ADAPTIVE EVOLUTION FOR THE STUDY OF COMPLEX PHENOTYPES IN MICROBIAL SYSTEMS

A Dissertation

by

LUIS HUMBERTO REYES BARRIOS

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

DOCTOR OF PHILOSOPHY

Chair of Committee,	Katy Kao
Committee Members,	Arul Jayaraman
	M. Nazmul Karim
	Deborah Siegele
Head of Department,	M. Nazmul Karim

August 2013

Major Subject: Chemical Engineering

ABSTRACT

Microbial-based industrial production has experienced a revolutionary development in the last decades as chemical industry has shifted its focus towards more sustainable production of fuels, building blocks for materials, polymers, chemicals, etc. The strain engineering and optimization programs for industrially relevant phenotypes tackle three challenges for increased production: optimization of titer, productivity, and yield. The yield of production is function of the robustness of the microbe, generally associated with complex phenotypes.

The poor understanding of complex phenotypes associated with increased production poses a challenge for the rational design of strains of more robust microbial producers. Laboratory adaptive evolution is a strain engineering technique used to provide fundamental biological insight through observation of the evolutionary process, in order to uncover molecular determinants associated with the desired phenotype.

In this dissertation, the development of different methodologies to study complex phenotypes in microbial systems using laboratory adaptive evolution is described. Several limitations imposed for the nature of the technique were discussed and tackled. Three different cases were studied. Initially, the n-butanol tolerance in *Escherichia coli* was studied in order to illustrate the effect of clonal interference in microbial systems propagated under selective pressure of an individual stressor. The methodology called Visualizing Evolution in Real Time (VERT) was developed, to aid in mapping out the adaptive landscape of n-butanol tolerance, allowing the uncovering of divergent mecha-

nisms of tolerance. A second case involves the study of clonal interference of microbial systems propagated under several stressors. Using VERT, *Saccharomyces cerevisiae* was evolved in presence of hydrolysates of lignocellulosic biomass. Isolated mutants showed differential fitness advantage to individual inhibitors present in the hydrolysates; however, some mutants exhibited increased tolerance to hydrolysates, but not to individual stressors. Finally, dealing with the problem of using adaptive evolution to increase production of secondary metabolites, an evolutionary strategy was successfully designed and applied in *S. cerevisiae*, to increase the production of carotenoids in a short-term experiment. Molecular mechanisms for increased carotenoids production in isolates were identified.

DEDICATION

To my parents (Luis Humberto Reyes and Nelcy Barrios de Reyes), my brother (Fabio David Reyes Barrios), my sister (Carolina del Pilar Reyes Barrios), my little niece (Ana Sofía Reyes Barrios) and my love (María Priscila Almario Falla), for their endless love, encouragement and support.

ACKNOWLEDGEMENTS

I would like to sincerely thank Dr. Katy Kao for their support and guidance throughout the course of my doctoral studies. Your passion for teaching and for research helped me to mold different aspects of my professional career. I would also like to thank to my labmates for their friendship and criticism when necessary.

TABLE OF CONTENTS

Page

Page

Page

Page

Page

LIST OF FIGURES

Page

Page

Page

Page

LIST OF TABLES

Page

Page

1. INTRODUCTION

Strain engineering and optimization is a fundamental problem in biotechnology applications. Genetic manipulation of microorganisms to create a strain that is superior to the wild type with regard to a particular phenotype is the goal of any strain improvement program. Biological systems are capable of a wide range of chemical reactions and have several advantages over chemical synthesis and petrochemical-based chemical production, including the ability to use sustainable feedstocks and the capability to produce a wide variety of products (1). However, microbial-based production of fuels and chemicals intrinsically has many challenges, such as the optimization of the rate and yield of production, and the final achievable titer (2, 3). Since many of the biochemical pathways are well characterized, the rate and yield of production can generally be readily addressed by metabolic engineering (4-10). On the other hand, the maximum final titer achievable is a function of the robustness of the microbial system in the production environment (11). Robustness is defined as the property of microbial systems to maintain their biological functions despite external and internal perturbations. The molecular mechanisms associated with such robustness are complex (involving multiple genetic determinants) and are not well understood, imposing a challenge for the rational engineering of microbial producers.

Advances in metabolic engineering and synthetic biology have led to the creation of several reverse engineering tools in order to better understand underlying molecular mechanisms indispensable for the rational engineering of strains. Some of such method-

ologies include: enrichment of genomic libraries (12-14), transcriptome analyses (15-17), trackable multiplex recombineering (TRMR) (18), global transcription machinery engineering (gTME) (19), ultra high throughput genomic sequencing (UHTS) coupled with laboratory adaptive evolution (20-22), among others. These techniques will be described in detail in the following chapter (Chapter 2).

Laboratory adaptive evolution has been broadly used for the engineering of microbial system for tolerance to inhibitors and operational conditions (21, 22). Characterization of molecular changes in isolated adaptive mutants from *in vitro* laboratory adaptive evolution experiments provides insights into the adaptive landscape for the phenotype of interest. Hence, determination of the adaptive landscape will drastically improve our knowledge on the important parameters underlying complex phenotypes needed for the rational engineering of strains. A complete description of laboratory adaptive evolution, its limitations and techniques can be found in Chapter 2.

In this dissertation, the development of diverse methodologies to study complex phenotypes in microbial systems using laboratory adaptive evolution is described. Several limitations of the traditional use of the laboratory adaptive evolution are discussed and tackled. Three different cases were studied: *i.* The study of clonal interference of microbial systems propagated under selective pressure of an individual stressor, *ii*. The study of clonal interference of microbial systems propagated under several stressors, and *iii*. The improvement of the production of secondary metabolites in yeast using adaptive evolution.

1.1. Dissertation Objectives

The central objective of this dissertation lies in the development of efficient methodologies for the study of complex phenotypes in microbial systems, using adaptive laboratory evolution as a platform for strain engineering, in conjunction of diverse molecular biology techniques, including, but not limited to transcriptional analysis and genomics, as well as strain engineering techniques such as genome shuffling. The elucidation of the molecular mechanisms involved in complex traits obtained after laboratory adaptive evolution facilitates the rational engineering of industrial strains, generally by finding improvements by means that could not have been predicted using classical metabolic engineering.

1.2. Specific Aims

To fulfill the above-mentioned general objective, the following specific aims were pursued:

- Develop "Visualizing Evolution in Real-Time" (VERT), a methodology based laboratory adaptive evolution, used to analyze complex phenotypes in microbial systems, to aid the mapping of adaptive landscapes in evolving microbial population. In addition, this method facilitated the isolation of adaptive mutants when effects of clonal interference are significant (see Chapter 2).

- Demonstrate the applicability of using VERT to study different industrially relevant complex phenotypes, as the n-butanol tolerance in *Escherichia coli* and the tolerance to hydrolysates of lignocellulosic biomass in *Saccharomyces cerevisiae.*
- Study of cross-tolerance and antagonistic pleiotropy between different industrially relevant stressors in evolved strains; further demonstrating the usefulness of VERT.
- Rationally design of experimental conditions (selective pressure) necessary for the use of laboratory adaptive evolution experiments for the improvement of production of secondary metabolites.

1.3. Thesis Organization

This dissertation is focused on showing diverse applications using a developed methodology to study complex phenotypes in microbial systems based on laboratory adaptive evolution. A background on different metabolic engineering and synthetic biology tools are presented in Chapter 2, including an ample emphasis on laboratory adaptive evolution theory. Chapter 3 and Chapter 4 are focused on the evolutionary study of *Escherichia coli* under n-butanol stress, using continuous cultures (chemostats) to carry out the experimentation. VERT was used to help map-out the evolutionary dynamics of the population. These chapters describe the effects of clonal interference under selective pressure of an individual stressor, as well as divergent mechanisms of stress response under the presence of the solvent. Chapter 5 describes the use of VERT for the evolution of yeast (*Saccharomyces cerevisiae*) in the presence of hydrolysates of lignocellulo-

sic biomass. In this case, the microorganisms were exposed to a mixture of different stressors (hydrolysates of lignocellulosic biomass contain several inhibitors formed during the pretreatment process), and the rising of several mechanisms of tolerance to the individual stressors were identified. Chapter 6 exemplify how appropriate selection of the experimental conditions can be used to improve the production of secondary metabolites using laboratory adaptive evolution. The production of carotenoids was improved significantly in a short-term experiment. Finally, Chapter 7 is a bonus chapter where genomic library enrichments were used to study the n-butanol tolerance in *E. coli*. Finally, Chapter 8 contains the Conclusions and Recommendations.

1.4. Contributions of the Dissertation

The main contributions of the dissertation are the following:

- A novel methodology to study complex phenotypes using microbial systems called Visualizing Evolution in Real-Time (VERT) was developed. Using fluorescently marked strains, VERT facilitates the map-out of the evolutionary dynamics and the rational selection of adaptive mutants.
- Novel mechanisms of n-butanol tolerance in *E. coli* were identified using VERT, which may otherwise be missed due to effects of clonal interference.
- Cross-tolerance and antagonistic pleiotropy were identified in populations evolved under n-butanol stress. Divergent mechanisms of n-butanol tolerance in *E. coli* were described.

- Several mechanisms of increased tolerance to hydrolysates of lignocellulosic biomass were identified using VERT. Isolated mutants showed differential tolerances to individual inhibitors present in the hydrolysates; however, some mutants exhibited increased tolerance to hydrolysates, but not to individual stressors.
- Successful design of a selective pressure was applied to increase the production of the natural pigments, carotenoids, in *S. cerevisiae* using laboratory adaptive evolution. The production of carotenoids was increased more than 200% using a short-term experiment. Molecular mechanisms for increased production of carotenoids in yeast were identified.

2. BACKGROUND*

2.1. Microbial Systems in Industry

Traditional industrial microbiology has been implemented for thousand of years. Some examples can be seen in the preservation of milk and vegetables, bread, beer and wine production, vinegar, among others. A revamping in the use of biotechnology occurred in the 20th century, with the large fermentation processes used for the production of solvents, vitamins, enzymes, and other products, as well as the discovering and developing of antibiotics as penicillin and streptomycin. But it was not until the introduction of the recombinant DNA technology, that the biotechnology industry was propelled in the U.S and worldwide.

Prokaryotic cells such as *Escherichia coli*, and microbial eukaryotic cells as filamentous fungi and yeast, have been the symbol of the production of diverse valuable chemicals in the last twenty years. The implementation of microbial systems has been especially successful in the pharmaceutical, food, chemical, health care and biofuel industries. Several reasons exist for the increased use of microbial systems as production platforms, compared to alternative chemical routes, or the use of plants and animals systems: (*i*) the intrinsic capacity to convert simple substrates into a wide variety of complex molecules; (*ii*) the higher ratio of surface area/volume increases the nutrients uptake

* Reprinted with permission from Luis H. Reyes, James Winkler, and Katy C. Kao, (2012). "Visualizing evolution in real-time method for strain engineering". *Frontiers in microbiology*, *3*, 198. DOI: 10.3389/fmicb.2012.00198. Printed with permission of Frontiers in Microbiology

necessary to support high rates of metabolism and biosynthesis; (*iii*) simple genetic manipulation to enable the production of chemicals that not occur naturally, (*iv*) Simplicity in the screening techniques, allowing the analysis of several variables in a timely manner, and (*v*) the wide diversity of microorganisms available.

Industrial applications of bioprocess become commercially viable when either of these two requirements are met: it is not feasible using chemical synthesis or better economical performance at large scale compared with traditional chemical synthesis. The ability to manipulate the biocatalyst to express desired complex traits in industrial settings is possible as a result of the integration of the tools of "classical" and "modern" strain engineering approaches and the remarkable resilience of microbial cells to such non-natural interventions. Optimization of industrial strains is generally tackled from three different aspects: increments in the rate and yield of production, as well as improvements in final titer achievable. Important advances have been achieved through the use of metabolic engineering and synthetic biology to overcome issues related with rate and yield of production (5, 8, 10, 23). Optimization of the final titer achievable is generally a function of the robustness of the microbial system, a complex phenotype that is usually not well understood. The main cause of low yield of production in microbial systems is high toxicity of the produced compound and process conditions that differ significantly from natural environments. Several approaches have been used in pursuing higher tolerance levels of microbial hosts (12, 13, 21, 22), and they will be described in the following pages. Figure 2.1 summarizes a typical flow chart for the engineering of complex phenotypes in industry (24).

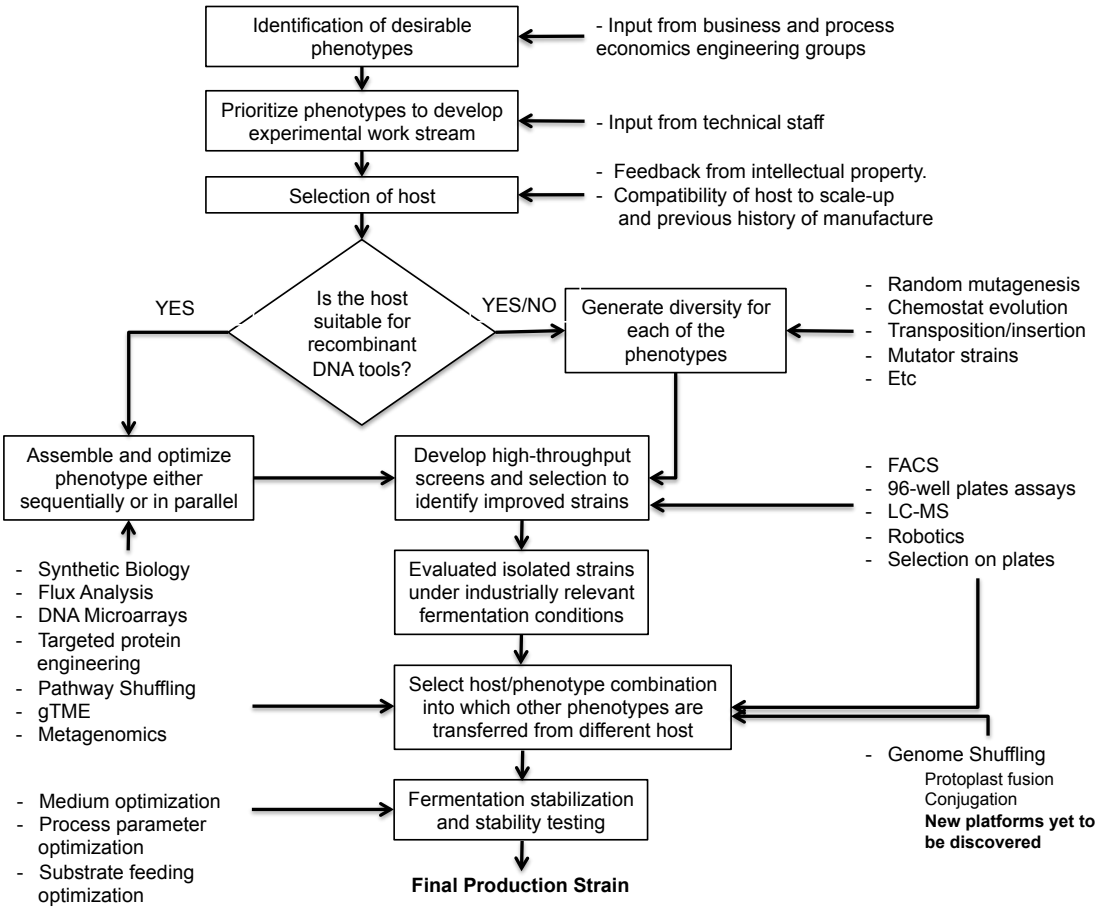


Figure 2.1. Strain engineering of complex phenotypes in industry. gTME: global Transcription Machinery Engineering, FACS: Fluorescent Activated Cell-Sorter, LC-MS: Liquid Chromatography – Mass Spectroscopy.

2.2. Methods for the Strain Engineering of Microorganisms

2.2.1. Classical strain engineering

Microorganisms can be adapted to the to the requirements of industrial processes by mutation and selection. A classical approach for strain engineering includes the creation of genetic variation using diverse random mutagenesis techniques, followed by different screening techniques to select the phenotypes of interest. Random mutagenesis

creates mutations at undefined sites and the knowledge of sequence or function is not required for its use. Random mutagenesis can be introduced in a number of ways including: chemical mutagenesis, error-prone PCR, UV irradiation, mutator strains or DNA shuffling. Some of them are discussed below.

2.2.1.1. Chemical mutagenesis

Chemical mutagenesis induces diverse types of random DNA damage in living cells across the entire genome. This method selects for non-lethal mutations because the cells must replicate for the changes to be observed. Due to the harshness of the procedure, the viability of microbial cultures is normally reduced to orders of 5% for efficient mutagenesis. The mutagenic process is followed for a tedious selection of the survivals, usually by plating in rich medium, with further screening for the desired phenotype. The complete program normally happened in a reiterative way. Disadvantages of chemical mutagenesis include accumulation of deleterious mutations, extreme toxicity of the reagents used, and hot spots, as in the case of ethyl methanesulfonate (EMS) where only transitions GC to AT or AT to GC are obtained (25).

2.2.1.2. Error-prone PCR

Error-prone PCR is the standard method that researchers use to create libraries of mutations within single genes. The experiment is simple and it is used in procedures that examine a small number of mutations. By altering the buffer conditions, the error rate of the Taq DNA polymerase is increased up to 1 – 3 base pairs substitutions per

kilobase of amplified DNA. The main disadvantage of this approach is that size of the library is limited by the efficiency of the cloning step. A technique that eliminate the ligation step limiting the library size is called rolling circle error-prone PCR. It is a variant of error-prone PCR in which the wild type sequence is first cloned into a plasmid, and then the whole plasmid is amplified under error-prone conditions. However, the amplification of the whole plasmid is less efficient than amplifying the coding sequence alone.

2.2.1.3. Mutator strains

In this approach, the sequence to be mutated is cloned into a plasmid and transformed into a mutator strain. A mutator strain is a microbe with deficiency in primary DNA repair pathways, such as XL1-Red® (Stratagene), an *E. coli* strain with mutations in *mutS*, *mutD* and *mutT*, creating errors during replicate of its DNA, including the cloned plasmid. One advantage of mutator strains is that a wide variety of mutations can be incorporated including substitutions, deletions and frame-shifts. However, with this method the strain becomes progressively less viable as it accumulates mutations in its own genome. By using inducible promoters to control the expression of DNA repair mechanisms, it is possible to cycle cells between mutagenic and normal periods of growth.

