to improve complex phenotypes, such as biofuel tolerance, which have been shown to incorporate a wide variety of mutations as tolerance improves over time [217]. Future development of this system will focus on creating an inducible mating system to control when recombination occurs and to ameliorate the sensitivity to membrane irritants triggered by the absence of surface exclusion.

We have harnessed an existing conjugation system capable of mating in well-mixed liquid cultures to engineer a practical *in situ* recombination system for *E. coli* for the purposes of evolutionary engineering. The developed strains are genetically stable and can be used for evolution experiments easily without any need to

interrupt adaptive evolution to allow for recombination, representing a significant improvement over conventional genome shuffling. Recombination, further enhanced by the disruption of surface exclusion from the Hfr-2xSFX-, proved beneficial in improving tolerance to inhibitors (streptomycin and chloramphenicol) that have complex tolerance mechanisms by allowing beneficial mutations to combine into a single genetic background through conjugation rather than sequential mutation. Whole cell mutagenesis strategies may also be employed to increase the initial diversity of a recombination-proficient population before adaptive evolution, increasing the likelihood of recombination between different mutant lineages. Phenotypes involving a small number of mutations with large fitness effects, such as trimethoprim resistance, are unlikely to benefit from recombination. This system should prove useful to evolutionary engineers seeking to develop strains with improved industrial phenotypes such as furfural, acid, and acetate resistance or product tolerance.

4. UNRAVELING OSMOTOLOLERANCE IN *ESCHERICHIA COLI* USING A PSEUDOSEXUAL MODEL SYSTEM

4.1 Summary

Biocatalyst robustness toward stresses imposed during fermentation is important for efficient bio-based production. Osmotic stress, imposed by high osmolyte concentrations or dense populations, can significantly impact growth and productivity. In order to better understand the osmotic stress-tolerance phenotype, we evolved sexual (capable of *in situ* DNA exchange) and asexual *Escherichia coli* strains under sodium chloride (NaCl) stress. Isolates from each population had significantly improved growth under selection and could grow in up to 0.85 M NaCl, a concentration that completely inhibits the growth of the unevolved parental strains. Whole genome resequencing revealed frequent mutations in genes controlling n-acetylglucosamine catabolism (*nagC*, *nagA*), cell shape (*mrdA*, *mreB*), osmoprotectant uptake (*proV*), and surface fibers (*fimA*). Possible epistatic interactions between *nagC*, *nagA*, *fimA*, and *proV* deletions were also detected when reconstructed as defined mutations. Transcriptome analysis revealed significant changes in *ompACGL* porin synthesis, and activation of sulfonate uptake systems in the adaptive mutants, providing additional insight into the genotype-phenotype connection for these strains. These findings expand our current knowledge of the osmotic stress phenotype and will be useful for the rational engineering of osmotic tolerance into industrial strains in the future.

4.2 Introduction

Escherichia coli, an important industrial microorganism for the production of a wide variety of fine chemicals, fuels, and proteins, has been extensively targeted to

improve its suitability as a biofactory. Strain development efforts have focused on improving tolerance of feedstocks containing toxic compounds [139, 216, 288, 5] or products [283, 14, 217, 99, 269]. Many environmental variables, including osmotic pressure, can negatively impact biocatalyst performance [293]. Use of non-conventional waste streams ,such as waste glycerol from biodiesel production or brackish water sources, to support microbial growth can also reduce process costs [243, 159, 94, 281] while reducing pressure on fresh water resources; however, these carbon and water sources generally contain high concentrations of salt that may be inhibitory to microbial growth. Previous studies have attempted to engineer improved osmotic tolerance in *E. coli* [183, 54, 280, 259], but overall, knowledge of the genetic mechanisms that confer tolerance of osmotic stress in general or to specific osmolytes remains limited. A detailed analysis of *E. coli* tolerance to osmolytes would therefore shed new insight into the molecular mechanisms underlying this complex phenotype.

Adaptive laboratory evolution [255, 96] is a promising approach to identify potentially novel osmotic tolerance mechanisms, as this technique requires no assumptions about the underlying genotype-phenotype relationship. Complex phenotypes such as enhanced resistance to biofuels [127, 14, 217, 269], lignocellulosic hydrolysates [319, 353, 5], antibiotics [320, 346, 277], and environmental conditions [273, 24, 71, 315, 97] have all been successfully characterized using this approach. In this study, sodium chloride (NaCl) was selected as the osmotic inhibitor. A recent evolutionary study aimed at characterizing cross-adaptation between several different stressors detected several potential mechanisms in a single evolved NaCl-tolerant isolate [97], but due to the possible existence of multiple adaptation mechanisms, additional information is needed to better understand the genetic bases of osmotolerance.

In a typical evolution experiment, isogenic biological replicate populations are propagated under gradually increasing selective pressure (in this case, osmotic stress)

to identify mutants with improved growth rates under these conditions. These mutants can then be analyzed using phenotypic, genetic, and transcriptional approaches to uncover the causal mutations conferring the phenotype of interest. In order to expedite this process, we utilize our previously developed evolution system that leverages horizontal gene transfer (HGT) within evolving *E. coli populations* to increase their rate of adaptation to various stressors [346]. This type of sexual exchange (or recombination) reduces competition between beneficial clones in a population by unlinking beneficial alleles from a single genome [92, 128] and permits additional adaptive paths on some fitness landscapes [339, 220]. The use of a sexual evolution system may therefore allow for faster improvement in osmotic tolerance compared to standard asexual *E. coli*, as has been shown for other phenotypes empirically[271, 72, 346, 181] and theoretically [222] in the past. Asexual populations tend to lose adaptive mutants through drift and competition [111, 300], whereas sexual populations can incorporate mutations from multiple competing lineages into a single genetic background to avert this outcome [74].

In order to identify novel adaptive mechanisms for osmotic (NaCl) tolerance, we have utilized adaptive evolution to generate osmotic tolerant mutants of three distinct *E. coli* strains: two capable of in situ recombination with different rates of mating [346] to reduce competition between beneficial clones and another, completely asexual strain. After being propagated for approximately 150 generations in the presence of increasing concentrations of NaCl, osmotolerant mutants were isolated, characterized, and sequenced to identify any genetic changes that occurred during evolution. The elucidated resistance mechanisms were then explored phenotypically to better understand their potential impact on *E. coli* physiology. Transcriptomic analyses of several mutants were subsequently conducted to better characterize the genotype-phenotype connection that resulted in enhanced osmotolerance.

4.3 Methods and Materials

4.3.1 Bacterial strains and growth media

All strains used for evolution in this study (Table 4.1) are previously developed BW25113 derivatives [346]. Briefly, Hfr-2xSFX(+,-) are conjugation proficient Hfr strains with an operon of F transfer proteins integrated at the *trp* locus [324]; surface exclusion genes are either present (+) or disrupted (-) in the strains, respectively. 2xOriT, an F- strain, was used as an asexual control. BW25113 was used to test the effect of overexpressed or disrupted genes on osmotic tolerance. Minimal M9 medium supplemented with 0.5% (w/v) glucose, 50 μg/ml tryptophan [324, 346] was used for routine cultivation and growth assays. Luria-Bertani (LB) broth and agar plates were used for strain isolation, transformation, and other analyses, as indicated. Sodium chloride (JT Baker) was utilized to adjust the osmotic strength of the medium during the evolution and for subsequent growth assays.

4.3.2 Adaptive evolution

Adaptive laboratory evolution was conducted via serial batch transfer experiments to improve the osmotic stress tolerance of Hfr-2xSFX-, Hfr-2xSFX+, and 2xOriT in parallel. Six replicate populations for each strain were inoculated from independent colonies to initiate the evolution experiment. Every 24 to 48 hours, a proportion (typically 1-3%, based on cell density) of each replicate population was diluted into fresh medium to ensure that each population underwent approximately 6-7 generations per transfer. Sodium chloride concentrations were increased in a step-wise manner over time based on observed population fitness increases. Each population replicate underwent approximately 150 generations over the course of the experiment. The fitness (S, Equation 4.1) of the evolving populations relative to their ancestral parents in 0.6-0.65 M NaCl was determined every 24 generations using

Table 4.1: Partial list of strains, plasmids used in this study

Strain or Plasmid	Genotype	Source
BW25113	$\Delta(araD\text{-}araB)567$, $\Delta lacZ4787(::rrnB\text{-}3)$, λ^-, *rph*-1, $\Delta(rhaD\text{-}rhaB)568$, *hsdR*514	CGSC[†]
2xOriT	BW25113 $\Delta mbhA::oriT$ $\Delta hyfC::oriT$	[346]
Hfr-2xSFX-	BW25113 $\Delta mbhA::oriT$ $\Delta hyfC::oriT$ *trp*::F[$\Delta traST$] (gen^R)	[346]
Hfr-2xSFX+	BW25113 $\Delta mbhA::oriT$ $\Delta hyfC::oriT$ *trp*::F (gen^R)	[346]
pCA24N	Cm^R, used as a control recipient for mating	[170]
G1	Isolate from Hfr-2xSFX- Population 1	This work
G2	Isolate from Hfr-2xSFX- Population 2	This work
G3	Isolate from Hfr-2xSFX- Population 3	This work
G4	Isolate from Hfr-2xSFX- Population 4	This work
G5	Isolate from Hfr-2xSFX- Population 5	This work
G6	Isolate from Hfr-2xSFX- Population 6	This work
A1	Isolate from 2xOriT Population 1	This work
A2	Isolate from 2xOriT Population 2	This work
A3	Isolate from 2xOriT Population 3	This work
A4	Isolate from 2xOriT Population 4	This work
A5	Isolate from 2xOriT Population 5	This work
A6	Isolate from 2xOriT Population 6	This work
H1	Isolate from Hfr-2xSFX+ Population 1	This work
H2	Isolate from Hfr-2xSFX+ Population 2	This work
H3	Isolate from Hfr-2xSFX+ Population 3	This work
H4	Isolate from Hfr-2xSFX+ Population 4	This work
H5	Isolate from Hfr-2xSFX+ Population 5	This work
H6	Isolate from Hfr-2xSFX+ Population 6	This work

[†]: Coli Genetic Stock Center, Yale University. Note that BW25113 was used as the host for all overexpression plasmids and the control vector pCA24N.

growth assays in microtiter plates to track their rates of adaptation; a logarithmic model (Equation 4.2) was then fitted to the fitness measurements for each population and used to calculate their overall rate of improvement throughout the adaptive evolution experiment. The constant α is a shape parameter for the logarithmic curve. The choice of this model is informed by the fact that fitness improvements

of this form are often observed during adaptive evolution [24]. Potential external and cross contamination of the experiment was monitored using PCR genotyping using the *mbhA*::*oriT* locus (present in all 3 strains), and growth on LB gentamicin agar to discriminate between the gentimicin resistant Hfr strains and the sensitive 2xOriT strain. Culture samples from each population were periodically plated onto non-selective LB to detect any unusual colony morphologies. Mating assays were performed weekly as previously described [346] to detect conjugation competence and any cross contamination between replicates by identifying any mating competent subpopulations in nominally asexual replicates.

$$S = \frac{\mu_i}{\mu_{\text{2xOriT}}} - 1 \tag{4.1}$$

$$S(t) = \alpha \log(t) \tag{4.2}$$

4.3.3 Mutant isolation and screening

One clonal isolate was randomly obtained from each evolved population at the end of the evolution experiment after streaking the evolved populations onto LB agar for single colonies. Isolates from the Hfr-2xSFX-, Hfr-SFX+, and 2xOriT populations are prefaced with G, H, and A respectively. All isolates were propagated in the absence of NaCl for at least 10 generations prior to any phenotypic analysis to ensure that any observed phenotypes were the result of mutation, rather than transient transcriptional or epigenetic adaptation. Following an initial fitness screen in M9 supplemented with 0.5% (w/v) glucose and 0.65 M NaCl at 37 °C to ensure that all isolates exhibited enhanced resistance to NaCl compared to the parental strains, the fitness of all isolates was analyzed in other osmotic or general stress conditions (excess

glucose [54 g/L], n-butanol [0.8% v/v], low pH [pH = 6.0], or elevated temperature [42 °C]). Two independent biological replicates were used for each experiment. All growth assays were performed in 96 well microtiter plates using a plate shaker and incubator (TECAN Infinite M200) at 37 °C (except for thermal stress assays) with four technical replicas per strain per condition. Relative mutant fitness (S) and improvement (RI) were respectively calculated for each condition (if necessary) using Equations 4.1 and 4.3.

$$RI = \frac{\mu_{i,C}/\mu_{i,U}}{\mu_{\text{2xOriT},C}/\mu_{\text{2xOriT},U}} - 1 \tag{4.3}$$

4.3.4 Fitness distribution analysis

Six randomly isolated clones were obtained by streaking from each of the evolved Hfr-2xSFX- and 2xOriT replicate populations (for a total of 36 random isolates per strain). Each colony was inoculated into 2 ml glucose minimal media supplemented with tryptophan and allowed to grow for 24 hours. The fitness of each isolate (relative to 2xOriT) was then determined using growth in a microtiter plate under 0.65 M NaCl challenge, as described previously. Four technical replicates were used per screened isolate.

4.3.5 Mutation rate under osmotic stress

The mutation rate of Hfr-2xSFX- and 2xOriT were measured using a standard fluctuation test [356] under osmotic stress to determine if the strains have unequal mutation rates under osmotic stress, which would influence their relative adaptation rates. Eight biological replicates of each strain were grown in unchallenged conditions (glucose minimal media with tryptophan) overnight and diluted 10^3-fold to reduce the cell concentration to several thousand cells per ml, reducing the possibility of

jackpot mutants. 1% of these diluted cultures were inoculated into the same media with 0.55 M NaCl and incubated for 72 hours at 37° C with shaking until reaching stationary phase. All but 10 μl of each replicate was then pelleted, resuspended in LB, and plated on LB plates with 20 μg/ml nalidixic acid as done previously [346]; the remaining 10 μl were diluted 10^6-fold and plated onto LB to measure CFU/ml. After 24 hours of incubation at 37° C, colonies on each plate were counted. FALCOR [134] was then used to estimate the mutation rate for 2xOriT and Hfr-2xSFX-. For 2xOriT, the mutation rate is 1.43 mutants per 10^9 cells, 95% confidence interval: [0.72, 2.31]; for Hfr-2xSFX-, the mutation rate is 1.52 mutants per 10^9 cells, 95% confidence interval: [0.71, 2.54]. The difference in mutation rate between the strains is not statistically significant.

4.3.6 Hyperosmotic shock tests

Single colonies of each isolate (G1-G6, A1-A6, and H1-H6) and the parental controls were inoculated into glucose minimal media and allowed to grow overnight. The stationary phase cultures were then diluted in fresh medium and propagated until mid-exponential phase (OD $\approx$ 0.3-0.6). The cultures were then normalized to ensure equivalent optical densities, pelleted, and resuspended in glucose minimal medium supplemented with 5.45 M NaCl and incubated at 37 °C for 2 hours. Each sample was then serially diluted up to 10000-fold in minimal media, spotted on LB plates, and incubated overnight at 37 °C. The number of colonies in each 1000-10000X dilution were then counted. Each assay was performed in duplicate with independent biological replicates.

4.3.7 Genome sequencing and verification

The evolved Hfr-2xSFX- (G1-G6) and 2xOriT (A1-A6) isolates, along with the unevolved parental strains, were sequenced to discover the genotype underlying the

observed sodium chloride tolerance. Genomic DNA was extracted from each isolate using the Zymo Fungal/Bacterial Genomic DNA miniprep kit and quantified using the NanoDrop spectrophotometer. Samples of genomic DNA were analyzed using gel electrophoresis to check for degradation and RNA contamination. Library preparation and sequencing were performed by The Texas A&M Genomics Center for sequencing on the Illumina HiSeq 2000 platform using 100 bp single-end reads. An average of 286-fold coverage was obtained for each isolate. Reads were assembled against the MG1655 reference genome and each mutant genome was compared to the parental sequences to identify any *de novo* mutations. The approach to mutation verification depended on the type of mutation; 11 SNPs and deletions were verified with Sanger sequencing and other large deletions were verified with junction-specific PCR.

4.3.8 Transcriptional analysis

Two biological replicates of A2, A4, G2, G3, G5, and G6 were used for microarray analysis. Two colonies of each strain along with the 2xOriT (A parent) and Hfr-2xSFX- (G parent) were inoculated into glucose minimal media supplemented with tryptophan and grown overnight at 37 °C with shaking. 500 μl of each overnight was diluted 50-fold (OD $\approx$ 0.02) into 250 ml baffled flasks containing 25 ml of glucose minimal media, supplemented with tryptophan and 0.55 M NaCl. Samples were grown at 37 °C and 225 rpm until reaching an OD $\approx$ 0.5, then harvested by rapid filtration using Nalgene Analytical filters and immediate resuspension in 10 ml of RNALater (Sigma). DNA-free total RNA was extracted using the Quick-RNA Miniprep kit (Zymo Research) according to the manufacturer's protocol. RNA was quantified using the Qubit RNA BR assay (Life Technologies), and quality was assessed using agarose gel electrophoresis. cDNA was then generated from 5 μg of sample RNA us-

ing Superscript III (Life Technologies) with added 300 μM amino-allyl dUTP (Fisher Scientific), 200 μM dTTP, and 500 μM of dATP, dCTP, and dGTP (Promega) to allow for indirect labeling. Following ethanol precipitation to remove unincorporated nucleotides, samples were labeled with Cy3 or Cy5 (GE Amersham) as required, and then hybridized onto *E. coli* V2 gene expression microarrays (Agilent Technologies). Microarray slides were washed and scanned (Axon 4200A) as specified by the manufacturer. Microarray data were then normalized using the LOWESS algorithm using MIDAS [285] and analyzed for statistical significance using the rank product method implemented in TM4 MeV [285]. Gene ontology enrichment analysis was performed using the DAVID software package [89] using the default settings. Microarray data are deposited in the Gene Expression Omnibus (GEO) under accession number GSE51611.

4.3.9 Screening genes of interest

Based on genomic data, 15 genes were selected for fitness assays in batch cultures to determine their benefit under conditions closely matching the evolution experiment. Overexpression plasmids were obtained from the ASKA GFP(-) collection [170], miniprepped from the AG1 host using the Zyppy miniprep kit (Zymo Research), and transformed into chemically competent BW25113 using standard procedures. For compensatory assays, NagA, NagC, ProV, and FimA overexpression plasmids were transformed into mutants with possibly inactivating lesions in these genes. Knockout strains were procured from the Keio collection [53] as needed. The kanamycin resistance (kan^R) marker on each knockout strain was flipped out by transformation with pCP20 and incubation at 43 °C overnight using the Datsenko and Wanner method [83] as needed. Strains with multiple defined knockouts were generated using serial *P1vir* transduction (donated generously by Dr. Deborah

Siegele) and flipping to allow for reuse of the kan^R marker. Fitness assays were conducted in screw-cap tubes with 5 ml of M9 minimal medium, supplemented with glucose, tryptophan and 0.55 M NaCl. Optical density readings were taken every two hours until exponential growth had been sustained for at least 3 doublings. Fitness of the overexpression and knockout strains relative to the appropriate references (empty vector controls for the overexpression strains, BW25113 for knockout strains) was then calculated using Equation 4.1. Three biological replicates per strain were analyzed.

4.3.10 Indole measurements

The G2, G3, G5, G6, A2, and A4 mutants, along with the Hfr-2xSFX- and 2xOriT parental strains, were grown in M9 medium supplemented with 0.55 M NaCl, 0.5% (w/v) glucose, and 50 μg/ml tryptophan until reaching stationary phase. Extracellular indole concentrations were measured using standard procedures with Kovac's reagent [95]. Briefly, 1 mL of each culture was pelleted and the supernatant transferred to a fresh tube. Kovac's reagent (10 g p-dimethylaminobenzaldehyde, 50 ml hydrochloric acid, and 150 ml amyl alcohol) was then added to the supernatant and allowed to react. 100 μl of the indole layer was then removed by pipetting and diluted into 900 μl of HCl-amyl alcohol mixture. The OD540 of each diluted mixture was read using a Tecan Infinite M200 reader. Indole concentrations were calculated using a standard curve measured at the same time. Two biological replicates were used for each measurement, and the experiment was performed twice for a total of 4 biological measurements.

4.4 Results and Discussion

4.4.1 Adaptive evolution under NaCl challenge

Six replicate populations of 2xOriT, Hfr-2xSFX+, and Hfr-2xSFX- were subjected to gradually increasing NaCl concentration (0.55 M up to 0.75 M) over the course of approximately 150 generations. Over this time course, significant fitness improvements were observed in all evolving populations (Figure 4.1) indicating the successful selection for adaptive mutants in each population. When compared to the observed rates of fitness improvement in the 2xOriT populations ($4.54 \cdot 10^{-3}$/gen), the improvement in rates in Hfr-2xSFX+ and Hfr-2xSFX- populations are significantly larger ($7.35 \cdot 10^{-3}$/gen and $6.70 \cdot 10^{-3}$/gen, $P < 0.003$, Student's t test). However, it is possible that due to the initial higher sensitivity of the Hfr strains to NaCl, mutants with larger fitness improvements tended to arise in the sexual populations as a result of stronger selection, leading to an apparent increase in the adaptation rate independent of recombination. Interestingly, many of mutant isolates from the Hfr-2xSFX- populations had higher relative fitness values than those from the 2xOriT populations, though on average, the difference between the mutant isolates was not significant. All populations reached similar phenotypic endpoints by the conclusion of the experiment.

Evolving populations are generally genetically heterogeneous and therefore contain competing mutant lineages that have arisen independently. Clonal interference will generally reduce this diversity, possibly resulting in a more narrow distribution of mutant fitnesses within asexual populations. To test this hypothesis, we isolated random clones from the Hfr-2xSFX- and 2xOriT populations to evaluate the degree of fitness heterogeneity within each population and to determine whether horizontal gene transfer had a detectable effect on the population structure. The fitness dis-

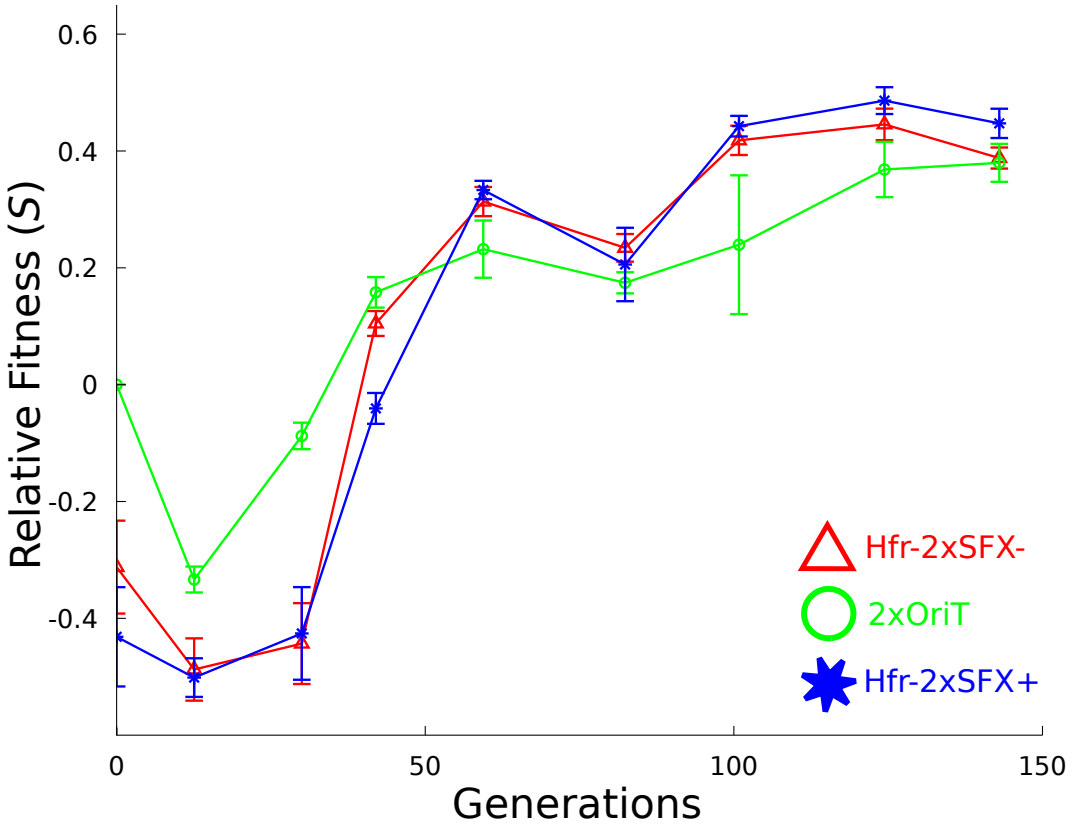

Figure 4.1: Average fitness improvements for the Hfr-2xSFX-, 2xOriT, and Hfr-2xSFX+ populations, relative to the 2xOriT parent strain during adaptive evolution in the presence of high sodium chloride concentrations. Fitness is defined as $S = \mu_{\text{pop}}/\mu_{\text{2xOriT}} - 1$ in this case. Error bars are 95% confidence intervals using the Student's t distribution.

tributions for these populations, shown in Figure 4.2, show that Hfr-2xSFX- lines tended to have slightly higher mean relative fitness but with significantly more variance than the 2xOriT replicates at a high level of significance ($P = 1.81 \cdot 10^{-4}$, Kolmogorov-Smirnov test). Genetic diversity within Hfr-2xSFX- lines may therefore be greater than that of the 2xOriT populations; a likely explanation is that horizontal gene transfer reduces the extinction of beneficial clones due to clonal interference and drift, resulting in a maintenance of heterogeneity that is less likely to occur in an asexual system [111, 300]. However, as all populations reached similar fitness end-

points during the experiment, further adaptation beyond that observed here could be mutation limited. Overall, all populations increased their growth rate in 0.65 M NaCl by approximately 40% by the final serial transfer, which compares favorably with previous examples of NaCl-tolerant *E. coli* [97]. No loss of mating competence was observed in the Hfr-2xSFX(+,-) populations over the course of the experiment.

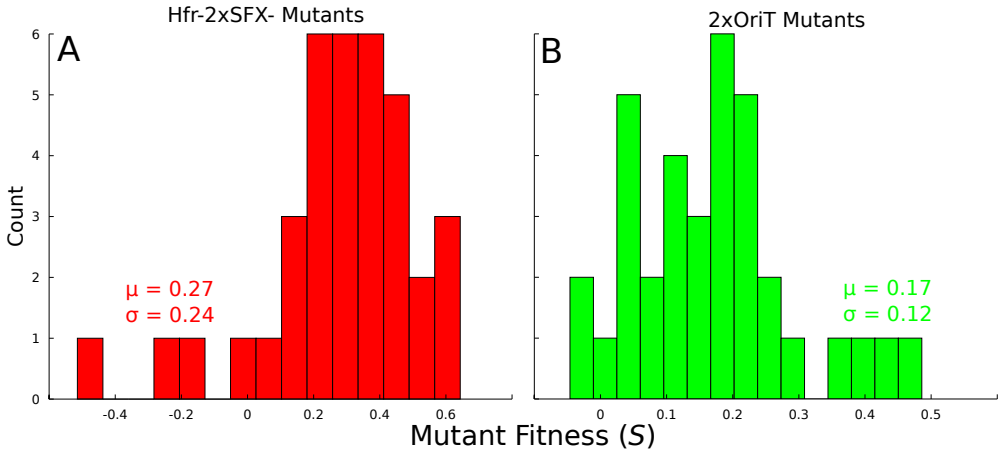

Figure 4.2: Histogram of relative fitness (to 2xOriT) in randomly screened isolates from six Hfr-2xSFX- and 2xOriT populations. All isolates were challenged with 0.65 M NaCl. Differences in the underlying fitness distributions in the Hfr-2xSFX- and 2xOriT populations are highly significant ($P = 1.81 \cdot 10^{-4}$, Kolmogorov-Smirnov test).

4.4.2 Mutant characterization

At the end of the evolution experiment, a single randomly isolated clone from each Hfr-2xSFX+ (H1-H6), Hfr-2xSFX- (G1-G6), and 2xOriT (A1-A6) population was subjected to detailed analyses to identify any novel phenotypes that arose during evolution. This isolation procedure ensures independence between the observed mutations. All mutants had significant improvements in relative fitness under NaCl (Figure 4.4A), though is it possible that mutations that enhance growth in minimal

medium [71] are responsible for the apparent tolerance increase. The relative fitness of each isolated mutant was therefore measured in minimum medium without NaCl challenge. Several isolates (G3, G4, H4, and H6) did exhibit small improvements in growth rate versus a 2xOriT reference in the absence of NaCl (data not shown). When the observed general fitness benefits were accounted for by calculating their relative improvements under NaCl stress in Figure 4.3, all four mutants remained significantly more tolerant than the unevolved 2xOriT strain. so these strains likely acquired mutations beneficial to both growth in minimum medium and in NaCl challenged conditions. In light of the improved NaCl tolerance of the mutants, we also examined their ability to withstand prolonged shocks under hyperosmotic conditions (5.5 M NaCl). While most mutants had no improvements in survival relative to 2xOriT under these conditions (data not shown), G3, G6, and H6 rapidly lost viability, exhibiting a 10-fold or more decrease in shock tolerance. It is possible that survival under extreme NaCl concentrations versus growth at lower concentrations requires divergent adaptive mechanisms. However, all of the mutants are capable of growth at 0.85 M NaCl, a concentration that completely inhibits growth of the 2xOriT and Hfr parent strains.

In an attempt to further improve the observed NaCl tolerance levels of the mutants, we utilized a short-term mating experiment where the G1-G6 isolates were mixed together to facilitate genetic transfer, and then propagated under NaCl selection for five days. Surprisingly, analysis of clonal isolates from the mixed G population revealed that the fittest clones were no better than the original fittest Hfr-2xSFX-mutant G6 (data not shown). One explanation for this apparent failure is that the NaCl tolerant Hfr isolates lost their conjugation proficiency during evolution, but all Hfr isolates were able to successfully conjugate a chromosomally integrated gentamicin resistance gene to a F- recipient. Genome sequencing (discussed in detail below)

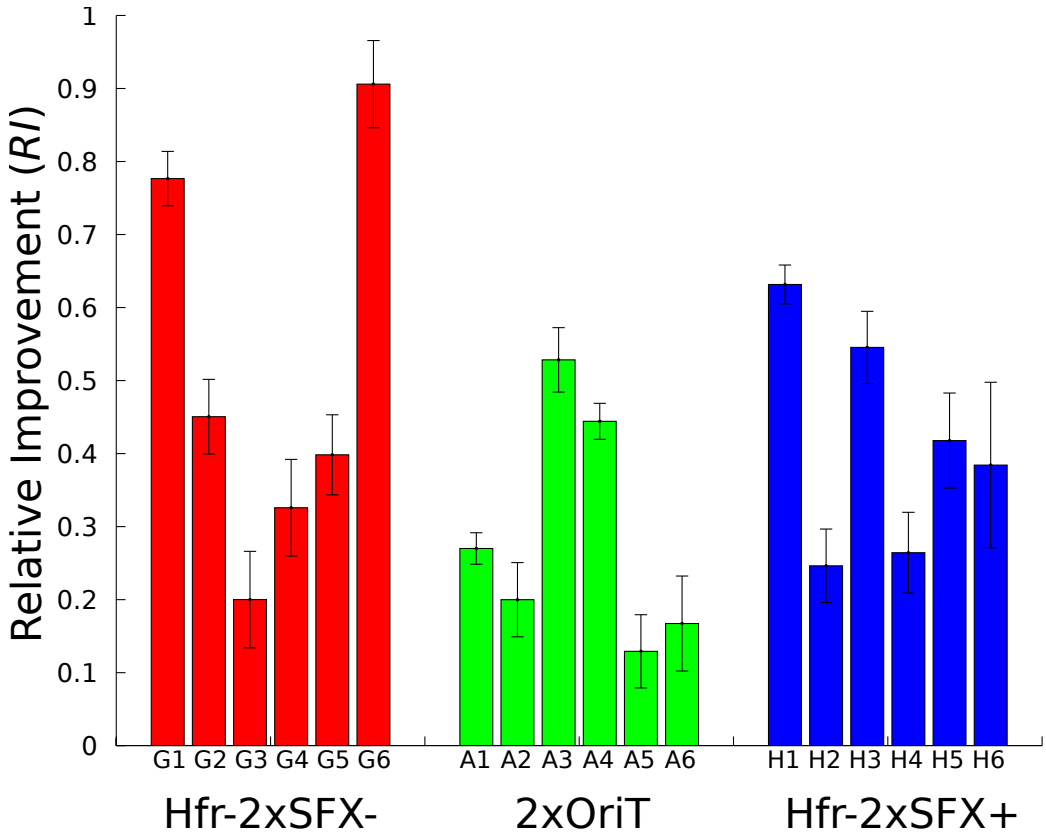

Figure 4.3: Improvement (ratios of inhibition) of Hfr-2xSFX-, 2xOriT, and Hfr-2xSFX+ isolates during growth under 0.65 M NaCl relative to the parental 2xOriT strain, after normalizing for difference in unchallenged growth rates. Error bars are 95% confidence intervals using the Student's t-distribution.

revealed that many of the G mutants have mutations in close proximity or actually overlapping in the case of the independent *nagC* mutations. Recombination events capable of combining these genotypes are therefore rare [112, 143, 119]. Negative epistasis between other mutations may have also prevented successful generation of recombinants.

4.4.3 Osmotic tolerance and other complex phenotypes

Cross adaptation is a phenomenon where a strain evolved for resistance to a specific stressor also exhibits improved growth in the presence of other inhibitors [97].

Resistance to osmotic stress is known to affect other phenotypes of industrial interest, such as n-butanol or low pH tolerance [97, 267] and growth at elevated temperatures [77]. Growth assays of the mutants growing in the presence of inhibitory levels of glucose, 0.8% n-butanol, low pH, and thermal stress (Figures 4.4B-E) revealed that the observed tolerance phenotypes are mostly specific to NaCl resistance alone, at least based on the tested stressors. The G3, G4, H4, and H6 isolates all exhibit increased tolerance of glucose levels known to trigger osmotic stress responses in *E. coli* [259]. However, G3, G4, and H6 are also significantly more sensitive to thermal stress. Glucose and thermal tolerance levels of the other isolates were generally similar to or slightly below that of the wildtype reference, so it is unlikely that there is a fundamental incompatibility between osmotic tolerance and these phenotypes in general. The acid tolerance of the mutants varied widely for the Hfr mutants, with G5 and G6 exhibited large (25-37%) declines in fitness under this stress condition. Significant but small decreases in relative fitness under acid stress were also observed in all A mutants. These results imply that there is some degree of antagonistic pleiotropy between tolerance of high osmotic pressures and acid stress, but additional investigation is needed to confirm this hypothesis.