2.2.1.4. DNA shuffling

DNA shuffling is a methodology in which members of a DNA library are randomly shuffled. By digesting the library with DNase I, and then rejoining the fragments using self-priming PCR, DNA shuffling can be applied to libraries produced by any of the aforesaid methods. This method allows the study of the effects of different combinations of mutations.

2.2.2. Metabolic engineering (ME)

Bailey (26) defined metabolic engineering (ME) as "the improvement of cellular activities by manipulation of enzymatic, transport, and regulatory functions of the cell with the use of recombinant DNA technology".

With classical strain engineering, the microbial genome is randomly changed, untargeted. Only a handful of generated mutants are chosen, displaying the improved phenotype. Modern recombinant DNA and synthetic biology permit targeted manipulation in the engineering of organisms, allowing rational strain engineering. These tools have advanced the field of constructing phenotypes at a much faster pace as compared with tools for identifying the underlying genetic basis of the phenotypes. However, ME requires well-defined systems, where the genetic basis for a given phenotype is adequately mapped. There are many examples demonstrating the applicability and success in the production of several products in a wide range of hosts (4, 27, 28), but they will not be covered in this dissertation.

Several factors play an important role in deciding when to use ME rather than classical strain engineering, such as the amenability of the host to the tools of recombinant DNA, the suitability of the host for commercial fermentation operation, and history with regulatory requirements (24) (Figure 2.1). ME strategies for increasing production rate, titer, and yield are highly dependent on the host organism. For this reason, most of the novel pathways are metabolically engineered in *E. coli,* yeast (*Saccharomyces cerevisiae*) and *Bacillus subtilis* (24).

2.2.3. Reverse strain engineering (RE)

Figure 2.2 presents a comparison between the approaches taken in ME and RE. ME follows a cycle of measurement/analysis/perturbation. The data provided from the measurements can be used to design mathematical models. These models can be analyzed to determine possible targets (hypotheses). After genetically manipulating the microorganisms, experiments are carried out to determine how the metabolic network has been affected. Using classical methods for ME, the identification of metabolic bottlenecks and rate-limiting steps are arduous problems to solve. Without a complete identification of the metabolic pathway to engineer, ME is likely to fail. RE is a tool to provide to the researcher an understanding about such mechanisms. The premise of RE is to use random approaches combined with genomic and molecular biology tools to identify the important parameters for a desired phenotype, thus providing a framework for understanding the cellular and metabolic response, creating an insight of the complex phenotypes for a rational strain engineering.

Knowledge of the metabolic pathways does not reveal the concentrations of intermediates, products and by-products of a pathway, or when the molecules start having toxic effects. For such reason, RE can be seen as a natural complement to methods of ME, and natural synergy exists between the two approaches.

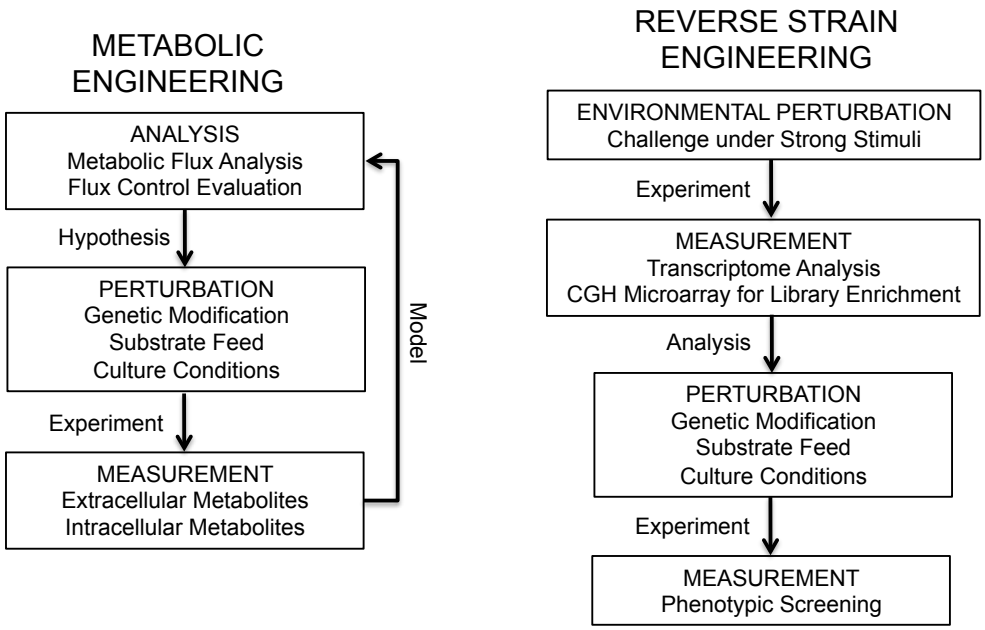


Figure 2.2. Metabolic engineering vs. reverse strain engineering. CGH: Comparative Genomic Hybridization.

2.2.3.1. Transcriptome profiling

The transcriptome is the complete set of transcripts in a cell, both in terms of type and quantity. Global transcriptome analysis is important in understanding how genetic variants contribute to complex phenotypes. Analysis of genome-wide differential RNA expression provides researchers with greater insights into biological pathways and molecular mechanisms in microbial systems.

Currently, the most widely used method to analyze global patterns of gene expression is the DNA microarray. Microarrays have been developed to monitor the gene expression after different perturbations take place in a microbial environment. In such microarrays, oligonucleotides corresponding to the genes whose expression has to be analyzed (the probes) are attached in an ordered fashion to a solid support, generally a glass slide. The relative abundance of the corresponding transcripts in a RNA preparation is measured by labeling the sample and a reference with different fluorescent markers, generally Cy3 (yellow-green marker) and Cy5 (red marker). The samples are hybridized on the slide, and the intensity of the hybridization signal is proportional to the relative abundance of the corresponding mRNA in the sample. The expression profiles or transcriptome refers to the complete collection of mRNAs present. Thus, comparing the hybridization signals for different mRNA samples allows changes in mRNA levels to be determined under the conditions tested for all the genes represented on the arrays. This method has been effective in the study of several complex phenotypes in different systems (15, 29-31).

However, since microarrays are hybridization-based, they cannot be used to detect RNA transcripts not included in the designed array or from repeated sequences. Furthermore, microarrays offer limited dynamic range to detect subtle changes in expression level of target genes, which is critical in understanding biological response to stimuli or environmental changes.

A recently developed technique to study the transcriptome is the Total RNA-sequencing (RNA-Seq) (32). This technique uses high-throughput sequencing to ana-

lyze the transcriptome at a much higher resolution than possible using microarray-based methods. Total RNA-Seq enables: (*i*) the detection of all known and novel RNAs present in biological samples, with no bias toward known RNA molecules as with probe-based technologies, (*ii*) identification of alternative splicing events, (*iii*) expressed SNPs (single nucleotide polymorphisms) or mutations, (*iv*) translocations and fusion transcripts, and (*v*) identification of allele specific expression patterns.

2.2.3.2. Library enrichments

A genomic library is a collection of cloned DNA fragments in a plasmid, or other vector, and then transformed into a host strain. Unlike cDNA libraries, used in transcriptome analysis and containing the sequence for genes that are transcribed by the organism, a genomic library contains all the genetic information, including non-coding regulatory sequences and other forms of "junk" DNA. Subsequent growth of the host under different stimuli allows the enrichment of clones overexpressing genes that confer a positive growth advantage under the tested experimental conditions. Genomic library enrichments are strictly limited to phenotypes that respond to overexpression of a single gene or an operon, however this traditional approach has proven successful studying several complex phenotypes, as n-butanol tolerance (12, 33), acid tolerance (13) and ethanol tolerance mechanisms (34), among others.

Once the phenotype-conferring fragment is isolated, further cloning is required to identify the protein products conferring the desired trait. The multiscale analysis of library enrichment, or SCALEs, is a method in which populations expressing genomics

libraries of different size of DNA fragment in plasmids are selected competitively under selected conditions, allowing a more analytical analysis of microarray signals, enabling parallel analysis of the entire library population (14, 35). An overview of the technique is shown in Figure 2.3.

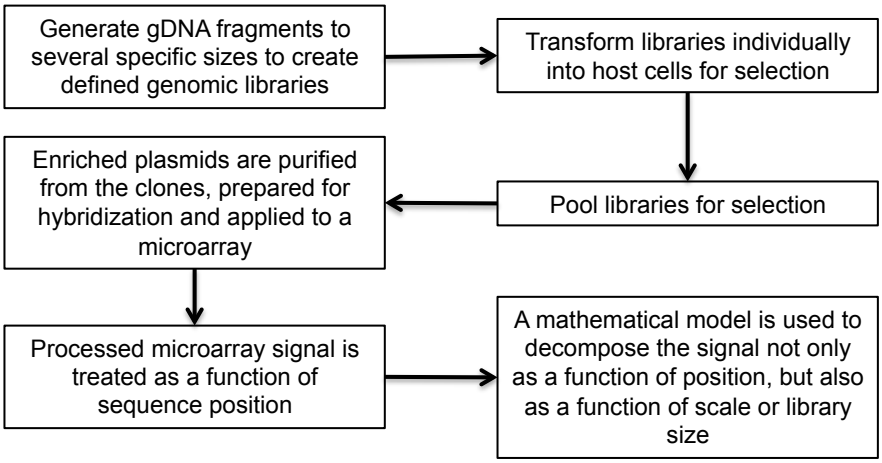


Figure 2.3. Overview of the multiscale analysis of library enrichment, or SCALEs

2.2.3.3. Metagenomics

The exploitation of microbial genomes found in nature has been proven to help in the engineering of complex phenotypes (36, 37). Metagenomics is the field of research that enables application of genomics-based tools to uncultivated microorganisms. The fact that more than 99% of microorganisms are thought to be uncultivable or difficult to culture in a laboratory (38), suggest that the unexplored genomic space is huge and many enzymes with desirable applications are there to be discovered. One approach to explore the genomic space of uncultivable microorganisms is the use of metagenomic libraries. In a metagenomic library, all genome sequences in a microbial environmental sample are

non-selectively cloned into a single library, allowing the study of genomes in an unbiased way. In theory, this technique allow a faster discovery of novel genes with industrial value, such as the case of chitinase, dehydrogenase, amylase, esterase, endoglucanase and cyclodextrinase (39-42). However, even if metagenomic libraries exploit options that other strain engineering techniques do not offer, many problems should be considered when the targets are being selected. Lacking of cofactors and prosthetic groups in host strain avoid proper protein folding, ignoring possible targets, as well as the lacking of appropriated sigma factors and transcriptional regulator, to promote the right expression of such targets. All this without counting the failure to produce clones representing these novel genes due primarily to their toxicity in the host strain.

2.2.4. Whole genome engineering

2.2.4.1. Genome shuffling

Even in well-characterized and mature industrial microbial systems, there is room for improvement in the rational and targeted strain engineering, mainly since most of the whole-cell phenotypes are by nature, complex, impeding the modeling and identification of the cellular machinery behind such phenotypes. Additionally, the asexual process has limitations associated to the poor accumulation of beneficial mutations and the difficulty to lose the deleterious ones in a linear fashion. Genome shuffling of selected strains is an efficient method for improvement of microbial strains with desirable phenotypes. This technique combines the advantage of multiparental crossing that is

allowed by DNA shuffling with the recombination of entire genomes normally associated with conventional breeding (43, 44). Genome shuffling allows the recombination of beneficial mutations acquired by different chemical mutagenesis techniques or natural evolution, and permits to clear out deleterious mutations accumulated in the engineered strains. Another advantage of protoplast fusion is that it does not require organisms that are deficient in recombination and restriction-modification systems.

The most common procedure for genome shuffling is protoplast fusion. In protoplast fusion, the cell wall of the microorganism is digested using a glycoside hydrolase, such as lysozyme. In case of gram-negative bacteria, the protoplast formation must be aided by the addition of EDTA (45). Protoplasts are maintained in buffers with high osmotic pressure, to avoid lysis of the protoplast due to the absence of cell wall. Protoplast fusants are created using polyethylene glycol to increase the osmotic pressure, promoting the protoplast fusion.

Genome shuffling has been applied in the study of several complex phenotypes such as the improvement in acid tolerance in *Lactobacillus* (44) and n-butanol tolerance in *E. coli* (43), as well as enhancing in the production of lactic acid in *Lactobacillus rhamnosus* (46).

2.2.4.2. Global transcription machinery engineering (gTME)

gTME is an approach that perturbs the whole cell transcriptional machinery in subtle ways, using a mutant transcriptional factor that enables expression of polygenic phenotypes, helping to uncover cellular phenotypes important for technological applica-

tions. Mutations present in sigma factors and transcription factors will affect the promoter preference of RNA polymerase, modulating the transcriptome at a global level (47). This technique is based in the generation of a mutant library of a gene (or genes) expressing transcription or sigma factor since it is difficult to predict *de novo* the interaction the mutation protein and its DNA sequence. Thus, it is possible to explore a large protein space, increasing significantly the screening effort. A general gTME methodology is depicted in Figure 2.4. gTME has been successful in the study of several complex phenotypes, including ethanol tolerance (47), lycopene production (47), and improving xylose fermentation (19).

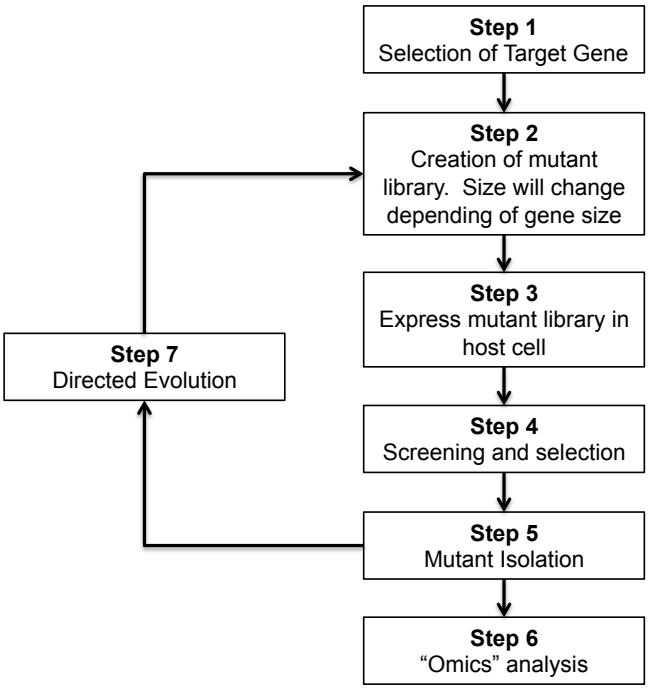


Figure 2.4. General gTME methodology. Depending of the goal of the project, steps 6 and 7 can change. Figure adapted from Lanza and Alper (48)

2.3. Laboratory Adaptive Evolution

Most of the industrially relevant complex traits in microbial systems involve multiple mechanisms, usually by expression of genes located in different loci. Since the identities of these genetic determinants are generally not known, making the rational engineering of strains for these complex phenotypes is a challenging process. Classical strain engineering for these traits generally involve several rounds of random mutagenesis followed by tedious and time-consuming selection. After successive rounds of induced mutagenesis, mutations that are deleterious or have negative epistatic effects tend to accumulate by hitchhiking with beneficial alleles, hindering the identification of beneficial mutations.

Laboratory adaptive laboratory evolution (LAE) allows the variation of genetic material, in companion with the natural selection of beneficial mutations in an unbiased fashion (49). LAE has been implemented in several applications, including activation of latent genetic functions (50), production of nonnative products (51, 52), increasing the metabolization rate of utilizable substrates (53, 54), and improvement of resistance to different chemicals (21), antimicrobials (55) and environmental conditions (56, 57). As aforementioned, LAE is based in the Darwinian principle of natural selection. The process of natural selection has four components:

- Genetic variation within the population, producing differences in diverse phenotypes. Mutations are the source of such genetic variation.
- Inheritance. The traits must be passed on to the next generation.

- An increment in population growth.
- A differential survival and reproduction, *a.k.a.* fitness advantage, must be provided for the acquired beneficial mutation(s).

If the desired trait can be coupled with growth, LAE can be used to improve the desired phenotype. This process is accomplished by applying a selective pressure so that beneficial mutants (mutants with increased fitness) can be obtained through the process of natural selection. The identities of the mutations residing in adaptive mutants obtained through natural selection or mutagenesis, and their subsequent effects on cellular processes, must be leveraged for further rational engineering. With advances in genomic tools, the genes and mechanisms involved can now be identified using combinations of whole-genome re-sequencing (58, 59), transcriptomics (60, 61), proteomics (62), and metabolomics (63, 64) studies. Detailed molecular characterization of adaptive mutants isolated from *in vitro* LAE experiments provides insights into the adaptive landscape for the phenotype of interest. Characterization of the adaptive landscape significantly enhances the knowledge on the important parameters underlying complex phenotypes needed for the rational engineering of strains.

Currently, most of the applications of LAE in strain engineering have used the workhorses of metabolic engineering: *S. cerevisiae* and *E. coli*; however, applications for other microorganisms are on the rise.

2.3.1. Adaptive landscape

In LAE, clones are typically isolated from the evolving population after an arbitrarily elapsed time or at the end of the experiment. However, since an evolving population is heterogeneous, interclonal competition (clonal interference (54, 65)) lead to the extinction of beneficial mutants. Depending on the population structure during the course of evolution, the random isolation of adaptive mutants may fail to identify some adaptive mutations that arise during the course of the evolution.

The idea of an adaptive landscape was first introduced as "surfaces of selective value" by Sewall Wright in his studies about the shifting balance theory in 1931 (66-68). The adaptive landscape is a multi-dimensional surface representation of the biological fitness of an organism in a particular environment. In an adaptive landscape map for a specific condition, each genotype is correlated with a fitness value (see Figure 2.5).

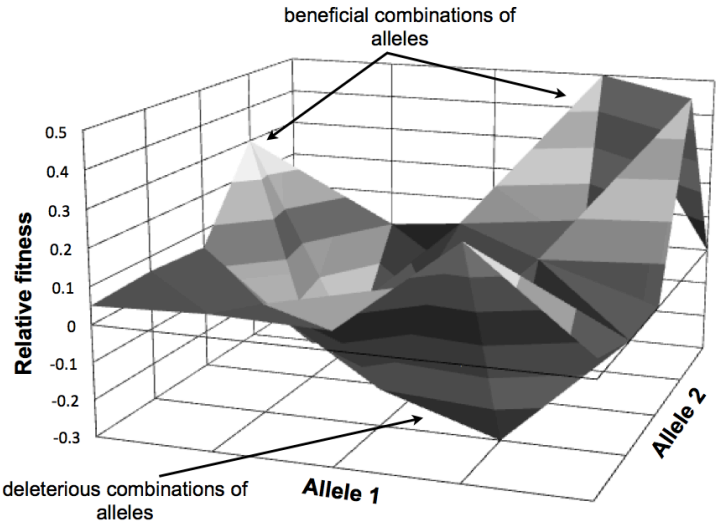


Figure 2.5. Simplified adaptive landscape for two alleles (for one background genotype in one condition). The figure depicts fitness values for beneficial (positive relative fitness values) and deleterious (negative relative fitness values) combinations of alleles.