Interestingly, no isolate had improved n-butanol tolerance in this case, contrary to previous examples of n-butanol-osmotic stress cross adaptation [269, 97]. The reasons for this apparent incompatibility are unclear, especially given that Dragosits et al. observed a slightly increased level of n-butanol tolerance in an isolate evolved under continuous 0.3 M NaCl stress [97]. Stronger NaCl selection could disfavor mutations that also improve n-butanol resistance, as it is reasonable to expect that different levels of osmotic stress select for distinct adaptive mechanisms. N-butanol tolerance is also not always associated with improved osmotic stress resistance in evolved mutants [267], so there are at least some adaptive paths on both the NaCl

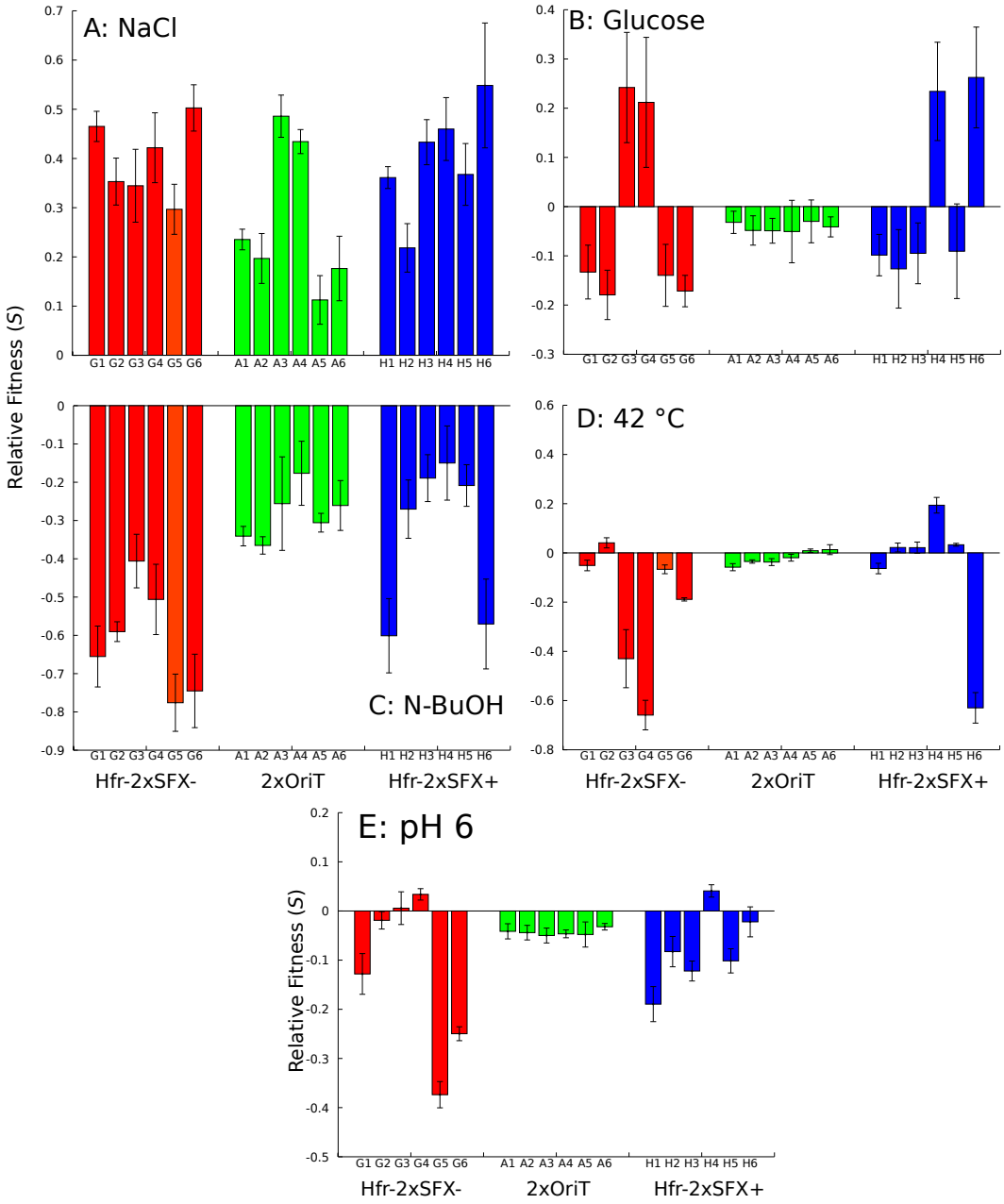

Figure 4.4: Relative fitness of Hfr-2xSFX-, 2xOriT, and Hfr-2xSFX+ isolates under several abiotic stressors, including A) 0.65 M NaCl, B) 0.3 M glucose, C) 0.8% n-butanol, D) growth at 42 °C, and E) pH 6. Error bars are 95% confidence intervals using the Student's t-distribution.

and n-butanol landscapes that lead to divergent tolerance phenotypes. The overall lack of significant cross adaptation for the isolates in this case indicate that specialist mutants with adaptations specific only to NaCl tolerance are favored under these evolutionary conditions.

4.4.4 Genetic patterns of adaptation

In order to better understand the genetic bases for the observed osmotolerance phenotypes and to compare adaptive mechanisms between the sexual and asexual populations, we selected the A1-A6 (2xOriT parent) and G1-G6 (Hfr-2xSFX- parent) isolates for genomic sequencing. A complete list of mutated genes, their putative functions, and structure of each mutation in the isolates is given in Table 4.2. Many isolates (8/12) harbor likely inactivating frameshifting mutations or large deletions within the n-acetylglucosamine (NAG) catabolic operon (*nag*), either in the gene encoding the transcriptional repressor *nagC* or the deacetylase *nagA*. Amino sugar catabolism may therefore have been altered in these strains. Though this adaptation has not been previously observed in strains evolved under continual osmotic stress, NAG forms a crucial component of peptidoglycan [330], and it is readily conceivable that adaptation to high osmotic stress would involve alterations to cell wall biosynthesis or peptidoglycan recycling. Disabling *nag* may be an effective way of increasing intracellular NAG pools for additional peptidoglycan synthesis. Mutations affecting glucosamine-6-phosphate biosynthesis have also been identified in evolved isobutanol tolerant *E. coli* [14], so this may be a common mechanism of adaptation to certain membrane-disrupting environmental conditions. Given that *nagA* is essential for the use of NAG as a carbon and energy source [146], we assayed the mutants along with defined *nagC* and *nagA* deletion strains from the Keio collection [17] for growth on minimal NAG agar. Mutants A1, A2, and A3, all containing mutations in *nagA*,

were unable to metabolize NAG as a sole carbon source, as expected from previous results in the literature. G4 unexpectedly failed to grow as well, despite the fact that the strain has no mutations affecting NAG genes; it is possible that this strain has a large genomic rearrangement affecting *nag* expression that was not detected in our analysis.

The independent mutations in *nagA*, *nagC*, and *nagCD*, along with the inability of G4 to metabolize NAG as a sole carbon source, point toward two distinct adaptive routes to improve osmotic tolerance that involve NAG metabolism. Figure 4.5 depicts the functional roles of NagA and NagC in controlling this pathway; NagD is a general ribonucleotide monophophatase that is not required for NAG catabolism despite being present in the *nag* operon. Following the transport of NAG into the cell via the phosphotransferase system (PTS), NagA deacetylates NAG-6-phosphate into glucosamine-6-phosphate that can then be used for anabolism or peptidoglycan synthesis [253, 146]. This reaction is absolutely required for the assimilation of NAG into peptidoglycan synthesis and central carbon metabolism, so the inability of G4, A1, A2, and A3 to grow on NAG as the sole carbon source confirms that NagA activity in these strains has been abolished.

In contrast to the critical role of NagA in assimilating NAG, NagC is the negative regulator of the *nag* operon [247]; it is not required for NAG catabolism, and its inactivation generally results in upregulation of the other *nag* operon members [252]. While transcriptional analysis of several of these mutants under osmotic stress (see below) did not detect increased expression of nag genes, the relative ubiquity of *nagC* or *nagCD* mutations in the A and G strains indicates that the inactivation through large deletions or small indels of *nagC* may have some fitness benefit under NaCl challenge. All of the *nagC* mutants (G2, G5, A4, A5, and A6) are capable of using NAG as a sole carbon source as well, unlike the NAG(-) mutants (G4, A1, A2,

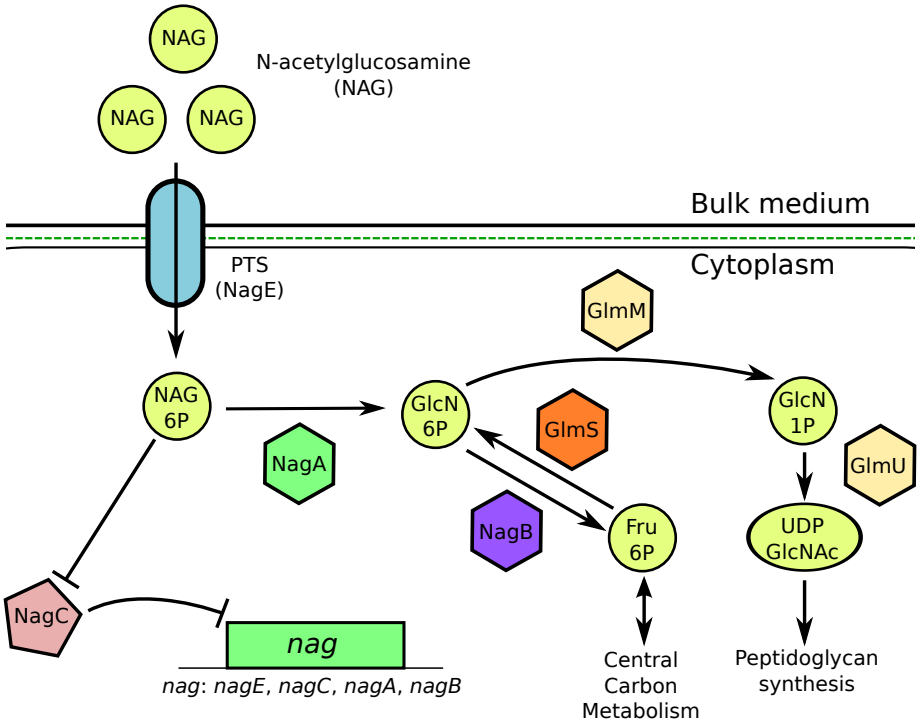

Figure 4.5: Summary of n-acetylglucoasmine (NAG) catabolism in *E. coli*. Pathway information adapted from Peri et al. [249], Yadav et al. [354], and Ecocyc [166]. Metabolite abbreviations: NAG, n-acetylglucosamine, NAG-6P, NAG-6-phosphate, GlcN-6P, glucosamine-6-phosphate, GlcN-1P, glucosamine-1-phosphate, UDP-GlcNAc, uridine diphosphate NAG, Fru-6P, fructose-6-phosphate. Protein abbreviations: NagE, NAG PTS permease, NagA, NAG-6 phosphate deacetylase, NagB, glucosamine-6-phosphate deaminase, NagC, DNA-binding transcriptional dual regulator, GlmS, L-glutamine:D-fructose-6-phosphate aminotransferase, GlmM, phosphoglucosamine mutase, and GlmU, fused NAG-1-phosphate uridyltransferase glucosamine-1-phosphate acetyltransferase. Arrows indicate chemical or transport reactions while bars indicate repression of transcription or activity.

and A3). Despite their opposite effects on NAG catabolism, both *nagA* and *nagC* deletions have been shown to affect surface fibers synthesis, including curli [23] and fimbriae [302] due to the accumulation of intracellular NAG-6-phosphate. Though fully elucidating the molecular mechanisms of each adaption is beyond the scope of this work, further investigation into both *nagC* and *nagA* mutations and their

Table 4.2: Mutations identified with genome sequencing

Strain	Gene(s)	Function	Type	Location
G1	*nmpC/essD*	outer membrane protein, holin	SNP	Intergenic
	cadB	lysine, cadaverine transporter	SNP	A171A
G2	*nagCD*	NAG metabolism	Del.	Coding (-715bp)
	yhdP	Unknown	SNP	L318R
G3	*rpoC*‡	β' subunit of RNAP	Dup.	Coding (2x84bp)
	treR	Trehalose regulator	SNP	S61S
G4	*rpoC*‡	β' subunit of RNAP	Dup.	Coding (2x84bp)
	hisC‡	Histidine biosynthesis	SNP	T170P
G5	*nagC*	NAG metabolism	Del.	Coding (-1bp)
G6	*mrdA*‡	Penicillin binding protein 2	SNP	Q51L
	rpsA‡	30S protein S1	SNP	Q421K
	ydjK	Predicted transporter	SNP	P17S
A1	*nagA*	NAG-6P Deacetylase	Del.	Coding (-1bp)
	bcr	Multidrug efflux transporter	SNP	S8A
	mreB‡	Actin homolog	SNP	I336L
A2	*nagA*	NAG-6P Deacetylase	Del.	Coding (-1bp)
	mreB‡	Actin homolog	SNP	T171S
A3	*nagA*	NAG-6P Deacetylase	SNP	A203E
	proV	Glycine betaine transporter	IS1	Coding (+9bp)
	msrB	Methionine sulfoxide repair	SNP	C118F
	fimA	Fimbrae A	IS186	Coding (+6bp)
A4	*nagCD*	NAG metabolism	Del.	Coding (-1570bp)
	proV	Glycine betaine transporter	IS1	Coding (+4bp)
	yejM‡	Predicted hydrolyase	Dup.	Coding (2x9bp)
A5	*nagC*	NAG metabolism	Del.	Coding (-2bp)
	mreB‡	Actin homolog	SNP	S185F
	fimA	Fimbrae A	IS186	Coding (+6bp)
A6	*nagC*	NAG metabolism	Ins.	Coding (+1bp)
	mreB‡	Actin homolog	SNP	K96Q
	bglB	Phospho β-glucosidase	Del.	Coding (-1bp)
	yobF/yebO	Stress, predicted protein	SNP	Integenic

List of mutations found in each strain (G1-G6, A1-A6) relative to their respective parental genotypes. NAG: n-acetyl-glucosamine. ‡: known essential genes (in general and in glucose minimal media supplemented with tryptophan specifically). The *rpoC* duplication in G3 and G4 duplicates amino acid 370 to 396 (KKMALELFKPFIYGKLELRGLATTIKA) in domain 2 of the protein [258].

relationship with osmotic tolerance would shed additional light on their roles in this context.

Besides alterations in NAG metabolism, genes encoding cell shape regulators are frequently mutated in the evolved strains. Non-synonymous single nucleotide polymorphisms (SNPs) in the cell shape regulating actin homolog *mreB* [327] and the peptidoglycan transpeptidase *mrdA* genes [157] were identified in five different mutants, suggesting that changes in cell morphology might also reduce osmotic stress on the cell. Microscope examination of the affected strains showed no observable difference in cell shape compared to parental controls (data not shown), however. It is difficult to speculate on how *nag* and *mreB* mutations might interact in the A1, A2, A5, and A6 mutants, as they affect related but distinct cellular processes; changes in *mreB*-chromosomal interactions may result in altered cell division or chromosomal segregation, which may in turn affect the amount of NAG precursor generated from peptidoglycan recycling. Based on the presence of mutations in these genes and NAG metabolic perturbations, changes in regulators of cell shape, and peptidoglycan synthesis may be important for improving osmotic tolerance at higher NaCl concentrations.

While both groups of mutants had similar mutation rates (see Methods), transposon insertions in *fimA* (type I fimbriae) and *proV* (the ATP binding cassette for the *proVXW* glycine betaine transporter) were only observed in the A strains. Fimbrial components, including *fimA*, are highly upregulated under osmotic stress [301], and a mutation that inactivates *fimA* presumably results in conservation of carbon and energy. The relationship between *nagA*, *nagC* and surface fiber expression may also play a role in the adaptive benefit of *fimA* inactivation, as the *fimA* insertions in A3 and A5 are associated with *nag* mutations as well. The inactivation of *proV* is more peculiar, given its extensively studied role in importing osmoprotectants into the cell

[202, 165]. Nonsense mutations affecting *proV* have been previously observed in an osmotolerant adaptive mutants [97], confirming that *proV* is under negative selection in hyperosmotic glucose minimal media. Several mutants do exhibit increased transcription of one or more *proVXW* genes relative to 2xOriT under osmotic stress (Table 4.3), in line with previous reports [337, 25]. Given the lack of glycine betaine in the evolution media, coupled with the inability of *E. coli* to synthesize this particular osmolyte without exogenous choline [311], expression of *proV* at a high level should be ineffective in reducing osmotic stress under these conditions. Interestingly, no mutations in *fimA* and *proV* were observed in any of the G strains. Gene expression data for A2 and A4 revealed high expression of genes encoding IS1, IS3, and IS186 transposases, which may account for the IS1 and IS186 insertions that were detected in these genes. At the same time, no significant changes in relative transposase transcript levels were detected in the G strains. While transposon activation is a well-known response to environmental stressors [50, 323], it is unclear why transposase upregulation would be limited to the A strains.

Mutations in several other genes, while not specifically known to affect osmotic tolerance, were also detected. The *rpoC* subunit of RNA polymerase, responsible for all gene transcription in *E. coli*, was found to have an 84 bp in-frame duplication in the G3 and G4 isolates of amino acids 370 to 396. The duplications occur in region 2 of the protein, which is responsible for RNA polymerization [258]. Given the role of *rpoC* in promoter recognition and sigma factor binding [166], this mutation should result in significant transcriptional and ultimately phenotypic alterations. Mutations in *rpoC* (along with *rpoB*) have been observed previously in other long-term evolution studies [71] as well, indicating that a wide-range of phenotypes can be improved via alteration of RNA polymerase components. A small in-frame duplication was also observed in the essential but uncharacterized hydrolase gene *yejM*, but the

effect of this mutation is unclear. A range of SNPs in intergenic regions and the coding sequences of various other genes were also detected, including a synonymous substitution in *treR*, negative regulator of trehalose biosynthesis [166].

In order to better understand the effect of these mutations on osmotolerance, we performed three tests to quantify the fitness effect of gene overexpression, compensation, and knockout on an unaltered host strain (BW25113) and the isolated evolved mutants. Eleven genes (*nmpC*, *yobF*, *ydjK*, *mreB*, *fimA*, *nagC*, *nagA*, *proV*, *treR*, *ydhP*, and *cadB*) were selected for overexpression studies due to their frequency of mutation or nearby intergenic or synonymous SNPs (results shown in Figure 4.6A). Of all genes tested, only *ydjK* overexpression conferred a statistically significant improvement in osmotolerance. YdjK is annotated as a putative metabolite transporter, and due to the inclusion of tryptophan in the evolution media, we reasoned that it might be an uncharacterized tryptophan transporter. However, a comparison between the growth rate of BW25113/pYdjK and an empty control vector with and without tryptophan (in the absence of NaCl) revealed no growth benefit for the overexpression strain ($P = 0.25$, Student's t-test). The true substrate of YdjK remains therefore unclear. As expected from their repeated inactivation during adaptive evolution, overexpression of *proV*, *nagC*, or *nagA* is somewhat deleterious under these growth conditions. Although *fimA* underwent transposon insertions in several independent mutants, its overexpression from the ASKA construct (via promoter leakage in the absence of IPTG induction) did not have a significant effect on host fitness under these conditions.

Likely inactivating mutations in a small set of genes, including *nagC*, *nagA*, *fimA*, and *proV*, were identified in several of the G and A mutants. Various combinations of these mutations were found in the adaptive mutants, suggesting the presence of potential epistatic interactions that affect osmotic tolerance. We tested this hypoth-

esis by systematically reconstructing double and triple mutants containing defined *nagA* or *nagC* disruptions, along with knockouts in *fimA* and *proV*, using the corresponding Keio collection strains [53]. Given their proximity on the genome, *nagA* and *nagC* could not be simultaneously disrupted by transduction from the appropriate Keio collection mutant [17]. Figure 4.6B shows the relative fitness of these knockout mutants versus an unmodified BW25113 strain. Of the single mutants, only the $\Delta(proV)$ strain is fitter than reference. Interestingly, *nagA* knockout has a large negative impact on strain fitness, suggesting that the insertion and deletion mutations identified in the sequenced mutants were not completely inactivating or the inactivation of *nagA* has a positive synergistic effect with other mutations in these mutants. As potential evidence for the presence of synergistic interactions between mutations, double knockouts of $\Delta(nagC)$ and $\Delta(proV)$, or $\Delta(nagA)$ and $\Delta(fimA)$ were found to be beneficial in the presence of NaCl challenge. Among the triple knockout mutants, only the $\Delta(nagC)$ $\Delta(fimA)$ $\Delta(proV)$ showed improved osmotolerance relative to BW25113. These combinations did not occur in the adaptive mutants, and it is possible structural differences between the defined knockouts and mutations in the osmotolerant isolates could influence these results. Overall, as was the case with overexpression analysis, these results confirm that these four genes affect the *E. coli* osmotolerance phenotype.

An additional way of confirming that the mutations in *nagA*, *nagC*, *proV*, and *fimA* play a role in modulating osmotolerance is to complement (via plasmid transformation) the mutations in the mutated strains, and then reassay their fitness under osmotic stress. The results of this test are shown in Figure 4.6C-D, and indicate that the effect of complementation depends heavily on the genetic background of the particular mutant. Relative fitness of only two *nagC* mutants (G5 and A4) out of five are decreased by *nagC* overexpression, whereas *nagA* complementation decreased os-

motic tolerance of A1-A3. The structures of the *nagC* mutations do vary significantly between the mutants, which may explain the lack of concordance in complementation results. Complementing the *proV* mutations in A3 and A4 had no significant fitness effect. While *fimA* overexpression decreased the fitness of A3, it also improved the fitness of A5 slightly, despite their identical IS186 insertions into the gene. It is likely that genetic differences between the various mutants and interactions between other mutations not screened in this assay also play a significant role in determining the fitness impact of gene complementation.

Though genome sequencing has revealed several novel loci involved with osmotolerance, the genotype-phenotype relationship remains unclear. To gain a more complete understanding of how these mutations translate into improved osmotic stress tolerance, we applied microarray technology to several A and G mutants with distinct underlying mutations to identify transcriptional perturbations that may result from their underlying mutations or altered stress responses.

4.4.5 Mutation-induced transcriptional perturbations

Six mutants were selected for transcriptomic analysis: G2, G3, G5, G6, A2, and A4; these mutants were chosen due to their distinct underlying genotypes and varied levels of osmotolerance, and therefore likely to have diverse adaptive mechanisms. The mutants and the parental references were grown under NaCl selection in order to specifically identify transcriptional changes associated with the osmotic stress response. A total of 432 overexpressed and 479 downregulated genes were identified in at least one mutant. Though each mutant has a unique expression profile, significant similarities exist between the differentially expressed genes in all strains. Analysis of gene ontology revealed significant enrichment and depletion of different membrane related genes ($P < 10^{-6}$), suggesting that membrane remodeling may contribute to

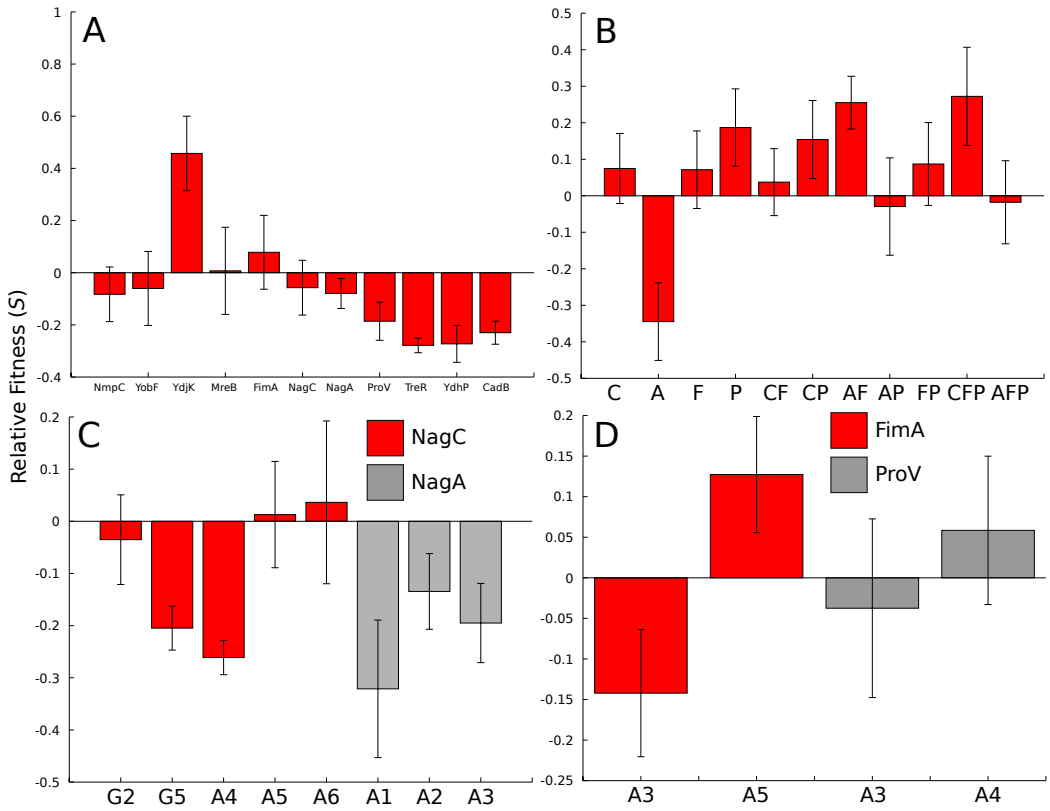

Figure 4.6: Batch growth screening of overexpression, knockout, and compensatory strains to identify their osmotolerance phenotypes in glucose minimal media supplemented with 0.55 M NaCl. A) Fitness of overexpression strains relative to the empty vector pCA24N control. Only *ydjK* expression results in a significant improvement of NaCl tolerance. B) Fitness of knockout strains relative to BW25113; strain genotypes are denoted as follows: $\Delta(nagC)$ = C, $\Delta(nagA)$ = A, $\Delta(fimA)$ = F, and $\Delta(proV)$ = P. C) Overexpression of NagC, NagA in mutants with possibly inactivating mutations in these genes; fitness is relative to the corresponding mutant with pCA24N. D) Overexpression of FimA, ProV in mutants with possibly inactivating mutations in these genes, with the empty vector as a reference. All error bars are 95% confidence intervals based on the Student's distribution.

osmotolerance. High expression levels of genes involved with cell envelope maintenance were also observed in G2, G3, G5, and G6, which may indicate an attempt to counteract NaCl-induced dessication. Because of the improved growth rates of the mutants under osmotic selection, ribosomal genes were overexpressed in each mutant

as a result of the known linkage between growth rate and rRNA abundance [228], and thus were excluded from further analysis.

Genes commonly upregulated in these mutants, along with those representing possibly novel adaptive mechanisms, are listed in Table 4.3. Several functional clusters possibly related to osmotic tolerance are immediately evident. The gene encoding AcrZ, a small protein associated with the AcrAB-TolC efflux pump complex known to affect AcrB substrate recognition [145], is upregulated in all mutants except A4. It is unclear how AcrZ may perturb AcrB activity under osmotic stress, but given the ubiquity of this response, it may be an important transcriptional adaptation. Sulfonate transport and metabolism genes (*tauABC*, *ssuEADC*) are also frequently upregulated in the mutants. Ordinarily these proteins are intended to scavenge sulfur from the environment, but under osmotic stress, they can also import osmoprotectants such as taurine [328]. This upregulation cannot be explained as simple sulfur starvation as the mutants and references were in exponential growth before RNA harvesting. The evolution media used in this study does not contain sulfonates, but it is possible that these transporters are capable of importing other metabolites into the cell. The *proVXW* operon, positively regulated by hyperosmotic conditions [202], is also overexpressed by G3 and G5, though *proX* and *proW* are also overexpressed in A4, G3, G5, and G6. This expression pattern may explain the fitness benefit of *proV* deletion found in A3 and A4 in terms of energy conservation, as discussed previously.

Though the 2xOriT and Hfr-2xSFX- parental strains are for the most part isogenic, the latter is a tryptophan auxotroph due to insertion of the *tra* operon into the *trp* locus [324]. These four G strains all exhibited overexpression of the tryptophan transporter *tnaB*, presumably in an effort to improve uptake of the amino acid from the bulk medium to compensate for their tryptophan auxotrophies. The tryptophan catabolism gene *tnaA* is also highly upregulated, as has been previously observed in

Table 4.3: Overexpressed genes of interest

Gene	COG	Function	Strain(s)
acrZ	-	Cell envelope stress response	A4,G2,G3,G5,G6
ssuD	C	Alkanesulfonate monooxygenase	G2,G3,G5,G6
proV	E	Glycine betaine transporter subunit	G3,G5
proW	E	Glycine betaine transporter subunit	A4,G6
proX	E	Glycine betaine transporter subunit	A4,G3,G5,G6
tnaA	E	Tryptophanase/L-cysteine desulfhydrase	G2,G3,G5,G6
tnaB	E	Tryptophan transporter of low affinity	G2,G3,G6
entC	HQ	Isochorismate synthase 1	G3
acs	I	Acetyl-CoA synthetase	G5,G6
ymdC	I	Cardiolipin synthase 3	G2
fabF	IQ	3-oxoacyl-[acyl-carrier-protein] synthase II	G6
nagC	KG	Repressor of NAG operon	A4,G2
arnC	M	Undecaprenyl phosphate-L-Ara4FN transferase	A2,A4
lpp	M	Murein lipoprotein	A2,A4,G2,G3
ompA	M	Outer membrane porin A	G2
ompC	M	Outer membrane porin C	G2,G3,G5,G6
ompG	M	Outer membrane porin G	A2,A4
ompL	M	Outer membrane porin L	A2,A4
ompX	M	Outer membrane protein X	A2,A4,G3,G5
feoB	P	Fused ferrous iron transporter	G5
fepG	P	Iron-enterobactin transporter subunit	G5
fiu	P	Catecholate siderophore receptor Fiu	G3
ssuA	P	Putative aliphatic sulfonate binding protein	G2,G3,G5
ssuC	P	Putative alkanesulfonate transporter subunit	A2,A4,G2,G3,G6
tauA	P	Taurine transporter subunit	G3,G6
tauB	P	Taurine transporter subunit	G2,G3,G6
tauC	P	Taurine transporter subunit	A4,G2,G3,G5,G6
entE	Q	Enterobactin synthase complex component	G3
entF	Q	Enterobactin synthase complex component	G3,G5
entH	Q	Thioesterase for efficient enterobactin production	G3
tauD	Q	Taurine dioxygenase, 2-oxoglutarate-dependent	G3,G6
fadM	R	Long-chain acyl-CoA thioesterase III	G2,G3,G6
ssuE	R	NAD(P)H-dependent FMN reductase	G2,G3,G5,G6
ydcH	S	Predicted protein	G2,G3,G5,G6
yodC	S	Predicted protein	G2,G3,G5,G6
csrA	T	Carbon source metabolism regulator	G2,G3,G5

Overexpressed genes of interest for A2, A4, G2, G3, G5, and G6 as identified with microarray analysis. The overexpression of these genes is statistically significant at a $P \leq 0.01$ level, as calculated by the rank product method developed by Breitling et al. [39].

media containing tryptophan [167]. TnaA converts tryptophan into pyruvate and indole, the latter being a potent signaling molecule involved with biofilm and cell cycle regulation [135, 188, 57, 91]. This common upregulation could simply be the result of the inherent tryptophan auxotrophy of the G mutants coupled with their increased metabolic activity under osmotically stressful conditions, though previous evolutionary engineering studies have identified *tnaA* mutants with improved tolerance to other conditions as well [14, 63].

When the indole concentration in stationary phase culture is measured for these strains (Figure 4.7), only G2, G5, and G6 exhibit statistically significant increases ($P < 0.05$, Student's t-test) in indole accumulation compared to the ancestral Hfr-2xSFX- parent. A4 produces significantly less indole ($P < 0.045$, Student's t-test) than its 2xOriT ancestor as well. Despite the fact that G3 has roughly equivalent upregulation of *tnaA* transcription to these strains, indole production is similar to wildtype; it could be that the unusual in-frame duplication in *rpoC* in this strain results in increased *tnaA* transcription, but the resulting mRNA is poorly translated into protein or degraded rapidly due to post-transcriptional regulation. A reconstructed prototrophic *rpoC* mutant does have a significantly increased growth rate in the presence of tryptophan, which suggests that G3 and G4 have acquired mutations that increase their tryptophan uptake rates during adaptive evolution. We also tested whether exogenous indole supplementation affected the growth of BW25113 in the presence of 0.55 M NaCl. No statistically significant change in growth rate was found when indole(-) and indole(+) conditions were compared in this case (data not shown), so indole accumulation in G2, G5, and G6 may be a secondary consequence of osmotic adaption or improved growth, not a primary driver of fitness improvement.

Beyond metabolite transporters, several genes involved with the iron uptake

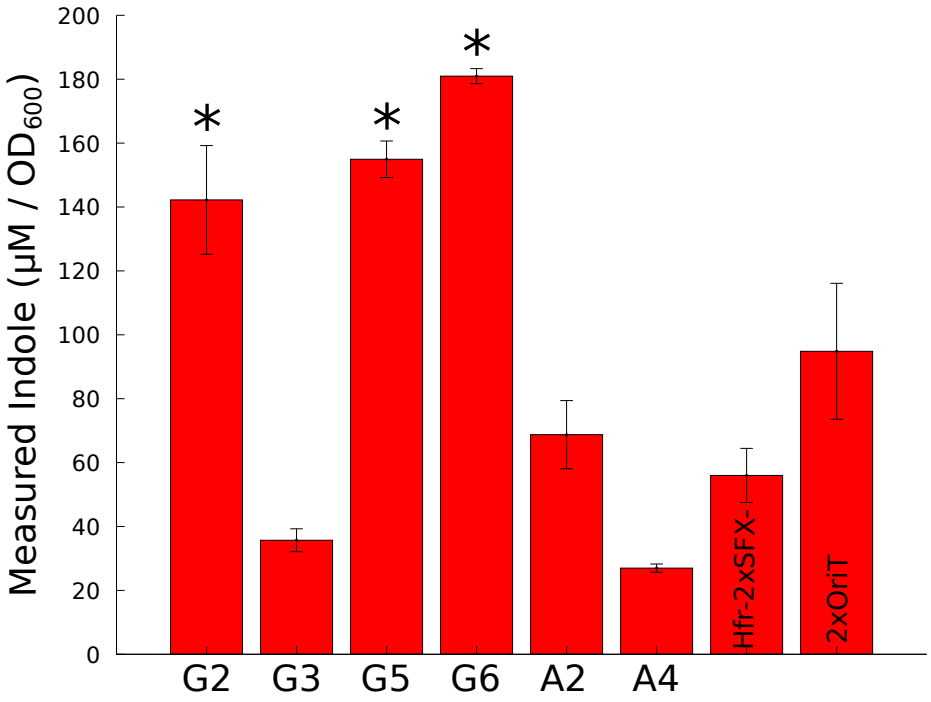

Figure 4.7: Levels of indole, normalized by biomass density, for G2, G3, G5, G6, A2, and A4 mutants, along with the G parent Hfr-2xSFX- and A parent 2xOriT. Error bars are standard deviations. G2, G5, and G6 have statistically significant increases in indole accumulation versus Hfr-2xSFX- ($P < 0.05$, Student's t-test).

(*entCEFH*, *feoB*, *fepG*, *fiu*) were overexpressed in G3 and G5 as well. Increased expression of iron transport and metabolism genes have been found in evolved mutants with improved osmotic or n-butanol tolerance [269, 97]. Perturbation of iron metabolism may therefore be a contributor to osmotic tolerance, though interestingly, the gene encoding the siderophore receptor *fiu* is also downregulated in G2 and G6. Other perturbed genes related with metabolism include the upregulation of *csrA*, a global regulator for carbon metabolism [316], which may have significant effects on central carbon metabolism and glycogen accumulation. Genes involved in membrane composition were also upregulated in many mutants, particularly certain outer membrane porins (*ompACGL*) along with the *ompX* gene. These transcriptional

disturbances may represent an attempt by the mutants to change membrane-sieving properties in order to reduce osmotic pressure on the cell. The peptidoglycan-outer membrane tether gene, *lpp* (murein lipoprotein), was also overexpressed in four different mutants (A2, A4, G2, G3), which may indicate that the peptidoglycan wall is more strongly attached to the outer membrane in these mutants to harden the cell against high external osmotic pressure. Several genes that encode hypothetical or predicted proteins (*ydcH*, *yodC*) are consistently overexpressed in the G mutants, but no information about their biochemical roles or relationship with osmotolerance is known.

In contrast to the relative similarities in gene overexpression under osmotic stress between these six mutants, less similarity is observed for genes repressed relative to the parental references (Table 4.4). The few commonly repressed genes include those coding for several hypothetical or conserved proteins (*yegR*, *ydeMN*, *rtcB*) and the transcriptional activator *ydeO*, known to regulate acid resistance in concert with EvgA [206]. Their repression may be associated with the reduced osmotic stress experienced by the mutants compared to their parents under these conditions, as these genes are not known to be directly regulated by osmotic stress. Downregulation of several siderophores and enterobactin transporters was observed in G2, G5, and G6, including *fepA*, which was found to be strongly overexpressed in a previously evolved osmotolerant *E. coli* strain [97]. Iron metabolic perturbations are likely to be distinct in this study due to the much higher osmotic stress used during this evolution experiment (0.6-0.75 M versus 0.3 M), so it is unlikely that the adaptive mechanisms favored at lower osmotic pressures would be identical to those observed here. Transcription of RNA polymerase components (*rpoB*, *rpoC*) is also reduced in A2 and A4, despite their faster growth rates (ordinarily associated with increased rRNA and mRNA synthesis) under osmotic stress.

Table 4.4: Repressed genes of interest

Gene	COG	Description	Strain(s)
yegR	-	Predicted protein	A4,G2,G3,G5,G6
rfaZ	-	Lipopolysaccharide core biosynthesis protein	A2,A4
nagB	G	Glucosamine-6-phosphate deaminase	A4,G2,G3,G5
nagE	G	NAG PTS enzyme: IIC, IIB, IIA components	G2,G3,G5
nagA	G	N-acetylglucosamine-6-phosphate deacetylase	G2,G3
ydeO	K	Transcriptional activator for mdtEF	G2,G3,G5,G6
rpoB	K	RNA polymerase, beta subunit	A2,A4
rpoC	K	RNA polymerase, beta prime subunit	A2,A4
rpoE	K	RNA polymerase, sigma 24 (sigma E) factor	A2,A4
kdsB	M	3-deoxy-manno-octulosonate cytidylyltransferase	A2,A4
mltD	M	Predicted lytic murein transglycosylase D	A2,A4
ompF	M	Outer membrane porin 1a (Ia;b;F)	G3,G6
wbbK	M	Lipopolysaccharide biosynthesis protein	A2,A4
acrA	M	Multidrug efflux system	G6
flgJ	MNO	Muramidase	G5
fimC	NU	Chaperone, periplasmic	G5,G6
fimA	NU	Major type 1 subunit fimbrin (pilin)	G6
fimD	NU	Puter membrane usher protein	G5
fimG	NU	Minor component of type 1 fimbriae	G5
ydeN	P	Conserved protein	A2,A4,G3,G5
fiu	P	Catecholate siderophore receptor Fiu	G2,G6
cirA	P	Catecholate siderophore receptor CirA	G6
efeO	P	Inactive ferrous ion transporter EfeUOB	G6
fecD	P	Iron-dicitrate transporter subunit	G5
fepA	P	Iron-enterobactin outer membrane transporter	G6
ydeM	R	Conserved protein	G2,G3,G5
rtcB	S	Conserved protein	G3,G5,G6

Repressed genes of interest for A2, A4, G2, G3, G5, and G6. The repression of these genes is statistically significant at a $P \leq 0.01$ level, as calculated by the rank product method developed by Breitling et al. [39].

As was the case for upregulated genes, repression of genes involved with peptidoglycan and membrane biosynthesis were significantly enriched in the mutant transcriptomes according to gene ontology analysis. NAG catabolism genes, *nagA*, *nagB*, and *nagE*, were repressed in A4 and several G mutants, potentially resulting in increased availability of NAG for peptidoglycan biosynthesis, or at least accumu-

lation of NAG-6-phopshate in the cytoplasm [23]. The expected effect of increased *nag* operon expression due to *nagA* or *nagC* inactivation [252] was not observed in this case. Repression of several other aspects of peptidoglycan or outer membrane related genes was observed in A2 and A4, particularly those genes involved with cell wall maturation (*mltD*, *murC*) or lipopolysaccharide synthesis (*rfaZ*, *kdsB*, *wbbK*). Membrane remodeling may therefore be a significant adaptive response under these conditions. Reflecting the improved tolerance of the A mutants, the cell envelope stress sigma factor *rpoE* was also downregulated in both A2 and A4 compared to the 2xOriT parent, suggesting that level of osmotic stress experienced by these strains has lessened as a result of their adaptations. Though the precise functional importance of these membrane related changes is unclear at this time, the enrichment of membrane and peptidoglycan related genes in both the genomic and transcriptomic data indicates that membrane remodeling is a significant part of the adaptive response to constant osmotic stress. Fimbriae synthesis and assembly genes were also repressed in G5 and G6, lending credence to the hypothesis that *fimA* inactivation may be part of an energy conservation response during laboratory evolution. Porin synthesis also is perturbed in G3 and G6 by simultaneous repression of *ompF* and overexpression of *ompC*, as is typically observed under high osmotic pressure [282, 80].

4.5 Conclusions

These results demonstrate that our previously developed *in situ* recombination method is suitable for improving industrially relevant strain characteristics. On a population level, the sexual populations acquired osmotolerance more quickly than the asexual controls, though the final phenotypic end points for the populations were similar. All isolates exhibited high levels of osmotolerance, as expected, and an unex-

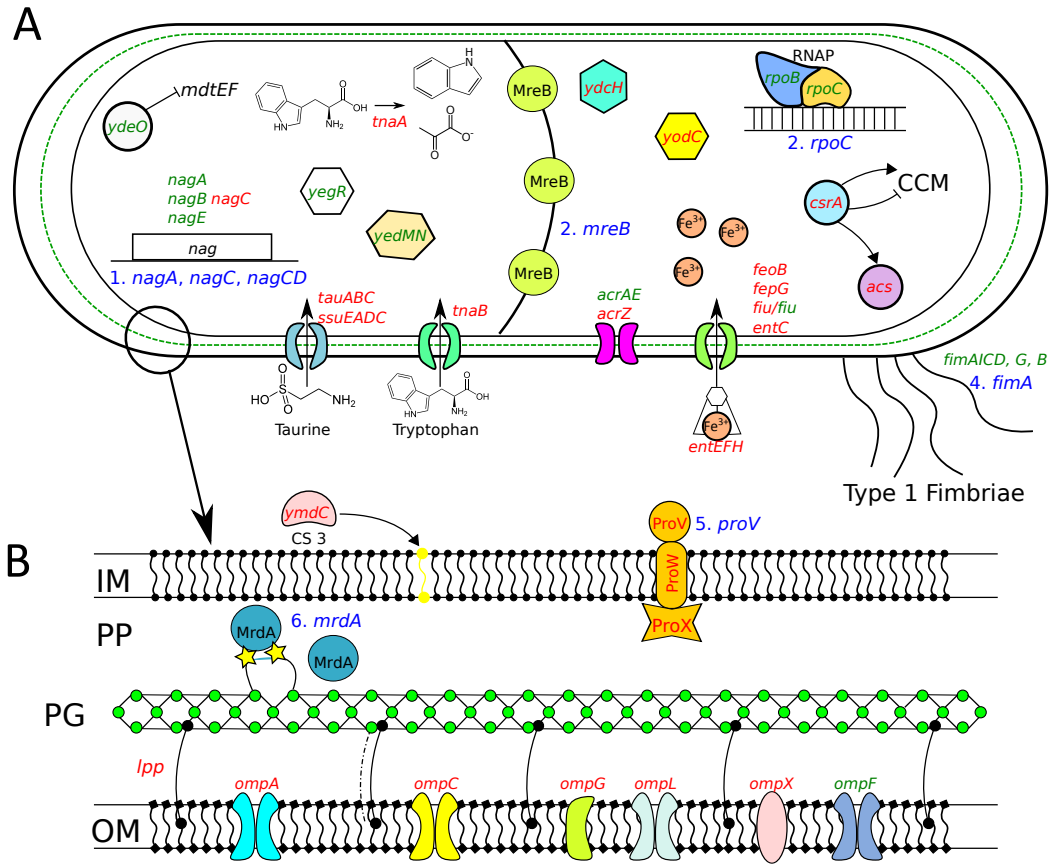

Figure 4.8: Summary of observed genetic and transcriptional responses to continual osmotic stress in the A) cytoplasm and B) the membrane. Overexpressed and repressed genes are denoted with red and green text, respectively, while genes with identified mutations are numbered. A detailed examination of how these responses relate to the observed osmotolerance phenotype is given in the text. Overall, significant changes in outer membrane protein expression coupled with altered amino-sugar and osmoprotectant metabolism, control of cell shape, and sulfonate transporter expression appear to be major adaptive mechanisms for the mutants. Abbreviations: NAG, n-acetyl-glucosamine, CCM, central carbon metabolism, IM, inner membrane, PP, periplasm, PG, peptidoglycan layer, and OM, outer membrane.

pected lack of cross-adaptation to other stressors such as excess glucose, n-butanol, low pH, and thermal stress. Genomic sequencing of 6 Hfr-2xSFX- and 6 2xOriT isolates revealed novel mutations in genes related to n-acetylglucosamine catabolism,

cell shape regulation, uptake of osmoprotectants, and global regulators such as *rpoC*; transcriptional analysis of a select group of mutants allowed for exploration of the genotype-phenotype connection as well. A complete overview of the observed mutations and transcriptional responses of the adaptive mutants is shown in Figure 4.8. Membrane and peptidoglycan synthesis, porin expression, sulfonate uptake, and iron metabolism are all significantly perturbed in various adaptive mutants relative to their parental references, and might be targets of interest for future studies of this phenotype. A series of overexpression, knockout, and compensation assays confirmed that many of genes of interest discovered using genetic affect osmotic tolerance in an unevolved host. Overall, this study further expands our current knowledge of the *E. coli* osmotolerance and should contribute significantly to future rational or evolutionary engineering of this phenotype into industrially relevant strains. Future work will aim to reconstruct the observed mutations in the A and G isolates in industrial *E. coli* strains to determine their effectiveness under typical cultivation conditions.

5. TRANSCRIPTIONAL RESPONSE OF *LACTOBACILLUS BREVIS* TO N-BUTANOL AND FERULIC ACID STRESSES*

5.1 Summary

The presence of anti-microbial phenolic compounds, such as the ubiquitous cell wall component ferulic acid, in biomass hydrolysates poses significant challenges to the widespread use of biomass in conjunction with whole-cell biocatalysis or fermentation. Currently, these inhibitory compounds must be removed through additional downstream processing or sufficiently diluted to create environments suitable for most industrially important microbial strains. Product toxicity must also be overcome to allow for efficient production of next generation biofuels such as n-butanol, isopropanol, and others from low-cost feedstocks. This study explores the high ferulic acid and n-butanol tolerance in *Lactobacillus brevis*, a lactic acid bacterium often used in fermentation processes, by global transcriptional response analysis. The transcriptional profile of *L. brevis* reveals that the presence of ferulic acid triggers the expression of currently uncharacterized membrane proteins, possibly in an effort to counteract ferulic acid-induced changes in membrane fluidity and ion leakage. In contrast to the ferulic acid stress response, n-butanol insult induces genes within the fatty acid synthesis pathway and reduces the proportion of 19:1 cyclopropane fatty acid within the *L. brevis* membrane. Both inhibitors also trigger generalized stress responses. Separate attempts to alter fatty acid synthesis in *Escherichia coli* by overexpressing acetyl-CoA carboxylase subunits and deleting the gene cyclopropane

fatty acid synthase (*cfa*) both failed to improve n-butanol tolerance in *E. coli*, indicating that additional components of the stress response are involved in n-butanol resistance. Overall, this study identified several promising routes for understanding both ferulic acid and n-butanol tolerance based on the *L. brevis* gene expression data. These insights may guide future engineering of model industrial organisms for better tolerance of these inhibitors.

5.2 Introduction

Lactobacillus brevis, a fastidious heterofermentative lactic acid bacterium often used in (and contaminating) fermentation processes [351], possesses multiple industrially advantageous complex phenotypes, including tolerance to short-chain alcohols such as ethanol and n-butanol [175, 200, 125], to aromatic organic compounds [133], and to hops [287]. However, because *L. brevis* has only recently begun to be extensively characterized [209, 210, 199, 123, 168, 198], little is known about the molecular mechanisms responsible for the array of environmental tolerances displayed by *L. brevis*. Of particular interest is tolerance of *L. brevis* to phenolic compounds generated during the preparation of lignocellulosic biomass for downstream use [240, 109, 251, 216, 224]. Several studies have examined the high phenolic acid tolerance of lactic acid bacteria involved in wine making or other fermentation processes [49, 326, 306, 266, 276] and have consistently identified resistance to these compounds as important factors. *L. brevis*, in particular, has been shown to possess superior tolerance of phenolic compounds [306, 133]. Improved understanding of phenolic compound tolerance in *L. brevis* would therefore be an important step forward in the industrial use of cellulosic biomass.

Because of its toxicity and abundance in plant biomass [177, 213], ferulic acid is often used to assess phenolic compound tolerance [133, 62, 105]. Chemical or

enzymatic treatments of lignocellulosic biomass release ferulic acid as a byproduct of processing [174, 8]. Ferulic acid is one of the most toxic hydroxycinnamic acids [276, 266], causing complete inhibition of *Clostridia beijerinckii* growth at 2 g/L [106]. Phenolic acids may also damage intracellular hydrophobic sites and cause ion leakage by altering membrane permeability [75, 105, 48]. The use of cellulosic biomass consequently requires either microbes with intrinsically higher tolerance of phenolic compounds or extensive downstream processing [201, 223, 286]. Tolerance of ferulic acid can thus potentially be a useful indicator of how a given organism will tolerate the phenolic compounds generated during biomass processing.

Although biomass hydrolysates can be used as the feedstock for many biotechnological processes, their use in the production of ethanol, n-butanol, and others is attracting much attention because of low costs and derivation from crops unsuited for human consumption [331, 190, 67]. The toxicity of the hydrolysates and the various fuel alcohols themselves renders efficient production of these compounds from unprocessed biomass challenging [309]. Production of n-butanol is particularly desirable because of its superior energy content compared to ethanol and compatibility with existing fuel infrastructure [100, 15]. However, n-butanol tends to partition into lipid membranes due to its low partition coefficient P_{ow} [329, 158], triggering changes in membrane fatty acid composition [152, 153] and interfering with cell metabolism due to its chaotropic properties [329]. Previous studies have shown that the 3% (v/v) n-butanol tolerance of *L. brevis* exceeds that of most other strains [175]; thus, understanding the mechanisms that confer n-butanol tolerance in *L. brevis* may simplify efforts to engineer this phenotype into industrial strains. A strain tolerating both biomass inhibitors and n-butanol simultaneously would be advantageous for the economical production of n-butanol from inexpensive feedstocks.

Numerous studies have attempted to address these roadblocks to n-butanol pro-

duction by characterizing n-butanol tolerance limits and mechanisms in *Escherichia coli* [283], *Clostridium acetobutylicum* [38, 20, 318, 317, 36], and other organisms [226, 347]. Given its multiplicity of nutrient auxotrophies and slow growth rate, *L. brevis* itself is unlikely to serve as a cost-effective host for most bioprocesses despite having been successfully engineered for n-butanol production [31]. However, understanding the basis of *L. brevis* tolerance to phenolic compounds and n-butanol could provide insights on engineering these desirable characteristics in organisms more amenable for industrial usage.

While *L. brevis* transcriptional responses to these stressors may identify numerous potential genomic loci that may affect inhibitor tolerance, there is no convenient to screen these loci to determine their role in influencing tolerance. Transferring these genes to other, more industrially tractable, organisms also poses significant challenges because of differences in codon usage and promoter structure. Previous studies have shown that it is possible to expand the *E. coli* transcriptional repertoire can be expanded to confer recognition of non-native promoters by expressing heterologous sigma factors [9, 307, 29], an approach that may viable option for rapidly screening *Lactobacillus* genes using a genomic library approach. The sigma factor complexed RNA polymerase controls initial promoter recognition [130, 352, 132, 43, 46], so this approach directly modifies the host transcriptional program to recognize promoters from the original organism. Most microbes contain numerous sigma factors expressed under a variety of stress or environmental conditions [141], although expression of genes during exponential growth generally depends on the housekeeping sigma factor σ^{70} [239]. It has been previously shown that *Lactobacillus* sigma factors are functional in *E. coli* [360], but the extent to which they can supplant the native sigma factor is unknown.

This work presents the first transcriptional analysis of *L. brevis* in response to

the stressors phenolic acid and n-butanol. It also describes an attempt to engineer the transcriptional machinery of *E. coli* to use native *Lactobacillus* promoters. Given that *L. brevis* is poorly suited to industrial bioprocesses for the reasons previously outlined, the primary goal of this study is to identify possible mechanisms for n-butanol and phenolic acid tolerance for subsequent study in other bacteria that are more amenable to engineering efforts. The transcriptional profiles associated with each inhibitor share elements of a generalized stress response, including the production of protein chaperones, increased transcription of genes involved in energy metabolism, and a general repression of growth-related functions. Several mechanisms to explain the transcriptional responses are proposed and, in the case of n-butanol tolerance, tested in two *E. coli* strains with modified fatty acid synthesis pathways. In an effort to expedite the screening of *Lactobacillus* genes in *E. coli*, housekeeping sigma factors from two different *Lactobacillus* strains were expressed and *E. coli* and found to alleviate temperature-sensitive defects in the native σ^{70}, indicating that heterologous sigma factor expression may be an effective tool for screening heterologous genomic libraries from other bacteria in *E. coli*.

5.3 Materials and Methods

5.3.1 Sigma factor identification, cloning, and expression

For sigma factor expression in *E. coli*, putative σ^{70} growth factors were identified on the basis of sequence annotation (*Lactobacillus plantarum*: LP-1926, *L. brevis*: LVIS-0756) [7] for study. Alignment using ClustalW2 program (http://www.ebi.ac.uk/Tools/) revealed highly conserved regions between these putative σ-factors and the native *E. coli* σ^{70} (σ^{70}_{Ec}) gene. Heterologous sigma factors were placed under the control of P_{lac} with the native *rpoD* ribosome binding site to ensure equivalent translation efficiency. Primers for these sequences were designed from the *E. coli*-MG1655, *L.*

brevis ATCC 367, and *L. plantarum* WCFS1 strains using Primer-BLAST [7]. *Lactobacillus* promoters are known to similar to those of *E. coli* [215], with the -35 site consensus sequence: TTgaca and -10 site consensus sequence: TAtAAT. A 17-20 bp spacer separates the sites. Capitalized letters are conserved more in *Lactobacillus*.

The low copy number pCL1920 plasmid [193] was used to express the σ^{70}-factors from *L. plantarum* and *L. brevis* in *E. coli*. First, the 300 base pair region upstream of the native *E. coli rpoD* gene containing the native RBS was introduced into pCL1920 to create pCL1920-EcS70 (pCLE). Note that the pCLE plasmid does not contain the *rpoD* gene. Subsequently, genes encoding housekeeping sigma factors from *L. brevis* and *L. plantarum* were amplified directly from the corresponding genomic DNA and purified using the Clean and Concentrator kit (Zymo Research). The fragments were digested with the appropriate restriction enzymes and ligated into the digested pCLE plasmid using T4 DNA ligase (New England Biolabs) to create pCLE-σ^{70}_{Lb} (pCLE-Lb) and pCLE-σ^{70}_{Lp} (pCLE-Lp). These expression plasmids were then introduced into *E. coli* (BW25113 and 285c) by electroporation (BioRad GenePulser XL) and plated onto selective media. All PCR cloning steps were performed using the Phusion polymerase (Finnzymes) and the constructs verified by sequencing (MCLAB Inc).

5.3.2 Bacterial strains and growth conditions for sigma factor expression

E. coli (BW25113 and 285c), *Lactobacillus brevis* (ATCC 367), and *Lactobacillus plantarum* (WCFS1) were used. *E. coli* BW25113 cultures were grown in LB broth at 37° C; 50 μg/mL of streptomycin was added as necessary. The *E. coli* mutant strain 285c with a temperature-sensitive defect in its σ^{70} protein and the parent strain P9OA5c [156, 137, 196] were generously provided by Robert Calendar (UC Berkeley) and cultured at 30 °C in LB media and 25 μg/ml streptomycin (if needed) to prevent selection of reversion in the gene encoding the heat shock sigma factor

σ^H [131]. For the complementation assay, the *E. coli* 285c strain was transformed with empty pCLE vector or plasmids containing *L. brevis* or *L. plantarum* σ-factors. Cultures were grown overnight at 30° C in LB media with 25 μg/ml streptomycin. The growth of the P9OA5c, 285c/pCLE, 285c/pCLE-Lp, and 285c/pCLE-Lb strains was monitored spectrophotomerically at 42 °C using a TECAN plate reader to assess the complementation effect of the heterologous sigma factors. Batch studies of growth at 43.5-46°C were conducted in a shaking water bath at 200 rpm. M9 minimal media with 0.5% glucose and 0.52 mM arginine (due to an *arg*$^-$ mutation in the 285c strain) was used to support microbial growth for the transcriptional studies.

5.3.3 Bacterial strains and growth conditions for transcriptional analysis

L. brevis ATCC367 (American Type Culture Collection) and *E. coli* BW25113 (CGSC) were used in this study. *L. brevis* cultures were grown at 30 °C and 100 rpm unless otherwise noted. *E. coli* BW25113 *cfa*::*kan* was obtained from the Keio collection [17]. Overnight cultures used to initiate the time-series inhibition experiments were grown in baffled 250 ml flasks with 25 ml MRS media (Difco) to $A_{600} \sim$ 1.5-1.8. Three 500 ml baffled flasks with 125 ml MRS media were then each inoculated with 3 ml of overnight culture each and grown to $OD_{600} \approx 0.4$ (mid-exponential phase) for the ferulic acid and n-butanol challenges. Each culture of *L. brevis* was challenged with 0.6 g ferulic acid (24 mM) or 1-2% (v/v) n-butanol. Subsequently, culture samples were harvested by filtration and resuspended in RNALater (Ambion) after 15, 75, and 135 minutes to track the short-term response and any long-term adaptations to either inhibitor. Prior to the addition of ferulic acid, a pooled control culture that was not exposed to chemical insult was created, harvested, and stored for use as a reference. Cell samples were stored at -80 °C for subsequent analysis. Three biological replicates were used in this study for the ferulic acid and the 1%

n-butanol (v/v) stressed cultures, while two biological replicates were used for the 2% (v/v) n-butanol stressed cultures.

5.3.4 Extraction of total RNA

Extraction of total RNA was performed with the ZR Fungal/Bacterial RNA MiniPrep kit (Zymo Research Corp) as follows: For each time point, 1.5 ml samples stored in RNALater were pelleted at 16000xg for 20 minutes, and the supernatant removed by aspiration. The bacterial pellet was then processed according to the manufacturer's protocol except that one volume of ice-cold ethanol was used to assist the RNA precipitation. DNAse I treatment was performed in-column as specified by the manufacturer. The resulting RNA was quantified using the Qubit fluorometer (Invitrogen). Gel electrophoresis was also used to confirm RNA quality. If necessary, samples were concentrated using ethanol precipitation and resuspended in 14 μl of molecular biology grade water.

5.3.5 Labeled cDNA generation and microarray hybridization

The SuperScript indirect cDNA labeling system (Invitrogen) was used to generate cDNA incorporating amino-allyl dUTP. Cy3 and Cy5 (GE Healthcare) or Alexa Fluor (Invitrogen) dyes were used to label the cDNA samples. Custom cDNA microarrays (Agilent) containing 15,209 probes (excluding positive and negative controls) were designed using the software package Picky [64] to maximize probe specificity and sensitivity under hybridization conditions with the following parameters. All probes are sixty base pair oligomers with 100% similarity to their corresponding target sequence. The minimum acceptable ΔT_m between the probe target and other sequences was set at 15° C, with a GC content range between 30-70%. The salt concentration in the hybridization media was set to 750 mM (personal communication with Agilent). The arrays contain at least three probes per ORF over 100 bp in length (2,157 ORFs

total) and 1 probe for every 1050 bp of the *L. brevis* genome [7]. Probes for the 33 protein coding genes located on the *L. brevis* megaplasmids were not included in this study; the majority of these genes are related to plasmid maintenance, mobilization, or partitioning and are therefore not expected to be relevant for stress tolerance. Labeled cDNA incubated with the cDNA arrays for 18 hours, washed with Agilent Wash Buffers 1 and 2, and then immediately scanned using a GenePix 4200A reader according to the manufacturer's protocol as described for Agilent two-color prokaryotic microarrays.

5.3.6 Microarray data analysis

The local background intensity was subtracted from the recorded signal from each array spot. Arrays were then subjected to LOWESS normalization individually using the MIDAS software package (TM4) [284]. The arithmetic average of the replicate sample and reference signals were used for downstream analysis [283]. Differentially expressed (DE) genes were identified for each time point using the rank product method with a critical p-value $P < 0.01$ [39]. The currently known functional annotations for *L. brevis* genes were obtained from the US Department of Energy Joint Genome Institute, and the number of differentially expressed hypothetical membrane proteins was identified by screening all differentially expressed hypothetical genes with the trans-membrane hidden Markov model TMHMM algorithm [176]. If required, BLAST [7] was utilized to determine *L. brevis-E. coli* homologous gene pairs. The statistical significance of the differentially expressed gene functional distributions were assessed using a hypergeometric distribution method [98]. The MeV (TM4) microarray analysis software was used for clustering and other expression profile analysis. The raw microarray data is available from the Gene Expression Omnibus (accession no. GSE24944).

5.3.7 Transcriptional target screening

To evaluate the effect of upregulated fatty acid synthesis on n-butanol tolerance, *E. coli* strain HB101(DE3) [305], kindly provided by Zhilei Chen (Texas A&M University), was transformed with the pMSD8 plasmid containing the AccABCD synthetic operon (generously provided by J. Cronan) [84] via electroporation in a Gene Pulser XL (Bio Rad). The growth rate of the HB101(DE3)/pMSD8 strain was then evaluated in M9 minimal media with 0%-2% n-butanol using an TECAN Infinite M200 plate reader (TECAN) over a period of 24 hours.

5.3.8 Fatty acid methyl ester analysis

L. brevis cultures were grown to $A_{600} \sim 0.4$ for fatty acid content analysis. Control cultures were pelleted and the supernatant removed immediately, followed by storage at -80 °C until processing. Experimental cultures were exposed to 2% n-butanol for 75 min and then stored in an analogous manner. Two biological replicates were used for each condition. Fatty acid analysis was later performed by Microbial ID Inc. to identify statistically significant changes in membrane composition following the n-butanol challenge.

5.4 Results and Discussion

The principal goal of this study was to identify mechanisms that contribute to ferulic acid and n-butanol tolerance in *L. brevis* through the application of novel gene expression microarrays. Cultures in mid-exponential growth were separately challenged with both inhibitors, and the transcriptional response was monitored to identify genes that respond immediately to each insult or that show long-term adaptation over the experimental time course. Comparing the transcriptional responses allows for the identification of genes that are expressed in immediately and long-

term response to general stressors (e.g. protein chaperones) and those genes that are expressed in response to each stress condition.

5.4.1 Transcriptional response to ferulic acid stress

Following the addition of 24 mM ferulic acid, the transcriptional data revealed an immediate stress response (summarized in Figure 5.1) in the *L. brevis* cultures. The combined functional distribution of upregulated genes (corresponding cluster of orthologous genes (COGs)) at 15 and 135 minutes demonstrated a marked departure from that of the reference samples. Genes involved in carbohydrate transport and metabolism (COG G), transcriptional regulation (COG K), and amino acid transport and metabolism (COG E) underwent severe pertubations as *L. brevis* adjusts its transcriptional program to respond to the ferulic acid challenge. The relative change in expression level of genes that appear most critical to the stress response are presented in Table 5.1. Many up-regulated genes encode proteins involved in the citric acid cycle or sugar utilization, such as malate dehydrogenase, fumarase, catabolite control protein A (*ccpA*), or sugar transporters. Expression of a NADPH-quinone oxidoreductase may be an attempt to control excess O_2^- levels [76] generated within an aerobic environment. The phenolic acid decarboxylase (LVIS-0213) of *L. brevis* is by far the most overexpressed gene, as would be expected, given the role of this enzyme in degrading phenolic acids. A up-regulated heat shock protein (LVIS-0112), conserved in many lactic acid bacteria [7], may participate in folding the heat labile phenolic acid decarboxylase [136, 88, 276, 182]; however, this chaperone is also expressed in response to n-butanol stress so it is likely a general stress response to protein mis-folding.

Many of the up-regulated genes detected encode hypothetical proteins with poor or no functional annotation. However, screening these *L. brevis* hypothetical ORFs

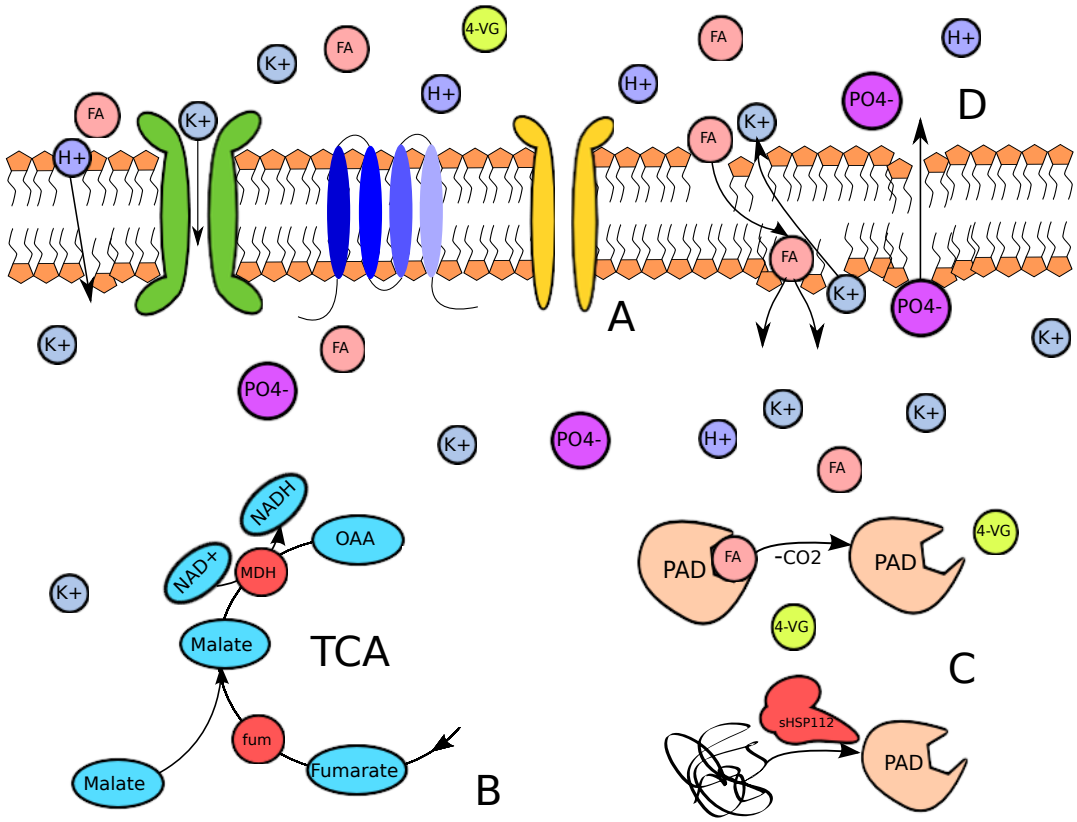

Figure 5.1: The essential *L. brevis* stress response to ferulic acid. A. Expression of many uncharacterized, multi-domained membrane proteins that may function as ion transporters, ferulic acid, or 4-vinylguaicol exporters, among several possibilities. B. Upregulation of fumarase (fum), malate dehydrogenase (mdh), and malate permases involved in the citric acid cycle. NADH dehydrogenase and glutathione reducatase may also play a role in NAD+/NADH recycling and superoxide generation. C. Expression of a small HSP to ensure proper folding of heat labile phenolic acid decarboxylase (PAD) which converts ferulic acid to 4-vinylguaicol [182]. **D.** Representation of membrane damage and ion leakage triggered by ferulic acid induced lipid disruption.

using the trans-membrane hidden Markov model TMHMM [176] revealed that 28 of the 55 hypothetical genes up-regulated at either time point possess one or more transmembrane helices. These data suggest that the proteinaceous content of the

membrane changes to counteract ferulic acid toxicity. These proteins may act as membrane stabilizers or ion transporters to counteract the membrane disruption triggered by ferulic acid, as exposure to n-butanol did not alter expression of these genes compared to the reference culture. Their expression changes specifically in response to ferulic acid stress or general phenolic compound stress. Two uncharacterized major facilitator superfamily permeases (LVIS-1730 and LVIS-1917) with unknown substrate specificity were also overexpressed following ferulic acid insult; these exporters may be involved in the active export of ferulic acid or its decarboxylated product 4-vinylguaicol. These uncharacterized membrane proteins and transporters are promising targets for additional studies to unravel the precise genetic basis of ferulic acid tolerance in *L. brevis*, especially given their unique occurrence in the ferulic acid or phenolic compound stress profile.

In contrast to these shifts in energy metabolism and transcriptional regulation, genes involved in membrane biogenesis (COG M) comprise only 4% of the 195 genes, indicating that *L. brevis* may not appreciably alter the lipid composition of cytoplasmic membrane or the thickness of the outer peptidoglycan layer in response to ferulic acid. However, the consistent upregulation of β-ketoacyl-(acyl-carrier-protein) reductase (*fabG*, LVIS-0378), a key enzyme in type II fatty acid synthesis [340], suggests that membrane lipid abundance may be altered as a possible defense mechanism against ferulic acid induced membrane fluidity. This hypothesis is supported by observed changes in the membrane of the closely related *Lactobacillus plantarum* when challenged with caffeic and ferulic acids [279]. No other genes in the fatty acid synthesis pathway are upregulated. Given the similarity between the organisms, membrane composition changes of *L. brevis* in response to ferulic acid are very likely to reflect those observed in phenolic acid-tolerant *L. plantarum*.