The key point is that adaptive landscapes can differ in the ruggedness. The resulting landscape can be smooth with a single optimum where the evolving population is required to acquire a specific set of mutations, or can be rugged where the accessible local optima will depend on the starting point within the landscape. It has been demonstrated that bacteria encounter both types of landscapes in evolution experiments (69, 70). Natural selection usually drives a population to the closest local optimum, but not necessarily the global optimum. Evolving populations tend to be trapped in suboptimal solutions in asexual systems (71). Thus, to reach the global optimum, processes that allow for large "jumps" in the adaptive landscape, such as recombination and horizontal DNA transfer, are necessary to reach new regions of the adaptive landscape in a semi-rational manner. Recombination allows the combination of beneficial mutations with positive synergy and the removal of deleterious mutations acquired in the evolutionary process while horizontal gene transfer allows the acquisition of new functions.

2.3.2. Theories governing population structure during asexual evolution

Numerous theories have been proposed for the population structure in *in vitro* laboratory adaptive evolution experiments. Several factors, including the selective pressure, size of the population, rate of mutations, frequency of beneficial mutations, and relative fitness of beneficial mutants, are involved in determining the population structure during evolution. In the simplest case, a well-adapted mutant rises in the population, and due to its increased fitness compared to the background population, the genotype will expand and eventually replace the parental population. This population struc-

ture is applicable to situations where: (i) the evolution is mutation-limited, (ii) the population size is small, and (iii) the time between the establishments of successive mutations is much larger than the time it takes for a beneficial mutant to fix in the population (strong positive selection). This theory, called *clonal replacement* (also called succession-fixation regime or strong-selection weak-mutation regime), implies that only one mutation can become fixed at a time, leading to successive complete selective sweeps (depicted in Figure 2.6).

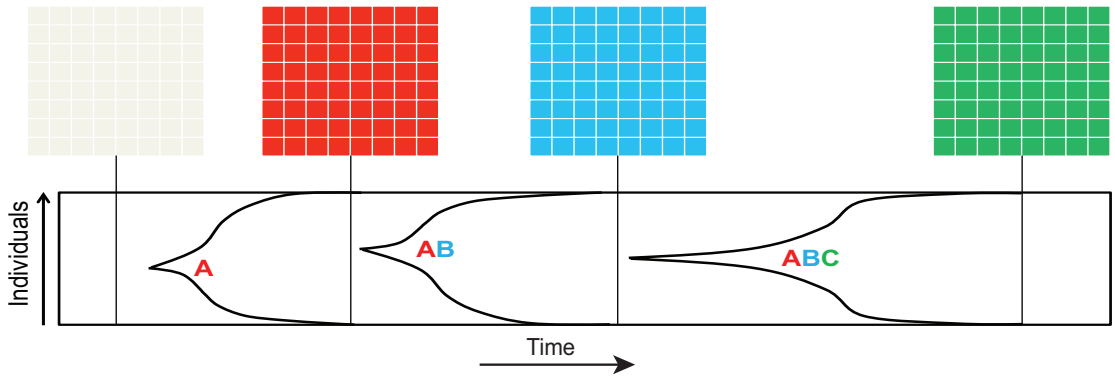


Figure 2.6. Clonal replacement model, theory governing population structure during asexual evolution where successive sweeps and fixation of different beneficial mutations take place in a small population; snapshots of the genotypes at different elapsed times show that the population is homogeneous except when the beneficial mutant is sweeping through the population. The graph represents the population structure as a function of time during asexual evolution. The capital letters represent different beneficial mutations in the population. The gridded boxes represent a snapshot of the frequency of different genotypes in the population at that one point in time.

The resulting population can be assumed to be homogeneous except during the periods when the beneficial mutant is sweeping through (72). However, when the mutations are established at a faster rate than the rate of fixation, multiple mutant lineages can coexist and compete for resources until one with the largest fitness advantage outcom-

petes all the other genotypes and become the next founding genotype for subsequent evolution. This theory, known as *clonal interference* (or one-by-one clonal interference), assumes that a single mutation can be fixed at a time, producing heterogeneous populations except immediately after the sweeping of the fittest mutant (depicted in Figure 2.7); this theory focuses on the competition between different mutations with positive relative fitness (73-78).

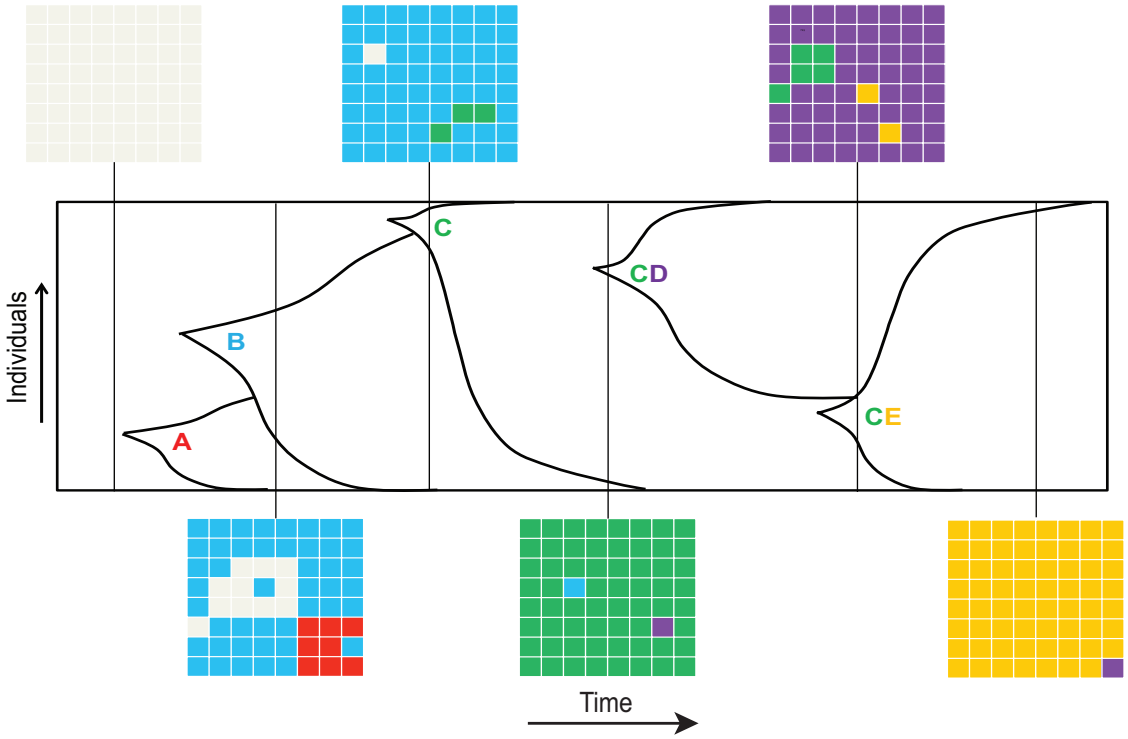


Figure 2.7. Clonal interference model, where different adaptive mutants compete until one with the largest fitness advantage sweeps through and becomes the founding genotype for subsequent evolution (e.g., mutations A, B, and C compete until C completely takes over the population). The graph represents the population structure as a function of time during asexual evolution. The capital letters represent different beneficial mutations in the population. The gridded boxes represent a snapshot of the frequency of different genotypes in the population at that one point in time.

The two theories described above assume that only one beneficial mutation can be fixed at a time. However, if the population size is large enough or the rate of mutation is high enough, multiple mutations can occur in the same lineage before fixation, leading to the multiple-mutation model (79) (Figure 2.8). The importance of this third theory on population structure has been demonstrated in several theoretical and experimental studies (65, 80, 81). In general, the population size in laboratory conditions is large enough that either one-by-one clonal interference or multiple mutations models shape the population structure.

2.3.3. Factors influencing population dynamics

As mentioned above, factors such as mutation rate, relative fitness advantage, population size, and rate of beneficial mutations are important in shaping the population dynamics during evolution. Since the evolution dynamics is dependent on the mutation rate, one would assume higher mutation rate to be advantageous for speeding up evolution by generating mutational diversity. However, an increase in mutation rates does not necessarily accelerate the pace of adaptation (82). While a low mutation rate would result in a slow discovery of beneficial mutations, prolonged exposure to high mutation rate (such as the use of a mutator strain) increases the occurrence and accumulation of deleterious mutations as well as the hitchhiking of apparent "silent" mutations during the course of evolution, increasing the genetic load (83-85). This is evidenced by the rarity of mutator strains in Lenski's long-term laboratory adaptive evolution experiment with *E. coli*; where mutators were found only after thousands of generations of evolution

(82, 86-89) and the fitness advantage conferred by the mutator strains is most likely a result of overcoming a mutation-limited bottleneck during the evolution. Mutagens are often used to increase genetic diversity in evolution experiments. However, since it is not convenient to periodically mutagenize the evolving population, a controllable mutator system can be used, where the expression of mutator alleles can be induced only when needed (90).

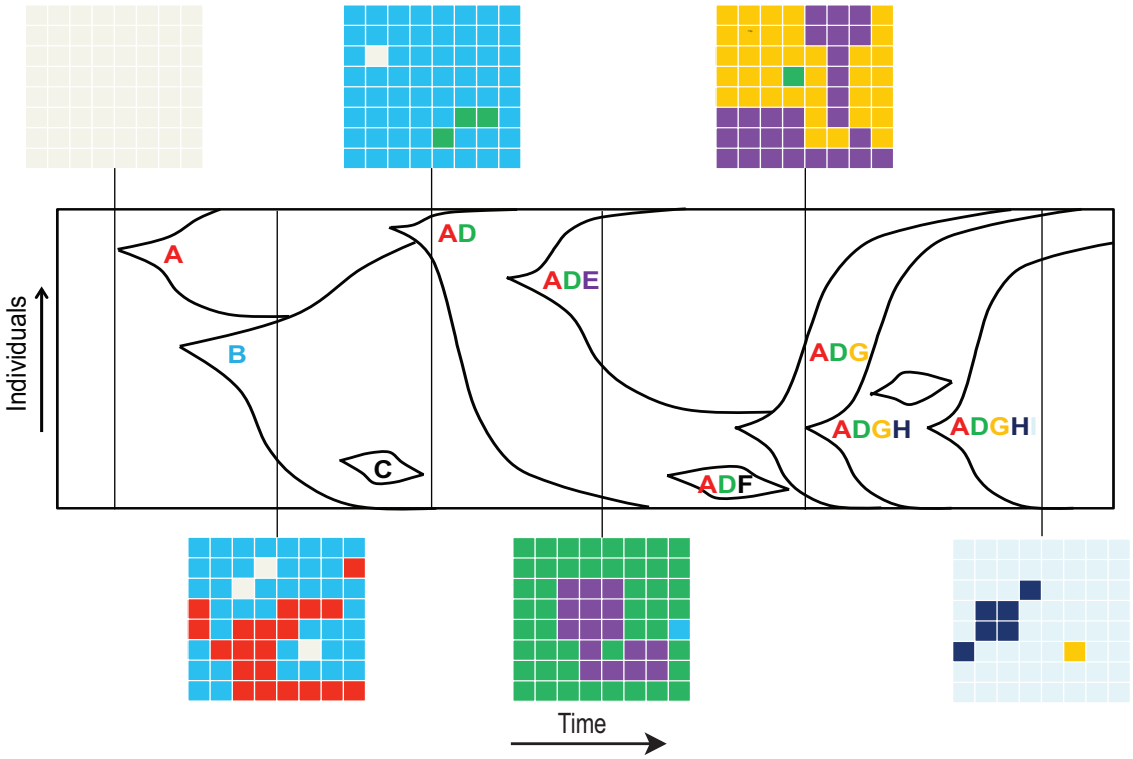


Figure 2.8. Multiple mutations model, where multiple mutations occur in the same lineage before fixation. In the latter two population structures, some adaptive mutations are lost from the population, and depending on when adaptive mutants are isolated, some mutants (and thus the underlying molecular mechanisms for adaptation) may not be identified. The graph represents the population structure as a function of time during asexual evolution. The capital letters represent different beneficial mutations in the population. The gridded boxes represent a snapshot of the frequency of different genotypes in the population at that one point in time.

The time it takes a beneficial mutation to become the majority in the population is called the fixation time and is an important factor in determining the population dynamics during evolution. This fixation time depends mainly on two factors, genetic drift and the fitness advantage of the beneficial mutation in comparison with the background, and is inversely proportional to the relative fitness advantage of the beneficial mutant (91). A beneficial mutation with a 10% relative fitness advantage will become the majority of the population after approximately 250 generation in serial batch transfer experiments (83) and 100 generations in continuous culture experiments (84). Genetic drift is defined as the probability that a beneficial mutation survives extinction (92). In *in vitro* laboratory adaptive evolution experiments, the main source of drift is genetic bottleneck due to random sampling. This phenomenon takes place when a significant amount of the population suddenly vanishes, as occurs when a fresh batch culture is inoculated from an overnight culture. The survival probability of an allele carrying a beneficial mutation that arose in the culture will depend on its proportion in the culture at the time of transfer and the amount of inoculum transferred; therefore there is a chance that it could be completely lost due to the stochasticity of sampling. In evolution experiments using serial batch transfers, genetic bottlenecks between transfers affect heterogeneity by transferring a small fraction of the population. A reduction of the effect of genetic bottleneck could be achieved by using continuous culture systems such as chemostats or turbidostats (20), where a much smaller genetic bottleneck is present.

2.3.4. Specificity of adaptation

Propagation of microbial systems commonly shows tradeoffs in their relative fitness across different environments. Evolution under a homogenous set of conditions tends to promote the fixation of specialist strains, microorganisms that can only thrive in a small set of environmental conditions (Figure 2.9). On the other hand, if the evolution is carried out under a mix of environments, individuals could acquire a more generalist behavior, being able to thrive in a wide variety of growth conditions, however, such benefit is sub-optimal in any growth condition.

Two different mechanisms have been associated with this underlying cost of adaptation. The first is antagonistic pleiotropy. Antagonistic pleiotropy makes reference to mutations that are beneficial in one condition, being deleterious another (93). A second mechanism is the accumulation of mutations by genetic drift. Such mutations are neutral in the selection environment but harmful elsewhere. Tradeoffs are very common in nature, however the underlying mechanisms are usually unknown (94-98).

2.3.5. Visualizing evolution in real-time (VERT)

As stated above, the population sizes in most *in vitro* evolution experiments are large enough to result in heterogeneous populations due to the effects of clonal interference and multiple mutations. Thus, laboratory adaptive experiments can significantly benefit from a more systematic isolation of adaptive mutants and ramping-up schedules for selective pressures. The VERT system was developed to address these limitations in traditional adaptive evolution experiments (21, 55, 99, 100). The basis for VERT is the

use of isogenic, but differentially labeled (typically with fluorescent proteins) strains to seed the initial evolving population. As a beneficial mutant arises and expands in the population, the colored subpopulation that it belongs is expected to increase in proportion. Using fluorescent activated cell sorting (FACS), the relative proportions of each of the colored subpopulations at each point in time can be measured. Each sustained expansion in the proportion of a colored subpopulation is called an "adaptive event". Thus, the tracking of the different colored subpopulations can serve as a tool for determining when a fitter mutant arises in the population.

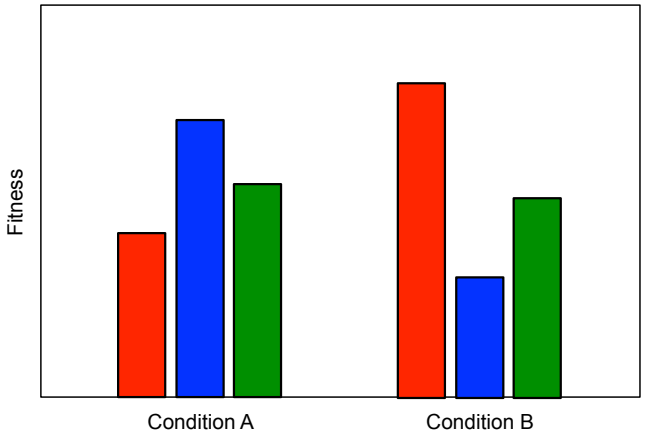


Figure 2.9. Specificity of adaptation. Hypothetical scheme of three genotypes under two different experimental conditions. The columns indicates the fitness associated with every genotype under the specified conditions. The blue and red bars show specialists strains evolved to the conditions A and B, respectively. The green column indicates a generalist strain with relatively high fitness under both growing conditions. Figure adapted from Elena and Lenski (101).

The relative subpopulation frequency data collected throughout the course of adaptive evolution represent the history of the population. The observed increase in the relative proportion of a colored subpopulation from consecutive data points is assumed

to be the result of the expansion of an adaptive mutant. Therefore, adaptive mutants can be isolated from samples based on the observed expansions and contractions, by sorting out the colored subpopulation that is expected to contain the adaptive mutant of interest. Since experimental data can suffer from noise, the identification of adaptive events may be challenging. Visual inspection of the data to identify adaptive events is a crude, but relatively effective method (54, 55, 99). However, since small changes in relative frequencies may be difficult to distinguish from noise, computational methods will provide less biased annotation of adaptive events; our group recently developed a supervised learning method for analysis of VERT data (102).

The basic feature of the VERT system, the number of labeled subpopulations, is the aspect that can most readily be manipulated directly by the experimentalist, but is somewhat restricted by the available equipment and properties of the labels themselves. The number of fluorescent markers used represents distinct subpopulations that can be visualized during the course of an evolution experiment. VERT labels must have distinguishable emission spectra and preferably have no significant fitness effect in the condition of interest. Widely used fluorescent proteins such as GFP, YFP, and RFP can be detected on most FACS machines and usually have little effect on the physiology of their host strains. At a minimum, two labeled subpopulations are trivially required to observe population dynamics. Three subpopulations, employing RFP, GFP, and YFP labeled strains, have been used successfully (54, 55) in fungal systems, and a two-color labeled system in *E. coli* (21). Additional subpopulations can be included if suitable

equipment is available. Simulated evolution may prove a useful tool for unraveling the connection between adaptive event discovery and initial population diversity.