As expected, the slower growth rate of *L. brevis* in the presence of ferulic acid

Table 5.1: Expression of select upregulated genes during ferulic acid stress

Locus	Gene Description	COG(s)	15 min	135 min
LVIS-0076	NADPH-quinone oxidoreductase	C	0.12	**2.56**
LVIS-0714	Fumarate hydratase	C	0.95	**2.15**
LVIS-2203	Malate dehydrogenase	C	1.54	**2.24**
LVIS-1730	MFS permease	G	**2.88**	**3.41**
LVIS-1917	MFS permease	G	**2.06**	**2.74**
LVIS-2255	Sugar transport system, permease	G	**1.39**	**2.48**
LVIS-2256	Sugar transport system, periplasmic	G	**1.54**	**2.99**
LVIS-0113	Amino acid transporter	E	**2.10**	**3.83**
LVIS-1879	Amino acid transporter	E	**1.51**	**2.78**
LVIS-1951	Glycine cleavage system H protein	E	0.28	**2.60**
LVIS-0213	Phenolic acid decarboxylase	Q	**5.75**	**4.19**
LVIS-0378	FabG	IRQ	**1.52**	**2.48**
LVIS-0924	Biotin operon repressor	H	**1.77**	0.05
LVIS-0086	Peroxiredoxin	O	**2.51**	**2.22**
LVIS-0112	Molecular chaperone (sHSP)	O	**4.18**	**2.62**
LVIS-2147	Catabolite control protein A	K	**1.51**	1.27
LVIS-0031	Hypothetical protein LVIS-0031	S	**1.34**	1.57
LVIS-0155	Hypothetical protein LVIS-0155	S	0.19	**2.13**
LVIS-0157	Hypothetical protein LVIS-0157	S	0.93	**2.61**
LVIS-0212	Hypothetical protein LVIS-0212	S	**4.26**	**3.93**
LVIS-0262	Hypothetical protein LVIS-0262	S	1.24	**2.36**
LVIS-0305	Hypothetical protein LVIS-0305	S	0.86	**2.81**
LVIS-0314	Hypothetical protein LVIS-0314	S	0.44	**2.58**
LVIS-0445	Hypothetical protein LVIS-0445	S	**1.90**	**2.23**
LVIS-0552	Hypothetical protein LVIS-0552	S	**1.32**	**2.54**
LVIS-0712	Hypothetical protein LVIS-0712	S	**1.87**	0.73
LVIS-1831	Hypothetical protein LVIS-1831	S	**1.48**	1.73
LVIS-1880	Hypothetical protein LVIS-1880	S	**2.84**	**2.91**
LVIS-1881	Hypothetical protein LVIS-1881	S	**2.64**	**2.12**
LVIS-2013	Hypothetical protein LVIS-2013	S	0.24	**2.22**
LVIS-2118	Hypothetical protein LVIS-2118	S	0.66	**2.17**
LVIS-2201	Hypothetical protein LVIS-2201	S	0.53	**2.99**
LVIS-0472	ABC-type Mn^{2+}/Zn^{2+} permease	P	0.56	**2.08**

COG definitions: C: Energy production and conversion, G-Carbohydrate transport and metabolism, I-Lipid transport and metabolism, K-Transcription, O-Posttranslational modification, protein turnover, chaperones, P-Inorganic ion transport and metabolism, Q-Secondary metabolites biosynthesis, transport and catabolism, and S-Function Unknown. COG S includes hypothetical proteins with trans-membrane domains identified by TMHMM. Bolded $\log_2$ scores indicate statistically significant overexpression during that time point.

is reflected in the functional roles of the down-regulated genes over the course of the challenge. An overview of the genes repressed by ferulic acid stress is given in Table 5.2. Genes encoding proteins responsible for amino acid and carbohydrate metabolism, transcription, and translation are strongly repressed, indicating that a metabolic shift from normal growth has occurred. Repression of several H^+ antiporters may be due to a disruption of normal cellular proton gradient caused by ferulic acid-induced membrane damage [51, 295]. A total of 35 genes encoding proteins without known functional roles were down-regulated, but only 8 of these proteins appear to contain trans-membrane segments. On the whole, genes repressed during the ferulic acid challenge are similar to those seen during the n-butanol challenges (see below), implying that growth and certain types of nutrient assimilation are inhibited within stressful environments. This result is also consistent with the repression of two key proteins involved in cell division (LVIS-0848, LVIS-1402) and the observed growth defects under both stress conditions.

5.4.2 Transcriptional response to n-butanol stress

The transcriptional response of *L. brevis* to n-butanol insults was tracked over time to reveal unique resistance mechanisms that could be combined with the insights gained from the ferulic acid challenges to produce a superior industrial strain. Unlike the expression pattern observed with ferulic acid challenge, the addition of 1-2% n-butanol induced profound perturbations in lipid biosynthesis, protection against excessive oxidative stress and synthesis of protein chaperones. Given the similarity of the transcriptional profiles in response to 1% and 2% n-butanol, only the latter will be discussed here. A survey of the *L. brevis* n-butanol transcriptional response is provided in Tables 5.3 and 5.4. Putative chaperones or chaperone components such as *dnaJ* (LVIS-1328), chaperone LVIS-0112, and ATP-binding subunits (LVIS-762,

Table 5.2: Expression of select downregulated genes during ferulic acid stress

Locus	Gene Description	COG(s)	15 min	135 min
LVIS-0199	Na^+/H^+ antiporter	C	**-2.19**	**-4.62**
LVIS-2211	Na^+/H^+ antiporter	C	-0.67	**-4.08**
LVIS-2141	Na^+/H^+ and K^+/H^+ antiporter	P	**-2.01**	-0.51
LVIS-0848	Cell division initiation protein	D	**-1.76**	-1.32
LVIS-1455	Cell division protein MraZ	S	-0.39	**-2.97**
LVIS-1587	Amino acid transporter	E	**-1.95**	**-3.15**
LVIS-2023	Carbamate kinase	E	**-1.84**	-1.09
LVIS-2024	Transaminase	E	**-1.94**	-0.85
LVIS-2025	Amino acid transporter	E	**-2.23**	-1.31
LVIS-2212	Amino acid transporter	E	-1.01	**-4.18**
LVIS-2213	Glutamate decarboxylase	E	-0.74	**-3.70**
LVIS-0934	Acyl carrier protein	IQ	-0.24	**-2.54**
LVIS-0936	Transcriptional regulator	K	-0.58	**-2.66**
LVIS-0937	FabZ	I	-0.59	**-2.59**
LVIS-0954	Glycerol-3-phosphate acyltransferase	I	-0.65	**-2.24**
LVIS-0064	Hypothetical protein	S	**-1.97**	**-2.29**
LVIS-0422	Hypothetical protein	S	**-2.56**	**-2.52**
LVIS-0899	Hypothetical protein	S	**-1.74**	**-2.68**
LVIS-1369	Hypothetical protein	S	**-1.68**	**-1.97**
LVIS-1834	Hypothetical protein	S	**-1.77**	-0.83
LVIS-1895	Hypothetical protein	S	-0.74	**-2.46**
LVIS-2216	Hypothetical protein	S	**-2.25**	**-4.64**

COG definitions: C: Energy production and conversion, D-Cell cycle control, cell division, chromosome partitioning, P-Inorganic ion transport and metabolism, and S-Function Unknown. COG S includes hypothetical proteins with trans-membrane domains identified by TMHMM. Bolded $\log_2$ scores indicate statistically significant repression during that time point.

LVIS-1554, LVIS-1700) are significantly upregulated during the entire experimental time course. This pattern of chaperone upregulation was similar to that observed during exposure to ferulic acid, suggesting that these proteins are expressed in response to general environmental stress instead of being evoked by a specific chemical stressor. There are several signs of oxidative stress as well; the upregulation of two peptide methionine sulfoxide reductases (LVIS-0809 and LVIS-0810), which repair oxidative

damage to methionine residues in conjunction with thioredoxin, NADPH-quinone oxidoreductase (LVIS-0076), and one thioredoxin (LVIS-1216). This response is consistent with previous reports of n-butanol-induced oxidative stress in *E. coli* [283]. Another component of the *L. brevis* oxidative stress response is the accumulation of manganese [11, 147], as shown by the high expression of the Mn^{2+}/Zn^{2+} ion transport system (LVIS-0451 and LVIS-0452) over the time series.

Expression of genes whose products are involved with amino acid, carbohydrate, and ion transport and metabolism are significantly altered during the n-butanol stress response. Several unannotated amino acid transporters are weakly downregulated only 15 minutes after n-butanol addition (LVIS-0619, LVIS-1789), but no change in the metabolism of amino acids known to function as osmoprotectants [93] is seen. Other metabolic functions involved in amino acid biosynthesis (LVIS-2023-LVIS-2027) are highly repressed as well. The expression of genes encoding enzymes responsible for the uptake and metabolism of carbohydrates (particularly pentoses), including phosphopentomutase (LVIS-1594), L-ribulokinase (LVIS-1742), and β-galactosidase (LVIS-2259) increases continually throughout the experiment, perhaps as part of a concerted effort to support the energy demanding stress response. The strong downregulation of a Na^+/H^+ antiporter (LVIS-2211) is consistent with with the observed effects of n-butanol on H^+ gradients in other organisms [51, 295]. In addition to these metabolic changes, expression of genes whose products are involved cell wall synthesis is also downregulated immediately following the n-butanol challenge, though this repression lessens over time.

Although these changes in expression pattern are consistent with a reduced growth rate during an energy-consuming stress response, repression of genes encoding proteins possibly involved in competing phenotypes is also seen. Genes encoding the putative homologs of several proteins involved in acid tolerance in *E. coli*, in-

Table 5.3: Expression of select upregulated genes during 2% n-butanol stress

Locus	Description	COG	15 min	75 min	135 min
LVIS-0076	NADPH-quinone oxidoreductase	CR	**3.05**	**3.06**	**2.24**
LVIS-0320	NADH dehydrogenase	C	1.457	**2.02**	**2.19**
LVIS-0811	ADP-ribose pyrophosphatase	F	1.32	**2.02**	**2.46**
LVIS-1594	Phosphopentomutase	G	-0.97	1.14	**1.82**
LVIS-1742	L-ribulokinase (putative)	G	**2.13**	0.77	1.00
LVIS-0187	Acetoin reductase	IQ	-0.13	**1.80**	**1.84**
LVIS-0925	Enoyl-(ACP) reductase	I	0.49	**1.91**	**1.82**
LVIS-0926	Acetyl-CoA carboxylase α	I	0.29	0.54	**1.68**
LVIS-0927	Acetyl-CoA carboxylase β	I	1.26	1.33	**1.88**
LVIS-0928	Biotin carboxylase	I	0.77	**1.85**	**2.41**
LVIS-0929	FabZ	I	1.049	**1.79**	**2.23**
LVIS-0930	Biotin carboxyl carrier protein	I	-0.22	1.19	**1.43**
LVIS-0931	FabF	IQ	0.62	**1.61**	**3.20**
LVIS-0932	FabG	IQR	0.68	**2.36**	**2.98**
LVIS-0933	(ACP) S-malonyltransferase	I	0.74	**2.15**	**2.76**
LVIS-0934	Acyl carrier protein	IQ	0.36	**2.08**	**2.59**
LVIS-0935	FabH	I	**1.55**	**2.23**	**2.62**
LVIS-0937	FabZ	I	1.203	**2.40**	**2.78**
LVIS-0112	Molecular chaperone (sHSP)	O	**1.96**	**2.01**	**1.65**
LVIS-0762	ATP-binding subunit of DnaK	O	**3.51**	**2.78**	**1.69**
LVIS-0809	Methionine sulfoxide reductase	O	1.14	**1.97**	**2.36**
LVIS-0936	Transcriptional regulator	K	0.91	**2.37**	**2.48**
LVIS-1216	Thiol-disulfide isomerase	O	1.18	**1.6**	**1.52**
LVIS-1328	DnaJ-like molecular chaperone	O	**1.87**	**2.21**	**1.87**
LVIS-1554	ATP-binding subunit of DnaK	O	1.50	**1.73**	0.89
LVIS-1700	ATP-binding subunit of DnaK	O	**1.66**	**1.56**	**1.57**
LVIS-1701	Repressor of class III stress genes	K	**2.01**	**1.63**	**1.56**
LVIS-2091	RNAP sigma subunit, σ^{24}-like	S	0.67	-0.52	**2.25**
LVIS-0116	Cation transport ATPase	P	**1.64**	**1.57**	0.79
LVIS-0471	Mn/Zn transporter ATPase	P	**1.64**	**1.56**	0.69
LVIS-0472	Mn^{2+}/Zn^{2+} transporter	P	1.44	**1.59**	0.85
LVIS-1844	Aldo/keto reductase	R	**2.30**	**2.85**	**2.24**

COG definitions: C: Energy production and conversion, D-Cell cycle control, cell division, chromosome partitioning, G-Carbohydrate transport and metabolism, I-Lipid transport and metabolism, K-Transcription, O-Posttranslational modification, protein turnover, chaperones P-Inorganic ion transport and metabolism, Q-Secondary metabolites biosynthesis, transport and catabolism, R-General function prediction only, and S-Function Unknown. Each $\log_2$ score is bolded if the gene is significantly overexpressed during that time point.

cluding glutamate decarboxylase (LVIS-2213) the glutamate acid resistance system *gadC* (LVIS-0078) [52], and cyclopropane fatty acid synthase (LVIS-2047) [56], are downregulated for 2 time points. Other elements of the *Lactobacillus* acid stress response, such as the arginine deiminase pathway (LVIS-2023, LVIS-2026, LVIS-2027) [85], are also strongly repressed. These responses are curiously similar to the observed antagonism between ethanol tolerance and acid tolerance phenotypes in *E. coli* [127], suggesting that the same relationship may exist between the n-butanol and acid tolerance phenotypes.

Along with oxidative and general stress adaptations that are common to most organisms exposed to n-butanol, an upregulation of the entire fatty acid synthesis pathway is clearly evident in the transcriptional data. The FA synthesis pathway itself is summarized in Figure 5.2. Increased expression of the genes encoding acetyl-CoA carboxylase subunits (*accABCD*) shunts additional acetyl-CoA to malonyl-CoA for dedicated use in the FAS pathway [207, 340, 84]. Subsequently, after the initial condensation step by catalyzed by FabH, FabFGZI homologs identified with BLAST in *L. brevis* (see Table 5.3) continue the process of fatty acid elongation. These fatty acids are then generally converted into membrane phospholipids by the glycerolphosphate acyltransferase system [358]. The expression of the *plsX* (LVIS-0954) and *plsC* (LVIS-1355) genes does not change significantly over the course of the experiment, so it is possible that expression of enzymes is not limiting for membrane lipid synthesis, that their activities are regulated post-transcriptionally, or that the fatty acid flux is diverted elsewhere in the cell. To test whether increased flux through the FAS pathway affects n-butanol tolerance, the *accABCD*-overexpression strain developed by Davis et al. [84] was subjected to a n-butanol challenge. No significant change in n-butanol tolerance or growth rate was observed with this strain, suggesting that the increased production of fatty acids per se is insufficient to confer a protective effect.

Table 5.4: Expression of select downregulated genes during 2% n-butanol stress

Locus	Gene Description	COG	15 min	75 min	135 min
LVIS-0514	L-lactate dehydrogenase	C	-0.35	-0.24	**-1.49**
LVIS-1558	NAD-dependent dehydrogenase	R	**-1.38**	**-1.29**	**-1.52**
LVIS-2211	Na^+/H^+ antiporter	C	**-2.07**	**-1.70**	**-1.31**
LVIS-0078	Glu-aminobutyrate antiporter	E	-0.95	**-1.73**	**-1.30**
LVIS-1712	Amino acid transporter	E	0.24	**-1.47**	**-1.18**
LVIS-1781	Amino acid transporter	E	-1.21	**-1.26**	-0.77
LVIS-2023	Carbamate kinase	E	**-2.10**	**-1.93**	**-1.33**
LVIS-2024	Transaminase	E	**-2.52**	**-1.85**	**-1.18**
LVIS-2025	Amino acid transporter	E	**-2.49**	**-1.59**	-0.86
LVIS-2026	Ornithine carbamoyltransferase	E	**-2.66**	**-1.86**	-0.33
LVIS-2027	Arginine deiminase	E	**-3.70**	**-2.04**	-0.89
LVIS-2049	Branched-chain AA permease	E	**-1.29**	-1.03	-0.71
LVIS-2213	Glutamate decarboxylase	E	**-2.27**	**-1.78**	**-1.17**
LVIS-0413	Glycerate kinase	G	-1.07	-1.05	**-1.56**
LVIS-0661	Glyceraldehyde-3-P dehydrogenase	G	**-1.61**	-1.02	**-1.19**
LVIS-0689	Glucosamine-6-P isomerase	G	**-1.71**	**-1.54**	**-1.38**
LVIS-1417	Cell wall-associated hydrolase	M	**-1.78**	-1.11	-0.43
LVIS-1419	Cell wall-associated hydrolase	M	**-1.30**	-0.93	-0.99
LVIS-1496	Amidase	M	**-1.58**	**-1.27**	**-1.28**
LVIS-1548	Integral membrane protein	S	0.08	-0.96	**-1.18**
LVIS-1575	Glycosyltransferase-like	M	**-1.78**	-1.04	-0.84
LVIS-1809	Cell wall-associated hydrolase	M	**-1.48**	-0.33	-0.42
LVIS-2047	Cyclopropane fatty acid synthase	M	**-1.48**	-1.14	-0.76
LVIS-1641	Cytochrome bd quinol oxidase	C	**-1.30**	-0.89	-0.26

COG definitions: C: Energy production and conversion, D-Cell cycle control, cell division, chromosome partitioning, G-Carbohydrate transport and metabolism, I-Lipid transport and metabolism, K-Transcription, O-Posttranslational modification, protein turnover, chaperones P-Inorganic ion transport and metabolism, Q-Secondary metabolites biosynthesis, transport and catabolism, R-General function prediction only, and S-Function Unknown. Each $\log_2$ score is bolded if the gene is significantly repressed during that time point. AA: amino acid, Glu: glutamate.

n-Butanol and other solvents alter the ratio of saturated and unsaturated membrane lipids in many organisms [152, 153, 262]. Altered gene expression within the fatty acid synthesis regulon may therefore be partially responsible for the n-butanol

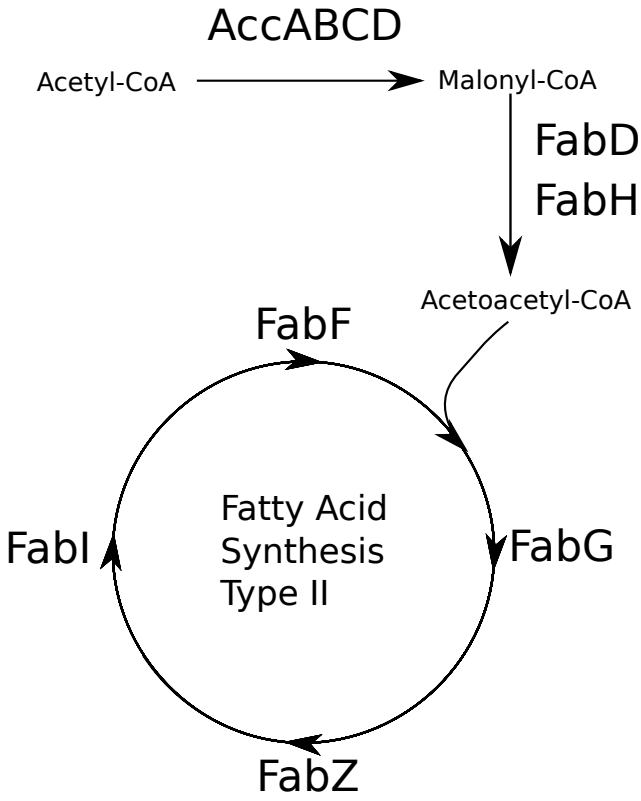

Figure 5.2: A visualization of the Type II fatty acid synthesis pathway found in most bacteria [340]. The AccABCD proteins carboxylate acetyl-CoA to form malonyl-CoA, followed by condensation of malonyl-CoA with acetyl-CoA to form acetoacetyl-CoA by FabH. Other proteins including FabG, FabZ, FabI and FabF elongate the acetoacetyl-CoA by two carbons every pass through the cycle. Transcription of each FAS gene is only slightly upregulated immediately following n-butanol addition but increases significantly after 75 and 135 minutes.

tolerance of *L. brevis*. Direct assessment of membrane fatty acid composition of *L. brevis* following 75 min of 2% (v/v) n-butanol stress revealed a significant 21.6% decrease in the abundance of 19:1 cyclopropane fatty acid (19:1-cfa) compared to unchallenged control cultures (Student's one tailed t-test, $p < 0.04$). This fatty acid comprised only 5.67% f all fatty acids found in the *L. brevis* membrane during exponential growth in the unchallenged controls versus 4.46% in cultures subjected to n-butanol insult. Given the known role of cycloproprane fatty acids in conferring acid

tolerance [42, 56], this result supports the inference that n-butanol and acid tolerance are are antagonistic phenotypes. In order to test whether this situation carries over to other organisms, the growth of a *E. coli* strain with a *cfa* knockout strain in 0-2% n-butanol was compared to the growth of the *E. coli* BW25113 parent strain. The growth rate and maximum n-butanol tolerance of the two strains were very similar. This result could be explained by the lack of cyclopropane fatty acids in the *E. coli* membrane during exponential growth, so that knocking out the *cfa* gene would have little effect on the actual membrane composition in that growth phase. A decrease in cyclopropane fatty acid content is also observed in several *E. coli* strains evolved for isobutanol tolerance [217]. Additional work is needed to determine the role that cyclopropane fatty acids play in modulating n-butanol resistance and to understand opposing patterns of gene expression associated with n-butanol tolerance and acid tolerance.

Numerous studies have examined the transcriptional programs of *C. acetobutylicum* and *E. coli* in response to n-butanol challenge. We chose to compare the n-butanol-induced transcriptional responses of *L. brevis* and *E. coli*, as *E. coli* is widely used for a variety of applications in the biotechnology industry. The recent study by Rutherford and coworkers [283] identified several features of the n-butanol stress response in *E. coli*: increased expression of genes involved with oxidative phosphorylation (*nuo*, *cyo*, and *sdh* operons), an oxidative stress response involving *sodA* (superoxide dismutase) and *yqhD* (alcohol dehydrogenase), perturbed amino acid and carbohydrate transport, and an extracytoplasmic stress response as indicated by *cpxP*, *degP*, *spy*, and *rpoE* expression. Although the lack of known functions for many *L. brevis* genes hinders a direct comparison between these organisms, the expression of oxidative stress genes and the disturbances in *L. brevis* metabolism (in terms of transport and energy demands to support the n-butanol stress response) agree well with those seen

in *E. coli*. While *L. brevis* may also be under significant extracytoplasmic stress, the *L. brevis rpoE*-like sigma factor, which most likely promotes the transcription of genes whose products respond to this form of cellular stress, is not significantly upregulated at any time point, though a σ^{24}-like protein is upregulated during late 2% n-butanol stress. No other regulators are apparent from BLAST comparisons with *E. coli* genes; however, the *dnaJ* and LVIS-0112 chaperones along with several other protein chaperones are upregulated as expected. Strikingly, no statistically significant upregulation of the *E. coli* FAS genes was observed in the Rutherford et al. study [283], in contrast to the increased expression of FAS genes in *L. brevis* following n-butanol induction.

5.4.3 Developing a high-throughput screening tool for Lactobacillus genes

One of the main limitations on screening the identified, stress-responsive *L. brevis* genes is the lack of a convenient expression method in *E. coli*. Previous studies have shown that it is possible to expand the *E. coli* transcriptional repertoire using heterologous sigma factor expression [290, 313, 211, 13, 307, 29], which may be a viable option for rapidly screening *Lactobacillus* genes using a genomic library approach. To this end, we asked if the expression of the housekeeping σ^{70} proteins from both *L. brevis* and *Lac plantarum* could allow for improved transcription of *Lactobacillus* promoters in *E. coli*. *L. plantarum* is a close relative of *L. brevis* with many of the same tolerance phenotypes [173], and so this strain could also be used as a source of genetic material to improve the tolerance of *E. coli* to stress.

Our initial attempts to clone the genes encoding *Lactobacillus* sigma factors using pUC-based high copy vectors failed to produce stable constructs. Large deletions were observed in several clones, indicating that expressionof the plasmid-borne genes were toxic, even in the absence of exogenous inducer. The sigma factor constructs

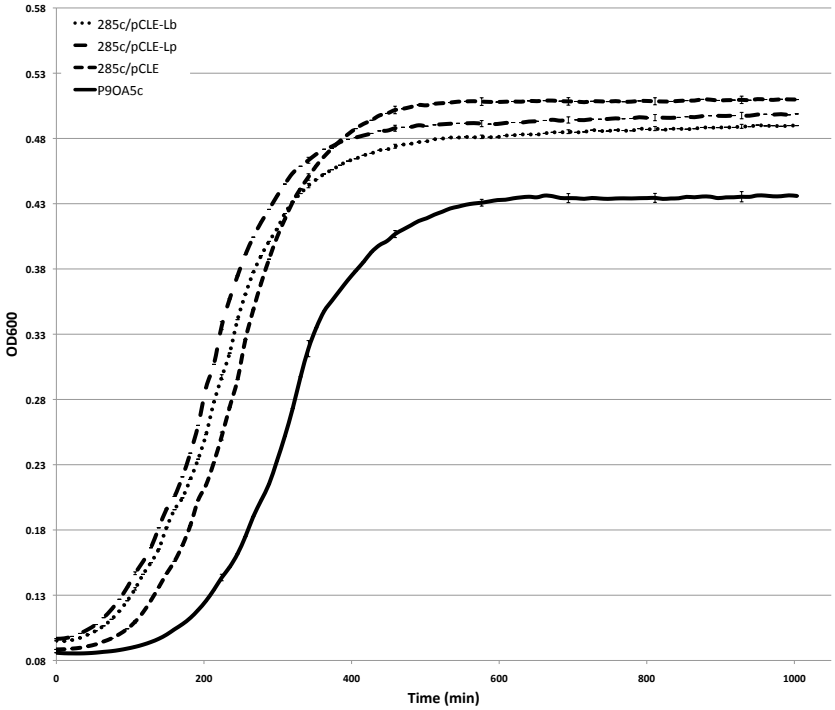

Figure 5.3: Growth of all 285c strains and the parent strain at 30 °C. Each strain was inoculated from overnights into LB with 25 μg/ml streptomycin (antibiotic omitted for P9OA5c) and cultured at 30 °C to determine the growth kinetics at the permissive temperature. As can be seen, the growth rates and the yields of the strains are similar.

were less toxic at lower expression levels, however, and low-copy plasmids containing the *Lactobacillus* sigma factors could be stably maintained over time. These low-copy plasmids containing the *Lactobacillus* sigma factors were introduced into *E. coli* 285c (see Methods and Materials) to test their ability to complement temperature sensitivity arising from a defective native *rpoD* gene. The growth kinetics of the 285c/pCLE-Lb, 285c/pCLE-Lp, 285c/pCLE, and the parent strains (P9OA5c) as shown as a function of temperature in Figure 5.3 are quite similar at the permissive temperature (30 °C), indicating that there is no defect in $\sigma^{70}_{Ec,ts}$ activity at low temperatures. However, growth at the non-permissive temperature (42 °C) revealed

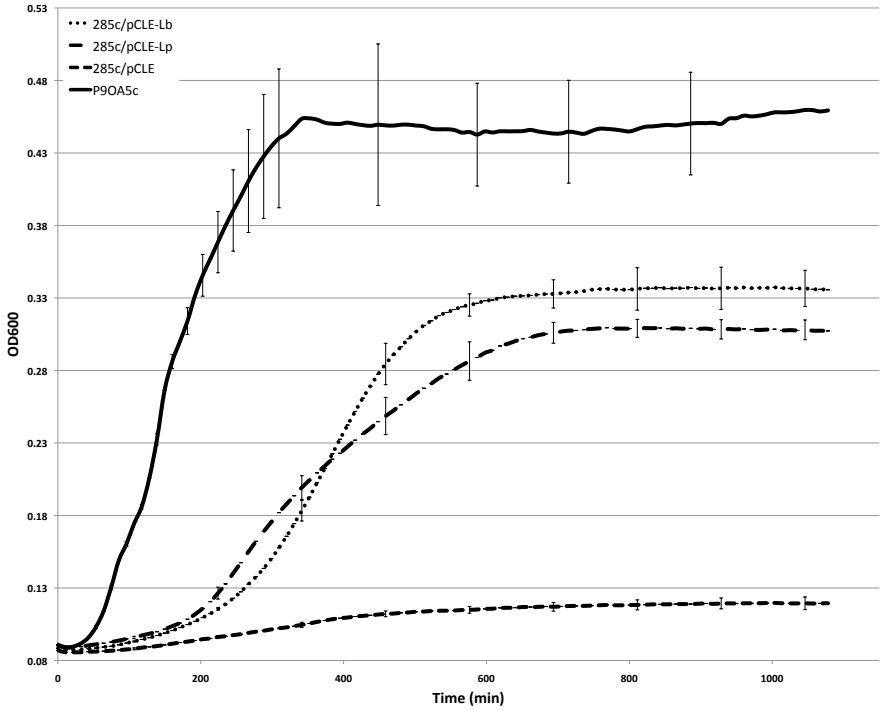

Figure 5.4: Complementation challenge at 42 °C. Each strain was inoculated from overnights into LB with 25 μg/ml streptomycin (antibiotic omitted for P9OA5c) and cultured at 42 °C to determine if the heterologous sigma factors could complement a temperature sensitive defect in the native σ^{70} protein. Robust growth is observed for 285c/pCLE-Lb and 285c/pCLE-Lp while the 285c/pCLE strain is not viable at the elevated temperature.

significant differences in the growth between the strains (Figure 5.4). Little to no growth is seen in the 285c/pCLE strain, while the 285c/pCLE-Lb and 285c/pCLE-Lp strains containing the heterologous sigma factors are not only viable, but grow fairly robustly. The doubling time of the 285c/pCLE-Lb and 285c/pCLE-Lp strains are significantly greater than that of the parent strain (∼255 min versus 108 min for P9OA5c), suggesting that σ^{70}_{LB} and σ^{70}_{LP} can restore viability at elevated temperatures, whereas the growth rate of 285c is severely impaired. These results strongly support the conclusion that the *L. brevis* and *L. plantarum* sigma factors can bind to the *E. coli* RNA polymerase and initiate transcription *in vivo*. A recent study

by Zingaro et al. [360] has also demonstrated that recognition of *L. planatarum* promoters is improved if σ^{70}_{Lp} is also expressed, though this study is the first, to the best of our knowledge, to demonstrate that *Lactobacillus* sigma factors are able to complement a defective native *rpoD*.

5.5 Conclusions

This study was aimed at revealing components of the *L. brevis* stress response to ferulic acid and n-butanol, two distinct inhibitors of microbial growth that are of great industrial importance. The results presented here, combined with previous studies on the mechanisms of phenolic acid toxicity, suggest that alterations in membrane structure and fluidity play an important role in maintaining cell membrane integrity in *L. brevis* exposed to ferulic acid. The diverse range of membrane proteins overexpressed by *L. brevis* in response to ferulic acid are likely involved in ameliorating abnormal ion flux across the membrane while enhancing rigidity, though molecular characterization of the gene products is needed for confirmation. Introduction of efflux pumps to expel ferulic acid and its decarboxylation product, along with increased levels of ion pumps to maintain intracellular K^+, H^+, and Na^+ at appropriate concentrations in the presence of ferulic acid, may be promising steps towards improving the tolerances to phenolic acids in other organisms. A novel method for expanding promoter recognition in *E. coli* using heterologous expression of sigma factors from two *Lactobacillus* strains was also pursued, with the initial data indicating that the *Lactobacillus* sigma factors are compatible with the *E. coli* RNA polymerase core enzyme.

In contrast to its response to ferulic acid, *L. brevis* responds to n-butanol stress by increasing defenses against oxidative damage, increasing carbon uptake, altering ion transport, and upregulating of fatty acid synthesis. Direct analysis of *L. brevis*

membrane composition revealed a significant decrease in the abundance of 19:1 cyclopropane fatty acid. However, the n-butanol tolerance of an *E. coli cfa* knockout strain, which is unable to synthesize cyclopropane fatty acids, was not improved compared to wild-type, indicating that additional elements of the stress response are required for tolerance. Neither *E. coli* nor *C. acetobutylicum* overexpress the FAS genes, suggesting that increased fatty acid synthesis in response to n-butanol insult contributes to n-butanol tolerance in *L. brevis*. Upregulated fatty acid synthesis may restore membrane integrity to oppose n-butanol-induced fluidization. Additional biochemical characterization of how *L. brevis* regulates its FAS response is crucial to incorporating this desirable phenotype into organisms better suited for industrial use. The apparent inverse correlation between the acid and alcohol tolerance *L. brevis* phenotypes, seen in the transcriptional data and fatty acid composition analyses, suggests that deletions in acid resistance pathways may further improve n-butanol tolerance.

6. COMPUTATIONAL IDENTIFICATION OF ADAPTIVE MUTANTS USING THE VERT SYSTEM*

6.1 Summary

The population dynamics of evolving populations can now be visualized using the Visualizing Evolution in Real Time (VERT) system, in which isogenic strains expressing different fluorescent proteins compete during adaptive evolution and are tracked using fluorescent cell sorting to construct a population history over time. Mutations conferring enhanced growth rates can be detected by observing changes in the relative proportions of each different fluorescent marker. Using data obtained from several VERT experiments, we constructed a hidden Markov-derived model to detect these adaptive events in VERT experiments. Analysis of annotated data revealed that the model achieves consensus with human annotation for 85-93% of the data points. A method to determine the optimal time at which to isolate adaptive mutants is also introduced. The model offers a new way to monitor adaptive evolution experiments without the need for external intervention, thereby simplifying efforts to track population frequencies. Future efforts to construct a fully automated system to isolate adaptive mutants may find the algorithm a useful tool.

6.2 Introduction

Strain development to improve the utility of microbial strains has been a focus of industry for decades. Numerous methods to improve the characteristics of

*Reprinted with permission from "Computational Identification of Adaptive Mutants using the VERT system" by J. Winkler and K.C. Kao, 2012. Journal of Biological Engineering, Volume 6(3), pp. 1-8, Copyright 2012 by J. Winkler.

biocatalysts have been developed, such as random mutagenesis [3, 171], genetic recombination [3, 246, 61, 21], serial transfers in the presence of various inhibitors [15], and others [310, 191, 6, 172, 334, 66]. A novel method to identify the occurrence and expansion of adaptive mutants within an evolving population was recently described by Kao and Sherlock [163]; the population dynamics of strains expressing different fluorescent proteins competing for the limiting carbon source in a chemostat system were monitored using fluorescence activated cell sorting (FACS). This approach (VERT, Visualizing Evolution in Real Time) has been applied to elucidate the population dynamics of *Candida albicans* in the presence of an antifungal agent [149] and to generate *Escherichia coli* mutants tolerant of n-butanol [269]. The use of fluorescent labels improves the ability to track various subpopulations in real time compared to microarrays [41] or quantitative PCR [336]. The VERT method is therefore ideal for identifying adaptive events more quickly than other techniques for monitoring adaptive dynamics.

A key aspect of VERT system and other population tracking methods involves analysis of population dynamics to detect adaptive events, which are subpopulation expansions triggered by novel adaptive mutants with growth-enhancing mutations. For example, if a growth-enhancing mutation (such as one that confers drug resistance or more efficient nutrient uptake) arises in a labeled subpopulation, the subpopulation in which that adaptive event occurred will increase in frequency at the expense of other competing, differently labeled subpopulations. An algorithm for analyzing data on population history method will be consistent and reliable than methods that visually detect these events. A simple, yet robust, method that can identify adaptive episodes automatically is the hidden Markov model (HMM) [260, 261], which involves the computation of the unknown state sequence that is most likely to produce the observed output (emissions) from the process in question. This technique

can be applied to determine whether each subpopulation is undergoing an adaptive expansion by examining the subpopulation ratios and then computing the probability of an adaptive event based on the model training data. A HMM based approach will also be flexible enough to accommodate variations between experiments arising from species-specific dynamics, data-quality issues, and other factors.

In this work, we introduce a population state model (PSM) that employs a hidden Markov model to identify likely adaptive events for several types of chemostat evolution experiments that employed the VERT tracking system. After showing that the PSM predictions are comparable to those obtained from human annotation, properties of several VERT experiments for different species are quantified. Several utilities have been developed that allow the PSM to analyze raw data and generate predictions concerning experimental evolutionary dynamics. Finally, the ability of the PSM to process other types of evolutionary experiments is discussed.

6.3 Methods and Materials

6.3.1 Experimental procedures

The specific experimental procedures for the VERT experiments used in this study are detailed elsewhere [163, 149]. The first requirement is that the construction of strains with chromosomally integrated genes encoding fluorescent proteins (such as RFP, GFP, YFP). The labeled strains must then be assayed to ensure that expression of fluorescent proteins has no effect on growth rates. Once the neutrality of expression of fluorescent labels has been established, equal proportions of each strain are inoculated into a continuous culture system (chemostats) or batch flasks and sampled daily using a FACS machine to determine the relative frequency of each labeled subpopulation. The complete series of FACS measurements for a VERT experiment (see Figure 6.1) can be interpreted as a quantitative measurement of

population dynamics. These data form the basis of the population state model developed in this work.

Figure 6.1: Population dynamics from an yeast population (KK-Large1-2007) selected for growth in glucose limited media.

6.3.2 Computational procedures

All software was implemented in MATLAB R2010a without additional toolboxes on the Mac OS X 10.6 operating system. Data for model training were annotated and stored as comma-separated value files. Experimental data were also stored in a similar format without annotations.

6.4 Results and Discussion

The first step in developing a model to analyze VERT population history is the examination of the population data to develop a method that can determine

if the observed proportion of population j at time point i represents a statistically significant change compared to point $i-1$. A simple statistical classifier based on data obtained from neutrality (e.g. no adaptive events) experiments is developed to answer this question. This classifier is then utilized to determine emission sequences that represent the statistical significance of changes in population proportion for the entire set of VERT data. A hidden Markov-based model, trained with human-annotated data, is then applied to determine whether or not a subpopulation has undergone an adaptive event based on these emissions. Finally, the error rate, behavior, and possible alternative applications of the model are considered.

6.4.1 Statistical classification of population dynamics data

We sought to analyze the population dynamics that arise during a chemostat evolution experiment. In this type of system, a constant-volume bioreactor is inoculated with several isogenic microbial populations, each marked with a different fluorescent protein (or equivalent unique label) and grown for hundreds of generations in the presence of the desired selective pressure. Adaptive mutants from each labeled subpopulation that arise during the course of the evolution experiment trigger an observable increase in the size of the labeled subpopulation, as shown in Figure 6.1. FACS devices are typically used to track the proportion of each fluorescent strain in the evolving population over time in a series a discrete measurements (typically 1 measurement/day); obtaining continuous data is usually not possible due to experimental and technical limitations. In this case, we utilize population-dynamics data obtained from evolving yeast and *Escherichia coli*.

The population-state model utilizes the rate of population expansion for the j^{th} subpopulation at time point i ($r_{pe,ij}$) as the measured variable to detect adaptive events from FACS data. Since adaptive events are characterized by rapid expansions

in population frequency, we designed our model to examine rates of subpopulation frequency change to automatically detect these events. This property may be calculated directly from FACS data for each time point as follows. First, the proportion of each labeled subpopulation j of J total subpopulations at time i (P_{ji}) is computed from each subpopulation:

$$P_{ji} = \frac{x_{ji}}{x_{j,0} \sum_{j \in J} \vec{x}_i} \tag{6.1}$$

where the summation $\sum_{j \in J} \vec{x}_i$ represents the total FACS reading (counts) at the i^{th} time point. This proportion is divided by $x_{j,0}$ to set $P_{j,0} = 1.0$ for all subpopulations, regardless of their initial proportion of the inoculum. Because the time interval between samples is not necessarily constant, let t_i represent the number of generations that have occurred by the i^{th} sample. Then, $\forall t_i > t_1, r_{pe,ij}$:

$$r_{pe,ij} = \frac{P_{ji} - P_{j,i-1}}{t_i - t_{i-1}} \tag{6.2}$$

The actual time-derivative $\dot{R}_j(t)$ can used in place of R_{ij} if continuous measurements are available, as the former contains much more information concerning the process dynamics and will allow more accurate detection of adaptive events.

Estimates for the mean $r_{pe,ij}$ (subsequently μ_r), representing a collection of slope measurements for one subpopulation, and its standard deviation (σ_r) of the same collection for metastable populations, are needed to draw inferences about which fluctuations in population proportions are significant. Calibration data in the form of neutrality experiments, where adaptive events are unlikely to occur, can be leveraged to obtain these data. In an ideal case, with a perfectly accurate FACS device and populations with exactly equal fitness, $\mu_r = \sigma_r = 0$ over the entire dataset; the

population proportions are fixed. In reality, fluctuations affecting both parameters tend to arise because of jackpot mutations, random stochasticity in the populations, or technical issues that generate noise in the data. The neutrality datasets are therefore used to calculate mean and variance of the slope. The values obtained for these parameters indicated that $\mu_r \in [-0.005, 0.004]$ and $\sigma_r = 0.018$ for 64 neutral measurements. The parameter μ_r also serves as an indicator of population stability and is, as expected, indistinguishable from zero at a 95% confidence level.

Generally, μ_r will be zero for fluorophores that have no fitness effect on their host strains. Some fluorescent proteins, such as *tdTomato*, have been observed to decrease strain fitness (data not shown), resulting in negative values of μ_r. The parameter values used here may therefore be unique to specific fluorophores and should be recomputed for each physically distinct setup.

These properties can be applied to construct a statistical test that will identify when populations begin to expand or contract more rapidly than is expected under a neutral regime. In formal terms, we compare the observed slopes with a random variable $R_{pe,ij}$ drawn from the t-distribution with estimated mean μ_r and standard deviation σ_r. A t-test can be used to ascertain whether there is a significant difference between the observed slope and the mean neutral measurement (alternative hypothesis, Equation 6.4) or if a population is stable (null hypothesis, Equation 6.3). A Gaussian distribution may also be used in place of the t-distribution if desired; however, if the number of samples is small (less than 30), the t-distribution is more appropriate. The statistic $T = \frac{r_{pe,ij} - \mu_r}{\sigma_r / \sqrt{n}}$ is used to determine if the difference between the observed and expected slopes is statistically significant.

$$H_o : r_{pe,ij} - \mu_r = 0 \tag{6.3}$$

$$H_a : r_{pe,ij} - \mu_r \neq 0 \tag{6.4}$$

Each subpopulation of a VERT experiment is analyzed to determine when to reject the null hypothesis. For slopes that are unlikely to be explained by the null hypothesis ($P < \alpha$), the sign of the slope is examined to determine if it represents a population increase (positive slope, P) or a contraction (negative slope, N). Slopes that fail to meet the significance threshold, in either direction, are recorded as zero (Z) slopes. The p-value threshold for significance, selected by empirical observation and based on model performance, was $\alpha = 0.10$ unless otherwise stated. These slope classifications were subsequently used in the population state model described below.

6.4.2 Definition of the population state model

The basic outline of the population-state model (hereafter PSM) exploits the statistical classifier to detect when one subpopulation of labeled cells undergoes expansion so that the initiation and termination of the expansion can be pinpointed. The mutant is assumed to have reached its highest frequency at the latter time point, allowing easier isolation of the desired mutant. The model itself utilizes two hidden states: "N," which indicates that a subpopulation is not undergoing a population expansion, and "A," to indicate that the subpopulation has experienced an adaptive event. Annotated training data from 8 multicolored yeast chemostats were used to calculate state transition probabilities within and between the states ($P_{AA}, P_{NN}, P_{AN}, P_{NA}$), and the emission probabilities of each symbol (Z, N, and P) in the respective states ($e_A(S)$ and $e_N(S)$, where $S \in \{Z, N, P\}$ as defined by the statistical classifier). This process was performed automatically by the model, allowing for the facile incorporation of additional data into the training dataset to improve model accuracy. Training data were not used in any subsequent analyses. Numeric

values for each of these parameters are calculated only from the training data and are shown in Table 6.1. State-transition probabilities are adjusted to account for contiguous positive slopes (C_P) or negative and zero slopes ($C_{!P}$) through the use of an exponential decay penalty function:

$$P_{AN} = P^{\circ}_{AN}(exp(-C_P)) \tag{6.5}$$

$$P_{NA} = P^{\circ}_{NA}(exp(-C_{!P})) \tag{6.6}$$

where P°_{AN} and P°_{NA} represent the nominal value of each state transition probability. Accordingly, $P_{NN} = 1 - P_{NA}$ and $P_{AA} = 1 - P_{AN}$ as well. These contiguous counts are reset to zero when symbols outside the considered set (i.e., Z, N for C_P) are encountered in the data. This modification represents a divergence from the traditional formulation of a hidden Markov model, where the state at position i only depends on position $i-1$. We used this approach to account for the fact that adaptive events, once they occur and survive initial drift, expand in a non-random fashion over time. The exponential decay function represents the decreasing probability of a transition out of an ongoing change in population proportion (i.e. a long adaptive expansion or continual decline); many possible forms for this function exist, but the exponential function seems to correlate well with the observed population dynamics. This formulation allows for the explicit consideration of the current population state in the chemostat and dramatically improves the accuracy of the model.

A total of 19 long-term chemostat experiments for *E. coli* [269], *S. cerevisae* [163], and *C. albicans* [149] were analyzed using the PSM. For a given chemostat experiment k, the emission sequence O_{kj} is generated for each of the j subpopulations using the statistical classifier at significance level $\alpha = 0.10$ (single-tailed). The most likely set

Table 6.1: Population state model parameters

Parameter	Value
State Transition	$P^{\circ}_{AN} = 0.154$, $P^{\circ}_{NA} = 0.079$
Adaptive Emission	$P_N = 0.102$, $P_Z = 0.150$, $P_P = 0.748$
Non-adaptive Emission	$P_N = 0.434$, $P_Z = 0.337$, $P_P = 0.229$

An overview of the Markov parameters used by the population state model. The emission probabilities in the non-adaptive state reflect the symmetry of the slope distribution in the control data and the adaptive emissions are heavily biased towards positive slopes as expected. In addition, the state transition probabilities indicate that entry into and exit from adaptive events is relatively uncommon in the training data.

of hidden states for the j^{th} subpopulation in the k^{th} chemostat (X_{kj}) can then be decoded using the Viterbi algorithm [261] in an iterative fashion:

$$X_{kj} = \{argmax(P_{ll} \cdot e_l(O_{k,i}), P_{lm} \cdot e_m(O_{k,i}))\forall i\} \tag{6.7}$$

where l denotes the previous hidden state and m the alternative state (e.g. $A \rightarrow A\ or\ N$). This process is shown graphically in Figure 6.2. Given that not all populations are expand immediately after chemostat inoculation, it is assumed that all populations are in state N at $i = 0$. In addition, the final predictions for adaptive states are translated back one time point (i.e. $i \rightarrow i - 1$) based on the empirical observation that doing so improves model accuracy. Model validation was accomplished by comparing the predicted hidden state sequences to human annotation of the 19 chemostats and then computing the number of true positives ($\mathrm{A}_{mod} = \mathrm{A}_{ann}$), true negatives ($\mathrm{N}_{mod} = \mathrm{N}_{ann}$), false positives ($\mathrm{A}_{mod} = \mathrm{N}_{ann}$), and false negatives ($\mathrm{N}_{mod} = \mathrm{A}_{ann}$) within the computational predictions. Despite the use of true and false designations, the human annotations may not always be accurate representations of each chemostat population. These error rates can be more accurately interpreted as

the difference between PSM and human annotations.

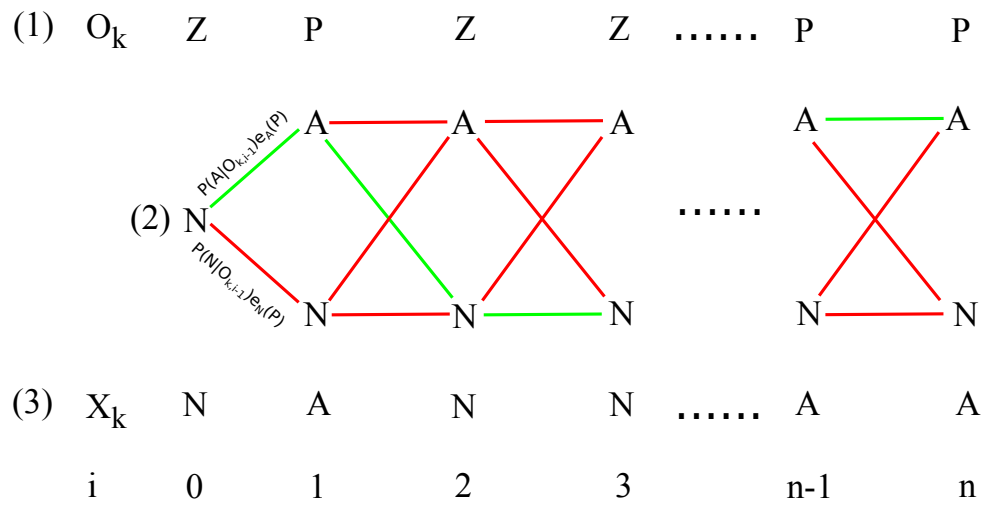

Figure 6.2: Decoding of the hidden Markov states for each labeled subpopulation occurs as follows. (1) The set of emission symbols O_k for a subpopulation is generated from the statistical classifier for all n measurements. (2) The forward Viterbi decoder generates the most likely set of hidden states by choosing the path of maximum likelihood through the system trellis (green lines) based upon the known Markov parameters and O_k. (3) The output set X_k is assembled from these predictions for all observations.

The use of a supervised learning approach, although allowing for relatively straightforward development and training of the PSM, introduces bias into what is considered an adaptive event. This effect, in turn, affects the model parameters computed from the annotated training set. An alternative approach to HMM training involves the use of unsupervised learning, in which the estimated state-transition and emission probabilities are computed automatically using algorithms such as Baum-Welch [32]. In essence, this type of HMM training computes the expected number of state transitions and the emission probabilities (in each state) that best fit the provided emission symbols and updates the model parameters accordingly. This iterative process con-

tinues until the change in HMM performance is below the user threshold. This type of training will be explored in future versions of the population-state model.

6.4.3 Properties of the population state model

Using the procedure outlined previously, the PSM is trained using an annotated dataset from *S. cerevisae* glucose-limited chemostats [163]. Depending on the species studied, length of the evolution, and conditions (mutagenic versus non-mutagenic), different estimates of the Markov parameters given in Table 6.1 may be obtained depending on the dataset used for model training. However, the calculated probabilities seem reasonable in light of the experimental population dynamics. Non-adaptive events typically have slopes that are close to zero ($P > 0.10$), and the remaining events are split evenly between positive and negative slopes ($P < 0.10$). Adaptive events are expected to be weighted towards producing measurements with positive slopes. The behavior of the PSM is affected most by the state-transition properties P°_{AN} and P°_{NA} as these parameters control how quickly the model responds to changes in chemostat dynamics.

In order to quantify the error rate of the model more precisely, the PSM was used to generate hidden state predictions for a collection of chemostat evolution experiments for *E. coli*, *S. cerevisae*, and *Candida albicans*. These predictions were then compared to human annotations. As can be seen in Table 6.3, the model achieves a prediction accuracy of 85% to 93% for the data examined. Discrepancies between the model and the annotated states typically arise from the inability of the statistical classifier to register positive slopes that do not meet the statistical threshold for significance; slow adaptive events (subpopulation growth rate < 0.0025 gen^{-1} at $\alpha = 0.10$) may therefore be missed by the model. Although these events are relatively rare and therefore do not substantially impact the accuracy of the PSM,

slow adaptive events may harbor new lineages or additional mutations relevant to the condition being evaluated. However, even with this limitation, the chemostat properties in Table 6.2 calculated using the PSM are not significantly different from those obtained from human annotation. In addition to these continuous culture systems, the PSM was also able to annotate VERT data obtained during a batch serial transfer experiment (data not shown).

Table 6.2: Analysis of population dynamics

System	AE/gen-color	Rate of PEX†	AE Length (s)
Human Annotation			
C. albicans	0.015 (0.007)	0.0058 (0.016)	3.26 (2.12)
E. coli	0.017 (0.005)	0.0065 (0.009)	1.80 (0.96)
S. cerevisae	0.008 (0.005)	0.005 (0.005)	4.124 (3.47)
Model predictions			
C. albicans	0.016 (0.009)	0.010 (0.015)	3.83 (2.79)
E. coli	0.013 (0.010)	0.005 (0.004)	2.46 (1.62)
S. cerevisae	0.009 (0.005)	0.005 (0.005)	4.33 (3.43)

Properties of adaptive events (AE) are calculated from the human annotated data and the predictions of the PSM to highlight differences between the annotation methods. The average value and standard deviation (in parentheses) are provided for each parameter of interest. There are no statistically significant differences between organisms (i.e. *E. coli* versus yeast) at the $\alpha = 0.05$ level. †PEX: *p*opulation *ex*pansion, is defined as $\Delta P_{ij}/\Delta t$ (t: generations).

6.4.4 Example application: analysis of a yeast chemostat

An example of the PSM prediction is shown for a yeast chemostat (Large1-KK-2007) in Figure 6.1. In this system, three fluorescent strains were competing for access to limited glucose; adaptive events occurred as individual cells acquired mutations that affected the rate of glucose transport into the cell. Upon visual inspection of the raw data, an experienced VERT user would likely conclude that adaptive

Table 6.3: Population state model error analysis

System	Description	$TP_{A=A}$	$TN_{N=N}$	$FP_{A=N}$	$FN_{N=A}$
C. albicans	Fluconazole challenge	0.213	0.598	0.108	0.082
E. coli	Butanol challenge	0.167	0.683	0.043	0.108
S. cerevisae	Glucose limitation	0.216	0.720	0.044	0.020

Error rates were calculated by comparing the set of hidden states generated by the PSM to human annotation and then applying the translation method discussed in the description of the model. Model parameters were calculated with $\alpha = 0.10$ for the statistical classifier. Slow adaptive events in the *E. coli* chemostats account for the increased proportion of false negatives generated by the model.

events (expansions) occurred several times in each subpopulation and that the mutations conferring the greatest fitness advantage occurred in the yellow population. Analyzing these population dynamics using the PSM produces the adaptive event predictions shown in Figure 6.3 as shaded regions within each subpopulation.

Once adaptive events have been identified, adaptive mutants must be isolated from the chemostat population. Preserved population samples stored at -80° C may be regrown in the selective media, plated, and analyzed to determine which clonal isolate contains the adaptive mutation. Since any sample can potentially contain the mutant of interest, an additional tool based on the emission sequence generated by the statistical classifier and the hidden-state data from the PSM was developed to guide sampling efforts to identify the sample with the highest proportion of the adaptive mutants. First, the endpoints of each contiguous series of adaptive events ("A" states) can be identified using the PSM output. Then, for each distinct adaptive event the emission sequence for that subpopulation should be examined until a "N" symbol (statistically significant negative slope) is found at point i. The sampling suggestion is then set to $i-1$ as the time point likely contains the highest proportion of the mutant cells Applying this procedure to the data in Figure 6.1 yields the sampling

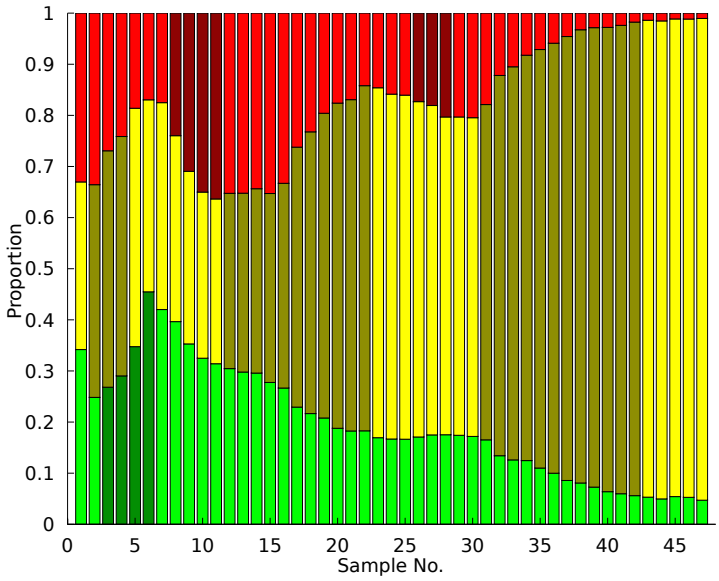

Figure 6.3: Using the experimental dynamics in 6.1 and the PSM, the timing of each adaptive event in the chemostat is calculated and displayed for the user as shaded time points.

predictions highlighted in dark blue in Figure 6.4. The identified sampling points are either immediately adjacent to each adaptive expansion (if followed shortly by another expansion in a different subpopulation) or, in the case of the final, high fitness yellow mutant, some distance away from the calculated adaptive event endpoint. The latter estimate arises from the fact that the yellow subpopulation essentially overran the chemostat environment, so that the optimum sampling point coincided with the final population measurement. Quantitative PCR measurement of the allele frequency in each population supported this sampling scheme [163]. Altogether, these sampling suggestions provide an useful and accurate tool to optimize the VERT experiment and to minimize unnecessary mutant isolation.

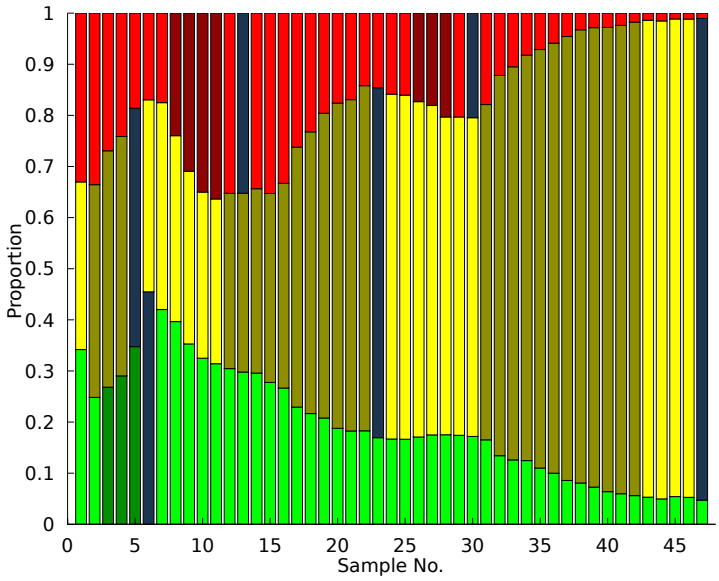

Figure 6.4: Following the identification of adaptive events, estimates of optimal sampling points, as described in the text, are then computed to assist in mutant isolation.

6.4.5 Distribution of adaptive events

In addition to the adaptive events themselves, it is also of interest to determine how these events are distributed between the various evolving subpopulations and whether there are differences in the initial seed populations or fitness effects of the fluorescent labels. If one label has a significant detrimental impact upon fitness, it is unlikely that many adaptive events will be identified in that particular subpopulation. The PSM was utilized to calculate the number of adaptive events, weighted by length of the detected adaptive event, per subpopulation for the entire set of available data (Figure 6.5). A consistent bias toward adaptive events in a particular subpopulation for chemostats seeded from the same initial inoculum may indicate the presence of a beneficial mutant that arose prior to exposure to the selective pressure in question (a

jackpot). A statistical method for identifying this type of biased population dynamics will be developed to investigate this phenomenon in a rigorous manner.

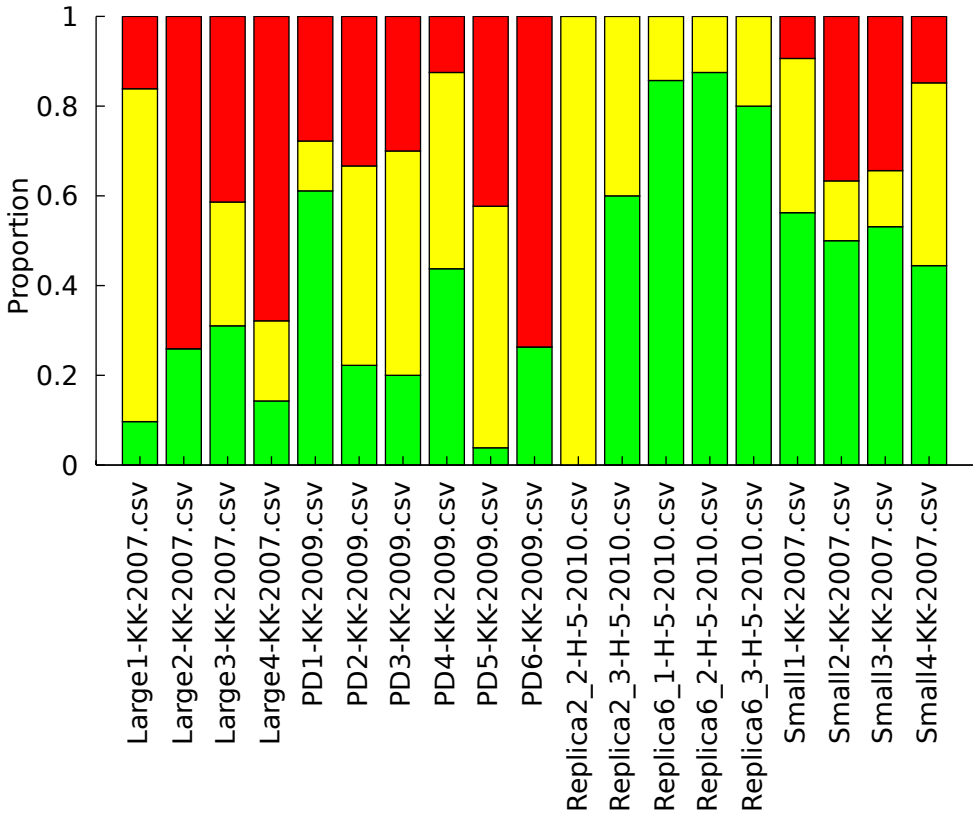

Figure 6.5: The relative proportions of adaptive events in each subpopulation, calculated using the PSM, in the three chemostat systems considered here. The neutrality of expression of the fluorescent protein implies that there should be no consistent bias of adaptive events toward any particular color, and this assumption holds here for all chemostats. Statistically significant differences in the abundance of adaptive events between the labeled populations would imply the presence of jackpot mutants.

6.4.6 Application to other evolution systems

Despite using of the VERT system and chemostat data in developing the PSM, there is no explicit dependence of the PSM on VERT data. Any method that can generate similar population histories over time (e.g., microarray or qPCR methods) can also be integrated into the PSM. The only requirement is that comparable neutrality experiments and annotated experimental data must be generated using the proposed alternative so that the PSM can estimate the required HMM parameters. The current implementation of the PSM automatically calculates all of the necessary parameters, except for μ_r and σ_r, for the new measurements, both of which must be determined by the end-user, as described previously. After this calibration procedure, the PSM should be able to analyze population histories obtained by alternative methods.

Another potential application of the PSM is the construction of an automated system (e.g. *autoVERT*) for the observation and isolation of adaptive mutants. Both batch and continuous culture systems are increasingly automated, and with the addition of an online FACS device, population histories can be collected continuously. The second task of an automated system is to identify when adaptive events occur so that samples of the population can be saved (on solid media or as frozen stocks) for later analysis. Given that the PSM has been shown to be effective in accomplishing this task, it may be possible to adapt the model to construct such a system. Additional work is needed to optimize the PSM for this new type of data forecasting as the model was primarily constructed for retrospective analysis of VERT experiments.

6.5 Conclusions

The population-state model offers the ability to detect adaptive events within fluorescent microbial populations easily and without the need for user intervention

automatically. A variety of VERT experimental properties may also be determined, enabling a quantitative comparison between the evolutionary dynamics of different VERT experiments. Comparison to human analysis of VERT experiments revealed that the PSM produced highly accurate predictions for adaptive events and sampling time points. This algorithm represents an important new tool for the analysis of population dynamics over time and will be integral to any VERT system capable of automatic identification of adaptive mutants.

7. HARNESSING FLUX BALANCE ANALYSIS FOR SIMULATING BIOLOGICAL EVOLUTION*

7.1 Summary

Current evolution simulators rely on artificial fitness landscapes that represent a collection of genotypes generated through random mutations that arise during population growth. Despite the many successes of these tools and other theoretical approaches that make similar approximations, a representative genome that allows for direct linkage between metabolism *in silico* and *in vivo* as the basis for simulated evolution would provide a powerful new method to predict mutations of interest prior to experimental evolution. As a first step towards addressing this issue, we have developed a genetic algorithm that represents individual cells using metabolic flux models, providing a direct linkage between genotype and metabolic behavior of all mutants during *in silico* evolution. After demonstrating the functionality of the simulator with an example model, gene-expression data for *Escherichia coli* growing under n-butanol or isobutanol stress is utilized to identify possible compensatory, non-trivial metabolic changes that occurred during adaptive evolution. Previously collected experimental data were found to agree with the simulator predictions, demonstrating the effectiveness of this approach for strain engineering purposes. These predictions can be used to develop methods to increase biofuel tolerance.

*"Harnessing Flux Balance Analysis for Simulating Biological Evolution," by J. Winkler and K.C. Kao (2013), under review at Artificial Life (MIT Press).

7.2 Introduction

Evolutionary engineering is a powerful tool for strain improvement, especially for complex phenotypes that are poorly understood [244, 289, 334, 333]. Many attempts have been made to model the evolutionary process in various contexts using computational simulators [341, 144, 194, 19] in order to explore properties of adaptation in different types of populations. These models can also be applied alongside theoretical approaches [314, 300, 111, 234, 233]. Algorithm for genetic optimization are of particular interest because of their use of defined "genomes" that represent all possible allelic combinations individuals may possess [22, 115] and their success in approaching complex optimization problems in a variety of fields. In essence, these algorithms apply random mutation to a collection of individuals with genomes of specified length and initial allele frequencies to generate new mutant individuals; the resulting population is then subjected to a selection to identify the fittest individuals. Additional rounds of mutation and selection can be performed until the desired level of performance is achieved. Avida [231] employs this approach to model digital lifeforms, and it has been quite useful for the study of evolution *in silico*. Despite its utility as a tool for simulating evolution, applying genetic programming to biological evolution is difficult because of challenges in developing biologically relevant simulated genomes, and the difficulty in linking mutant genotypes to fitness in a realistic manner. If both problems can be addressed, investigators will be able to evolve organisms computationally and then extract biologically meaningful data regarding adaptation mechanisms and the population dynamics during the evolution experiment [102].

Metabolic flux models, by virtue of representing both genomes and metabolic networks in a single package [257, 265, 229], are a promising choice to model the

underlying genotype for an evolution simulator. These models, available for many organisms [232, 235, 208, 294, 138], consist of a reaction network defined by a stoichiometric matrix, reaction constraints derived from thermodynamics and experimental measurements, and genes that produce required enzymes. The utility of these models is demonstrated by the wide range of metabolic engineering tools built upon constraint-based modeling to design strains that overproduce various biochemicals of interest [275, 73, 274, 263, 204, 45]. Constraint-based models have also been shown to predict optimal evolutionary outcomes for adaptation to different carbon sources [113, 114], although those studies did not attempt to simulate population dynamics during experimental evolution. Estimates of optimal growth rates, obtained by maximizing the biomass formation objective function in each model, have therefore been shown to be quantitative [101, 325].

Strain fitness, in the form of biomass flux or growth rate, can be determined directly from the structure of the flux model and informed reaction constraints without recourse to arbitrary fitness assignments that are inherent to synthetic fitness landscapes. Metabolic perturbations due to unfavorable environmental conditions can also be captured by incorporating experimental flux, gene expression, or proteomic data into the models [195, 362, 294, 161, 186, 34, 355, 150, 264] so that the inhibitory selective pressures, such as biofuels or lignocellulosic biomass inhibitors, can be incorporated into the simulation of evolution. Assuming that the necessary computational resources are available, the use of flux models to represent genomes as the basis for evolution simulators allows for the determination of the genetic mechanisms of adaptation directly, while also observing changes in population composition over time to facilitate analysis of evolution experiments.

This work represents a synthetic genome that incorporates a validated metabolic flux model into a genetic algorithm to simulate laboratory evolution. In this manner,

a well-defined fitness landscape, based directly on the physiology of the organism of interest, is coupled with random mutation of reaction constraints to identify optimal genotypes and track their population dynamics in several types of simulated bioreactors. After introducing the basic algorithms for mutation, recombination, and fitness assignment, a simple scenario describing adaptation to a novel environment under batch and continuous cultivation is described using the simulator. Evolution under n-butanol or isobutanol stress is then explored by incorporating gene-expression data into the metabolic flux model.

7.3 Methods and Materials

7.3.1 Software packages and data sources

All evolution simulations were performed in Matlab R2011b (Mathworks) on a Mac OS X 10.6.8. The COBRA toolbox [292] with the default glpk solver and *E. coli* core model without regulation [291] were used for all FBA calculations except when noted. Otherwise, the latest *E. coli* (iJO1366) [235] full metabolic models were used to generate estimates of growth rate. Any valid COBRA model should be useable with the simulator, however, although the use of complex models containing large numbers of reactions (thousands or more) may be computationally prohibitive to assess if the mutation rate is high. Microarray data used in this study were obtained from GEO records of Brynildsen and Liao [44] and Rutherford et al. [283] (195 minute time point); only genes with statistically significant changes in expression for wild-type *E. coli* under biofuel stress were used as data inputs. When necessary, the rank product method [39] implemented in TM4-MeV microarray analysis software was used to identify significant differences in expression [285]. The GiMME algorithm [27] was used to integrate high-throughput gene expression data into metabolic flux models. However, instead of removing repressed genes and their dependent reactions

from the model entirely, as is normally done by GiMME, the flux constraints of deactivated reactions was multiplied by $\alpha = 0.01$ or $\alpha = 0.001$ to represent stress-related flux perturbations. Genes that are expressed at 50% or less of the wild-type level under each condition of interest are considered repressed in this case as an initial approximation. Other expression thresholds may also be used depending on the effect of the stressor in question. It may be possible to identify the optimal threshold using metabolic flux and gene expression data obtained under the same stress conditions, but we did not attempt to make such a determination here.

7.3.2 Genotype definition

COBRA models are used here as a simplified representation of an actual genome. We specifically utilized individual reaction constraints, which limit forward and reverse flux through particular reactions, as targets for mutation and recombination. Individual genotypes were represented as the collections of mutations that affect reaction boundaries. The genome size corresponding to an individual genotype is therefore equal to the number of reactions included in the model of interest multiplied by two, with upper and lower bounds for each reaction. It was assumed that mutations affect all reaction constraints with equal probability (*i.e.*, no mutational hotspots) and that all genotypes can be stably recombined into a single strain using a technique such as conjugation [230] or protoplast fusion [308, 126]. Optimizations that preferentially targeted reactions with altered flux or positive reduced costs [257] after constraint with gene expression data were also utilized where noted. Following the generation of a new genotype by mutation or recombination, the growth rate of the new mutant was calculated by applying the specified mutations to the ancestral flux model and then calculating the optimal rate of biomass formation. Finding the flux distribution that optimizes biomass formation for each mutated COBRA model

typically requires computing time on the order of milliseconds; simulator runs required approximately 10 minutes and 3 hours for the core and full *E. coli* models, respectively.

7.3.3 Evolution simulation

Both a turbidostat with a proportional-derivative (PD) controller and batch serial-transfer cultivation systems were implemented for evolution experiments. The initial population in each simulation was assumed to be genetically homogenous except for N_{label} neutral markers. Growth in the former continuous system obeys the standard mass balance for a turbidostat, and it is assumed that nutrients are present in excess to facilitate continued exponential growth; otherwise, the functional dependence of growth rate on nutrient and inhibitor concentrations can be readily included by expanding the mass balance. In the case of batch simulations, the logistic growth equation was used with carrying capacity N_{pop} to incorporate the changes in growth rate seen over the course of typical batch cultivations. Batch serial transfers are performed when the population is greater than 95% of N_{pop}. As mentioned previously, the growth rate of each genotype was determined by optimizing the corresponding flux model, incorporating any mutations associated with the genotype, for the formation of biomass. Growth of each genotype was simulated using the forward Euler method with a time-step of $\Delta t = 10$ minutes, although other, more accurate numerical methods can be used as well. Simulations were run for approximately 2260 steps with 200,000 mutants generated over the simulation and three initial labels in the population. The results of each simulation were stored in a data structure containing the genotype, growth rate, label (if desired), and population frequency of every mutant over the course of the simulation. These data were used for subsequent analysis.