Visualizing evolution in real-time-based *in vitro* adaptive evolution experiments can be used in either serial batch transfer or continuous culture systems. Provided that the different fluorescently marked strains show no significant fitness bias, then equal proportions of each strain maybe used to seed the population for evolution. Samples are then withdrawn and quantified using FACS every few generations to track the population dynamics. It is typically assumed that the adaptive mutant will expand until a fitter mutant arises in another subpopulation and expands sufficiently to impede its' expansion. It is further assumed that the generation at which the expanding subpopulation has reached a maximum proportion will contain the largest fraction of the adaptive mutant responsible for the expansion, simplifying the isolation of the mutant considerably.

In traditional adaptive evolution experiments, selective pressure is generally ramped-up at arbitrarily chosen time intervals. An alternative to this approach, based on using a feedback controller to maintain selective pressure so that the overall population growth rate approaches a user-defined set point, was recently introduced by Toprak *et. al.* in 2012 (59). Since the use of VERT allows the users to readily identify when adaptive events occur, it can be used to design a more systematic ramp-up schedule. For example, an increase in selective pressure could be applied when a minimum of 2 adaptive events are observed. The optimal frequency of ramp-up as a function of observed number of adaptive events may differ depending on the adaptive landscape for the phenotype of interest and needs to be investigated.

The isolated adaptive mutants can be further characterized to elucidate the molecular mechanisms of resistance in the selective pressure of interest. Whole-genome re-sequencing, transcriptomics, proteomics, and metabolomics analyses can be used to elucidate the evolutionary trajectories during the process of adaptive evolution. The availability and cost of whole-genome re-sequencing has improved significantly, but in most cases is still more expensive than transcriptome analysis using DNA microarrays. VERT tracks the individual subpopulations, making it easier to distinguish whether genome-wide perturbations observed in the transcriptional regulation found in different isolates arose independently or transitively without whole-genome re-sequencing data (if the isolates come from different colored subpopulations). Since not all the observed perturbations are involved in the complex phenotype of interest, common perturbations observed in independent lineages provide a level of confidence for their involvement. The potential adaptive mechanisms identified can serve as targets for further strain engineering.

The original development of VERT used the yeast *S. cerevisiae* evolving under glucose-limited conditions; a three-colored VERT system was used to seed eight parallel populations (54). The VERT data from one of the populations is shown in Figure 2.10; generations and subpopulations from which adaptive mutants were isolated from are indicated. Detailed genotypic and transcriptome analyses of the isolated adaptive mutants showed convergence in the perturbation of the protein kinase A regulatory network in independent lineages (54). Subsequent development and application of a two-colored

VERT system in *E. coli* for *n*-butanol tolerance revealed previously undiscovered resistance mechanisms (21).

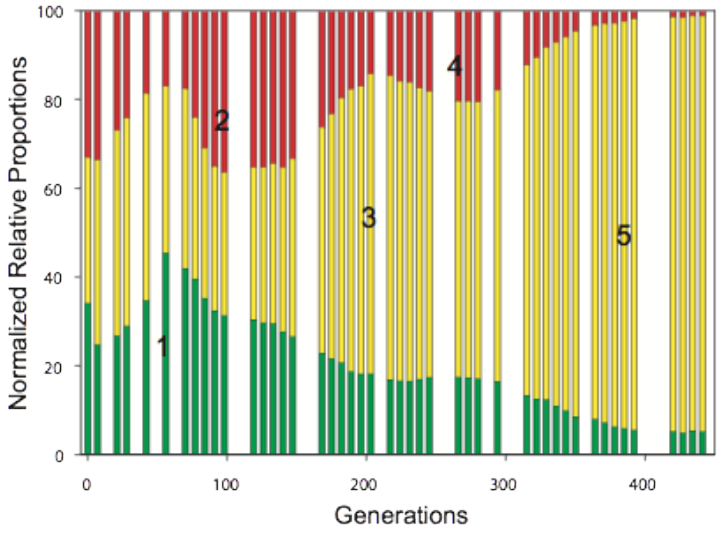


Figure 2.10. Example population dynamics from a three-colored VERT system (adapted from Kao and Sherlock (54)). The colored bars represent the relative proportions of each colored subpopulation. An increase in the relative proportion of a colored subpopulation is indicative of the occurrence and expansion of an adaptive mutation in that subpopulation, and is defined as an adaptive event. Under the assumption that the adaptive mutant responsible for the specific adaptive event is at its' highest proportion at the end of the sustained expansion, the adaptive mutants are isolated from the expanding subpopulation from the generation at the end of each expansion. The generation and colored subpopulation from which adaptive mutants were isolated are numbered 1–5.

3. VISUALIZING EVOLUTION IN REAL-TIME TO DETERMINE THE MOLECULAR MECHANISMS OF N-BUTANOL TOLERANCE IN *ESCHERICHIA COLI*[*]

3.1. Summary

Toxicity of products or feedstock components poses a challenge in the biocatalyst-based production of fuels and chemicals. The genetic determinants that are involved in increased resistance to an inhibitor form the adaptive landscape for the phenotype; so in order to engineer more robust biocatalysts, a better understanding of the adaptive landscape is required. Here, we used the adaptive laboratory evolution method called Visualizing Evolution in Real Time (VERT) to help map out part of the adaptive landscape of *Escherichia coli* tolerance to the biofuel n-butanol. VERT enables identification of adaptive events (population expansions triggered by adaptive mutants) via visualization of the relative proportions of different fluorescently labeled cells. Knowledge of the occurrence of adaptive events allows for a more systematic isolation of adaptive mutants while simultaneously reducing the number of missed adaptive mutants (and the underlying adaptive mechanisms) that result from clonal interference during the course of in vitro evolution. Based on the evolutionary dynamics observed, clonal interference

[*] Reprinted with permission from Luis H. Reyes, M. Priscila Almario, James Winkler, Margarita M. Orozco, and Katy C. Kao, (2012). "Visualizing evolution in real time to determine the molecular mechanisms of n-butanol tolerance in Escherichia coli". *Metabolic engineering*, 14(5), 579–590. DOI: 10.1016/j.ymben.2012.05.002. Copyright 2012 ELSEVIER

was found to play a significant role in shaping the population structure of *E. coli* during exposure to n-butanol, and VERT helped to facilitate the isolation of adaptive mutants from the population. We further combined adaptive laboratory evolution with genome shuffling to significantly enhance the desired n-butanol tolerance phenotype. Subsequent transcriptome analysis of the isolated adaptive mutants revealed different mechanisms of n-butanol resistance in different lineages. In one fluorescently marked subpopulation, members of the Fur regulon were upregulated; which was not observed in the other subpopulation. In addition, genome sequencing of several adaptive mutants revealed the genetic basis for some of the observed transcriptome profiles. We further elucidated the potential role of the iron-related genes in n-butanol tolerance via overexpression and deletion studies and hypothesized that the upregulation of the iron-related genes indirectly led to modifications in the outer membrane that contributed to enhanced n-butanol tolerance.

3.2. Introduction

Biochemical and biofuel production using microbial systems have been the renewed focus of innumerable research in the last decade. Among the many accomplishments in this area, metabolic engineering efforts have significantly improved the rate and yield of production (5-10, 23); however, the maximum final titer achievable is a function of the robustness of the microbial system in the production environment (11). The genetic determinants underlying tolerance form the adaptive landscape of the phenotype of

interest, are generally complex, and involve interactions between multiple genes (103). Further understanding of the adaptive landscape for these complex phenotypes of interest is necessary for the rational engineering of strains. However, these complex phenotypes are largely not well understood, and therefore pose a challenge for the rational design of more robust microbial producers.

Advances in genomic tools have led to the development of reverse engineering tools for the determination of the underlying genetic determinants (points on the adaptive landscape) and molecular mechanisms associated with complex phenotypes; such as transcriptome analysis (16, 17), genomic enrichment (12, 33), SCALEs (14), TRMR (18), and coupling high throughput genomic technology with the use of adaptive laboratory evolution (20-22). In adaptive laboratory evolution, adaptive mutants (strains with increased fitness) arise spontaneously and expand in a population under a specific selective pressure (22, 50, 104-108). The adaptive mutants harbor genetic mutations that render them fitter than the background population, allowing them to expand in the population. Identification of the underlying genetic mutations and their relative fitness effects (mapping the adaptive landscape) allows researchers to use adaptive evolution as a tool to identify important parameters for the rational engineering for the specific complex phenotype. In addition, the evolutionary trajectories can be elucidated using adaptive evolution to shed additional insight into the fundamental evolutionary processes involved in the acquisition of complex phenotypes. Typically, mutants are isolated from evolving populations for further analysis after an arbitrarily chosen elapsed time and/or at the end of the evolution experiment. However, depending on the population size, the

mutation rate, the rate of adaptive mutations, and the relative fitness associated with beneficial mutations, the population may consist of multiple beneficial mutants that co-exist and compete in a phenomenon known as clonal interference (109-113). Clonal interference results in a heterogeneous population and creates complex evolution dynamics during in vitro evolution, where beneficial mutations that arise in the population may be outcompeted and lost through the course of asexual evolution. Thus, depending on when samples are isolated, some beneficial mutations may not be identified, thus imposing a limit on our ability to map the adaptive landscape and in the determination of the trajectories involved during the evolution for the desired phenotype.

To help alleviate the limitation of traditional evolutionary engineering approaches, here, we demonstrate the use of a recently developed adaptive evolution method, called visualizing evolution in real-time (VERT) (54, 55), combined with genome shuffling, for the engineering and characterization of complex phenotypes in microbial systems. Using different fluorescently marked (but otherwise isogenic) strains, VERT can be used to readily determine when adaptive events occur (the expansion of adaptive mutants in the population) and to facilitate the isolation of those mutants from the population. Additionally, during adaptive evolution, genetic recombination between bacteria does not occur, and therefore beneficial mutations arising in different lineages cannot be combined synergistically to create a superior genotype. To overcome this limitation, we complemented our adaptive laboratory evolution experiment with genome shuffling to allow for recombination between mutant linages. Using this approach, we were able recombine beneficial mutations arising in different individuals in a manner analogous to

sexual evolution. Based on prior work by us and others, we hypothesize that butanol tolerance in *E. coli* is a complex phenotype, where the use of adaptive evolution will result in complex population dynamics with multiple beneficial lineages competing as a result of clonal interference, and the use of VERT combined with genome shuffling will help to enhance n-butanol tolerance and aid in the identification of additional molecular mechanisms involved.

3.3. Materials and Methods

3.3.1. Bacterial strains and plasmid construction

The *E. coli* K-12 strain, BW25113 (F^-, Δ(*araD-araB*)*567*, Δ*lacZ4787*(::*rrnB*-3), *lambda*$^-$, *rph*-1, Δ(*rhaD-rhaB*)568, *hsdR514*), was used in this study. The conditional replication, integration, and modular (CRIM) system (114) was used to integrate the plasmids into the genome. Overnight cultures were growth in Luria-Bertani (LB) medium or on solid LB agar plates supplemented with the appropriate antibiotics and incubated at 37°C. For the construction of the fluorescent-marked strains, the pBAD promoter on the plasmid pTB108ext (115) was replaced by the growth-rate regulated promoter *rrnB* P1 (116, 117), by PCR amplification from the *E. coli* genome using the forward primer 5'-GGC CAA GCT TCT ATA CAA ATA ATA ACT GCA GCC AAG-3' and the reverse primer 5'-GGC CAA GCT TGA GCT CGG CCG TTG CTT CGC AAC-3'. The engineered promoter has a decreased activity by approximately 50 fold in comparison with the unaltered promoter (-150 to +1 region), by using only the -50 to +1 re-

gion (117). The DNA sequences corresponding to the *rrnB P1* promoter, the GFP protein and *rrnB* terminator were cloned into the CRIM plasmid vector pAH144 in the multi-cloning-site. The CRIM plasmid was integrated into the genome using the HK022 phage attachment site, following the procedure described in Haldimann and Wanner (114). The gene encoding the yellow fluorescent protein (YFP) was amplified from the plasmid pKKGS5, used to construct GSY31137 (54), using the forward and reverse primers 5'-CCG GTC TAG ATC AAA GAT GAG TAA AGG AGA AGA ACT T-3' and 5'-CCG GCT GCA GGC GGC CGC CTA TTT GTA TAG-3' respectively, and cloned into the plasmid pAH144. The plasmid was integrated as described above.

3.3.2. Adaptive evolution experiment

The adaptive evolution experiments were carried out in two chemostat bioreactors for the n-butanol challenged experiments in M9 minimal medium supplemented with 5 g/L of D-glucose under aerobic conditions. The system was maintained at steady state with a constant volume of 30 ml, flow rate of 7 ml/h and temperature of 37°C. The concentration of n-butanol in the feed was increased in a step-wise fashion with an initial concentration of 0.5% (v/v) and raised to 1.3% (v/v) (approximate steady-state concentrations inside the bioreactor) during the course of the evolution experiment. The concentration of n-butanol was increased arbitrarily every 20 to 30 generations. At generation 88 the chemostats were restarted from glycerol stocks due to mechanical problems in the feeding peristaltic pump. During the re-start, the concentration of n-butanol was decreased while the chemostats reached steady state. At regular intervals (approximate-

ly every eight generations), samples were taken from the populations and a portion was stored at -80ºC in 17% glycerol for future analysis. The relative proportions of each of the two fluorescent populations were monitored using the fluorescent activated cell sorter (FACS) (BD FACScan^{TM}).

3.3.3. Genome shuffling

Protoplasts of *E. coli* were generated using a modified protocol of Dai *et al* (45). *E. coli* cells were grown at 37°C in LB medium until mid-exponential phase, harvested by centrifugation, washed three times using ice-cold 0.01 M Tris/HCl (pH 8.0) and resuspended in 0.01 M Tris/HCl pH 8.0 supplied with 0.5 M sucrose. Removal of outer membrane was initiated by using EDTA to a final concentration of 0.01 M. Cells were washed using SMM buffer (0.5 M sucrose, 20 mM sodium maleate monohydrate, 20 mM $MgCl_2$, pH 7.1) and resuspended in SMM buffer with 2 mg/ml lysozyme to remove the cell membrane, generating the protoplasts. The resulting protoplasts of *E. coli* were treated with DNase I (Promega) to digest DNA to prevent transformations due to the release of DNA from lysed cells, incubating at room temperature for 10 minutes. The protoplasts were harvested by centrifugation and resuspended in PEG buffer (SMM + 40% v/v PEG 6000, 10 mM $CaCl_2$, 5% v/v DMSO), and incubated for 6 minutes to allow the fusion of the protoplasts. The protoplasts were harvested again and resuspended in SMM buffer. Cells were plated in LB plates, supplied with 0.5M sucrose, at various dilutions and incubated at 37°C for 3 days. All cells were recovered from the plates, and used to initiate a new set of chemostats.

3.3.4. Isolation of mutants

To isolate the adaptive mutants from each of the observed adaptive events, eight clones were isolated from the expanding colored-subpopulation and pairwise competition experiments were carried out in chemostats fed with M9 minimal medium supplemented with 5 g/L D-glucose and 0.8% (v/v) n-butanol at 37ºC to measure the relative fitness coefficient of each clone against the previous adaptive mutant (the clones isolated from the first adaptive event were competed against a wild type expressing a different fluorescent protein). Figure 3.1 shows the flow of the procedure. The clone with the largest relative fitness coefficient was selected as the adaptive clone from the expanding subpopulation.

For each pairwise competition experiment, equal numbers of the two strains (expressing different fluorescent proteins) to be competed were used to seed each chemostat. The relative proportions of each colored subpopulation were tracked and measured approximately every 4 generations using FACS. The fitness coefficient was calculated as shown in Equation 3.1, where P_i is the relative proportion of strain *i* in the population and t_j is the generation at which the sample was analyzed.

Equation 3.1

$$s = \frac{\ln\left(\frac{P_{\text{Adaptive mutant @ } t_2}}{P_{\text{Adaptive mutant @ } t_1}} \cdot \frac{P_{\text{Reference @ } t_1}}{P_{\text{Reference @ } t_2}}\right)}{t_2 - t_1}$$

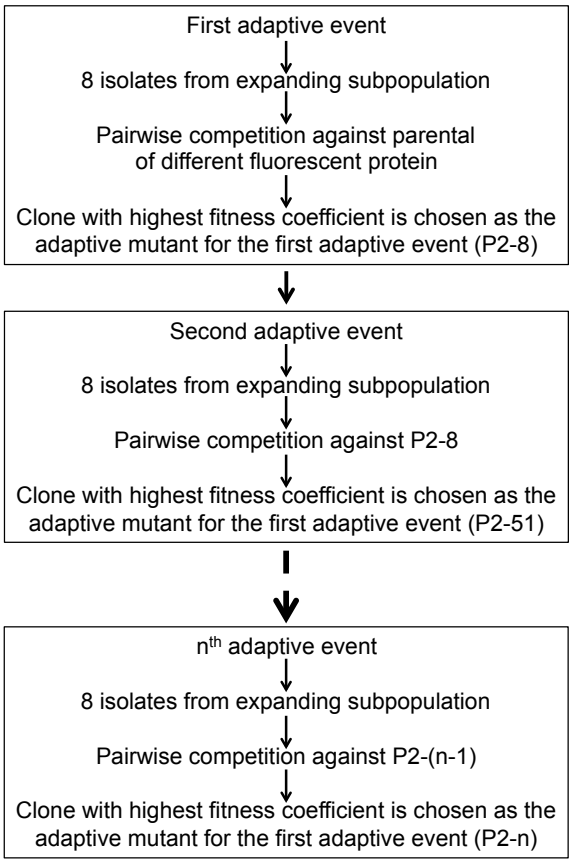


Figure 3.1. Procedure used for the identification of adaptive mutants

3.3.5. n-Butanol-shock experiment

An n-butanol-shock experiment was performed in 2% n-butanol (v/v). Cells isolated from each adaptive event were grown overnight at 37°C in M9 minimal medium supplemented with 5 g/L D-glucose, then diluted at a ratio of 1:100 into fresh M9 minimal medium, and grown at 37°C to mid-exponential phase (OD_{600} 0.6 ~1.0). The OD_{600} of each culture were normalized to 0.5 using fresh M9 minimal medium and n-butanol was added to a final concentration of 2% (v/v). Samples were incubated for 1 hour at 37°C and 220 rpm. After incubation, the samples were serially diluted in fresh M9 min-

imal medium and plated on LB plates. After 18 hours, pictures of the plates were taken to assess relative survival rates between the different strains.