Mutant genotypes were generated by random mutation of reactions in the parental genotype, except targets forbidden by the user. Three types of mutations were modeled: boundary null (flux boundary set to zero) occurred in 34% of mutations, boundary decrease (flux constraint multiplied by $r \sim U(0,1)$) occurred in 33% of mutations, and boundary increase (flux constraint multiplied by $r \sim U(1,1.5)$) for the remainder. Certain simulations with iJO1366 utilized boundary increase limits up to 500% ($r \sim U(1,5)$). Other distributions of mutational effect can also be used if sufficient data is available or if theoretical distributions of fitness are of interest [104]. If one reaction constraint (lower or upper) was set to zero, then the non-zero constraint was altered; if neither constraint is zero, then the mutation target was chosen at random. Recombinant genotypes are generated by randomly selecting two parents from within the evolving population and then constructing a new genotype that contains the unique mutations from both parents. In both cases, the resulting models are solved such that they maximize biomass formation using a modified version of the *optimizeCbModel* function in the COBRA toolbox that does not perform standard checking of model integrity.

7.4 Results and Discussion

7.4.1 Simulator properties

The simulator used in this work combines a genetic algorithm that perturbs the "genome" (set of reaction flux constraints) to enhance biomass production with simulated growth in turbidostat or batch culture. The number of alleles in the genome for a flux model is simply twice the number of reactions, as each reaction has both an upper and lower constraint on its flux. Random mutations during the simulation create new genotypes by altering the flux constraints on any reaction in the network, except those excluded by the user, and may increase, reduce, or eliminate flux

through the targeted reaction; the optimal biomass formation rate is then computed for each newborn metabolic model, given its new constraints, to determine the effect of the new mutation. Successful genotypes have high rates of biomass formation, allowing them to increase in frequency relative to the background population over the simulated time-course. Mathematical descriptions of these processes and cultivation systems are given in the *Materials and Methods* section.

Selection of model parameters, such as the population size, simulation length, and mutation rate are key parameters that control the results of each simulation. Population size affects the time required for population sweeps by new, fitter mutants, and it should be significantly larger than the number of mutants produced per time step. Similarly, the optimal timespan of each simulation depends on the expected fitness effect of adaptive mutations. The mutation rate, actually representing a fixed number of mutations generated per simulation, is perhaps the most important parameter because it will dramatically influence the observed population dynamics and the rate of simulated adaption. In this case, the frequency of simulated mutation is not perfectly analogous to a biological mutation rate, since the mutational space for a metabolic flux model is far smaller than that for a real organism. For the following examples, approximately two thousand mutations per reaction were generated in the course of each simulation, using the 95-reaction core *E. coli* model. This mutational depth is probably not feasible for more complete models because the sheer number of mutant genotypes that must be stored; in these cases, only a few mutants are generated per reaction, reducing the discovery of potential adaptive mutations.

The most practical method for overcoming this problem is to exclude all reactions from perturbation if they are inactive in the base model or represent trivial adaptations that are not of interest, such as mutations that affect carbon or nitrogen uptake fluxes, and to otherwise simplify the model of interest as much as possible.

Several means of limiting the sequence space (number of reactions) open to mutation are discussed below. Potential adaptive mutations can also be identified with a smaller model that represents the core metabolism of an organism and then verified using the model of the full organism through direct modification. However, given that this work focuses on predicting population dynamics during adaptive laboratory evolution, these optimizations are not employed unless otherwise stated.

7.4.2 Evolution in non-optimal media

Some of the most common, though sometimes incidental, cases of adaptive evolution appear when an organism grows in a new medium for which it is ill-adapted. Over time, such a strain acquires mutations that improve its ability to assimilate the specific cocktail of nutrients in the medium, to tolerate the cultivation temperature, and possibly to alter the relative amount of time spent in different growth phases. This process can be simulated by generating a maladapted version of the *E. coli* core metabolic network through random perturbation of reaction flux constraints and then evolving the perturbed network until it achieves the original rate of biomass production of the unaltered core model. An alternative approach to model construction is to assess the gene expression changes triggered by non-optimal media and then constrained the metabolic network according to these data (see below). Here, however, we elect to use a randomly altered model, because of its simplicity as a test case to demonstrate the functionality of the simulator. Unless otherwise mentioned, the perturbed model is identical to the *E. coli* core model in network connectivity except that 25 reactions are constrained to 1/100th their original flux in both directions, resulting in a fitness of 25% relative to the parent network.

The population dynamics during evolution of the perturbed model in continuous (turbidostat) and batch culture, with and without recombination between genotypes,

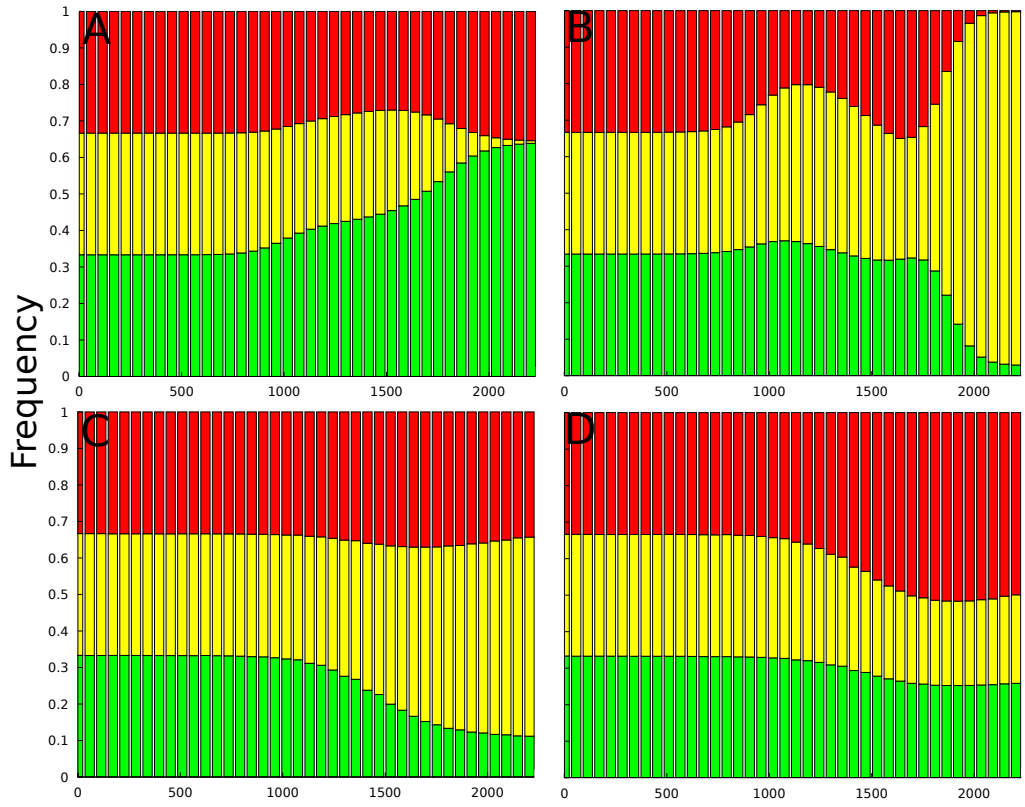

Figure 7.1: A) Turbidostat without recombination, B) Turbidostat with recombination, C) Serial batch without recombination, and D) Serial batch with recombination. All simulations were run for 2260 iterations (equivalent to 15 days of experimental evolution).

are shown in Figure 7.1. All reactions except the requirement for ATP maintenance can be altered via random mutation or recombination (if applicable). We use the same representation as VERT [345, 269, 163] to represent the population dynamics as changes in label frequency within the population. Overall, the speed and frequency of adaptive events (increases in label proportion) are relatively low, indicating that there are only a few mutations that have a positive effect on growth rate. A detailed analysis of the acquisition of mutations can be performed using the exported simulation data if it is desirable to determine the exact order of appearance, effect,

and frequency of each mutation of interest. This approach enables investigators to identify single or multiple mutations that collectively contribute to improved strain fitness. This capability is especially important when simulating evolution on complex fitness landscapes with multiple possible routes of adaptation [178].

Another useful feature of the simulator is its ability to generate pedigrees (Figure 7.2), a directed network that shows the genotype history for all mutants that reach at least a minimum frequency in the population (typically $> 1\%$). The pedigrees for the populations simulated without recombination tended to demonstrate clonal interference, whereas enabling recombination resulted in more complex patterns of inheritance and a stable coexistence of multiple genotypes descended from differentially labeled strains, matching theoretical expectations [346, 111, 248, 72].

7.4.3 Evolution under stress conditions

Although a randomly perturbed model is adequate to test the capabilities of the simulator, predictions concerning adaptive paths under real selective pressures must be made for the simulator to be truly useful. One possible approach to modeling the effects of arbitrary inhibitors on cellular metabolism is to utilize high-throughput gene expression or proteomic data to modify the network architecture of the metabolic model using actual expression levels under the condition of interest. Explicit flux measurements can be used for similar purposes, but they are more difficult and costly to obtain and will therefore be available less frequently.

In this study, the GiMME algorithm [27], which removes reactions from the metabolic network based on input data while accounting for model consistency, was selected for integrating gene expression data into metabolic models because of its simplicity and because of it is already implemented in the COBRA toolbox. In this case, instead of being constrained to zero flux, reactions are reduced to a fraction of

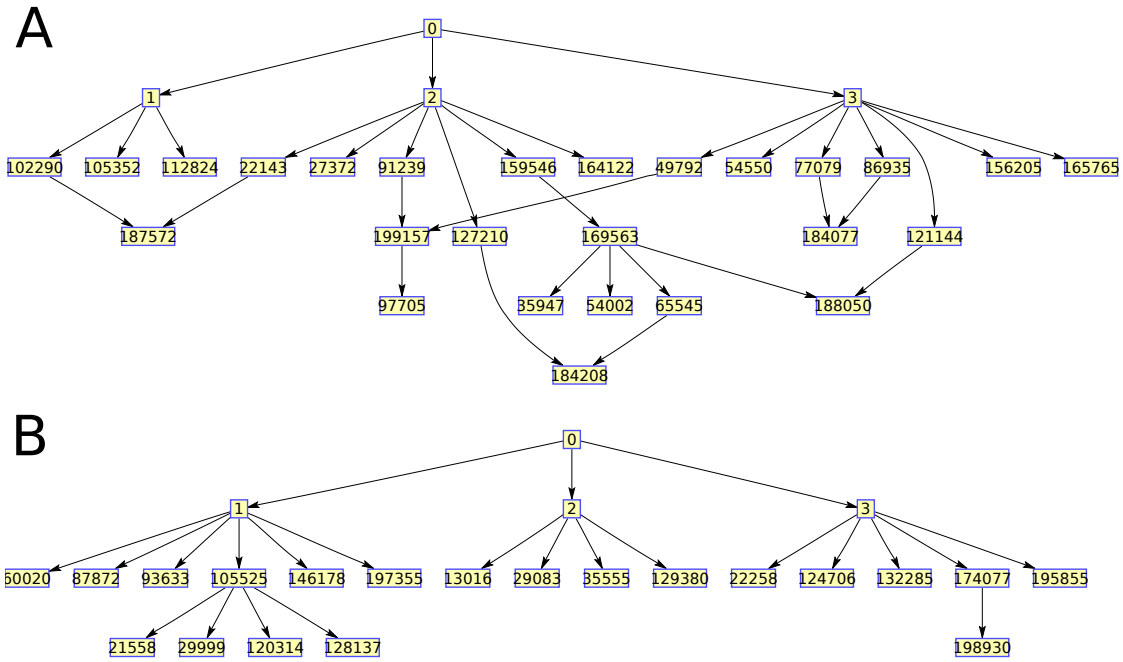

Figure 7.2: Representative inheritance patterns in turbidostats undergoing evolution A) with and B) without recombination. Directed edges connect parent to child genotypes. All genotypes are descended from a single initial genotype, and each simulation is seeded with a specified number (three, in this case) of labeled but otherwise isogenic strains. Note that clonal interference is attenuated when recombination is permitted between different mutant lineages.

their original value by multiplying the upper and lower flux constraints with the parameter $\alpha \in (0, 1)$. Fluxes were constrained to 1% ($\alpha = 0.01$) unless otherwise noted. This change to the typical GiMME implementation allows for mutations to restore flux throughout the network in order to compensate for stress-induced metabolic perturbations.

Given the interest in understanding the complex phenotypes underlying biofuel tolerance, we selected two gene-expression datasets under isobutanol [44] and n-butanol [283] selection for integration into the *E. coli* core model and subsequent simulated evolution. The full *E. coli* model (iJO1366) [235] was utilized as noted for purposes of comparison. Based on their respective gene expression datasets, GiMME constrained 13 reactions in the core model under isobutanol stress (44% fitness relative to the core model), and 8 reactions under n-butanol selection (56% fitness relative to the core model). In both cases, glucose uptake and ATP maintenance reactions were removed from the set of mutable reactions to ensure that mutants were adapting to biofuel-related network changes. Growth of the altered metabolic models in a turbidostat revealed consistent patterns of mutation among independent simulations for both biofuels, as discussed below. Population dynamics over the course of the simulations, summarized in Figure 7.3, populations under isobutanol adapted gradually, whereas simulations incorporating n-butanol stress revealed mutations that conferred large benefits that resulted in rapid adaption. This result implies that for two initially isogenic populations propagated under isobutanol or n-butanol stress, the n-butanol populations will be more heterogenous if mutations for optimizing the metabolic network are the primary drivers of phenotype improvement.

In the presence of toxic levels of n-butanol (approximately 0.8% (v/v) in minimal media), GiMME predicted that only a small number of core metabolic reactions would be inactivated, including succinate and 2-oxoglutarate dehydrogenases,

Figure 7.3: Representative population dynamics during evolution of the *E. coli* models after being constrained with n-butanol and isobutanol gene expression data. A turbidostat without recombination was used for all simulations. A) Core model, isobutanol stress; and B) core model, n-butanol stress.

formate metabolic reactions (transport and pyruvate-formate lysase), and ATP synthase. Our initial expectations were that random mutation in the simulator would act to reverse these deactivations, but the key route for improving fitness appears to be the overproduction of cytochrome oxidase *bd*, perhaps as a result of a redox imbalance. This reaction was not in the set initially constrained based upon the n-butanol expression data.

Enhanced flux through the cytochrome *bd* reaction was observed in the fittest mutant produced in every simulation, and it was an extremely common mutation within the evolving population as a whole. The operon encoding this multicomponent enzyme (*cyo*) is, indeed, significantly upregulated following exposure to n-butanol [283], confirming that this modification plays an important role in ameliorating the physiological impact of n-butanol stress. A recent analysis of n-butanol tolerance identified *nuoI* overexpression as improving n-butanol tolerance as well [267]. However, mutations affecting cytochrome expression were not detected in a recent study

of n-butanol tolerant *E. coli* mutants [269], suggesting that other routes of adaptation might be more important *in vivo*.

A range of other mutations occurred in the fittest mutants within the population in conjunction with cytochrome *bd* overexpression, including increased glutaminase and glutamate transport activity, overexpression of citric acid cycle proteins such as isocitrate dehydrogenase, and malate dehydrogenase, and several proteins in the pentose phosphate shunt or glycolytic pathways, including transketolase, ribulose-5-phosphaste 3-epimerase, and phosphofructokinase. Overall, these mutations appear to counteract the reduction of metabolic activity, especially carbon assimilation and amino acid synthesis, typically observed during biofuel stress [109, 99].

There also appears to be a single common mutation that significantly improves metabolic performance under 1% (v/v) isobutanol stress. Most of the genes repressed under isobutanol stress, as expected, are involved in the citric acid cycle. In contrast to the observed adaptations to n-butanol, a strong selection for enhanced flux through the NADH dehydrogenase reaction was observed in all simulations, along with occasional enhancements of fluxes through the ATP synthase, isocitrate lysase, and glucose 6-phosphate dehydrogenase reactions. All of these reactions were initially constrained based on the available data for gene expression, indicating that a reversal of these effects is necessary to restore strain fitness. Mutations affecting the expression of NADH dehydrogenase were not observed in either of the recent long-term evolutionary studies for isobutanol tolerance of *E. coli* [14, 217], although increased transcription of the *nuo* operon (encoding NADH dehydrogenase-I) was induced by n-butanol stress [283]. The same mutations improve fitness of the full iJO1366 model after it is constrained with the same *E. coli* gene-expression data obtained under isobutanol stress. One possible explanation for these results is that altered activity or deletion of global regulators under these conditions results in

increased expression of NDH-1 or NDH-2 [44, 35].

Because the core *E. coli* model contains only a small fraction of the entire *E. coli* metabolic network, we also used the full metabolic model (iJO1366) in order to ascertain whether our results were an artifact of the core model. Unexpectedly, the full model predicted no drop in growth rate when it was constrained according to changes in gene expression seen under 0.8% n-butanol stress. This finding contrasts with previous measurements in *E. coli* [175], and could indicate that the effects of butanol toxicity on the metabolic network are overweighted by the core model because of its lack of redundancy.

Constraining iJ10366 with the isobutanol expression data also resulted in only a 8% decrease in the expected growth rate when the same parameter values that were used to constrain the core model. Setting $\alpha = 0.001$ and re-applying the gene expression constraints with GiMME revealed larger fitness decrements of 40% for isobutanol and 3.4% for n-butanol stresses. The discrepancy in fitness costs at different values of α is most likely a result of the increased flexibility of iJO1366 network compared to the comparatively simple core model. Thus, users will be advised to test a range of α values in more complex flux models in order to constrain the models adequately. Attempts to simulate evolution under the isobutanol stress condition with $\alpha = 0.001$ failed to identify adaptive mutants because of the significantly expanded mutation space associated with the 2261 reactions in the iJO1366 network. The distribution of fitness effects caused by mutation may also have been inadequate to generate adaptive mutants with the initial α flux constraint. Using an equivalent number of mutations per reaction with iJO1366, to explore potential tolerance-conferring genotypes more fully proved to be computationally infeasible.

7.4.4 Reducing the target sequence space

Given the large number of reactions in the full *E. coli* model, reducing the size of the sequence space available for mutation is required to increase discovery of potential adaptive genotypes. In contrast to experimental adaptive evolution, where mutations occur randomly throughout the genome without dependence on local sequences, the use of a flux-based model in the simulation affords additional control over which reactions can be perturbed by random mutation. A simple, yet logically appealing, heuristic is to constrain reaction flux limits with GiMME and then target all these reaction for random mutation. This reaction set includes both the reactions constrained based on the gene-expression data as well as any other reactions perturbed elsewhere in the metabolic network; approximately 40% of the reactions in the core model were excluded from the sequence space by this approach when n-butanol or isobutanol gene expression data is used to constrain the flux distribution. The mutational outcomes observed under n-butanol or isobutanol selection are identical to those previously found when any reaction constraint in the flux model genome could be altered, indicating that this approach does not bias the *in silico* evolution results except by avoiding neutral mutations. The practical consequence of this approximation would be to increase the frequency of adaptive events.

The reduction in sequence space is significantly larger with the iJO1366 model, with 79% and 74% of the modeled reactions exhibiting no change in flux when the full metabolic network is constrained by the n-butanol and isobutanol gene expression data. Although this set of 542 to 672 reactions is large compared to the 95 reaction core model, the number of mutations per reaction increased by a factor of 3.8 to 4.8, representing a significant improvement over usage of the entire set of reactions for the simulated genome. Applying this approximation and simulating evolution with

iJO1366 constrained with isobutanol gene expression data identified increase flux though cytochrome *bo3* and NADH dehydrogenase reactions as improving fitness, as one mutation or the other is present in the fittest strains arising in each simulation. However, the maximum fitness increase in these mutants is quite small, representing only a 0.77-0.78% increase in growth averaged over all simulations.

Given the tight constraints placed on the networks by setting $\alpha = 0.001$, it may be that mutations cannot relax these constraints sufficiently to permit more than marginal improvements in biomass production. To test this hypothesis, the potential upper bound constraints were increased from a maximum of 50% to 500% to facilitate more rapid adaptation. With this change, we found that mutations that resulted in increased flux though the malate dehydrogenase (*mdh*) and malate:quinone oxidoreductase (*mqo*) reactions became more frequent, in addition to overexpression of NADH dehydrogenase. Surprisingly, few mutations increased flux through the cytochrome *bo3* reaction, although manually increasing the maximum permitted flux through that reaction does improve biomass production in the isobutanol-constrained iJO1366 model. These mutations likely compensate for disruptions in quinone metabolism in iJO1366, but the identification of *mdh* as a potentially important locus for isobutanol tolerance contradicts the consistent pattern of *mdh* deactivation observed by Minty et al. [217] in response to continual isobutanol stress. The reason for this apparent discrepancy between the evolved full model and experimental evolution data is unclear, but it indicates that additional validation of the genotype-phenotype predictions generated from the simulator is needed. However, we expect that continued refinement of the *E. coli* metabolic model and the development of more holistic simulations of cell behavior will also increase the accuracy of the simulator by accounting for a greater range of environmental and metabolic perturbations.

In order to determine the set of reactions that could potentially improve biomass production in the isobutanol-constrained full *E. coli* model, we examined the set of reactions that were limiting biomass formation by identifying those with positive reduced variable costs [257]. Reactions with negative or zero reduced variable costs are produced in excess or are exactly optimal, respectively, for biomass production. The set of reactions with positive reduced variable costs therefore represents the sequence subspace in which adaptive mutations can occur. Although the simulator can easily target these reactions for mutation during simulated evolution, there is little correspondence between such simulations and evolution in the laboratory because the vast majority of mutations are deleterious or neutral with regard to fitness [104]. This approximation would effectively introduce mutational hot spots in the flux model genome that are not observed in actual organisms. An approach more explicitly focused on designing tolerant strains using a genetic algorithm could make use of this approximation to enhance selection for improved genotypes. For example, only a small number of reactions, including those previously mentioned along with formate dehydrogenase, catalase, 2-oxogluterate dehydrogenase, and three succinate antiporters for succinate-aspartate, succinate-fumarate, and succinate-D-tartrate have positive reduced variable costs in the constrained iJO1366 system. This restriction means that the simulations are effectively searching for only a small number of potentially beneficial genotypes. Because our goal in this study was to demonstrate a more realistic evolution simulator rather than to design an explicit strain improvement tool, we did not explore this type of optimization further.

7.4.5 Targets for further development

Despite the success of the simulator in identifying biologically relevant adaptive mutations, there are limitations inherent to the use of flux-based models that pre-

vent the discovery of all potential adaptive mechanisms. One problem is that *E. coli* biomass composition varies with growth rate [256] a factor that is not accounted for during these evolution simulations. Although this effect could alter the relative fitness of competing genotypes, and therefore perturb the dynamics during simulated evolution, it is likely to be important only for phenotypes for which large and sudden changes in fitness are possible (e.g., antibiotic resistance). It should have less effect on the gradual adaptation observed with most complex inhibitors such as biofuels, furans, or weak acids. Another issue is that constraint based models focus on simulating metabolism rather than the totality of cellular activities, unlike the holistic cell model recently developed for *Mycoplasma genitalium* [164]. Thus, mutations are limited to known metabolic activities. An example of a potentially important non-metabolic adaptation route is the enhancement in biofuel tolerance through the overexpression of protein chaperones [361]. This effect could not have been predicted solely using flux-base models. In addition, metabolic reconstructions, though they make accurate quantitative predictions about cellular growth phenotypes, are not yet complete even for the best-characterized model organisms, like *E. coli*. As flux-based models undergo further refinement and more whole-cell models become available, the obstacles to accurate prediction of evolutionary dynamics that encompass both metabolic and non-metabolic adaptations should diminish. Isotopic measurements of metabolic flux under different stress conditions would also help constrain and improve the abilities of the simulation to predict population dynamics.

7.5 Conclusions

Combining metabolic flux models, which provide well-defined fitness landscapes that accurately represent real organisms, with an efficient genetic algorithm to explore the mutational space of interest, represents a new approach to predicting geno-

typic outcomes in laboratory evolution. This approach was used here to identify key metabolic changes that occur under isobutanol and n-butanol stresses by incorporating *E. coli* gene-expression data into the core *E. coli* metabolic model. In the case of isobutanol tolerance, core model mutations were found to improve growth of the full *E. coli* model constrained using gene-expression data for isobutanol stress, indicating that the core model is suitable as a base genome for evolution. Previous results from the literature support the adaptation mechanisms identified in these simulations, confirming that the identified mutations are biologically relevant. Taken together, the results indicate that the simulator can be applied effectively and quickly to generate predictions of evolutionary dynamics and potential adaptive mutations that may arise during laboratory evolution.

The approach used in this study should be generally suitable for the analysis of any organism for which a metabolic model has been developed. Because high-throughput analyses of transcript or protein abundance are routinely generated for purposes of strain characterization, their incorporation into the metabolic model of interest should normally be practical. These altered models can also be used as inputs to the evolution simulator to predict adaptive mutations that are likely to occur during evolution, providing a computational method to complement experimental tools that are currently used to identify and engineer complex phenotypes like biofuel tolerance. Despite these advantages, the simulator has several limitations that stem from its reliance on metabolic-flux models. Foremost of these is the inherent restriction that adaptive mutations can only be identified for components of the encoded metabolic network. It is likely that this constraint relax over time as more complete and validated whole-cell models are developed to facilitate analysis of the complex population structures of the organisms of interest.

8. CONCLUSIONS AND RECOMMENDATIONS

8.1 Summary of Results

This work demonstrates the possible advantages of an evolution system for *E. coli* that exploits reciprocal genetic exchange to generate increased diversity in the genome under selection. Adaptation of an Hfr strain that can undergo reciprocal genetic exchanges agrees closely with theoretical predictions of the benefit of sexual recombination on complex (polygeneic) and simple (single allele) fitness landscapes. This mating system was tested for its ability to improve more industrially relevant tolerance phenotypes such as osmotic tolerance. The latter study significantly expanded knowledge of osmotic tolerance determinants in *E. coli* and identified the *nag* operon and genes that regulate cell wall synthesis as being key sources of mutations to improve tolerance. Work using conjugation to combine tolerance phenotypes from tolerant donors into multi-tolerant recombinants is ongoing and holds significant promise for conducting evolution and adaptation in parallel.

An analysis of *L. brevis* transcriptional responses to n-butanol and ferulic acid was also conducted. Hundreds of genes that were differentially expressed after chemical insult were identified, and provide new insight into the stress responses of *L. brevis*. Steps were also taken toward the development of a method for screening *L. brevis* genes with their native promoters by verifying that the housekeeping sigma factor from two *Lactobacillus* species can complement temperature-sensitive defects in a*E. coli* σ^{70} mutant. Two new computational tools were developed to simplify studies of population dynamics during evolution and to simulate possible evolutionary outcomes using a flux balance analysis of fitness landscapes.

8.2 Recommendations for Further Research

One of the main benefits of a sexual evolution system, in addition to a faster rate of adaptation across complex fitness landscapes, is the theoretical ability of donors to transfer selected phenotypes to arbitrary recipients. Using this approach, evolution experiments could be executed in parallel, followed by mixing of characterized isolates or populations to generate recombinants with multiple desirable phenotypes. Efforts in this area have not been successful to date, but several key areas that require improvement have been identified. These involve adjusting the level of selection pressure to apply during recombinant selection, choosing time and format (in liquid or on a solid surface) of the mating between evolved strains, and selecting against spontaneous parental mutants during the procedure. Given the potential reduction in time, effort, and cost that this type of parallel evolution offers, this area promises significant rewards for the biotechnology field.

Because of the decreasing cost of deep population sequencing, it may soon be feasible to compare genetic dynamics between sexual and asexual populations quantitatively without restricting the investigation to a small number of loci. A study of this type would directly test theoretical assumptions concerning allele fate in each type of system and could provide a new way of testing theories of population genetics. The large number of variables in each evolution experiment (type of challenge, selection strength, cultivation system, and others) can be manipulated to see how *in situ* recombination affects population dynamics in these regimes. The results would represent a significant contribution to evolutionary biology and could also be used to optimize evolutionary engineering protocols to maximize the retention of novel adaptive mutants.

The work presented utilized *E. coli* not only to test the mating system but as

the ultimate host for any novel genetic determinants for selected phenotypes. As our knowledge of non-model organisms improves, developing a version of the mating system that can work in other prokaryotic hosts is of interest. Although the F plasmid is transfers DNA extremely efficiently under typical cultivation conditions, other conjugative plasmids in the R family have a much broader host range, although with a lower transfer efficiency, than F. There are substantial challenges to developing an R-based mating system, such as the requirement for mating to take place on solid surfaces, the lack of genetic tools to integrate the conjugative plasmid into host genomes, and a more limited understanding of the biology of R plasmid compared to the F plasmid. Overcoming these obstacles will require combining fundamental molecular biology research on R, the development of new techniques to generate and characterize Hfr strains in a variety of species, and extensive evolutionary engineering to verify that newly sexual bacterial species derive the expected adaptive benefit of increased genetic recombination. However, the resulting mating systems would be an invaluable tool for an evolutionary engineering, especially to expedite adaptation of slower-growing organisms that possess metabolic capabilities of high industrial interest.

REFERENCES

[1] M. Achtman, N. Kennedy, and R. Skurray. Cell–cell interactions in conjugating *Escherichia coli*: role of *traT* protein in surface exclusion. *Proceedings of the National Academy of Sciences*, 74(11):5104, 1977.

[2] S. Ackerman, A. R. Kermany, and D. A. Hickey. Finite populations, finite resources, and the evolutionary maintenance of genetic recombination. *Journal of Heredity*, 101(suppl 1):S135–S141, 2010.

[3] J. Adrio and A. Demain. Genetic improvement of processes yielding microbial products. *FEMS Microbiology Reviews*, 30(2):187–214, 2006.

[4] E. Alejandre-Duran, D. de Andres-Cara, and A. Porras. Simulation of the effect of population size on the evolution of the recombination fraction. *Journal of Theoretical Biology*, 128(4):399–405, 1987.

[5] M. P. Almario, L. H. Reyes, and K. C. Kao. Evolutionary engineering of *Saccharomyces cerevisiae* for enhanced tolerance to hydrolysates of lignocellulosic biomass. *Biotechnology and Bioengineering*, 110(10):2616–2623, 2013.

[6] H. Alper and G. Stephanopoulos. Global transcription machinery engineering: a new approach for improving cellular phenotype. *Metabolic Engineering*, 9(3):258–267, 2007.

[7] S. Altschul, W. Gish, W. Miller, E. Myers, and D. Lipman. Basic local alignment search tool. *Journal of Molecular Biology*, 215(3):403–410, 1990.

[8] W. Anderson and D. Akin. Structural and chemical properties of grass lignocelluloses related to conversion for biofuels. *Journal of Industrial Microbiology and Biotechnology*, 35(5):355–366, 2008.

[9] L. Aniskovitch and H. Winkler. *Rickettsia prowazekii* sigma factor σ^{73} can be overexpressed in *Escherichia coli* and promotes RNA polymerase binding and transcription. *Microbiology*, 142(4):901, 1996.

[10] K. Anthony, C. Sherburne, R. Sherburne, and L. Frost. The role of the pilus in recipient cell recognition during bacterial conjugation mediated by F-like plasmids. *Molecular Microbiology*, 13(6):939–953, 1994.

[11] F. Archibald and I. Fridovich. Manganese, superoxide dismutase, and oxygen tolerance in some lactic acid bacteria. *Journal of Bacteriology*, 146(3):928, 1981.

[12] S. J. Arnold, M. E. Pfrender, and A. G. Jones. The adaptive landscape as a conceptual bridge between micro-and macroevolution. *Genetica*, 112(1):9–32, 2001.

[13] M. Asayama, H. Suzuki, A. Sato, T. Aida, K. Tanaka, H. Takahashi, and M. Shirai. The *rpoD1* gene product is a principal sigma factor of RNA polymerase in *Microcystis aeruginosa* K-81. *Journal of Biochemistry*, 120(4):752, 1996.

[14] S. Atsumi, T. Wu, I. Machado, W. Huang, P. Chen, M. Pellegrini, and J. Liao. Evolution, genomic analysis, and reconstruction of isobutanol tolerance in *Escherichia coli*. *Molecular Systems Biology*, 6(1):449, 2010.

[15] S. Atsumi, A. F. Cann, M. R. Connor, C. R. Shen, K. M. Smith, M. P. Brynildsen, K. J. Chou, T. Hanai, and J. C. Liao. Metabolic engineering of *Escherichia coli* for 1-butanol production. *Metabolic Engineering*, 10(6):305–311, 2008.

[16] E. Baake and W. Gabriel. Biological evolution through mutation, selection, and drift: an introductory review. *Annual Reviews of Computational Physics*,

7:203–264, 2000.

[17] T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner, and H. Mori. Construction of *Escherichia coli K-12* in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, 2(1):2008, 2006.

[18] A. Babić, A. Lindner, M. Vulić, E. Stewart, and M. Radman. Direct visualization of horizontal gene transfer. *Science*, 319(5869):1533, 2008.

[19] D. Bachtrog and I. Gordo. Adaptive evolution of asexual populations under muller's ratchet. *Evolution*, 58(7):1403–1413, 2007.

[20] S. Baer, H. Blaschek, and T. Smith. Effect of butanol challenge and temperature on lipid composition and membrane fluidity of butanol-tolerant *Clostridium acetobutylicum*. *Applied and Environmental Microbiology*, 53(12):2854, 1987.

[21] P. Bajwa, D. Pinel, V. Martin, J. Trevors, and H. Lee. Strain improvement of the pentose-fermenting yeast *Pichia stipitis* by genome shuffling. *Journal of Microbiological Methods*, 81(2):179–186, 2010.

[22] W. Banzhaf, J. Koza, C. Ryan, L. Spector, and C. Jacob. Genetic programming. *Intelligent Systems and their Applications, IEEE*, 15(3):74–84, 2000.

[23] M. M. Barnhart, J. Lynem, and M. R. Chapman. GlcNAc-6P levels modulate the expression of Curli fibers by *Escherichia coli*. *Journal of Bacteriology*, 188(14):5212–5219, 2006.

[24] J. E. Barrick, D. S. Yu, S. H. Yoon, H. Jeong, T. K. Oh, D. Schneider, R. E. Lenski, and J. F. Kim. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature*, 461(7268):1243–1247, 2009.

[25] A. Barron, G. May, E. Bremer, and M. Villarejo. Regulation of envelope protein composition during adaptation to osmotic stress in *Escherichia coli*. *Journal of Bacteriology*, 167(2):433–438, 1986.

[26] T. Basso, S. de Kok, M. Dario, J. do Espirito-Santo, G. Müller, P. Schlölg, C. Silva, A. Tonso, J. Daran, A. Gombert, A. van Maris, J. Pronk, and B. Stambuk. Engineering topology and kinetics of sucrose metabolism in *Saccharomyces cerevisiae* for improved ethanol yield. *Metabolic Engineering*, 13(6):694–703, 2011.

[27] S. Becker and B. Palsson. Context-specific metabolic networks are consistent with experiments. *PLoS Computational Biology*, 4(5):e1000082, 2008.

[28] L. Becks and A. F. Agrawal. The evolution of sex is favoured during adaptation to new environments. *PLoS Biology*, 10(5):e1001317, 2012.

[29] D. Beier, G. Spohn, R. Rappuoli, and V. Scarlato. Functional analysis of the *Helicobacter pylori* principal sigma subunit of RNA polymerase reveals that the spacer region is important for efficient transcription. *Molecular Microbiology*, 30(1):121–134, 1998.

[30] G. Bell. The masterpiece of nature: the evolution and genetics of sexuality. *Croom Helm Applied Biology Series*, 1982.

[31] O. Berezina, N. Zakharova, A. Brandt, S. Yarotsky, W. Schwarz, and V. Zverlov. Reconstructing the clostridial n-butanol metabolic pathway in *Lactobacillus brevis*. *Applied Microbiology and Biotechnology*, 87(2):635–646, 2010.

[32] J. Bilmes. A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. *International*

Computer Science Institute, 4:126, 1998.

[33] J. A. Birdsell and C. Wills. The evolutionary origin and maintenance of sexual recombination: a review of contemporary models. In R. J. Macintyre and M. T. Clegg, editors, *Evolutionary Biology*, pages 27–138. Springer US, New York, 2003.

[34] A. Blazier and J. Papin. Integration of expression data in genome-scale metabolic network reconstructions. *Frontiers in Physiology*, 3(299), 2012.

[35] J. Bongaerts, S. Zoske, U. Weidner, and G. Linden. Transcriptional regulation of the proton translocating NADH dehydrogenase (*nuoA-N*) of *Escherichia coli* by electron acceptors, electron donors and gene regulators. *Molecular Microbiology*, 16(3):521–534, 2006.

[36] J. Borden and E. Papoutsakis. Dynamics of genomic-library enrichment and identification of solvent tolerance genes for *Clostridium acetobutylicum*. *Applied and Environmental Microbiology*, 73(9):3061, 2007.

[37] N. Bouck and E. Adelberg. Mechanism of action of nalidixic acid on conjugating bacteria. *Journal of Bacteriology*, 102(3):688, 1970.

[38] L. Bowles and W. Ellefson. Effects of butanol on *Clostridium acetobutylicum*. *Applied and Environmental Microbiology*, 50(5):1165, 1985.

[39] R. Breitling, P. Armengaud, A. Amtmann, and P. Herzyk. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Letters*, 573(1):83–92, 2004.

[40] M. A. Brockhurst, N. Colegrave, and D. E. Rozen. Next-generation sequencing as a tool to study microbial evolution. *Molecular Ecology*, 20(5):972–980, 2011.

[41] E. Brodie, T. DeSantis, D. Joyner, S. Baek, J. Larsen, G. Andersen, T. Hazen, P. Richardson, D. Herman, T. Tokunaga, et al. Application of a high-density oligonucleotide microarray approach to study bacterial population dynamics during uranium reduction and reoxidation. *Applied and Environmental Microbiology*, 72(9):6288, 2006.

[42] J. Brown, T. Ross, T. McMeekin, and P. Nichols. Acid habituation of *Escherichia coli* and the potential role of cyclopropane fatty acids in low pH tolerance. *International Journal of Food Microbiology*, 37(2-3):163–173, 1997.

[43] D. Browning and S. Busby. The regulation of bacterial transcription initiation. *Nature Reviews Microbiology*, 2(1):57–65, 2004.

[44] M. Brynildsen and J. Liao. An integrated network approach identifies the isobutanol response network of *Escherichia coli*. *Molecular Systems Biology*, 5(1):277, 2009.

[45] A. P. Burgard, P. Pharkya, and C. D. Maranas. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnology and Bioengineering*, 84(6):647–657, 2003.

[46] S. Busby. More pieces in the promoter jigsaw: recognition of -10 regions by alternative sigma factors. *Molecular Microbiology*, 72(4):809–811, 2009.

[47] A. Cadiere, A. Ortiz-Julien, C. Camarasa, and S. Dequin. Evolutionary engineered *Saccharomyces cerevisiae* wine yeast strains with increased in vivo flux through the pentose phosphate pathway. *Metabolic Engineering*, 13(3):263–271, 2011.

[48] F. Campos, J. Couto, A. Figueiredo, I. Tóth, A. Rangel, and T. Hogg. Cell membrane damage induced by phenolic acids on wine lactic acid bacteria. *In-*

ternational Journal of Food Microbiology, 135(2):144–151, 2009.

[49] F. Campos, J. Couto, and T. Hogg. Influence of phenolic acids on growth and inactivation of *Oenococcus oeni* and *Lactobacillus hilgardii*. *Journal of Applied Microbiology*, 94(2):167–174, 2003.

[50] P. Capy, G. Gasperi, C. Biémont, and C. Bazin. Stress and transposable elements: co-evolution or useful parasites? *Heredity*, 85(2):101–106, 2000.

[51] C. P. Cartwright, M. J. Beavan, F. M. Ruby, S. M. De Morais, and A. H. Rose. Ethanol dissipates the proton-motive force across the plasma membrane of *Saccharomyces cerevisiae*. *Journal of General Microbiology*, 132(2):369–377, 1986.

[52] M. Castanie-Cornet, T. Penfound, D. Smith, J. Elliott, and J. Foster. Control of acid resistance in *Escherichia coli*. *Journal of Bacteriology*, 181(11):3525, 1999.

[53] L. Cavalli, J. Lederberg, and E. Lederberg. An infective factor controlling sex compatibility in *Bacterium coli*. *Journal of General Microbiology*, 8(1):89, 1953.

[54] S. T. Chambers, C. M. Kunin, D. Miller, and A. Hamada. Dimethylthetin can substitute for glycine betaine as an osmoprotectant molecule for *Escherichia coli*. *Journal of Bacteriology*, 169(10):4845–4847, 1987.

[55] D.-E. Chang, H.-C. Jung, J.-S. Rhee, and J.-G. Pan. Homofermentative Production of D- or L-Lactate in Metabolically Engineered *Escherichia coli* RR1. *Applied and Environmental Microbiology*, 65(4):1384–1389, 1999.

[56] Y. Chang and J. Cronan. Membrane cyclopropane fatty acid content is a major factor in acid resistance of *Escherichia coli*. *Molecular Microbiology*,

33(2):249–259, 1999.

[57] E. L. Chant and D. K. Summers. Indole signalling contributes to the stable maintenance of *Escherichia coli* multicopy plasmids. *Molecular Microbiology*, 63(1):35–43, 2007.

[58] L. Chao. Evolution of sex in RNA viruses. *Journal of Theoretical Biology*, 133(1):99–112, 1988.

[59] K. Cheah and R. Skurray. The F plasmid carries an IS3 insertion within *finO*. *Journal of General Microbiology*, 132(12):3269, 1986.

[60] W. Chen, K. Ohmiya, and S. Shimizu. Intergeneric protoplast fusion between *Fusobacterium varium* and *Enterococcus faecium* for enhancing dehydrodivanillin degradation. *Applied and Environmental Microbiology*, 53(3):542–548, 1987.

[61] X. Chen, P. Wei, L. Fan, D. Yang, X. Zhu, W. Shen, Z. Xu, and P. Cen. Generation of high-yield rapamycin-producing strains through protoplasts-related techniques. *Applied Microbiology and Biotechnology*, 83(3):507–512, 2009.

[62] D. Cho, Y. Lee, Y. Um, B. Sang, and Y. Kim. Detoxification of model phenolic compounds in lignocellulosic hydrolysates with peroxidase for butanol production from *Clostridium beijerinckii*. *Applied Microbiology and Biotechnology*, 83(6):1035–1043, 2009.

[63] H. Chong, H. Geng, H. Zhang, H. Song, L. Huang, and R. Jiang. Enhancing *E. coli* isobutanol tolerance through engineering its global transcription factor cAMP receptor protein (CRP). *Biotechnology and Bioengineering*, 111(4):700–708, 2014.

[64] H. Chou, A. Hsia, D. Mooney, and P. Schnable. Picky: oligo microarray design for large genomes. *Bioinformatics*, 20(17):2893, 2004.

[65] H. H. Chou and J. D. Keasling. Programming adaptive control to evolve increased metabolite production. *Nature Communications*, 4:2595, 2013.

[66] B. Chung, S. Selvarasu, C. Andrea, J. Ryu, H. Lee, J. Ahn, H. Lee, and D. Lee. Genome-scale metabolic reconstruction and in silico analysis of methylotrophic yeast *Pichia pastoris* for strain improvement. *Microbial Cell Factories*, 9(50):1–15, 2010.

[67] J. Clomburg and R. Gonzalez. Biofuel production in *Escherichia coli*: the role of metabolic engineering and synthetic biology. *Applied Microbiology and Biotechnology*, 86(2):419–434, 2010.

[68] P. Cocconcelli, L. Morelli, M. Vescovo, and V. Bottazzi. Intergeneric protoplast fusion in lactic acid bacteria. *FEMS Microbiology Letters*, 35(2):211–214, 1986.

[69] N. Colegrave. Sex releases the speed limit on evolution. *Nature*, 420(6916):664–666, 2002.

[70] N. Colegrave. The evolutionary success of sex. *EMBO Reports*, 13(9):774–778, 2012.

[71] T. M. Conrad, M. Frazier, A. R. Joyce, B.-K. Cho, E. M. Knight, N. E. Lewis, R. Landick, and B. Ø. Palsson. Rna polymerase mutants found through adaptive evolution reprogram *Escherichia coli* for optimal growth in minimal media. *Proceedings of the National Academy of Sciences*, 107(47):20500–20505, 2010.

[72] T. F. Cooper. Recombination speeds adaptation by reducing competition between beneficial mutations in populations of *Escherichia coli*. *PLoS Biology*, 5(9):e225, 2007.

[73] C. Cotten and J. L. Reed. Constraint-based strain design using continuous modifications (CosMos) of flux bounds finds new strategies for metabolic engineering. *Biotechnology Journal*, 8(5):595–604, 2013.

[74] J. F. Crow and M. Kimura. Evolution in sexual and asexual populations. *American Naturalist*, pages 439–450, 1965.

[75] C. Cueva, M. Moreno-Arribas, P. Martín-Álvarez, G. Bills, M. Vicente, A. Basilio, C. Rivas, T. Requena, J. Rodríguez, and B. Bartolomé. Antimicrobial activity of phenolic acids against commensal, probiotic and pathogenic bacteria. *Research in Microbiology*, 161(5):372–382, 2010.

[76] J. Cullen, M. Hinkhouse, M. Grady, A. Gaut, J. Liu, Y. Zhang, C. Darby Weydert, F. Domann, and L. Oberley. Dicumarol inhibition of NADPH: quinone oxidoreductase induces growth inhibition of pancreatic cancer via a superoxide-mediated mechanism. *Cancer Research*, 63(17):5513, 2003.

[77] A. J. Cullum, A. F. Bennett, and R. E. Lenski. Evolutionary adaptation to temperature. IX. Preadaptation to novel stressful environments of *Escherichia coli* adapted to high temperature. *Evolution*, 55(11):2194–2202, 2001.

[78] M. Dai and S. Copley. Genome shuffling improves degradation of the anthropogenic pesticide pentachlorophenol by *Sphingobium chlorophenolicum* ATCC 39723. *Applied and Environmental Microbiology*, 70(4):2391, 2004.

[79] M. Dai, S. Ziesman, T. Ratcliffe, R. T. Gill, and S. D. Copley. Visualization of protoplast fusion and quantitation of recombination in fused protoplasts of auxotrophic strains of *Escherichia coli*. *Metabolic Engineering*, 7(1):45–52, 2005.

[80] C. Darcan, R. Özkanca, and K. Flint. Survival of nonspecific porin-deficient mutants of *Escherichia coli* in black sea water. *Letters in Applied Microbiology*, 37(5):380–385, 2003.

[81] C. Darwin. The genetical theory of natural selection. *The Eugenics Review*, 22(2):127, 1930.

[82] K. Datsenko and B. Wanner. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences*, 97(12):6640, 2000.

[83] K. A. Datsenko and B. L. Wanner. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences*, 97(12):6640–6645, 2000.

[84] M. Davis, J. Solbiati, and J. Cronan. Overproduction of acetyl-CoA carboxylase activity increases the rate of fatty acid biosynthesis in *Escherichia coli*. *Journal of Biological Chemistry*, 275(37):28593, 2000.

[85] M. De Angelis and M. Gobbetti. Environmental stress responses in *Lactobacillus*: a review. *Proteomics*, 4(1):106–122, 2004.

[86] J. De Visser and D. Rozen. Clonal interference and the periodic selection of new beneficial mutations in *Escherichia coli*. *Genetics*, 172(4):2093, 2006.

[87] J. De Visser and S. Elena. The evolution of sex: empirical insights into the roles of epistasis and drift. *Nature Reviews Genetics*, 8(2):139–150, 2007.

[88] G. Degrassi, P. Polverino de Laureto, and C. Bruschi. Purification and characterization of ferulate and p-coumarate decarboxylase from *Bacillus pumilus*. *Applied and Environmental Microbiology*, 61(1):326, 1995.

[89] G. Dennis Jr, B. T. Sherman, D. A. Hosack, J. Yang, W. Gao, H. C. Lane, and R. A. Lempicki. DAVID: database for annotation, visualization, and integrated discovery. *Genome Biology*, 4(9):1–11, 2003.

[90] M. Desai, D. Fisher, and A. Murray. The speed of evolution and maintenance of variation in asexual populations. *Current Biology*, 17(5):385–394, 2007.

[91] P. Di Martino, R. Fursy, L. Bret, B. Sundararaju, and R. Phillips. Indole can act as an extracellular signal to regulate biofilm formation of *Escherichia coli* and other indole-producing bacteria. *Canadian Journal of Microbiology*, 49(7):443–449, 2003.

[92] X. Didelot and M. C. Maiden. Impact of recombination on bacterial evolution. *Trends in Microbiology*, 18(7):315–322, 2010.

[93] F. Diez-Gonzalez and Y. Karaibrahimoglu. Comparison of the glutamate-, arginine-and lysine-dependent acid resistance systems in *Escherichia coli* O157: H7. *Journal of Applied Microbiology*, 96(6):1237–1244, 2004.

[94] R. Dobson, V. Gray, and K. Rumbold. Microbial utilization of crude glycerol for the production of value-added products. *Journal of Industrial Microbiology and Biotechnology*, 39(2):217–226, 2012.

[95] J. Domka, J. Lee, and T. K. Wood. YliH (BssR) and YceP (BssS) regulate *Escherichia coli* K-12 biofilm formation by influencing cell signaling. *Applied and Environmental Microbiology*, 72(4):2449–2459, 2006.

[96] M. Dragosits and D. Mattanovich. Adaptive laboratory evolution–principles and applications for biotechnology. *Microbial Cell Factories*, 12(1):64, 2013.

[97] M. Dragosits, V. Mozhayskiy, S. Quinones-Soto, J. Park, and I. Tagkopoulos. Evolutionary potential, cross-stress behavior and the genetic basis of acquired

stress resistance in *Escherichia coli*. *Molecular Systems Biology*, 9(1):643, 2013.

[98] S. Drghici, P. Khatri, R. Martins, G. Ostermeier, and S. Krawetz. Global functional profiling of gene expression. *Genomics*, 81(2):98–104, 2003.

[99] M. J. Dunlop, Z. Y. Dossani, H. L. Szmidt, H. C. Chu, T. S. Lee, J. D. Keasling, M. Z. Hadi, and A. Mukhopadhyay. Engineering microbial biofuel tolerance and export using efflux pumps. *Molecular Systems Biology*, 7(1):487, 2011.

[100] P. Durre. Biobutanol: an attractive biofuel. *Biotechnology Journal*, 2(12):1525–1534, 2007.

[101] J. Edwards, R. Ibarra, and B. Ø. Palsson. In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nature Biotechnology*, 19(2):125–130, 2001.

[102] S. F. Elena and R. E. Lenski. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature Reviews Genetics*, 4(6):457–469, 2003.

[103] M. Ettayebi, S. Prasad, and E. Morgan. Chloramphenicol-erythromycin resistance mutations in a 23S rRNA gene of *Escherichia coli*. *Journal of Bacteriology*, 162(2):551–557, 1985.

[104] A. Eyre-Walker and P. Keightley. The distribution of fitness effects of new mutations. *Nature Reviews Genetics*, 8(8):610–618, 2007.

[105] T. Ezeji, C. Milne, N. Price, and H. Blaschek. Achievements and perspectives to overcome the poor solvent resistance in acetone and butanol-producing microorganisms. *Applied Microbiology and Biotechnology*, 85(6):1697–1712, 2010.

[106] T. Ezeji, N. Qureshi, and H. Blaschek. Butanol production from agricultural residues: Impact of degradation products on *Clostridium beijerinckii* growth

and butanol fermentation. *Biotechnology and Bioengineering*, 97(6):1460–1469, 2007.

[107] J. Felsenstein. The evolutionary advantage of recombination. *Genetics*, 78(2):737–756, 1974.

[108] M. Fernández-Sandoval, G. Huerta-Beristain, B. Trujillo-Martinez, P. Bustos, V. González, F. Bolivar, G. Gosset, and A. Martinez. Laboratory metabolic evolution improves acetate tolerance and growth on acetate of ethanologenic *Escherichia coli* under non-aerated conditions in glucose-mineral medium. *Applied Microbiology and Biotechnology*, 96(5):1291–1300, 2012.

[109] C. Fischer, D. Klein-Marcuschamer, and G. Stephanopoulos. Selection and optimization of microbial hosts for biofuels production. *Metabolic Engineering*, 10(6):295–304, 2008.

[110] D. S. Fisher. Asexual evolution waves: fluctuations and universality. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(01):P01011, 2013.

[111] C. A. Fogle, J. L. Nagle, and M. M. Desai. Clonal interference, multiple mutations and adaptation in large asexual populations. *Genetics*, 180(4):2163–2173, 2008.

[112] S. S. Fong, A. R. Joyce, and B. Ø. Palsson. Parallel adaptive evolution cultures of *Escherichia coli* lead to convergent growth phenotypes with different gene expression states. *Genome Research*, 15(10):1365–1372, 2005.

[113] S. S. Fong, J. Y. Marciniak, and B. Ø. Palsson. Description and interpretation of adaptive evolution of *Escherichia coli* K-12 MG1655 by using a genome-scale in silico metabolic model. *Journal of Bacteriology*, 185(21):6400–6408, 2003.

[114] S. S. Fong and B. Ø. Palsson. Metabolic gene–deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. *Nature Genetics*, 36(10):1056–1058, 2004.

[115] S. Forrest. Genetic algorithms-principles of natural selection applied to computation. *Science*, 261(5123):872–878, 1993.

[116] V. Francois, A. Conter, and J. Louarn. Properties of new *Escherichia coli* Hfr strains constructed by integration of pSC101-derived conjugative plasmids. *Journal of Bacteriology*, 172(3):1436, 1990.

[117] L. Frost, K. Ippen-Ihler, and R. Skurray. Analysis of the sequence and gene products of the transfer region of the F sex factor. *Microbiology and Molecular Biology Reviews*, 58(2):162, 1994.

[118] L. Frost and J. Manchak. F-phenocopies: characterization of expression of the F transfer region in stationary phase. *Microbiology*, 144(9):2579, 1998.

[119] M. Gac, T. F. Cooper, S. Cruveiller, C. Médigue, and D. Schneider. Evolutionary history and genetic parallelism affect correlated responses to evolution. *Molecular Ecology*, 22(12):3292–3303, 2013.

[120] M. Garcillán-Barcia and F. de la Cruz. Why is entry exclusion an essential feature of conjugative plasmids? *Plasmid*, 60(1):1–18, 2008.

[121] P. Gerrish and R. Lenski. The fate of competing beneficial mutations in an asexual population. *Genetica*, 102:127–144, 1998.

[122] H. Girgis, A. Hottes, and S. Tavazoie. Genetic architecture of intrinsic antibiotic susceptibility. *PLoS One*, 4(5):e5629, 2009.

[123] Y. Goh and T. Klaenhammer. Genomic features of *Lactobacillus* species. *Frontiers in Bioscience*, 14:1362–1386, 2009.

[124] D. Gokhale, E. Soo-Han, V. Srinivasan, and D. N. Deobagkar. Transfer of DNA coding for cellulases from Cellulomonas species to *Bacillus subtilis* by protoplast fusion. *Biotechnology Letters*, 6(10):627–632, 1984.

[125] R. Gold, M. Meagher, R. Hutkins, and T. Conway. Ethanol tolerance and carbohydrate metabolism in *Lactobacilli*. *Journal of Industrial Microbiology and Biotechnology*, 10(1):45–54, 1992.

[126] J. Gong, H. Zheng, Z. Wu, T. Chen, and X. Zhao. Genome shuffling: Pprogress and applications for phenotype improvement. *Biotechnology Advances*, 27(6):996–1005, 2009.

[127] H. Goodarzi, B. D. Bennett, S. Amini, M. L. Reaves, A. K. Hottes, J. D. Rabinowitz, and S. Tavazoie. Regulatory and metabolic rewiring during laboratory evolution of ethanol tolerance in *E. coli*. *Molecular Systems Biology*, 6(1):378, 2010.

[128] J. C. Gray and M. R. Goddard. Sex enhances adaptation by unlinking beneficial from detrimental mutations in experimental yeast populations. *BMC Evolutionary Biology*, 12(1):43, 2012.

[129] D. Greig, R. H. Borts, and E. J. Louis. The effect of sex on adaptation to high temperature in heterozygous and homozygous yeast. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 265(1400):1017–1023, 1998.

[130] C. Gross, C. Chan, A. Dombroski, T. Gruber, M. Sharp, J. Tupy, and B. Young. The functional and regulatory roles of sigma factors in transcription. In *Cold Spring Harbor Symposia on Quantitative Biology*, volume 63, page 141. Cold Spring Harbor Laboratory Press, 1998.

[131] A. Grossman, Y. Zhou, C. Gross, J. Heilig, G. Christie, and R. Calendar. Mutations in the *rpoH* (*htpR*) gene of *Escherichia coli* K-12 phenotypically suppress a temperature-sensitive mutant defective in the sigma 70 subunit of RNA polymerase. *Journal of Bacteriology*, 161(3):939, 1985.

[132] T. Gruber and C. Gross. Multiple Sigma Subunits and the Partitioning of Bacterial Transcription Space. *Annual Reviews in Microbiology*, 57(1):441–466, 2003.

[133] W. Guo, W. Jia, Y. Li, and S. Chen. Performances of *Lactobacillus brevis* for producing lactic acid from hydrolysate of lignocellulosics. *Applied Biochemistry and Biotechnology*, 161(1):124–136, 2010.

[134] B. Hall, C. Ma, P. Liang, and K. Singh. Fluctuation AnaLysis CalculatOR: a web tool for the determination of mutation rate using Luria–Delbrück fluctuation analysis. *Bioinformatics*, 25(12):1564–1565, 2009.

[135] T. H. Han, J.-H. Lee, M. H. Cho, T. K. Wood, and J. Lee. Environmental factors affecting indole production in *Escherichia coli*. *Research in Microbiology*, 162(2):108–116, 2011.

[136] T. Harada and Y. Mino. Some properties of p-coumarate decarboxylase from *Cladosporium phlei*. *Canadian Journal of Microbiology*, 22(9):1258, 1976.

[137] J. Harris, J. Heilig, I. Martinez, R. Calendar, and L. Isaksson. Temperature-sensitive *Escherichia coli* mutant producing a temperature-sensitive sigma subunit of DNA-dependent RNA polymerase. *Proceedings of the National Academy of Sciences*, 75(12):6177, 1978.

[138] B. Heavner, K. Smallbone, B. Barker, P. Mendes, and L. Walker. Yeast 5–an expanded reconstruction of the *Saccharomyces cerevisiae* metabolic network.

BMC Systems Biology, 6(1):55, 2012.

[139] D. Heer and U. Sauer. Identification of furfural as a key toxin in lignocellulosic hydrolysates and evolution of a tolerant yeast strain. *Microbial Biotechnology*, 1(6):497–506, 2008.

[140] J. M. Heffernan and L. M. Wahl. The effects of genetic drift in experimental evolution. *Theoretical Population Biology*, 62(4):349–356, 2002.

[141] J. Heimann. The extracytoplasmic function (ECF) sigma factors. *Advances in Microbial Physiology*, 46:47–110, 2002.

[142] C. Herring, A. Raghunathan, C. Honisch, T. Applebee, A. Joyce, T. Albert, F. Blattner, D. van den Boom, C. Cantor, and B. Palsson. Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. *Nature*, 38(12):1406–1412, 2006.

[143] M. D. Herron and M. Doebeli. Parallel evolutionary dynamics of adaptive diversification in *Escherichia coli*. *PLoS Biology*, 11(2):e1001490, 2013.

[144] A. Hintze and C. Adami. Evolution of complex modular biological networks. *PLoS Computational Biology*, 4(2):e23, 2008.

[145] E. C. Hobbs, X. Yin, B. J. Paul, J. L. Astarita, and G. Storz. Conserved small protein associates with the multidrug efflux pump AcrB and differentially affects antibiotic resistance. *Proceedings of the National Academy of Sciences*, 109(41):16696–16701, 2012.

[146] R. Holmes and R. Russell. Mutations affecting amino sugar metabolism in *Escherichia coli* K-12. *Journal of Bacteriology*, 111(1):290, 1972.

[147] M. Horsburgh, S. Wharton, M. Karavolos, and S. Foster. Manganese: elemental defence for a life with oxygen. *Trends in Microbiology*, 10(11):496–501, 2002.

[148] R. Hotchkiss and M. Gabor. Biparental products of bacterial protoplast fusion showing unequal parental chromosome expression. *Proceedings of the National Academy of Sciences*, 77(6):3553, 1980.

[149] M. Huang, M. McClellan, J. Berman, and K. Kao. Evolutionary dynamics of *Candida albicans* during in vitro evolution. *Eukaryotic Cell*, 10(11):1413–1421, 2011.

[150] D. Hyduke, N. Lewis, and B. Ø. Palsson. Analysis of omics data with genome-scale models of metabolism. *Molecular BioSystems*, 9(2):167–174, 2013.

[151] R. Ibarra, J. Edwards, and B. Palsson. *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. *Nature*, 420(6912):186–189, 2002.

[152] L. Ingram. Adaptation of membrane lipids to alcohols. *Journal of Bacteriology*, 125(2):670, 1976.

[153] L. Ingram. Microbial tolerance to alcohols: role of the cell membrane. *Trends in Biotechnology*, 4(2):40–44, 1986.

[154] M. Inui, M. Suda, S. Kimura, K. Yasuda, H. Suzuki, H. Toda, S. Yamamoto, S. Okino, N. Suzuki, and H. Yukawa. Expression of *Clostridium acetobutylicum* butanol synthetic genes in *Escherichia coli*. *Applied Microbiology and Biotechnology*, 77(6):1305–1316, 2008.

[155] K. Ippen-Ihler and E. Minkley Jr. The conjugation system of f, the fertility factor of *Escherichia coli*. *Annual Review of Genetics*, 20(1):593–624, 1986.

[156] L. Isaksson, S. Skold, J. Skjoldebrand, and R. Takata. A procedure for isolation of spontaneous mutants with temperature sensitive synthesis of RNA and/or protein. *Molecular and General Genetics*, 156(3):233–237, 1977.

[157] F. Ishino, W. Park, S. Tomioka, S. Tamaki, I. Takase, K. Kunugita, H. Matsuzawa, S. Asoh, T. Ohta, and B. Spratt. Peptidoglycan synthetic activities in membranes of *Escherichia coli* caused by overproduction of penicillin-binding protein 2 and RodA protein. *Journal of Biological Chemistry*, 261(15):7024–7031, 1986.

[158] S. Isken and J. de Bont. Bacteria tolerant to organic solvents. *Extremophiles*, 2(3):229–238, 1998.

[159] T. Ito, Y. Nakashimada, K. Senba, T. Matsui, and N. Nishio. Hydrogen and ethanol production from glycerol-containing wastes discharged after biodiesel manufacturing process. *Journal of Bioscience and Bioengineering*, 100(3):260–265, 2005.

[160] K. Jantama, M. Haupt, S. A. Svoronos, X. Zhang, J. Moore, K. Shanmugam, and L. Ingram. Combining metabolic engineering and metabolic evolution to develop nonrecombinant strains of *Escherichia coli* C that produce succinate and malate. *Biotechnology and Bioengineering*, 99(5):1140–1153, 2008.

[161] P. Jensen and J. Papin. Functional integration of a metabolic network model and expression data without arbitrary thresholding. *Bioinformatics*, 27(4):541–547, 2011.

[162] R. P. John, D. Gangadharan, and K. Madhavan Nampoothiri. Genome shuffling of *Lactobacillus delbrueckii* mutant and *Bacillus amyloliquefaciens* through protoplasmic fusion for L-lactic acid production from starchy wastes. *Bioresource Technology*, 99(17):8008–8015, 2008.

[163] K. Kao and G. Sherlock. Molecular characterization of clonal interference during adaptive evolution in asexual populations of *Saccharomyces cerevisiae*. *Nature Genetics*, 40(12):1499–1504, 2008.

[164] J. R. Karr, J. C. Sanghvi, D. N. Macklin, M. V. Gutschow, J. M. Jacobs, B. Bolival, N. Assad-Garcia, J. I. Glass, and M. W. Covert. A whole-cell computational model predicts phenotype from genotype. *Cell*, 150(2):389–401, 2012.

[165] B. Kempf and E. Bremer. Uptake and synthesis of compatible solutes as microbial stress responses to high-osmolality environments. *Archives of Microbiology*, 170(5):319–330, 1998.

[166] I. M. Keseler, A. Mackie, M. Peralta-Gil, A. Santos-Zavaleta, S. Gama-Castro, C. Bonavides-Martínez, C. Fulcher, A. M. Huerta, A. Kothari, and M. Krummenacker. EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Research*, 41(D1):D605–D612, 2013.

[167] A. B. Khodursky, B. J. Peter, N. R. Cozzarelli, D. Botstein, P. O. Brown, and C. Yanofsky. DNA microarray analysis of gene expression in response to physiological and genetic changes that affect tryptophan metabolism in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 97(22):12170–12175, 2000.

[168] J. Kim, S. Shoemaker, and D. Mills. Relaxed control of sugar utilization in *Lactobacillus brevis*. *Microbiology*, 155(4):1351, 2009.

[169] Y. Kim and H. A. Orr. Adaptation in sexuals vs. asexuals: clonal interference and the Fisher-Muller model. *Genetics*, 171(3):1377–1386, 2005.

[170] M. Kitagawa, T. Ara, M. Arifuzzaman, T. Ioka-Nakamichi, E. Inamoto, H. Toyonaga, and H. Mori. Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of *E. coli* K-12 ORF archive): unique resources for biological research. *DNA Research*, 12(5):291–299, 2006.

[171] D. Klein-Marcuschamer and G. Stephanopoulos. Method for Designing and Optimizing Random-Search Libraries for Strain Improvement. *Applied and Environmental Microbiology*, 76(16):5541, 2010.

[172] D. Klein-Marcuschamer, C. N. S. Santos, H. Yu, and G. Stephanopoulos. Mutagenesis of the bacterial RNA polymerase alpha subunit for improvement of complex phenotypes. *Applied and Environmental Microbiology*, 75(9):2705–2711, 2009.

[173] D. Klein-Marcuschamer and G. Stephanopoulos. Assessing the potential of mutational strategies to elicit new phenotypes in industrial strains. *Proceedings of the National Academy of Sciences*, 105(7):2319–2324, 2008.

[174] H. Klinke, A. Thomsen, and B. Ahring. Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass. *Applied Microbiology and Biotechnology*, 66(1):10–26, 2004.

[175] E. P. Knoshaug and M. Zhang. Butanol tolerance in a selection of microorganisms. *Applied Biochemistry and Biotechnology*, 153(1-3):13–20, 2009.

[176] A. Krogh, B. Larsson, G. Von Heijne, and E. Sonnhammer. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Journal of Molecular Biology*, 305(3):567–580, 2001.

[177] P. Kroon, G. Williamson, N. Fish, D. Archer, and N. Belshaw. A modular esterase from *Penicillium funiculosum* which releases ferulic acid from plant cell walls and binds crystalline cellulose contains a carbohydrate binding module. *European Journal of Biochemistry*, 267(23):6740–6752, 2003.

[178] E. Kussell. Evolution in Microbes. *Annual Review of Biophysics*, 42(1):493–514, 2013.

[179] M. Kuyper, M. J. Toirkens, J. A. Diderich, A. A. Winkler, J. P. Dijken, and J. T. Pronk. Evolutionary engineering of mixed-sugar utilization by a xylose-fermenting *Saccharomyces cerevisiae* strain. *FEMS Yeast Research*, 5(10):925–934, 2005.

[180] D. J. Kvitek and G. Sherlock. Whole Genome, Whole Population Sequencing Reveals That Loss of Signaling Networks Is the Major Adaptive Strategy in a Constant Environment. *PLoS Genetics*, 9(11):e1003972, 2013.

[181] J. Lachapelle and G. Bell. Evolutionary rescue of sexual and asexual populations in a deteriorating environment. *Evolution*, 66(11):3508–3518, 2012.

[182] J. Landete, H. Rodríguez, J. Curiel, B. de las Rivas, J. Mancheño, and R. Muñoz. Gene cloning, expression, and characterization of phenolic acid decarboxylase from *Lactobacillus brevis* RM84. *Journal of Industrial Microbiology and Biotechnology*, 37(6):617–624, 2010.

[183] B. Landfald and A. R. Strøm. Choline-glycine betaine pathway confers a high level of osmotic tolerance in *Escherichia coli*. *Journal of Bacteriology*, 165(3):849–855, 1986.

[184] G. I. Lang, D. P. Rice, M. J. Hickman, E. Sodergren, G. M. Weinstock, D. Botstein, and M. M. Desai. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature*, 500(7464):571–574, 2013.

[185] T. Lawley, W. Klimke, M. Gubbins, and L. Frost. F factor conjugation is a true type IV secretion system. *FEMS Microbiology Letters*, 224(1):1–15, 2003.

[186] D. Lee, K. Smallbone, W. Dunn, E. Murabito, C. Winder, D. Kell, P. Mendes, and N. Swainston. Improving metabolic flux predictions using absolute gene expression data. *BMC Systems Biology*, 6(1):73, 2012.

[187] H. Lee, C. Harvey, D. Cane, and C. Khosla. Improved precursor-directed biosynthesis in *E. coli* via directed evolution. *The Journal of Antibiotics*, 64(1):59–64, 2010.

[188] J.-H. Lee and J. Lee. Indole as an intercellular signal in microbial communities. *FEMS Microbiology Reviews*, 34(4):426–444, 2010.

[189] S. Lee, L. Frost, and W. Paranchych. FinOP repression of the F plasmid involves extension of the half-life of FinP antisense RNA by FinO. *Molecular and General Genetics*, 235(1):131–139, 1992.

[190] S. Lee, H. Chou, T. Ham, T. Lee, and J. Keasling. Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Current Opinion in Biotechnology*, 19(6):556–563, 2008.

[191] S. Lee, D. Lee, and T. Kim. Systems biotechnology for strain improvement. *Trends in Biotechnology*, 23(7):349–358, 2005.

[192] R. E. Lenski, J. A. Mongold, P. D. Sniegowski, M. Travisano, F. Vasi, P. J. Gerrish, and T. M. Schmidt. Evolution of competitive fitness in experimental populations of *E. coli*: what makes one genotype a better competitor than another? *Antonie van Leeuwenhoek*, 73(1):35–47, 1998.

[193] C. Lerner and M. Inouye. Low copy number plasmids for regulated low-level expression of cloned genes in *Escherichia coli* with blue/white insert screening capability. *Nucleic Acids Research*, 18(15):4631, 1990.

[194] B. Levin, V. Perrot, and N. Walker. Compensatory mutations, antibiotic resistance and the population genetics of adaptive evolution in bacteria. *Genetics*, 154(3):985–997, 2000.

[195] N. Lewis, B. Cho, E. Knight, and B. Palsson. Gene expression profiling and the use of genome-scale in silico models of *Escherichia coli* for analysis: providing context for content. *Journal of Bacteriology*, 191(11):3437–3444, 2009.

[196] H. Liebke, C. Gross, W. Walter, and R. Burgess. A new mutation *rpoD800*, affecting the sigma subunit of *E.coli* RNA polymerase is allelic to two other sigma mutants. *Molecular and General Genetics*, 177(2):277–282, 1980.

[197] Y.-L. Lin and H. P. Blaschek. Butanol production by a butanol-tolerant strain of *Clostridium acetobutylicum* in extruded corn broth. *Applied and Environmental Microbiology*, 45(3):966–973, 1983.

[198] M. Liu, J. Bayjanov, B. Renckens, A. Nauta, and R. Siezen. The proteolytic system of lactic acid bacteria revisited: a genomic comparison. *BMC Genomics*, 11(1):36, 2010.

[199] M. Liu, A. Nauta, C. Francke, and R. Siezen. Comparative genomics of enzymes in flavor-forming pathways from amino acids in lactic acid bacteria. *Applied and Environmental Microbiology*, 74(15):4590, 2008.