3.3.6. Cell harvest for RNA extraction

Cells were grown in chemostats with a constant volume of 30 ml, flow rate of 7 ml/h, fed with M9 minimal medium supplemented with 5 g/L D-glucose and 0.8% (v/v) n-butanol at 37ºC. After the system reached steady state (approximately 36 hours), the cells were harvested by filtration using NALGENE analytical test filter funnels (Nalgene) and immediately resuspended in 10 ml of RNA*later* (Sigma) and stored at -80ºC for later processing.

3.3.7. Total RNA extraction

The total RNA was extracted using the RNeasy Mini Kit (QIAGEN). For each of the isolated mutants, 1.5 ml of sample stored at -80°C in RNA*later* were centrifuged and the pellet was recovered. Total RNA was extracted by following manufacturer's instructions. The extracted RNA was treated with DNase I and purified using acid phenol/chloroform followed by ethanol precipitation. The RNA was quantified using the Qubit fluorometer using the RNA quantification kit (Invitrogen).

3.3.8. Microarray hybridization and data analysis

The SuperScript indirect cDNA labeling system (Invitrogen) was used to generate cDNA incorporating amino-allyl dUTP. cDNA was recovered using ice cold ethanol

precipitation. Cy3- and Cy5-dUTP dyes (GE Healthcare) were used to label cDNA samples by following manufacture's protocol. The labeled cDNA was hybridized to the *E. coli* Gene Expression Microarray (Agilent Technologies). The arrays were scanned using the GenePix 4100A Microarray Scanner and images analysis performed using GenePix Pro 6.0 Software (Molecular Devices).

The Microarray Data Analysis System (MIDAS) software was used to normalize the data using LOWESS based normalization algorithm (118). Differentially expressed (DE) genes were identified using the rank product method with a critical p-value $P <$ 0.01 (119). The MeV (TM4) (120) microarray analysis software was used for clustering and other expression profile analysis.

3.3.9. Calculation of growth kinetic parameters

The parameters "Percentage of Inhibition" and "Improvement in n-butanol tolerance" were calculated as shown in Equation 3.2 and Equation 3.3, respectively. These parameters were determined by measuring the maximum specific growth rate ($\mu_{max,i}$) of each strain (strain *i*) in M9 minimal medium (supplied with 0.5% (w/v) D-glucose) supplemented with 0% and 0.5% (v/v) n-butanol. The growth kinetics for each strain was measured using a TECAN Infinite M200 Microplate reader (TECAN). Four biological replicas were obtained per sample. Student's t-test was used to determine if there was a significant improvement in the n-butanol tolerance in overexpression or deletion strains.

Equation 3.2

$$\text{Inhibition} = 1 - \left(\frac{\mu_{\text{Strain @ 0.5\% n-Butanol}}}{\mu_{\text{Strain @ 0\% n-Butanol}}} \right)$$

Equation 3.3

$$\text{Relative increase in } \mu_{\text{Clone}} = \frac{\left(\frac{\mu_{\text{Clone @ 0.5\% n-Butanol}}}{\mu_{\text{Clone @ 0\% n-Butanol}}} \right) - \left(\frac{\mu_{wild-type \text{ @ 0.5\% n-Butanol}}}{\mu_{wild-type \text{ @ 0\% n-Butanol}}} \right)}{\left(\frac{\mu_{wild-type \text{ @ 0.5\% n-Butanol}}}{\mu_{wild-type \text{ @ 0\% n-Butanol}}} \right)}$$

3.3.10. Overexpression and deletion studies

The ASKA Clone (-) collection (121), which contains a set of plasmids overexpressing all predicted *E. coli* K-12 ORFs under the control of an IPTG-inducible promoter, was used as the source of all plasmid constructs for overexpression studies. *E. coli* K-12 strain AG1 (*recA1, endA1, gyrA96, thi-1, hsdR17 (r_k^- m_k^+), supE44, relA1*) is the host for the ASKA collection. Plasmids of the clones from the ASKA Clone (-) collection overexpressing the genes of interest were isolated and transformed into *E. coli* strain BW25113. All the overexpression analyses in this study were also performed in AG1 to reduce strain-specific effects. For deletion studies, the Keio collection (122), which contains single gene deletions for all non-essential genes in BW25113, was used.

3.3.11. Whole genome resequencing

Genomic DNA (5 μg) was isolated from single colonies for each of the isolated mutants to be sequenced using Solexa technology (Illumina Inc., San Diego, CA, USA), with multiplexing of 6 strains per lane. DNA library generation and sequencing was performed by the Borlaug Genomics and Bioinformatics core facility at Texas A&M University. The software CLC Genomics Workbench 4 (CLC bio, Germany) was used to

assemble the different sequences and to identify single-nucleotide polymorphisms (SNPs) and possible genome rearrangements. Verification of possible SNPs and translocations identified from the genome re-sequencing were performed using traditional Sanger sequencing (MCLAB, South San Francisco, CA, USA).

3.3.12. Allele tracking

Genomic DNA was extracted using DNeasy Blood & Tissue Kit (Qiagen). The CFX384 Real-Time PCR Detection System (Bio-Rad) was used to carry out the qRT-PCR experiments. GoTaq® qPCR Master Mix (Promega) was used for all the quantitative real-time PCR (qRT-PCR) experiments using 20 ng of total genomic DNA and 0.5 μM of primers. The primers used for the tracking of the different alleles are: A) the *rhoS82F* allele, For-Wild-Type: CCG GTC CTG ATG ACA TCT ACG TTT C, For-Mutation: CCG GTC CTG ATG ACA TCT ACG TTT T, Rev: TTA CCA GAC GCT GCA TCT CGG, B) the *feoA201::IS5* allele, For-Wild-Type: CCA CGG CGC GCC ATT AC, For-Mutation: CCA CGG CGC GCC ATT AT, Rev: TCA TCG CGT TCG GTG GGA, C) the *FBSmut* allele, For: ATG CAA TAC ACT CCA GAT ACT GCG TGG, Rev-Wild-Type: CTG CGA TGA GAT GGT GGT CAG AGA ATA, Rev-Mutation: CAA GGG GTT GAT GAA AGA CGA TAA CCA AC, D) the *nusAE212A* allele, For-Wild-Type: TCT TCG CCG ATT TCT GGC ACT T, For-Mutation: TCT TCG CCG ATT TCT GGC ACT G, Rev: TCC TGC GCG AAG ATA TGC TGC, and E) the *relA142::IS2* allele, For-Wild-Type: GCA ATA CGC TCC GCC AGT TTG AT, For-Mutation: GGA GAT TCA GGG GGC CAG TCT A, Rev: AGC TGA AAG CGA

CGC ACA CTG. The thermocycler protocol used is: 95°C for 3 min followed by 39 cycles of 95°C for 10 seconds followed by 60°C for 30 seconds. The allele frequency calculations were done as described by Kao and Sherlock (54).

3.4. Results

3.4.1. Confirmation of neutrality and definition of adaptive events

The primary utility of the VERT method is to identify adaptive events (population expansions) by tracking the relative proportions of different fluorescent subpopulations during the evolutionary time-course. As a beneficial mutant arises and expands in the population, the corresponding labeled subpopulation increases in proportion. By tracking the different subpopulations using FACS, it is possible to determine when mutants with increased relative fitness arise in the evolving population; allowing us to isolate adaptive mutants from the populations in a more rational manner. Thus, it is important that any fitness biases in the fluorescent proteins used are taken into consideration in the analysis in order to avoid isolation of false adaptive mutants. The basal variability of relative subpopulation proportions from FACS measurements of the different fluorescently marked strains were analyzed using neutrality tests, which typically follow the short time propagation (<~20 generations typically) of the labeled subpopulations to detect any fitness differences arising from different fluorescent protein expression, as beneficial mutations are unlikely to occur within such a short time scale (123). To assess the neutrality between the two fluorescently marked strains used for this work (GFP

and YFP-labeled), we used the early time points from control experiments, where we seeded parallel continuous culture experiments with approximately equal numbers of the two fluorescently marked strains in M9 minimal medium supplemented with glucose. The data for the first 24 generations after the populations reached steady-state (> 6 volume replacements) were used to assess the variation in fitness between the two fluorescently marked strains (Figure 3.2); a slight increase in the yellow subpopulation was observed in both populations, suggesting a potentially slight fitness benefit in the yellow subpopulation either due to fluorophore differences or due to jackpot mutations present in the overnight inoculum. The neutrality data was then used to measure the expected levels of measurement noise.

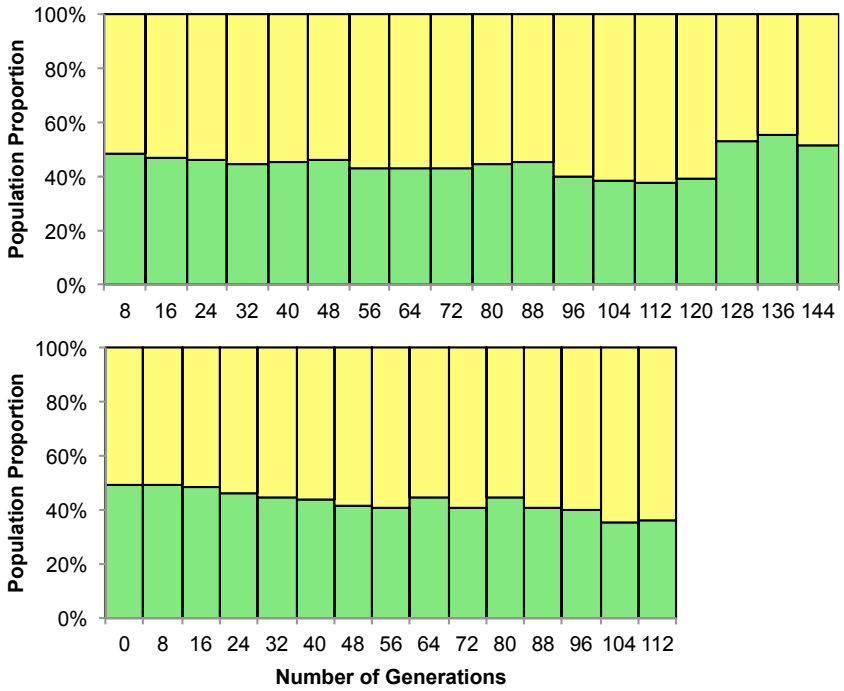


Figure 3.2. Neutrality test between GFP-marked and YFP-marked BW25113 conducted in the presence of M9 supplemented with 5g/L D-glucose in chemostats. Yellow bars: proportion of yellow subpopulation. Green bars: proportion of green subpopulation.

The analysis for estimating measurement noise focuses on the average change in population proportion per generation (μ_R) and its standard deviation (σ_R) within the neutrality dataset. Since no (or few) adaptive events occur during the first few generations of an evolution experiment, it is very likely that any change observed arises from random variations during measurement or the sampling procedure. Evolved populations can be also analyzed in a similar fashion using regions in which the subpopulations are in a meta-stable state (i.e. no net expansion or contraction). These meta-stable states may occur if there are no significant differences between the relative fitness of the adaptive mutants in the different colored subpopulations. If the fluorescent proteins are in fact neutral, the average slope should be close to zero ($\mu_R \approx 0$) to reflect equivalent fitness of the two colored sub-populations. These parameters together intuitively describe how much a given population proportion can be expected to fluctuate between subpopulations with equal fitness. Based on data generated during this study, $\mu_R = 0.001$ (95% CI: [0.0031, -0.0009]) and $\sigma_R = 0.004$ (N = 16) based on the assumption that deviations from the mean are normally distributed. These statistical inferences derived from the neutrality data may also be applied to determine if an arbitrary population expansion meets a statistically significant threshold (102) and will be used to identify adaptive events based on the evolutionary dynamics described below.

3.4.2. The evolutionary dynamics of *E. coli* during evolution for enhanced n-butanol tolerance

Approximately equal numbers of GFP and YFP marked strains were used to seed two n-butanol challenged populations (P1 and P2) in continuous cultures in chemostats with a starting concentration of 0.5% (v/v) n-butanol. The concentration of n-butanol in the chemostats was increased to 1.3% (v/v) in a step-wise manner over the course of approximately 144 generations (see Figure 3.3). The relative proportions of the two fluorescently marked subpopulations throughout the course of the evolution experiment were monitored using FACS (the evolutionary dynamics during the *in vitro* evolution are shown in Figure 3.3). An expansion in a colored subpopulation is indicative of the occurrence and expansion of an adaptive mutant in that subpopulation. Thus by tracking the expansions and contractions of the two colored-subpopulations, we were able to identify adaptive events occurring throughout the course of the *in vitro* evolution. The expansions and contractions in both colored-subpopulations are clear indications that clonal interference plays a role in shaping the population structure in our populations.

Upon completion of the adaptive evolution experiment, population P2 was arbitrarily chosen for further analysis. We identified seven adaptive events, where the expanding subpopulations reached their maximum proportion at generations 7, 32, 104, 112, 128, and 144, in this population via visual inspection. Between generation 32 and 80, a meta-stable region was observed, which led us to hypothesize that an adaptive event occurred in the green subpopulation that was able to impede the continued expansion of the yellow subpopulation at generation 32. Thus, generation 80 was also identi-

fied as the end of an adaptive event (in this case, the end of the meta-stable region). Using the computational algorithm, 4 adaptive events, reaching maximum proportion at generations 32, 104, 128, and 144, were identified (Figure 3.3A). We isolated adaptive mutants from all the potential adaptive events identified visually (all the computationally identified adaptive events were also identified visually).

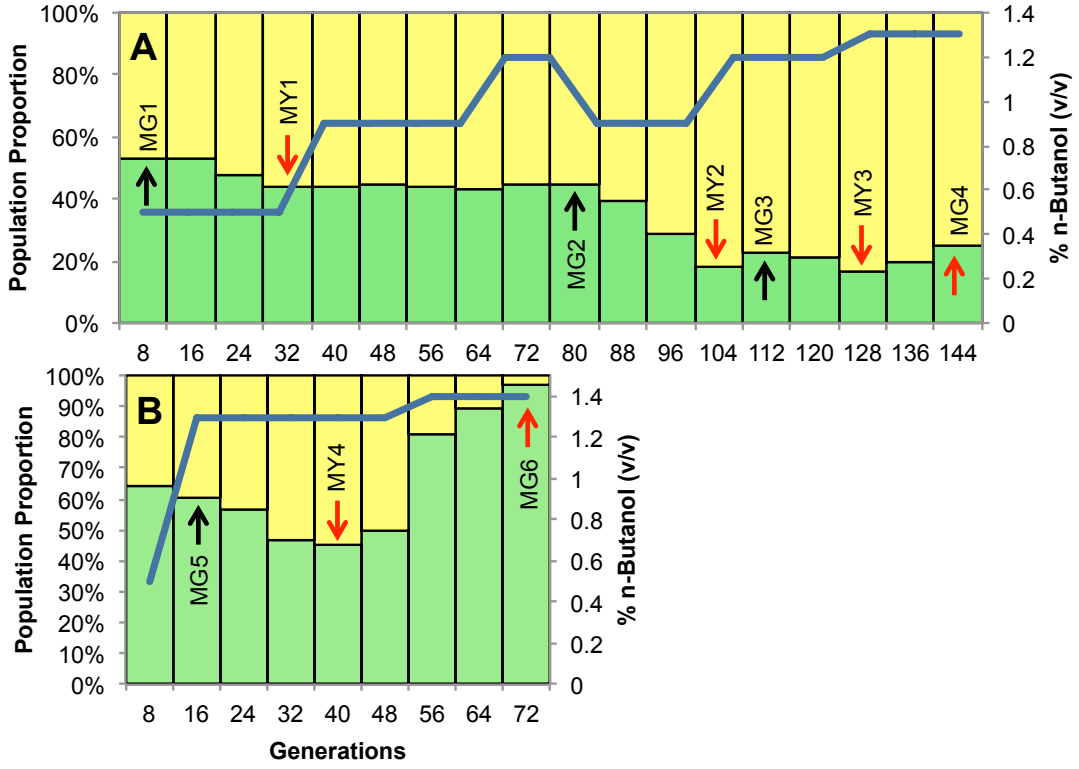


Figure 3.3. The evolutionary dynamics of the *in vitro* evolution for A) n-butanol challenged population P2, and B) the second round of evolution after genome shuffling in the presence of n-butanol. The yellow bars indicate the proportion of yellow subpopulation and the green bars indicate the proportion of the green subpopulation. The blue line indicates the approximate steady-state concentration of n-butanol in the chemostat. A visual inspection of the evolution dynamics allows us to determine different adaptive events, defined as an expanding subpopulation when such expansion reaches a maximum, marked with arrows. Isolated mutants from adaptive events are marked (MG: Mutant Green, MY: Mutant Yellow). The red arrows indicate the adaptive events determined using the computational algorithm.

Table 3.1. Isolated mutants from population P2.

Mutant	Fluorescence	Generation	Fitness Coefficient per generation vs previous isolated mutant	Fitness Coefficient per generation vs wild type strain
MG1	GFP	8	0.012 ± 0.004	0.011 ± 0.008
MY1	YFP	32	0.022 ± 0.005	0.013 ± 0.005
MG2	GFP	80	0.053 ± 0.006	0.008 ± 0.003
MY2	YFP	104	0.150 ± 0.001	0.051 ± 0.014
MG3	GFP	112	-0.117 ± 0.016	
MY3	YFP	128	0.193 ± 0.020	0.023 ± 0.020
MG4	GFP	144	-0.140 ± 0.076	
MG5	GFP	After Protoplast		0.210 ± 0.012
MY4	YFP	After Protoplast		0.162 ± 0.022
MG6	GFP	After Protoplast		0.332 ± 0.035

To isolate the adaptive mutants responsible for the observed expansions in population P2, eight clones were randomly picked for each of the observed adaptive events, from the expanding colored subpopulation at the generation when the relative proportions are at its maximum (or at the end of the meta-stable region) (marked by arrows in Figure 3.3A). Using pair-wise competition experiments in chemostats with 0.8% (v/v) n-butanol, the clone with the highest relative fitness coefficient against the previous adaptive mutant was selected as the adaptive mutant from that expanding subpopulation. These adaptive mutants are named according to their color and sequence in the evolution (labeled and shown in Figure 3.3A and Table 3.1). The relative fitness coefficient for each isolated adaptive mutant was also measured against a differentially marked parental strain (Table 3.1 and Figure 3.4). All but 2 (MG3 and MG4) of the isolated adaptive mutants showed positive or neutral fitness coefficients in the presence of 0.8% (v/v) n-butanol; MG3 and MG4 were not analyzed further.