[200] S. Liu and N. Qureshi. How microbes tolerate ethanol and butanol. *New Biotechnology*, 26(3-4):117–121, 2009.

[201] M. Lopez, N. Nichols, B. Dien, J. Moreno, and R. Bothast. Isolation of microorganisms for biological detoxification of lignocellulosic hydrolysates. *Applied Microbiology and Biotechnology*, 64(1):125–131, 2004.

[202] J. M. Lucht and E. Bremer. Adaptation of *Escherichia coli* to high osmolarity environments: Osmoregulation of the high-affinity glycine betaine transport system ProU. *FEMS Microbiology Reviews*, 14(1):3–20, 1994.

[203] G. Luli and W. Strohl. Comparison of growth, acetate production, and acetate inhibition of *Escherichia coli* strains in batch and fed-batch fermentations. *Applied and Environmental Microbiology*, 56(4):1004–1011, 1990.

[204] D. S. Lun, G. Rockwell, N. J. Guido, M. Baym, J. A. Kelner, B. Berger, J. E. Galagan, and G. M. Church. Large-scale identification of genetic design strategies using local search. *Molecular Systems Biology*, 5(1):296, 2009.

[205] M. Lynch, T. Warnecke, and R. Gill. SCALEs: multiscale analysis of library enrichment. *Nature Methods*, 4(1):87–93, 2006.

[206] Z. Ma, N. Masuda, and J. W. Foster. Characterization of EvgAS-YdeO-GadE branched regulatory circuit governing glutamate-dependent acid resistance in *Escherichia coli*. *Journal of Bacteriology*, 186(21):7378–7389, 2004.

[207] K. Magnuson, S. Jackowski, C. Rock, and J. Cronan Jr. Regulation of fatty acid biosynthesis in *Escherichia coli*. *Microbiology and Molecular Biology Reviews*, 57(3):522, 1993.

[208] R. Mahadevan, D. Bond, J. Butler, A. Esteve-Nunez, M. Coppi, B. Palsson, C. Schilling, and D. Lovley. Characterization of metabolism in the Fe (III)-reducing organism *Geobacter sulfurreducens* by constraint-based modeling. *Applied and Environmental Microbiology*, 72(2):1558–1568, 2006.

[209] K. Makarova, A. Slesarev, Y. Wolf, A. Sorokin, B. Mirkin, E. Koonin, A. Pavlov, N. Pavlova, V. Karamychev, N. Polouchine, V. Shakhova, I. Grigoriev, Y. Lou, D. Rohksar, S. Lucas, K. Huang, D. M. Goodstein, T. Hawkins, V. Plengvidhya, D. Welker, J. Hughes, Y. Goh, A. Benson, K. Baldwin, J.-H. Lee, I. Daz-Muiz, B. Dosti, V. Smeianov, W. Wechter, R. Barabote, G. Lorca, E. Altermann, R. Barrangou, B. Ganesan, Y. Xie, H. Rawsthorne, D. Tamir, C. Parker, F. Breidt, J. Broadbent, R. Hutkins, D. O'Sullivan, J. Steele,

G. Unlu, M. Saier, T. Klaenhammer, P. Richardson, S. Kozyavkin, B. Weimer, and D. Mills. Comparative genomics of the lactic acid bacteria. *Proceedings of the National Academy of Sciences*, 103(42):15611–15616, 2006.

[210] K. Makarova and E. Koonin. Evolutionary genomics of lactic acid bacteria. *Journal of Bacteriology*, 189(4):1199, 2007.

[211] J. Malakooti and B. Ely. Principal sigma subunit of the *Caulobacter crescentus* RNA polymerase. *Journal of Bacteriology*, 177(23):6854, 1995.

[212] F. Martani, T. Fossati, R. Posteri, L. Signori, D. Porro, and P. Branduardi. Different response to acetic acid stress in *Saccharomyces cerevisiae* wild-type and L-ascorbic acid-producing strains. *Yeast*, 30(9):365–378, 2013.

[213] S. Mathew and T. Abraham. Ferulic acid: an antioxidant found naturally in plant cell walls and feruloyl esterases involved in its release and their applications. *Critical Reviews in Biotechnology*, 24(2-3):59–83, 2004.

[214] J. Maynard Smith. What use is sex? *Journal of Theoretical Biology*, 30(2):319–335, 1971.

[215] A. McCracken, M. Turner, P. Giffard, L. Hafner, and P. Timms. Analysis of promoter sequences from *Lactobacillus* and *Lactococcus* and their activity in several *Lactobacillus* species. *Archives of Microbiology*, 173(5):383–389, 2000.

[216] T. Mills, N. Sandoval, R. Gill, et al. Cellulosic hydrolysate toxicity and tolerance mechanisms in *Escherichia coli*. *Biotechnology and Biofuels*, 2(11):1–11, 2009.

[217] J. Minty, A. Lesnefsky, F. Lin, Y. Chen, T. Zaroff, A. Veloso, B. Xie, C. McConnell, R. Ward, D. Schwartz, J. Rouillard, Y. Gao, E. Gulari, and X. Lin.

Evolution combined with genomic study elucidates genetic bases of isobutanol tolerance in *Escherichia coli*. *Microbial Cell Factories*, 10(1):18, 2011.

[218] D. Misevic, C. Ofria, and R. E. Lenski. Sexual reproduction reshapes the genetic architecture of digital organisms. *Proceedings of the Royal Society B: Biological Sciences*, 273(1585):457–464, 2006.

[219] D. Misevic, C. Ofria, and R. E. Lenski. Experiments with digital organisms on the origin and maintenance of sex in changing environments. *Journal of Heredity*, 101(suppl 1):S46–S54, 2010.

[220] D. Moradigaravand and J. Engelstädter. The effect of bacterial recombination on adaptation on fitness landscapes with limited peak accessibility. *PLoS Computational Biology*, 8(10):e1002735, 2012.

[221] H. J. Muller. Some genetic aspects of sex. *American Naturalist*, 66(703):118–138, 1932.

[222] R. A. Neher, B. I. Shraiman, and D. S. Fisher. Rate of adaptation in large sexual populations. *Genetics*, 184(2):467–481, 2010.

[223] N. Nichols, B. Dien, G. Guisado, and M. López. Bioabatement to remove inhibitors from biomass-derived sugar hydrolysates. In *Twenty-Sixth Symposium on Biotechnology for Fuels and Chemicals*, pages 379–390. Springer, 2005.

[224] S. Nicolaou, S. Gaida, and E. Papoutsakis. A comparative view of metabolite and substrate stress and tolerance in microbial bioprocessing: from biofuels and chemicals, to biocatalysis and bioremediation. *Metabolic Engineering*, 12(4):307–331, 2010.

[225] S. Nicolaou, S. Gaida, and E. Papoutsakis. Coexisting/Coexpressing Genomic Libraries (CoGeL) identify interactions among distantly located genetic

loci for developing complex microbial phenotypes. *Nucleic Acids Research*, 39(22):e152, 2011.

[226] D. Nielsen, E. Leonard, S. Yoon, H. Tseng, C. Yuan, and K. Prather. Engineering alternative butanol production platforms in heterologous bacteria. *Metabolic Engineering*, 11(4-5):262–273, 2009.

[227] K. Nishino and A. Yamaguchi. Analysis of a complete library of putative drug transporter genes in *Escherichia coli*. *Journal of Bacteriology*, 183(20):5803–5812, 2001.

[228] T. E. Norris and A. L. Koch. Effect of growth rate on the relative rates of synthesis of messenger, ribosomal and transfer RNA in *Escherichia coli*. *Journal of Molecular Biology*, 64(3):633–649, 1972.

[229] M. Oberhardt, B. Palsson, and J. Papin. Applications of genome-scale metabolic reconstructions. *Molecular Systems Biology*, 5(1):320, 2009.

[230] H. Ochman, J. G. Lawrence, and E. A. Groisman. Lateral gene transfer and the nature of bacterial innovation. *Nature*, 405(6784):299–304, 2000.

[231] C. Ofria and C. Wilke. Avida: a software platform for research in computational evolutionary biology. *Artificial Life*, 10(2):191–229, 2004.

[232] A. Oliveira, J. Nielsen, and J. Förster. Modeling *Lactococcus lactis* using a genome-scale flux model. *BMC Microbiology*, 5(1):39, 2005.

[233] H. Orr. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution*, pages 935–949, 1998.

[234] H. Orr. The genetic theory of adaptation: a brief history. *Nature Reviews Genetics*, 6(2):119–127, 2005.

[235] J. Orth, T. Conrad, J. Na, J. Lerman, H. Nam, A. Feist, and B. Palsson. A comprehensive genome-scale reconstruction of *Escherichia coli* metabolism: 2011. *Molecular Systems Biology*, 7(1):535, 2011.

[236] B. Otte, E. Grunwaldt, O. Mahmoud, and S. Jennewein. Genome shuffling in *Clostridium diolis* DSM 15410 for improved 1,3-propanediol production. *Applied and Environmental Microbiology*, 75(24):7610, 2009.

[237] S. P. Otto and T. Lenormand. Resolving the paradox of sex and recombination. *Nature Reviews Genetics*, 3(4):252–261, 2002.

[238] J. Ou. Role of surface exclusion genes in lethal zygosis in *Escherichia coli* K-12 mating. *Molecular and General Genetics*, 178(3):573–581, 1980.

[239] M. Paget and J. Helmann. The σ^{70} family of sigma factors. *Genome Biology*, 4(203):2003–4, 2003.

[240] E. Palmqvist and B. Hahn-Hagerdal. Fermentation of lignocellulosic hydrolysates. II: inhibitors and mechanisms of inhibition. *Bioresource Technology*, 74(1):25–33, 2000.

[241] E. T. Papoutsakis. Engineering solventogenic clostridia. *Current Opinion in Biotechnology*, 19(5):420–429, 2008.

[242] S. Parekh, V. Vinci, and R. Strobel. Improvement of microbial strains and fermentation processes. *Applied Microbiology and Biotechnology*, 54(3):287–301, 2000.

[243] R. Pate, G. Klise, and B. Wu. Resource demand implications for US algae biofuels production scale-up. *Applied Energy*, 88(10):3377–3388, 2011.

[244] R. Patnaik. Engineering complex phenotypes in industrial strains. *Biotechnology Progress*, 24(1):38–47, 2008.

[245] R. Patnaik, S. Louie, V. Gavrilovic, K. Perry, W. Stemmer, C. Ryan, and S. del Cardayré. Genome shuffling of *Lactobacillus* for improved acid tolerance. *Nature Biotechnology*, 20(7):707–712, 2002.

[246] R. Patnaik, S. Louie, V. Gavrilovic, K. Perry, W. P. Stemmer, C. M. Ryan, and S. del Cardayré. Genome shuffling of *Lactobacillus* for improved acid tolerance. *Nature Biotechnology*, 20(7):707–712, 2002.

[247] C. Pennetier, L. Domínguez-Ramírez, and J. Plumbridge. Different regions of Mlc and NagC, homologous transcriptional repressors controlling expression of the glucose and N-acetylglucosamine phosphotransferase systems in *Escherichia coli*, are required for inducer signal recognition. *Molecular Microbiology*, 67(2):364–377, 2008.

[248] L. Perfeito, L. Fernandes, C. Mota, and I. Gordo. Adaptive mutations in bacteria: high rate and small effects. *Science*, 317(5839):813–815, 2007.

[249] K. G. Peri, H. Goldie, and E. Waygood. Cloning and characterization of the n-acetylglucosamine operon of *Escherichia coli*. *Biochemistry and Cell Biology*, 68(1):123–137, 1990.

[250] R. Petri and C. Schmidt-Dannert. Dealing with complexity: evolutionary engineering and genome shuffling. *Current Opinion in Biotechnology*, 15(4):298–304, 2004.

[251] P. Pienkos and M. Zhang. Role of pretreatment and conditioning processes on toxicity of lignocellulosic biomass hydrolysates. *Cellulose*, 16(4):743–762, 2009.

[252] J. Plumbridge. Repression and induction of the nag regulon of *Escherichia coli* K-12: the roles of *nagC* and *nagA* in maintenance of the uninduced state. *Molecular Microbiology*, 5(8):2053–2062, 1991.

[253] J. Plumbridge and E. Vimr. Convergent pathways for utilization of the amino sugars n-acetylglucosamine, n-acetylmannosamine, and n-acetylneuraminic acid by *Escherichia coli*. *Journal of Bacteriology*, 181(1):47–54, 1999.

[254] A. Poon and L. Chao. Drift increases the advantage of sex in RNA bacteriophage Φ6. *Genetics*, 166(1):19–24, 2004.

[255] V. A. Portnoy, D. Bezdan, and K. Zengler. Adaptive laboratory evolutionharnessing the power of biology for metabolic engineering. *Current Opinion in Biotechnology*, 22(4):590–594, 2011.

[256] J. Pramanik and J. Keasling. Stoichiometric model of *Escherichia coli* metabolism: Incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. *Biotechnology and Bioengineering*, 56(4):398–421, 1997.

[257] N. Price, J. Reed, and B. Palsson. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nature Reviews Microbiology*, 2(11):886–897, 2004.

[258] M. Punta, P. C. Coggill, R. Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E. Sonnhammer, S. Eddy, A. Bateman, and R. Finn. The Pfam protein families database. *Nucleic Acids Research*, 40(D1):D290–D301, 2012.

[259] J. E. Purvis, L. Yomano, and L. Ingram. Enhanced trehalose production improves growth of *Escherichia coli* under osmotic stress. *Applied and Environmental Microbiology*, 71(7):3761–3769, 2005.

[260] L. Rabiner and B. Juang. An introduction to hidden Markov models. *ASSP Magazine, IEEE*, 3(1):4–16, 1986.

[261] L. Rabiner. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[262] J. Ramos, E. Duque, J. Rodríguez-Herva, P. Godoy, A. Hadour, F. Reyes, and A. Fernández-Barrero. Mechanisms for solvent tolerance in bacteria. *Journal of Biological Chemistry*, 272(7):3887, 1997.

[263] S. Ranganathan, P. F. Suthers, and C. D. Maranas. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Computational Biology*, 6(4):e1000744, 2010.

[264] J. Reed. Shrinking the Metabolic Solution Space Using Experimental Datasets. *PLoS Computational Biology*, 8(8):e1002662, 2012.

[265] J. Reed and B. Palsson. Thirteen years of building constraint-based in silico models of *Escherichia coli*. *Journal of Bacteriology*, 185(9):2692–2699, 2003.

[266] C. Reguant, A. Bordons, L. Arola, and N. Rozes. Influence of phenolic compounds on the physiology of *Oenococcus oeni* from wine. *Journal of Applied Microbiology*, 88(6):1065–1071, 2000.

[267] L. H. Reyes, A. S. Abdelaal, and K. C. Kao. Genetic determinants for n-butanol tolerance in evolved *E. coli* mutants: cross adaptation and antagonistic pleiotropy between n-butanol and other stressors. *Applied and Environmental Microbiology*, 79(17):5313–5320, 2013.

[268] L. H. Reyes, M. P. Almario, and K. C. Kao. Genomic library screens for genes involved in n-butanol tolerance in *Escherichia coli*. *PLoS One*, 6(3):e17678, 2011.

[269] L. H. Reyes, M. P. Almario, J. Winkler, M. M. Orozco, and K. C. Kao. Visualizing evolution in real time to determine the molecular mechanisms of n-butanol

tolerance in *Escherichia coli*. *Metabolic Engineering*, 14(5):579–590, 2012.

[270] L. H. Reyes, J. M. Gomez, and K. C. Kao. Improving carotenoids production in yeast via adaptive laboratory evolution. *Metabolic Engineering*, 21:26–33, 2014.

[271] W. R. Rice. Experimental tests of the adaptive significance of sexual recombination. *Nature Reviews Genetics*, 3(4):241–251, 2002.

[272] W. Rice. Requisite mutational load, pathway epistasis, and deterministic mutation accumulation in sexual versus asexual populations. *Genetica*, 102:71–81, 1998.

[273] M. M. Riehle, A. F. Bennett, and A. D. Long. Genetic architecture of thermal adaptation in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, 98(2):525–530, 2001.

[274] I. Rocha, P. Maia, P. Evangelista, P. Vilaça, S. Soares, J. Pinto, J. Nielsen, K. Patil, E. Ferreira, and M. Rocha. OptFlux: an open-source software platform for in silico metabolic engineering. *BMC Systems Biology*, 4(1):45, 2010.

[275] G. Rockwell, N. J. Guido, and G. M. Church. Redirector: Designing cell factories by reconstructing the metabolic objective. *PLoS Computational Biology*, 9(1):e1002882, 2013.

[276] H. Rodríguez, J. Curiel, J. Landete, B. de las Rivas, F. de Felipe, C. Gómez-Cordovés, J. Mancheño, and R. Muñoz. Food phenolics and lactic acid bacteria. *International Journal of Food Microbiology*, 132(2-3):79–90, 2009.

[277] A. Rodríguez-Verdugo, B. S. Gaut, and O. Tenaillon. Evolution of *Escherichia coli* rifampicin resistance in an antibiotic-free environment during thermal stress. *BMC Evolutionary Biology*, 13(1):50, 2013.

[278] I. Roy Curtiss and D. Stallions. Probability of F integration and frequency of stable Hfr donors in F+ populations of *Escherichia coli* K-12. *Genetics*, 63(1):27, 1969.

[279] N. Rozes and C. Peres. Effects of phenolic compounds on the growth and the fatty acid composition of *Lactobacillus plantarum*. *Applied Microbiology and Biotechnology*, 49(1):108–111, 1998.

[280] K. L. Rozwadowski, G. G. Khachatourians, and G. Selvaraj. Choline oxidase, a catabolic enzyme in *Arthrobacter pascens*, facilitates adaptation to osmotic stress in *Escherichia coli*. *Journal of Bacteriology*, 173(2):472–478, 1991.

[281] K. Rumbold, H. J. van Buijsen, K. M. Overkamp, J. W. van Groenestijn, P. J. Punt, and M. J. Van Der Werf. Microbial production host selection for converting second-generation feedstocks into bioproducts. *Microbial Cell Factories*, 8(1):1–11, 2009.

[282] F. D. Russo, J. M. Slauch, and T. J. Silhavy. Mutations that Affect Separate Functions of OmpR the Phosphorylated Regulator of Porin Transcription in *Escherichia coli*. *Journal of Molecular Biology*, 231(2):261–273, 1993.

[283] B. Rutherford, R. Dahl, R. Price, H. Szmidt, P. Benke, A. Mukhopadhyay, and J. Keasling. Functional genomic study of exogenous n-butanol stress in *Escherichia coli*. *Applied and Environmental Microbiology*, 76(6):1935–1945, 2010.

[284] A. Saeed, V. Sharov, J. White, J. Li, W. Liang, N. Bhagabati, J. Braisted, M. Klapa, T. Currier, and M. Thiagarajan. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques*, 34(2):374, 2003.

[285] A. I. Saeed, N. K. Bhagabati, J. C. Braisted, W. Liang, V. Sharov, E. A. Howe, J. Li, M. Thiagarajan, J. A. White, and J. Quackenbush. TM4 Microarray Software Suite. *Methods in Enzymology*, 411:134–193, 2006.

[286] S. Sakai, Y. Tsuchida, S. Okino, O. Ichihashi, H. Kawaguchi, T. Watanabe, M. Inui, and H. Yukawa. Effect of lignocellulose-derived inhibitors on growth of and ethanol production by growth-arrested *Corynebacterium glutamicum* R. *Applied and Environmental Microbiology*, 73(7):2349, 2007.

[287] M. Sami, H. Yamashita, T. Hirono, H. Kadokura, K. Kitamoto, K. Yoda, and M. Yamasaki. Hop-resistant *Lactobacillus brevis* contains a novel plasmid harboring a multidrug resistance-like gene. *Journal of Fermentation and Bioengineering*, 84(1):1–6, 1997.

[288] N. R. Sandoval, T. Y. Mills, M. Zhang, and R. T. Gill. Elucidating acetate tolerance in *E. coli* using a genome-wide approach. *Metabolic Engineering*, 13(2):214–224, 2011.

[289] C. Santos and G. Stephanopoulos. Combinatorial engineering of microbes for optimizing cellular phenotype. *Current Opinion in Chemical Biology*, 12(2):168–176, 2008.

[290] J. Scaife, J. Heilig, L. Rowen, and R. Calendar. Gene for the RNA polymerase sigma subunit mapped in *Salmonella typhimurium* and *Escherichia coli* by cloning and deletion. *Proceedings of the National Academy of Sciences*, 76(12):6510, 1979.

[291] J. Schellenberger, J. Park, T. Conrad, and B. Palsson. BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC Bioinformatics*, 11(1):213, 2010.

[292] J. Schellenberger, R. Que, R. Fleming, I. Thiele, J. Orth, A. Feist, D. Zielinski, A. Bordbar, N. Lewis, S. Rahmanian, J. Kang, D. Hyduke, and B. Palsson. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature Protocols*, 6(9):1290–1307, 2011.

[293] J. Shiloach and R. Fass. Growing *E. coli* to high cell density: a historical perspective on method development. *Biotechnology Advances*, 23(5):345–357, 2005.

[294] T. Shlomi, M. Cabili, M. Herrgård, B. Palsson, and E. Ruppin. Network-based prediction of human tissue-specific metabolism. *Nature Biotechnology*, 26(9):1003–1010, 2008.

[295] J. Sikkema, J. De Bont, and B. Poolman. Mechanisms of membrane toxicity of hydrocarbons. *Microbiology and Molecular Biology Reviews*, 59(2):201, 1995.

[296] R. Skurray, H. Nagaishi, and A. Clark. Molecular cloning of DNA from F sex factor of *Escherichia coli* K-12. *Proceedings of the National Academy of Sciences*, 73(1):64, 1976.

[297] G. Smith. Conjugational recombination in *E. coli*: myths and mechanisms. *Cell*, 64(1):19–27, 1991.

[298] J. M. Smith. Evolution in sexual and asexual populations. *American Naturalist*, pages 469–473, 1968.

[299] K. Smith and J. Liao. An evolutionary strategy for isobutanol production strain development in *Escherichia coli*. *Metabolic Engineering*, 13(6):674–681, 2011.

[300] P. D. Sniegowski and P. J. Gerrish. Beneficial mutations and the dynamics of adaptation in asexual populations. *Philosophical Transactions of the Royal*

Society B: Biological Sciences, 365(1544):1255–1263, 2010.

[301] J. A. Snyder, B. J. Haugen, E. L. Buckles, C. V. Lockatell, D. E. Johnson, M. S. Donnenberg, R. A. Welch, and H. L. Mobley. Transcriptome of uropathogenic *Escherichia coli* during urinary tract infection. *Infection and Immunity*, 72(11):6373–6381, 2004.

[302] B. K. Sohanpal, S. El-Labany, M. Lahooti, J. A. Plumbridge, and I. C. Blomfield. Integrated regulatory responses of *fimB* to N-acetylneuraminic (sialic) acid and GlcNAc in *Escherichia coli* K-12. *Proceedings of the National Academy of Sciences*, 101(46):16322–16327, 2004.

[303] P. Soucaille, G. Joliff, A. Izard, and G. Goma. Butanol tolerance and autobacteriocin production by *Clostridium acetobutylicum*. *Current Microbiology*, 14(5):295–299, 1987.

[304] B. Springer, Y. Kidan, T. Prammananan, K. Ellrott, E. Böttger, and P. Sander. Mechanisms of streptomycin resistance: selection of mutations in the 16S rRNA gene conferring resistance. *Antimicrobial Agents and Chemotherapy*, 45(10):2877–2884, 2001.

[305] D. Sriprapundh, C. Vieille, and J. Zeikus. Directed evolution of *Thermotoga neapolitana* xylose isomerase: high activity on glucose at low temperature and low pH. *Protein Engineering*, 16(9):683, 2003.

[306] D. Stead. The effect of hydroxycinnamic acids on the growth of wine-spoilage lactic acid bacteria. *Journal of Applied Microbiology*, 75(2):135–141, 2008.

[307] P. Steffen, S. Goyard, and A. Ullmann. The *Bordetella pertussis* sigma subunit of RNA polymerase confers enhanced expression of *fha* in *Escherichia coli*. *Molecular Microbiology*, 23(5):945–954, 1997.

[308] G. Stephanopoulos. Metabolic engineering by genome shuffling. *Nature Biotechnology*, 20(7):666–668, 2002.

[309] G. Stephanopoulos. Challenges in engineering microbes for biofuels production. *Science*, 315(5813):801, 2007.

[310] G. Stephanopoulos, H. Alper, and J. Moxley. Exploiting biological complexity for strain improvement through systems biology. *Nature Biotechnology*, 22(10):1261–1267, 2004.

[311] O. Styrvold, P. Falkenberg, B. Landfald, M. Eshoo, T. Bjørnsen, and A. Strøm. Selection, mapping, and characterization of osmoregulatory mutants of *Escherichia coli* blocked in the choline-glycine betaine pathway. *Journal of Bacteriology*, 165(3):856–863, 1986.

[312] M. Takata, M. S. Sasaki, E. Sonoda, C. Morrison, M. Hashimoto, H. Utsumi, Y. Yamaguchi-Iwai, A. Shinohara, and S. Takeda. Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells. *EMBO*, 17(18):5497–5508, 1998.

[313] K. Tanaka and H. Takahashi. Cloning and analysis of the gene (*rpoDA*) for the principal sigma factor of *Pseudomonas aeruginosa*. *Biochimica et Biophysica Acta (BBA)-Gene Structure and Expression*, 1089(1):113–119, 1991.

[314] O. Tenaillon, B. Toupance, H. Le Nagard, F. Taddei, and B. Godelle. Mutators, population size, adaptive landscape and the adaptation of asexual populations of bacteria. *Genetics*, 152(2):485–493, 1999.

[315] O. Tenaillon, A. Rodríguez-Verdugo, R. L. Gaut, P. McDonald, A. F. Bennett, A. D. Long, and B. S. Gaut. The molecular diversity of adaptive convergence.

Science, 335(6067):457–461, 2012.

[316] J. Timmermans and L. Van Melderen. Post-transcriptional global regulation by CsrA in bacteria. *Cellular and Molecular Life Sciences*, 67(17):2897–2908, 2010.

[317] C. Tomas, J. Beamish, and E. Papoutsakis. Transcriptional analysis of butanol stress and tolerance in *Clostridium acetobutylicum*. *Journal of Bacteriology*, 186(7):2006, 2004.

[318] C. Tomas, N. Welker, and E. Papoutsakis. Overexpression of *groESL* in *Clostridium acetobutylicum* results in increased solvent production and tolerance, prolonged metabolism, and changes in the cell's transcriptional program. *Applied and Environmental Microbiology*, 69(8):4951, 2003.

[319] E. Tomás-Pejó, M. Ballesteros, J. Oliva, and L. Olsson. Adaptation of the xylose fermenting yeast *Saccharomyces cerevisiae* F12 for improving ethanol production in different fed-batch SSF processes. *Journal of Industrial Microbiology and Biotechnology*, 37(11):1211–1220, 2010.

[320] E. Toprak, A. Veres, J.-B. Michel, R. Chait, D. L. Hartl, and R. Kishony. Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nature Genetics*, 44(1):101–105, 2011.

[321] L. Trut, I. Plyusnina, and I. Oskina. An experiment on fox domestication and debatable issues of evolution of the dog. *Russian Journal of Genetics*, 40(6):644–655, 2004.

[322] L. Trut, I. Oskina, and A. Kharlamova. Animal evolution during domestication: the domesticated fox as a model. *Bioessays*, 31(3):349–360, 2009.

[323] E. Twiss, A. M. Coros, N. P. Tavakoli, and K. M. Derbyshire. Transposition is modulated by a diverse set of host factors in *Escherichia coli* and is stimulated by nutritional stress. *Molecular Microbiology*, 57(6):1593–1607, 2005.

[324] A. Typas, R. Nichols, D. Siegele, M. Shales, S. Collins, B. Lim, H. Braberg, N. Yamamoto, R. Takeuchi, B. Wanner, M. Hirotada, J. Weissman, N. Krogan, and C. Gross. A tool-kit for high-throughput, quantitative analyses of genetic interactions in *E. coli*. *Nature Methods*, 5(9):781, 2008.

[325] P. Unrean and F. Srienc. Metabolic networks evolve towards states of maximum entropy production. *Metabolic Engineering*, 13(6):666–673, 2011.

[326] M. Vaquero, M. Alberto, and M. de Nadra. Antibacterial effect of phenolic compounds from different wines. *Food Control*, 18(2):93–101, 2007.

[327] P. Vats, Y.-L. Shih, and L. Rothfield. Assembly of the MreB-associated cytoskeletal ring of *Escherichia coli*. *Molecular Microbiology*, 72(1):170–182, 2009.

[328] A. Verheul, J. Wouters, F. Rombouts, and T. Abee. A possible role of ProP, ProU and CaiT in osmoprotection of *Escherichia coli* by carnitine. *Journal of Applied Microbiology*, 85(6):1036–1046, 1998.

[329] M. Vermue, J. Sikkema, A. Verheul, R. Bakker, and J. Tramper. Toxicity of homologous series of organic solvents for the gram-positive bacteria *Arthrobacter* and *Nocardia* Sp. and the gram-negative bacteria *Acinetobacter* and *Pseudomonas* Sp. *Biotechnology and Bioengineering*, 42(6):747–758, 1993.

[330] W. Vollmer, D. Blanot, and M. A. De Pedro. Peptidoglycan structure and architecture. *FEMS Microbiology Reviews*, 32(2):149–167, 2008.

[331] L. Wackett. Biomass to fuels via microbial transformations. *Current Opinion in Chemical Biology*, 12(2):187–193, 2008.

[332] C. Wang, M. Oh, and J. Liao. Directed evolution of metabolically engineered *Escherichia coli* for carotenoid production. *Biotechnology Progress*, 16(6):922–926, 2000.

[333] T. Warnecke, M. Lynch, A. Karimpour-Fard, M. Lipscomb, P. Handke, T. Mills, C. Ramey, T. Hoang, and R. Gill. Rapid dissection of a complex phenotype through genomic-scale mapping of fitness altering genes. *Metabolic Engineering*, 12(3):241–250, 2010.

[334] J. Warner, R. Patnaik, and R. Gill. Genomics enabled approaches in strain engineering. *Current Opinion in Microbiology*, 12(3):223–230, 2009.

[335] J. Warner, P. Reeder, A. Karimpour-Fard, L. Woodruff, and R. Gill. Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nature Biotechnology*, 28(8):856–862, 2010.

[336] K. Watanabe, S. Yamamoto, S. Hino, and S. Harayama. Population dynamics of phenol-degrading bacteria in activated sludge determined by *gyrB*-targeted quantitative PCR. *Applied and Environmental Microbiology*, 64(4):1203, 1998.

[337] A. Weber and K. Jung. Profiling early osmostress-dependent gene expression in *Escherichia coli* using DNA macroarrays. *Journal of Bacteriology*, 184(19):5502–5507, 2002.

[338] W. Wei, K. Wu, Y. Qin, Z. Xie, and X. Zhu. Intergeneric protoplast fusion between *Kluyveromyces* and *Saccharomyces cerevisiae*–to produce sorbitol from Jerusalem artichokes. *Biotechnology Letters*, 23(10):799–803, 2001.

[339] D. B. Weissman, M. W. Feldman, and D. S. Fisher. The rate of fitness-valley crossing in sexual populations. *Genetics*, 186(4):1389–1410, 2010.

[340] S. White, J. Zheng, Y. Zhang, and C. Rock. The structural biology of type II fatty acid biosynthesis. *Annual Review of Biochemistry*, 74:791–831, 2005.

[341] C. Wilke, J. Wang, C. Ofria, R. Lenski, and C. Adami. Evolution of digital organisms at high mutation rates leads to survival of the attest. *Nature*, 412(6844):331–333, 2001.

[342] N. Willetts. Mapping loci for surface exclusion and incompatibility on the F factor of *Escherichia coli* K-12. *Journal of Bacteriology*, 118(3):778, 1974.

[343] N. Willetts and R. Skurray. The conjugation system of F-like plasmids. *Annual Review of Genetics*, 14(1):41–76, 1980.

[344] G. C. Williams and J. B. Mitton. Why reproduce sexually? *Journal of Theoretical Biology*, 39(3):545–554, 1973.

[345] J. Winkler and K. Kao. Computational identification of adaptive mutants using the VERT system. *Journal of Biological Engineering*, 6:3, 2012.

[346] J. Winkler and K. Kao. Harnessing recombination to speed adaptive evolution in *Escherichia coli*. *Metabolic Engineering*, 14(5):487–495, 2012.

[347] J. Winkler, M. Rehmann, and K. Kao. Novel *Escherichia coli* hybrids with enhanced butanol tolerance. *Biotechnology Letters*, 32(7):915–920, 2010.

[348] J. Winkler, L. H. Reyes, and K. C. Kao. Adaptive Laboratory Evolution for Strain Engineering. In *Systems Metabolic Engineering*, pages 211–222. Springer US, New York, 2013.

[349] M. J. Wiser, N. Ribeck, and R. E. Lenski. Long-Term Dynamics of Adaptation in Asexual Populations. *Science*, 342(6164):1364–1367, 2013.

[350] H. Wisselink, C. Cipollina, B. Oud, B. Crimi, J. Heijnen, J. Pronk, and A. van Maris. Metabolome, transcriptome and metabolic flux analysis of arabinose fermentation by engineered *Saccharomyces cerevisiae*. *Metabolic Engineering*, 12(6):537–551, 2010.

[351] B. Wood. *Microbiology of Fermented Foods*. Blackie Academic & Professional, 1998.

[352] M. Wosten. Eubacterial sigma-factors. *FEMS Microbiology Reviews*, 22(3):127–150, 1998.

[353] J. Wright, E. Bellissimi, E. de Hulster, A. Wagner, J. T. Pronk, and A. J. van Maris. Batch and continuous culture-based selection strategies for acetic acid tolerance in xylose-fermenting *Saccharomyces cerevisiae*. *FEMS Yeast Research*, 11(3):299–306, 2011.

[354] V. Yadav, B. Panilaitis, H. Shi, K. Numuta, K. Lee, and D. L. Kaplan. N-acetylglucosamine 6-Phosphate Deacetylase (*nagA*) Is Required for N-acetyl Glucosamine Assimilation in *Gluconacetobacter xylinus*. *PloS ONE*, 6(6):e18099, 2011.

[355] K. Yizhak, T. Benyamini, W. Liebermeister, E. Ruppin, and T. Shlomi. Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. *Bioinformatics*, 26(12):i255–i260, 2010.

[356] K. Young. In vitro antibacterial resistance selection and quantitation. *Current Protocols in Pharmacology*, 2001.

[357] C. Zeyl and G. Bell. The advantage of sex in evolving yeast populations. *Nature*, 388(6641):465–468, 1997.

[358] Y. Zhang and C. Rock. Membrane lipid homeostasis in bacteria. *Nature Reviews Microbiology*, 6(3):222–233, 2008.

[359] Y. Zhang, K. Perry, V. Vinci, K. Powell, W. Stemmer, and S. del Cardayré. Genome shuffling leads to rapid phenotypic improvement in bacteria. *Nature*, 415(6872):644–646, 2002.

[360] K. A. Zingaro, S. Nicolaou, Y. Yuan, and E. T. Papoutsakis. Exploring the heterologous genomic space for building, stepwise, complex, multicomponent tolerance to toxic chemicals. *ACS Synthetic Biology*, 2014.

[361] K. A. Zingaro and E. Terry Papoutsakis. GroESL overexpression imparts *Escherichia coli* tolerance to i-, n-, and 2-butanol, 1, 2, 4-butanetriol and ethanol with complex and unpredictable patterns. *Metabolic Engineering*, 15:196–205, 2013.

[362] H. Zur, E. Ruppin, and T. Shlomi. iMAT: an integrative metabolic analysis tool. *Bioinformatics*, 26(24):3140–3142, 2010.