3.4.3. Recombination significantly enhances the desired phenotype

At generation 144, the n-butanol-challenged populations began to washout at 1.3% (v/v) n-butanol, marking the end of the first round of adaptive evolution experiments. At this point, we subjected samples from the two n-butanol-challenged populations to genome shuffling via protoplast fusion (43, 45) in an attempt to further enhance tolerance by recombining the beneficial and potentially synergistic mutations between different lineages. The resulting protoplast fusion library was used to seed a new chemostat in the presence of n-butanol for a second round of evolution (Figure 3.3B). An enhancement in n-butanol tolerance was detected immediately, as the population was stably maintained at 1.3% (v/v) n-butanol (whereas they were washing out prior to genome shuffling). Adaptive mutants were identified and isolated from this second round of evolution, and were named MG5 (generation 16 post genome shuffling), MY4 (generation 40), and MG6 (generation 72).

3.4.4. Enhanced survival upon n-butanol shock

The survival rate after short-term exposure to a higher concentration of n-butanol was conducted to determine the tolerance level of the adaptive mutants. The cells were exposed to 2% (v/v) n-butanol for 1 hour and their viabilities were assessed (see Figure 3.5). Most adaptive mutants (MG2, MY3, MG5, MY4 and MG6) showed at least a ten-fold increase in survival rate upon n-butanol exposure compared to the parental strains. The adaptive mutants isolated after the genome shuffling (MG5, MY4, and MG6) showed ~10-100-fold increase in survival rate compared to the parental strains. This

corresponded to the higher relative fitness coefficient in these strains compared to the parental strains in the presence of n-butanol. Similarly, mutant MG1 that showed little increase in its n-butanol tolerance also showed a lack of relative fitness coefficient gain relative to the wild type. MY2 is an exception: it showed a decreased viability after 2% (v/v) n-butanol exposure, but had a higher fitness coefficient relative to the previous adaptive mutant (Figure 3.4), suggesting that this particular mutant either only has a higher resistance to lower concentrations of n-butanol (*i.e.* the concentration in the bio-reactor that it was isolated from) or it may have mutation(s) that confer a fitness advantage for more efficient nutrient assimilation.

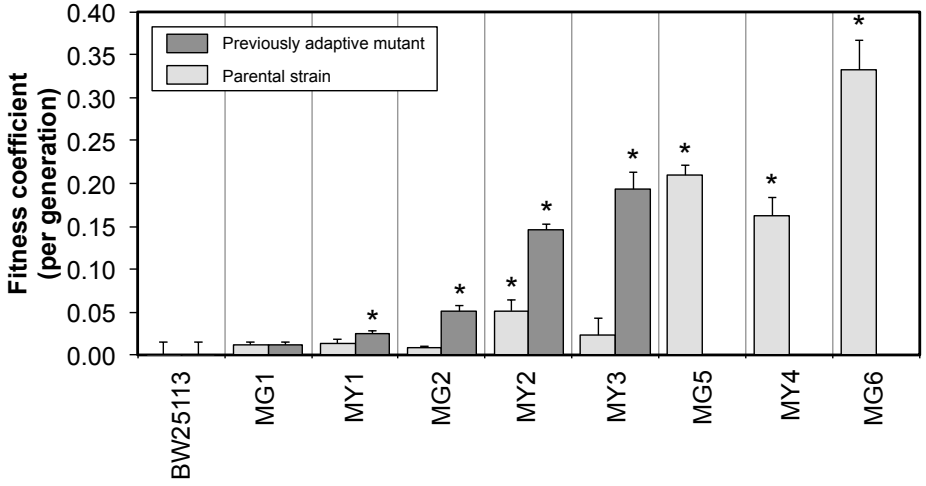


Figure 3.4. The relative fitness coefficients of isolated adaptive mutants from population P2 in the presence of 0.8% n-butanol. Light gray bars: the fitness coefficients compared to the previously isolated mutant as reference. Dark gray bars: the fitness coefficient compared to a differentially labeled parental strain. Measurements that are statistically significantly different from the wild type (student t-test with a p-value cut-off of 0.05) are marked by asterisks.

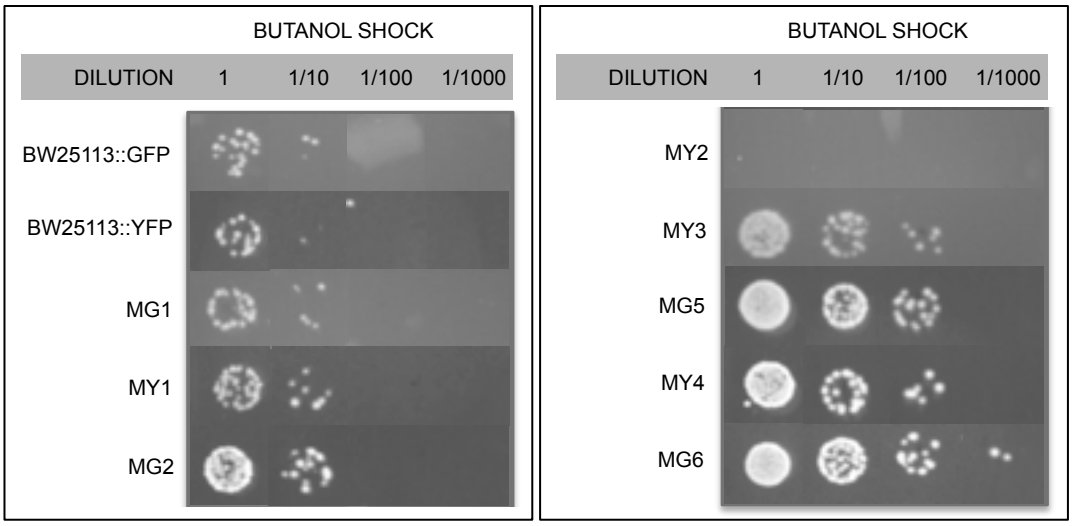


Figure 3.5. Relative cell viability of fluorescently marked wild type and isolated adaptive mutants after short-term exposure to 2% (v/v) n-butanol.

3.4.5. Transcriptome analysis

DNA microarrays were used to assess the transcriptional changes in the isolated adaptive mutants compared to the parental strains in response to n-butanol stress. The transcriptome data showed some similarities and also some major differences between the adaptive mutants isolated from the two lineages (GFP vs. YFP). Some of the key transcriptional perturbations and subsequent experimental validations are described below.

3.4.5.1. Changes in regulation of fatty acids and membrane composition

We found significant expression changes in genes involved in cell wall and membrane biosynthesis in every isolated mutant analyzed. Table 3.2 summarizes the differentially regulated genes related to fatty acids and membrane composition in each isolated mutant. Genes related to the biotin biosynthetic pathway were upregulated in

several adaptive mutants in both lineages (MY1, MG2 and MG5). Biotin is an essential cofactor for acetyl-CoA carboxylase, a key enzyme in the biosynthesis of fatty acids (124). The biotin-related genes were also found to be enriched in a recent study using an *E. coli* genomic library to identify genes involved in n-butanol resistance (12). Overexpression of the gene *bioA*, which encodes the 7,8-diaminopelargonic acid synthase enzyme in the biotin biosynthesis pathway in the wild type strain, showed significant improvements in its n-butanol tolerance (30±4%) (Figure 3.6).

The lipopolysaccharide (LPS) biosynthesis genes were also differentially regulated in all the isolated adaptive mutants (Table 3.2). The bacterial LPS is the main component of the outer membrane, composed of a hydrophobic domain (lipid A), a phosphorylated oligosaccharide, and an O-antigen; however, *E. coli* K-12 normally does not make O-antigen due to a mutation in the *rfb* gene cluster (125). Increase in the expression of genes involved in lipid A production were observed in adaptive mutants from both lineages (MY3, MG5, MY4 and MG6), suggesting a possible change in the hydrophobicity of the cell wall in these strains.

Changes in the expression of many enterobacterial antigen related genes were also observed (Table 3.2) and may also contribute to changes in cell wall hydrophobicity. Increase in the LPS content may lead to less hydrophobic cell surfaces, hindering the penetration of hydrophobic compounds into the cytosol, and increasing its tolerance to organic solvents (126).

Table 3.2. Selected genes involved in fatty acid or membrane composition that were upregulated (bold) or downregulated (non-bolded) in the isolated adaptive mutants.

Strain	Function						
	A*	B†	C‡	D§	E**	F††	G‡‡
MG1							***yhhK***
MY1		***rffM***		*plsB, mdoB, plsX*	***mraW, bamB***	***bioA***	
MG2		***rffH***, ***glmM***, ***basR***, *lpxB, lpxD, lpxH*	*mrdA, prc, amiA, mtgA, mltA*		***lplA, mraW, borD, yqeF***	***bioA***, ***bioB***, ***bioC***, ***bioF***	*panD, coaA*
MY2		*rffD, htrL,*	*ddlA, murD, oppA, oppC, oppD*	*ugpC, ugpE*			
MY3		***lpcA***, ***lpxK***	***bacA***,	*psd, plsX*	***ybfP, lplA,*** *yqeF*	*bioA*	
MG5	***cfa***	***lpxL***, *kdsA, glmS, wbbJ, rfaG, rfaP*	*prc, mrcA*	*ybgC*	***lplA, ycfL***, *csgG, bamB, smpA*	***bioA***	
MY4	***scpA, lipB, mhpE***	***glf, rfbD, rffA, rffM, lpxK,*** *gmm*	***murE***, ***murF***, ***amiB***, *murB, hipA, oppD*		***lplA,*** *ycfL*		
MG6	***cfa***	***lptA, basR,*** *gmm*			***lplA***		

* Fatty acid biosynthesis

† Lipopolysaccharide biosynthesis

‡ Peptidoglycan synthesis

§ Phospholipids biosynthesis

** Lipoprotein

†† Biotin biosynthesis

‡‡ Pantothenate synthesis

After genome shuffling, we saw an upregulation of genes involved in fatty acids biosynthesis, including the genes *cfa* (MG5 and MG6) and *scpA, lipB* and *mhpE* (in MY4). The cyclopropane fatty acyl phospholipid synthase gene (*cfa*) modifies the double bond of unsaturated fatty acids into their cyclopropane derivatives (127). This conversion, known as homeoviscous adaptation, occurs in different bacteria in response to environmental changes. This change in the membrane of some of the adaptive mutants was confirmed through Fatty Acid Methyl Ester (FAME) analysis and the results are shown in Table 3.3. This was further experimentally verified via overexpression of *cfa* in the wild type strain, which led to significant improvement in n-butanol tolerance (Figure 3.6**)**. Long-chain alcohols can affect membrane structure via insertion into the lipid bilayer in any orientation, increasing membrane fluidity. This phenomenon has been demonstrated by electron-spin resonance (128) and differential spin calorimetry (129). Modifications in the membrane that counteract these disruptions may be obtained by increasing the saturated/unsaturated fatty acids ratio or by increasing the length of the fatty acid (130). The results from FAME analysis revealed a decrease in the proportion of unsaturated and short-chain fatty acids in MY3 and MY4, whereas an increase of more than five times the amount of *cis*-10,11-Methylene-nonadecanoic acid was observed in MG6.

3.4.5.2. Iron-ion transport and metabolism

In several isolated adaptive mutants, we observed an unexpected transcriptional perturbation in the iron metabolic process (Table 3.4). Iron uptake is carried out by si-

derophores, organic ligands that sequester the insoluble Fe^{+3} into the cell (131). We found increased expression in genes related to the biosynthesis and transport of enterobactin and genes involved in different high affinity iron transport mechanisms, such as *feoA, feoB, ftnB, fiu* and *fieF* in mutants isolated from the green lineage (MG2, MG5 and MG6). Interestingly, the mutants from the yellow lineage exhibited the opposite response in iron ion related genes, suggesting the two lineages had evolved different mechanisms for enhanced tolerance.

Table 3.3. Results of fatty acid quantification via FAME analysis.

Fatty Acid	BW25113 Percent	MY3 Percent	MY4 Percent	MG6 Percent
10:0		0.0298		0.0555
12:0	3.1453	3.3583	3.4002	4.8576
13:0	0.1405	0.1432	0.0665	0.1832
12:0 3OH	0.0490	0.0464		0.0638
14:0 iso	0.1047	0.0600		0.1892
14:0	5.2076	5.4330	5.4132	6.5894
15:0 iso	0.1034	0.0679		0.2190
15:0 anteiso	0.1161	0.0637		0.2786
15:1 w8c				0.0502
16:0 iso	0.0876			0.0600
16:0 N alcohol				0.0503
16:1 w5c	0.2014	0.2161	0.2400	0.2306
16:1 w9c				0.2517
16:0	28.6642	28.4217	30.9490	32.6446
15:0 3OH	0.0500	0.0494		0.0534
17:1 w8c	0.2409	0.2425	0.1356	0.1236
17:0 cyclo	6.8178	5.9412	10.1498	18.0491
17:0	0.7621	0.7271	0.3089	0.6062
16:0 3OH				0.0487
18:1 w5c	0.1909	0.1979	0.1426	0.1244
18:0	0.7391	0.7372	0.5528	0.4125
19:0 iso	0.0678	0.0345	0.0832	0.0565
19:0 cyclo w8c	0.7875	0.6884	1.5416	5.4565
19:0	0.0952	0.0781	0.0884	0.0974
20:1 w7c	0.0597	0.0586		

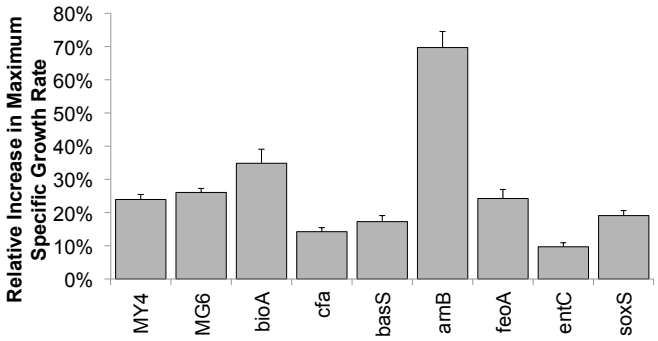


Figure 3.6. The relative increase in maximum specific growth rate via overexpression of individual genes compared to the wild type in the presence of 0.5% (v/v) n-butanol. From left to right: MY4, MG6, BW25113/pCA24N::bioA, BW25113/pCA24N::cfa, BW25113/pCA24N::basS, BW25113/pCA24N::arnB, BW25113/pCA24N::feoA, BW25113/pCA24N::entC and BW25113/pCA24N::soxS.

Table 3.4. Selected genes involved in iron transport and metabolism that were upregulated (bold) and downregulated (non-bolded) in the isolated adaptive mutants.

Strain	Function	Genes
MG1	Iron ion transport	***fecC, fiu, ftnB ftnB***
MG2	Iron ion transport	***ftnB, fiu, fepA, fhuE cirA***, ***basR***, ***basS*** *yaaA, fhuA,*
	Sideropore biosynthesis	***entA, entB, entC, entD, entE***
MY2	Iron ion transport	*fhuD, entF*
	Sideropore biosynthesis	*entE, entF*
MY3	Iron ion transport	*fecC, exbB*
MG5	Iron ion transport	***fhuF***, ***feoA***, ***feoB***, ***feoC***, *yaaA*
	Sideropore biosynthesis	***entB***, ***entC***, ***entD***, ***entE***
MY4	Iron ion transport	*fecR, fecI*
	Sideropore biosynthesis	*entA, entC, entS*
MG6	Iron ion transport	***basR, feoA***, ***feoB***, ***feoC***, ***fecA, fecR, fecI***
	Sideropore biosynthesis	***entC***, ***entE***

Fur is the main transcription factor involved in the regulation of iron-related genes. The increase in the expression of such a large number of these genes suggest a decrease in the activity of Fur (132) in these mutants. Using Network Component Analysis (NCA) (133, 134), Fur activity was identified to be significantly repressed in several green mutants (MG2, MG5, MG6) and significantly increased in several yellow mutants (MY2, MY3, MY4) (see Figure 3.7). Fur was the only transcription factor from the

NCA compliant set of transcription factors that showed such a significant difference in activity between the two different lineages.

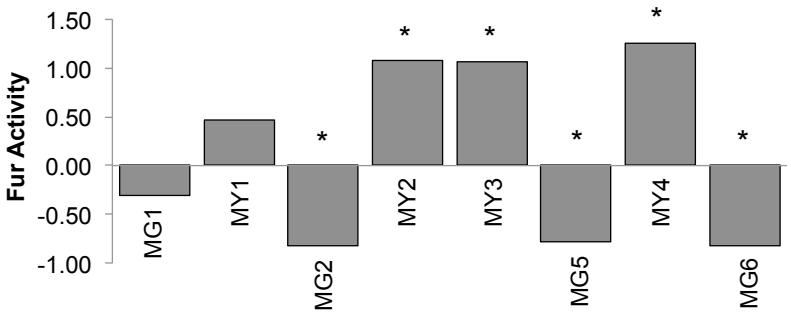


Figure 3.7. Fur activity in the isolated mutants determined using NCA. The asterisks indicate statistical significance with a p-value threshold of 0.05 using 1000 permutations.

3.4.5.3. Determining whether iron is the limiting growth condition

A series of experiments were conducted to determine whether the increase in gene expression of the iron uptake and transport genes identified in the adaptive mutants might be due to potential iron limitation in the media. The growth kinetics of the wild type strain and two isolated mutants (MY4 and MG6) were measured at various concentrations of iron in the media. The results indicate that iron is not the limited nutrient in our system (Figure 3.8).

3.4.5.4. Verification of iron-related genes via overexpression studies

If decreased Fur activity was responsible for increased n-butanol tolerance in the mutants in the green lineage, we would expect a decrease in tolerance if Fur was overexpressed in the mutant. To test this hypothesis, we overexpressed *fur* in MG6 and com-

pared it to wild type overexpressing *fur*. In the wild type strain, overexpression of *fur* does not significantly increase the inhibition by n-butanol (Figure 3.9). However, over-expression of *fur* in MG6 inhibited growth significantly in the presence of n-butanol, supporting the model that the activity of Fur is involved in n-butanol tolerance in this mutant.

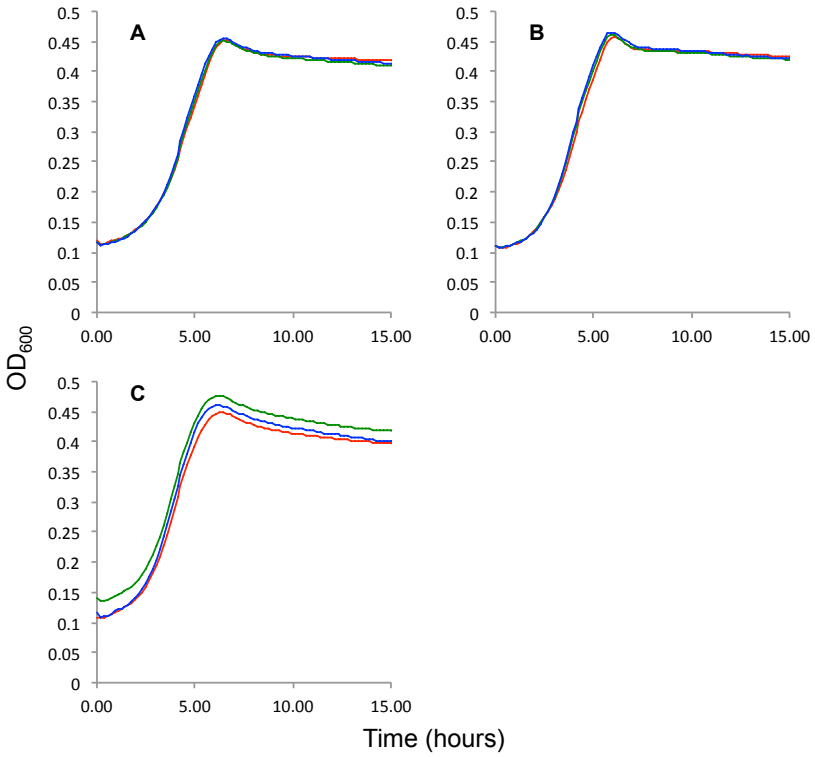


Figure 3.8. Growth kinetics at three different concentrations of iron for the strains, A) BW25113, B) MY4 and C) MG6. Red line: 18.5 µM, green line: 37 µM, and blue line: 74 µM of $FeCl_3.6H_2O$.

The contribution of iron-related genes in n-butanol tolerance was validated by overexpressing *entC* and *feoA* in the wild type strain. In the presence of 0.5% (v/v) n-butanol, the overexpression of *entC* and *feoA* genes improved growth compared to wild

type under the same conditions by 10±1% and 24±3%, respectively, demonstrating that the genes involved in enterobactin biosynthesis and transport are involved in n-butanol tolerance (see Figure 3.6). The genes *entC* and *feoA* were also enriched in a previous study of n-butanol tolerance in *E. coli* (12). The enhanced tolerance conferred by over-expression of the iron-related genes was not specific to BW25113, as an increase in tolerance was also achieved when overexpressed in strain AG1.

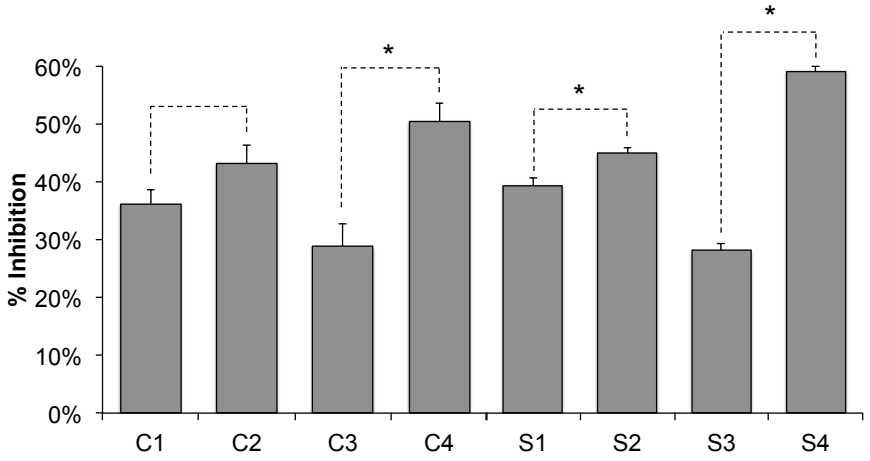


Figure 3.9. Growth inhibition in the presence of 0.5% (v/v) n-butanol. C1. BW25113/pCA24N, C2. BW25113/pCA24N::fur, C3. MG6/pCA24N, C4. MG6/pCA24N::fur. S1. BW25113/pCA24N, S2. BW25113ΔbasR/pCA24N, S3. BW25113/pCA24N::entC, and S4. BW25113ΔbasR/pCA24N::entC. Asterisks indicate measurements that are statistically significantly different between the two strains (determined using a student t-test with a p-value cut-off of 0.05).

The effects of the overexpression of iron-related genes on cell viability under n-butanol shock were also assessed (Figure 3.10). The overexpression of *entC* and *feoA* increased the viability of the strain to n-butanol exposure by more than 10-fold. Thus, the iron-related genes not only increase tolerance towards n-butanol in terms of fitness advantage, but also in terms of cell viability.

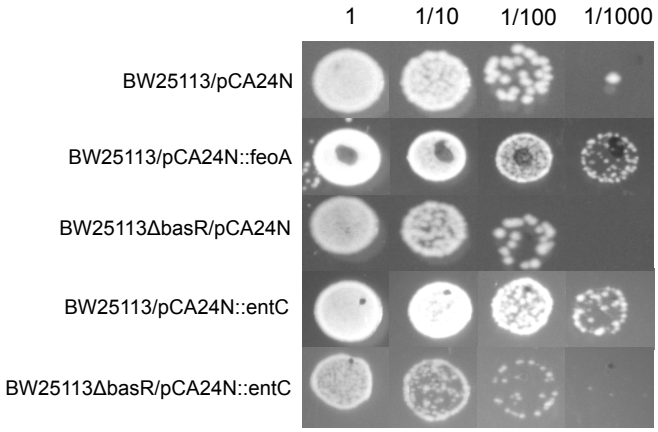


Figure 3.10. Cell viability after short-term exposure to 2% (v/v) n-butanol to determine the effects of Δ*basR* on the n-butanol tolerance of strains overexpressing iron-related genes.

3.4.5.5. Downregulation of Fur activity may have resulted in changes in lipid A composition

To the best of our knowledge, iron transport and metabolism has not been shown to play a role in n-butanol tolerance. Thus, to investigate the role of iron in n-butanol tolerance, the transcriptome data was analyzed for potential membrane modifications that result in changes in the cell wall or membrane, which is the most common mechanism of solvent tolerance. Both the MG2 and MG6 strains exhibited a general upregulation of iron-regulated genes including the *basR* gene, which encodes the bacterial adaptive response transcriptional regulatory protein. BasR belongs to the BasS/BasR two-component system, which responds to high concentration of Fe^{+3}. Activation of BasR via phosphorylation by BasS leads to increased expression of genes involved in modification of lipopolysaccharide to prevent excessive Fe^{+3} binding (135). We found that overexpression of *basS* increased n-butanol tolerance (Figure 3.6). Members of the BasSR regulon include the *arn* operon (*arnBCADTEF*) (135), which is involved in the

resistance of gram-negative bacteria to cationic antimicrobial peptides (CAMPs) via the addition of positively charged 4-amino-4-deoxy-l-arabinose to lipid A of the LPS (136). Even though the *arn* operon was not significantly upregulated in several other strains that showed overexpressions of the iron-related genes (MG2, MG5 and MG6), we decided to test whether increased expression of the *arn* operon would increase n-butanol tolerance. Overexpression of *arnB* increased n-butanol tolerance by approximately 70% in comparison to the wild type (Figure 3.6). This suggests that the potential LPS change resulting from the overexpression of the *arn* operon increases n-butanol resistance.

Based on the upregulation of the iron-related genes and *basR* and the increased n-butanol tolerance when we overexpressed *entC*, *feoA, basS* and *arnB* in the wild type strain, we hypothesize that the decreased activity of Fur led to an increase in activation of BasSR, which led to changes in the LPS, ultimately increasing n-butanol tolerance. To demonstrate that the increased n-butanol tolerance conferred by overexpression of iron-related genes may be due to the downstream effects of increased BasSR activity, we overexpressed *entC* in the Δ*basR* strain. This combination resulted in a significant increase in susceptibility to n-butanol compared to wild type, while overexpression of *entC* in the wild type strain reduced the susceptibility to n-butanol (Figure 3.9 and Figure 3.10), suggesting that the effects on n-butanol tolerance conferred by the iron-related genes are through the BasSR system. If our hypothesis were correct, then we would expect to see an increase in resistance of mutants MG2, MG5 and MG6 (GFP-labeled mutants exhibiting increased expression of iron-related genes) to the CAMP polymyxin B. Indeed, the data (Figure 3.11) agreed with the expected results, demonstrating that the

mutants were more resistant to the antibiotic compared to the wild type. On the other hand, the mutants that did not show an upregulation in the iron-related genes did not exhibit any increase in resistance to polymyxin B.

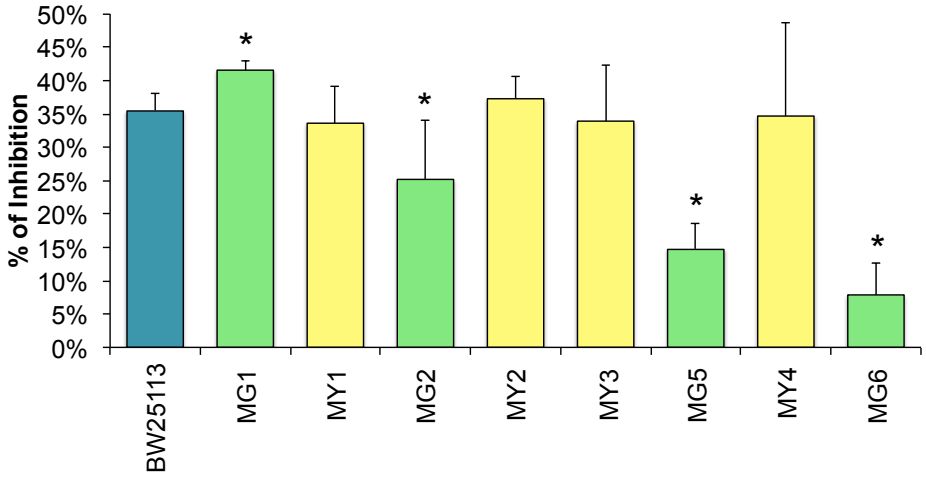


Figure 3.11. Growth inhibition of BW25113, MG1, MY1, MG2, MY2, MY3, MG5, MY4 and MG6 in the presence of 3 μg/ml of the cationic antibiotic peptide Polymyxin B. Asterisks indicate adaptive mutants that are affected significantly compared to the wild type (determined using a student t-test with a p-value cut-off of 0.05).

3.4.5.6. Genome sequencing

The genomes of all the isolated adaptive mutants were re-sequenced via ultra high throughput sequencing using 32 bp single-end sequencing for an average coverage of 30X. All confirmed mutations are shown in Table 3.5. Note that due to the short single-end sequencing, we were confident in our ability to identify SNPs, but not chromosomal rearrangements.

In the green mutants, we only identified mutations in the last two mutants. The genome sequence of the mutants MG5 and MG6 revealed a single-nucleotide polymor-

phism (a C → T transition (S82F)) in *rho*. A recent study using adaptive evolution in *E. coli* demonstrated that a point mutation in *rho* plays an important role in ethanol resistance (64). However, the defective termination factor *rho* imparts pleiotropic phenotypes due to the physiological role of this protein (137), making its effect unpredictable; further characterization is needed to elucidate the exact role this mutation plays on n-butanol tolerance.

Table 3.5. List of validated mutations detected using genome resequencing.

Strain	Mutation
MG1	No mutations validated
MY1	No mutations validated
MG2	No mutations validated
MY2	*relA142::IS2*
MY3	No mutations validated
MG5	*rhoS82F* *feoA201::IS5*
MY4	*nusAE212A*
MG6	*rhoS82F* *feoA201::IS5* *FBSmut*, a SNP (C→T) Fur binding site fes-ybdZ-entF-fepE (at 141 bp from fes) and fepA-entD (at 157 bp from fepA)

Even with the single-end reads we managed to identify and validate two chromosomal rearrangements. One is a translocation of the transposable insertion sequence IS5 within the 3' end of the *feoA* gene (*feoA201::IS5*) shared by both MG5 and MG6. Studies have demonstrated that the insertion of IS5 increases transcription of downstream genes in the same operon (138-140). Since the insertion of the IS5 is located at the end of the *feoA*, the genes *feoB* and *feoC* in the *feoABC* operon can potentially increase their transcript abundances by being transcribed from promoters located inside the transposa-

ble element, and bypass the repression by Fur. Indeed, in MG5 and MG6, we observed an upregulation of *feoB* and *feoC*.

The only genotypic difference between MG5 and MG6 is the SNP (C→T) located in the Fur binding site (*FBSmut*) of the operons *fes-ybdZ-entF-fepE* (at 141 bp from *fes*) and *fepA-entD* (at 157 bp from *fepA*) found in MG6 but not in MG5. The SNP was likely the cause of the higher transcriptional activation of the *fepA-entD* operon in MG6 compared to MG5. The data indicates that this mutation is beneficial since the fitness advantage of MG6 is greater than MG5 (as shown in Figure 3.4 and Figure 3.5). Furthermore, the data strongly suggest that the MG6 is direct progeny of MG5.

The other chromosomal rearrangement we identified and validated was a translocation of the 1331 bp transposable insertion sequence IS2 within the 3' end of the gene *relA* (*relA142::IS2*) in the yellow mutant MY2. RelA is the key enzyme involved in the activation of the stringent response. Studies have shown that mutant alleles of *relA* confer temperature-sensitive phenotypes (141). IS2 has been associated as a controller element, depending on the integration orientation (142). Inclusion of stringent response regulon in NCA confirmed that the stringent response was significantly perturbed in MY2 (data not shown). Further analysis is needed to elucidate the role of *relA* in the tolerance to this solvent.

In the mutant MY4 (yellow lineage) a point mutation was detected in *nusA* (*nusAE212A*). The allele *nusAE212K* has been identified to confer a slower growth phenotype at 30ºC (143). NusA interacts with many proteins involved with transcription and termination, including *rpoB*, *rpoC*, *rho*, and core RNAP (144). The presence of two

independent mutations affecting transcriptional termination implies that the mutated *rho* and *nusA* potentially increases the n-butanol tolerance phenotype. In addition, since mutations in both of these genes have been shown to confer temperature-sensitive phenotypes, there may be a correlation between temperature-sensitivity and n-butanol tolerance in *E. coli*.

No SNPs were identified in any of the other mutants. However, as stated earlier, there may be potential genome rearrangements present in these mutants that we were not able to detect.

3.4.5.7. Dynamics of allele frequencies during evolution

We subjected the genomic DNA extracted from population samples at different generations to qPCR in order to measure the allele frequencies throughout the course of the evolution. Calibration experiments for each set of allelic primers were conducted using known ratios of genomic DNA with wild type and mutant alleles; all sets of primers performed well (Figure 3.12).

The allelic data for the population samples are shown in Table 3.6. The pattern shown for the mutation *rhoS82F*, for samples isolated after the genome shuffling, correlates with the evolution dynamics traced using VERT for the green subpopulation (Figure 3.13). This result allows us to conclude that *rhoS82F* was fixed in the green subpopulation shortly after the initiation of the second round of evolution. Based on the fact that *rhoS82F* and *feoA201::IS5* were both found in MG5, we expected to see similar

frequencies of both alleles in the population samples; however as illustrated in Figure 3.13, the frequency of *feoA201::IS5* allele was constantly lower than *rhoS82F*.

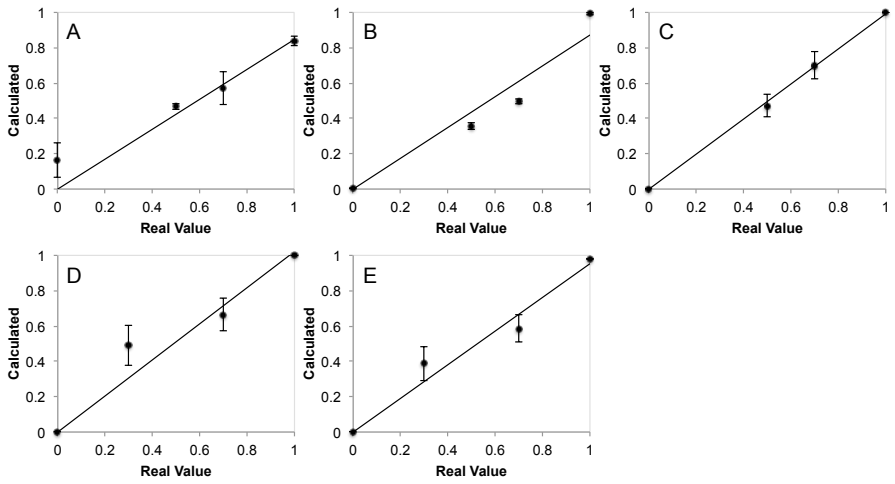


Figure 3.12. Calibration data for the primers used to determine the allelic frequency of: A) the *rhoS82F* allele, B) the *feoA201::IS5* allele, C) the *FBSmut* allele, D) the *nusAE212A* allele, and E) the *relA142::IS2* allele.

To determine if such difference was due to an actual reduced proportion of *feoA201::IS5* or inefficiency of the qPCR process (although the allelic primers performed well in the calibration experiments), the presence of *feoA201::IS5* in 16 randomly isolated colonies chosen from generation 72 was assessed.

We found that ~93.8% of the whole population contains such mutation (15 out of 16 isolated colonies), which is very close to the frequency of the *rhoS82F* (~92%) at generation 72. Both of these mutations were nearly fixed in the green subpopulation (~97%). However, the reason why a 50% reduction in the proportion of *feoA201::IS5* using the qPCR was seen is still unclear. In the case of the mutation *FBSmut*, it appeared in between generations 40 and 48 in the green subpopulation and rapidly became

the majority of the population. This result correlates with the higher fitness advantage observed in the mutant MG6 in comparison with MG5 (Figure 3.4).

Table 3.6. The mutant allele frequencies at various time points during evolution.

	BEFORE PROTOPLAST FUSION				
	rhoS82F	feoA201::IS5	FBSmut	relA142::IS2	nusAE212A
Gen 8	0.0±0.0%	0.0±0.0%	0.0±0.0%	0.0±0.0%	0.0±0.0%
Gen 40	0.0±0.0%	0.0±0.0%	0.0±0.0%	0.0±0.0%	0.0±0.0%
Gen 80	0.0±0.0%	0.0±0.0%	0.0±0.0%	0.0±0.0%	0.0±0.0%
Gen 120	0.0±0.0%	0.0±0.0%	0.0±0.0%	4.6±1.5%	0.0±0.0%
Gen 144	0.0±0.0%	0.0±0.0%	0.0±0.0%	18.7±7.1%	0.0±0.0%
	AFTER PROTOPLAST FUSION				
Gen 8	64.9±10.1%	22.9±3.4%	9.7±5.0%	0.0±0.0%	13.2±1.6%
Gen 24	47.9±8.0%	15.4±2.9%	7.7±1.3%	0.0±0.0%	17.1±2.2%
Gen 32	28.7±5.4%	5.7±3.9%	8.5±2.2%	0.0±0.0%	10.1±8.4%
Gen 40	23.4±4.6%	4.4±2.8%	9.7±4.1%	0.0±0.0%	13.7±5.7%
Gen 48	38.8±4.1%	7.4±2.8%	8.0±0.2%	0.0±0.0%	3.8±3.7%
Gen 56	53.0±0.1%	15.5±5.0%	58.0±13.4%	0.0±0.0%	2.0±0.4%
Gen 72	93.3±1.7%	49.2±2.6%	86.8±0.6%	0.0±0.0%	1.9±0.4%
	LAST GENERATION POPULATION P1				
P1	8.9±3.2%	22.0±12.4%	0.0±0.0%	0.0±0.0%	0.1±0.1%

To determine whether *feoA201::IS5* and *rhoS82F* were present in population P2 prior to the genome shuffling, we analyzed the genomic DNA from population samples at various time points prior to genome shuffling. Neither of these mutations was present in P2, indicating that they either came from population P1 or were generated during the genome shuffling process. A quick analysis of the population sample isolated from generation 144 in population P1 showed that the proportion of *rhoS82F* was 8.88±3.20% and *feoA201::*IS5 was 21.99±12.41% prior to genome shuffling.

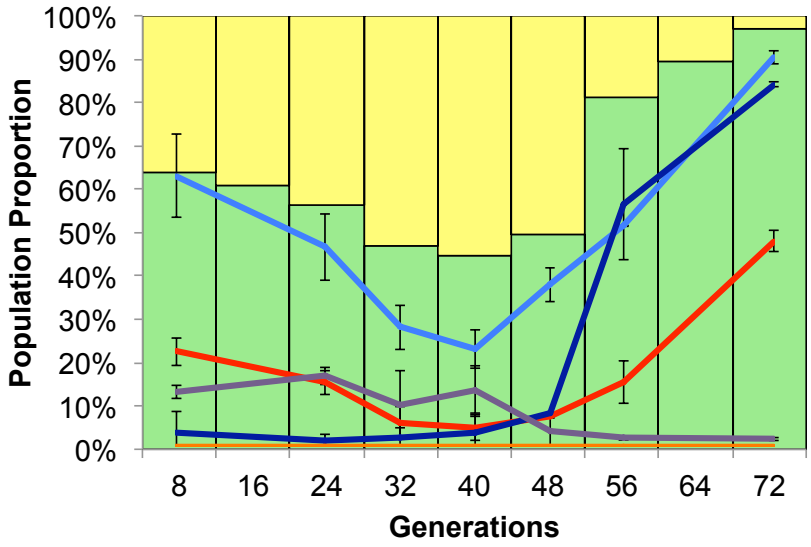


Figure 3.13. Mutant allele frequencies in the evolving population samples isolated after genome shuffling. Yellow bars: proportion of the yellow subpopulation. Green bars: proportion of the green subpopulation. Blue line: approximate steady-state concentration of n-butanol in the chemostat. Light blue line: frequency of the *rhoS82F* allele. Red line: frequency of the *feoA201::IS5* allele. Blue line: frequency of the *FBSmut* allele. Purple line: frequency of the *nusAE212A* allele. Orange line: frequency of the *relA142::IS2* allele.

The mutations identified in the yellow mutants were also tracked throughout the course of the evolution. The data showed that *relA142::IS2* (identified in MY2) appeared between generations 80 and 120 during the first round of evolution and was lost from the population after the genome shuffling process, and *nusAE212A* (identified in MY4) was not present in populations P1 or P2 prior to genome shuffling (Table 3.6). The *nusAE212A* mutation was either in too low of a frequency in the population to be detected using qPCR or may have arisen during the genome shuffling process; it remained in the population during the second round of evolution at a fairly constant pro-

portion until generation 40 and was rapidly lost due to the expansion of MG6 (Figure 3.13).

3.4.5.8. Additional important gene expression changes observed in the adaptive mutants

The *soxS* gene, which is known to protect the cell against antibiotics and organic solvents via lipopolysaccharide modification (145, 146), was found to be upregulated in strain MG5. The effect of this gene in n-butanol tolerance was validated by overexpression (Figure 3.6**)**. We found upregulation of *sodB* in the isolated mutants MY3 and MY4 (both from the yellow lineage). This gene codes a superoxide dismutase containing iron. This enzyme protects superoxide-sensitive enzymes under oxidative stress (147). The gene *cpxA* that is part of the CpxAR two-component signal transduction system, that activates degradation and folding of proteins in response to cell envelope damage (148, 149), was found to be downregulated in MG5 and MG6.

Genes involved in fimbriae and flagellar biogenesis were also found to be highly perturbed (Table 3.7). Perturbations in fimbriae biogenesis, specifically the downregulation of the operon *fimAICDFGH*, have been demonstrated to improve resistance to organic solvents (64). However, it is not clear whether these transcriptional changes are adaptive, since different responses were detected between different mutants in our data.

We found a reduction in the expression of the three initiation factors (IF-1, IF-2 and IF-3), encoded by the genes *infABC*, in all the strains in the green lineage (except MG1). This led to a strong downregulation of genes involved in tRNA aminoacylation.

Changes in initiation factor transcription have been found in other stress responses, such as cold-shock (150, 151). Many genes expressing heat shock proteins were found to be upregulated, such as *hslJ, htrA, hslU, clpB, htrC, ycaL* and *uspG.*

Table 3.7. Selected flagellar and fimbriar biosynthesis related genes that were upregulated (bold) or downregulated in the isolated adaptive mutants.

Strain	Flagellum Biosynthesis and Taxis	Fimbriae Biosynthesis
MG1	*cheB, flgD, flgG*	***fimC***
MY1	*cheA,* ***cheZ, flgK,*** *flhD, fliO,* ***fliT***	
MG2	*flgG, flhD, fliT, dsbB, flgC, flhA, fliH, hns*	***fimB, fimC, fimZ***
MY2	***cheA, fliT, flgF, fliL***	***fimI,*** *fimB*
MY3	*flgB*	*fimA, fimC, fimZ*
MG5	*csrA, flgF, fliJ, fliM*	*fimB, fimE,fimG, fimI*
MY4	***fliD***	***fimA, fimB***
MG6	*csrA, flgA, flgK, fliC*	***fimZ***

The *envR* gene was downregulated in MG6. EnvR negatively regulates the transcription of drug efflux pumps (152), and may be a mechanism in n-butanol tolerance in this strain. There were significant changes in the regulation of genes involved in the glyoxylate cycle (upregulation of *aceB, aceE, aceF, lpd* and downregulation of *sdhD, pykA* and *pykF*) in all the isolated mutants. Studies have demonstrated that under glucose-limited chemostats, changes in the glyoxylate cycle may play an important role in the adaptation under this selective pressure (153). This result suggests that in addition to n-butanol tolerance, the adaptive mutants were also adapting to the nutrient limitation.

Finally, we also found an upregulation of the cytochrome genes *cyoB* and *cyoD*, which are part of the oxidative stress response (154).

3.5. Discussion

3.5.1. Exploring evolution dynamics using VERT

In this study, we used a two-colored VERT system to track the evolutionary dynamics of *E. coli* under increasing concentrations of n-butanol. Unlike traditional evolutionary engineering methods, VERT allows adaptive events to be easily identified via visualization of the relative proportions of different fluorescently labeled cells. It can help to facilitate a more rational scheme for the isolation of adaptive mutants for further characterization and can be broadly applicable to the adaptive evolution of microbial systems. In addition, while it was not employed in this work, a formal schedule for the ramp-up of selective pressure can also be established by the use of VERT.

Since adaptive events are identified via significant changes in the relative proportion of the different colored subpopulation, we developed a computational approach to identify adaptive events and compared it with visual inspection. Adaptive events visually identified at generations 8, 80, and 112 in Figure 3.3 were not identified computationally and may therefore represent transient fluctuations. MG1 (isolated from generation 8) showed no significant increase in relative fitness compared to the wild type (and no genomic changes were identified from whole genome re-sequencing; see below), and is therefore likely a false positive. All 8 isolates from generation 112 showed negative fitness compared with both the wild type and the previous adaptive mutant; this may be due to differences in the conditions during the actual evolution and that in the pairwise competition, but does not preclude the possibility of it being a falsely identified adaptive

event. An adaptive event apparently occurred in the green subpopulation after generation 32 that gave rise to a mutant that was able to impede the expansion of the yellow subpopulation. This episode led to a meta-stable state, visible from generation 32 to 80, where there are negligible subpopulation proportion changes. Indeed, generation 80 was an actual adaptive event, as MG2 that was isolated from the expanding subpopulation at that generation showed an increase in relative fitness compared to the previous adaptive mutant (MY1). This result was not identified by the computational algorithm due to the lack of significant change in relative proportions of the two colored-subpopulations, which suggests that MY1 and MG2 may be equally fit under evolution conditions. Based on these results, our computational algorithm has the potential to be more sensitive and accurate compared to visual inspection, but may not be able to identify adaptive events where the fitness differential between populations is small; further development of the computational algorithm is underway to incorporate meta-stable regions into the analysis.

3.5.2. Role of Fur in n-butanol tolerance

Upon isolation of the adaptive mutants from the evolved populations, transcriptome studies and subsequent NCA analysis revealed differential activation of Fur in the yellow and the green lineages. Fur activity was down in several of the green mutants, and is a potentially newly discovered adaptive mechanism of *E. coli* for n-butanol tolerance. We determined that the decreased activation of Fur might have led to an increase in the activity of the BasS/BasR system, which ultimately led to changes in LPS compo-

sition. The hypothesis was further supported by the increased resistance to the CAMP, polymyxin B, in MG2, MG5 and MG6, which showed increased gene expression of many iron-related genes.

Some of our mutants exhibited an overexpression of *cfa*, which was further validated to lead to an increase in n-butanol tolerance. However, this result seems to contradict previous results in *E. coli* mutants isolated from populations evolved in isobutanol (22), where a downregulation in *cfa* was observed. This contradiction may be due to differences in membrane disruptions due to structural differences between the two isomers or the differences in other adaptive mechanisms present in the mutants.

It was clear from our data that certain potential adaptive mechanisms arose independently. For example, we found the upregulation of the biotin biosynthesis genes in both the green and the yellow lineages (MY1 (yellow) and MG2 (green)), suggesting that mutations that lead to an upregulation in biotin biosynthesis are frequently acquired. Unfortunately, we did not identify any SNPs in MY1 and MG2; longer paired-end reads will be needed to identify any potential genome rearrangements in these mutants. As mentioned previously, the upregulation of the iron-related genes were only observed in the green lineage; in fact, several genes related to iron transport were downregulated in two of the yellow adaptive mutants (MY2 and MY3). After the genome shuffling, adaptive mutants isolated from the green subpopulation exhibited increases in the iron-related genes. However, the adaptive mutant isolated from the yellow sub-population (MY4) showed a decrease in iron-related gene expression, suggesting that genetic interactions are potentially involved in the adaptive mechanism associated with Fur.

Even though several genes we verified to be involved in n-butanol tolerance were already upregulated in some of the earlier isolated green mutants, such as *basS*, *entC* and *bioA*, the tolerance levels of these earlier green mutants were not significantly higher than the wild type strain at 0.8% (v/v) n-butanol. It was not until after the genome shuffling did we observe a significant increase in n-butanol tolerance. These results further confirm the complex nature of n-butanol tolerance in *E. coli*; involvement of multiple genes and potential genetic interactions being involved in conferring the desired phenotype.

3.6. Conclusions

In this proof-of-principle study, our results showed that while subsequent adaptive mutants have fitness advantages over the previous adaptive mutant, they might not always be better than the wild type. This type of phenomenon was observed previously with yeast evolved under glucose-limitation (107), and is likely due to the heterogeneities in the evolving population and potential interactions between clones. Thus, this result further highlights the benefit of tracking the adaptive events in a population to isolate each adaptive mutant, as the last one isolated from the population may not necessarily be the best one. By isolating mutants from each observed adaptive event, we are able to identify tolerance mechanisms that may otherwise be missed due to the effects of clonal interference. In addition, by combining VERT with genome shuffling, we were able to significant enhance the rate of adaptation.

3.7. Data Availability

All raw data are MIAME compliant and have been deposited in the GEO database with accession number GSE30005.

4. GENETIC DETERMINANTS FOR N-BUTANOL TOLERANCE IN EVOLVED *E. COLI* MUTANTS. CROSS ADAPTATION AND ANTAGONISTIC PLEIOTROPY BETWEEN N-BUTANOL AND OTHER STRESSORS*

4.1. Summary

Cross-tolerance and antagonistic pleiotropy have been observed between different complex phenotypes in microbial systems. These relationships between adaptive landscapes are important for the design of industrially relevant strains, which are generally subjected to multiple stressors. In our previous work, we evolved *Escherichia coli* for enhanced tolerance to the biofuel, n-butanol, and discovered a molecular mechanism of n-butanol tolerance that also conferred tolerance to the cationic polymicrobial peptide polymyxin B in one specific lineage (GFP-labeled) in the evolved population. In this work, we aim to identify additional mechanisms of n-butanol tolerance in an independent lineage (YFP-labeled) from the same evolved population and to further explore potential cross-tolerance and antagonistic pleiotropy between n-butanol tolerance and other industrially relevant stressors. Analysis of the transcriptome data of the YFP-labeled mutants allowed us to discover additional membrane-related and osmotic stress related

* Reprinted with permission from Luis H. Reyes, Ali S. Abdelaal and Katy C. Kao (2013). "*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 (AEM). DOI: 10.1128/AEM.01703-13. Copyright 2013 American Society for Microbiology.

genes that confer n-butanol tolerance in *E. coli*. Interestingly, the n-butanol resistance mechanisms conferred by the membrane-related genes appear to be specific to n-butanol and are in many cases antagonistic with isobutanol and ethanol. Furthermore, the YFP-labeled mutants showed cross-tolerance between n-butanol and osmotic stress while the GFP-labeled mutants showed antagonistic pleiotropy between n-butanol and osmotic stress tolerance.

4.2. Introduction

n-Butanol, an industrial intermediate chemical, solvent and a potential biofuel, is naturally produced by Clostridium species. Several nonnative microbial systems have been engineered for its production, including *Escherichia coli* (4), *Lactobacillus brevis* (155), *Pseudomonas putida* (156), *Bacillus subtilis* (156) and *Saccharomyces cerevisiae* (157). However, this solvent is highly toxic to microorganisms, imposing a limit on the productivity of bio-based production, leading to the development of simultaneous fermentation and separation techniques to mitigate the toxic effects of the biofuel (158) and efforts to identify the genetic determinants and molecular mechanisms associated with n-butanol tolerance for reverse engineering of more robust strains (12, 17, 21, 22, 43, 159, 160). Prior strain engineering efforts include overexpression of GroESL in *Clostridium acetobutylicum* (resulting in a 50% improvement in total growth in 0.75% (v/v) n-butanol (160) and recently in *E. coli* (resulting in a 2.8-fold increase in total growth in 48 h cultures in 0.75% (v/v) n-butanol (159)), and overexpression of gene CAC1869 in *C.*

acetobutylicum (resulting in an 81% increase in total growth after 12 hours (33). Since n-butanol tolerance is a complex phenotype (12, 33) and the production environment involve multiple stressors, additional knowledge on the genetic determinants and molecular mechanisms involved and their effects under different stress conditions are essential for future strain engineering efforts.

We previously reported the use of an adaptive laboratory evolution-based method called Visualizing Evolution in Real-Time (VERT) to study n-butanol tolerance in *E. coli* (21). Using a two-color VERT system (with green fluorescent protein (GFP) labeled and yellow fluorescent protein (YFP) labeled cells, allowing the tracking of independent lineages), we isolated several n-butanol tolerant adaptive mutants throughout the evolution and used whole genome transcriptome profiling and resequencing analyses to identify the underlying n-butanol tolerance mechanisms. A reduced activity of the ferric uptake regulator Fur, leading to increased siderophore biosynthesis and transport, which ultimately led to membrane modifications, was identified to be a likely mechanism of enhanced n-butanol tolerance. The deactivation of Fur also led to cross-tolerance between n-butanol and the cationic antimicrobial peptide polymyxin B. However, this tolerance mechanism was only observed in mutants from the GFP-labeled subpopulation and not in the YFP-labeled subpopulation, suggesting a different route(s) of adaptation in the YFP-labeled adaptive mutants. In this study, we aim to identify the mechanisms of n-butanol tolerance in the YFP-labeled mutants and any additional cross-resistance and/or antagonistic pleiotropy between n-butanol and other stressors in the isolated adaptive mutants. Detailed analysis of the transcriptome profiles of the yellow-labeled sub-

population under n-butanol stress was performed to determine additional genetic determinants involved in tolerance to the solvent. Phenotypic analyses revealed divergent relative fitness profiles in different stressors between the two different lineages. Several genes related with membrane transporters and cardiolipin biosynthesis, an important component of bacterial membranes, were determined to be involved in resistance to n-butanol exclusively among the solvents tested.

4.3. Materials and Methods

4.3.1. Bacterial strains and plasmids

The *E. coli* K-12 strain, BW25113 (F^-, Δ(*araD-araB*)567, ΔlacZ4787(::*rrnB-3*), $lambda^-$, *rph-1*, Δ(*rhaD-rhaB*)568, *hsdR514*), was used in this study. Plasmids isolated from clones in the ASKA collection (121) were used for the overexpression studies in BW25113.

4.3.2. Growth conditions and maintenance

E. coli strains were routinely cultured aerobically in liquid Luria-Bertani (LB) medium at 220 rpm and 37°C and on agar-solidified LB at 37°C. When required, the medium was supplemented with 30 μg/mL of chloramphenicol. Frozen stocks were prepared from overnight cultures and stored in 17.5% glycerol at -80°C. Cells from a single colony were used to inoculate liquid cultures. Growth curves were carried out in M9

minimal media supplemented with 5 g/L of glucose, 0.01% (w/v) thiamine and appropriated antibiotic when required.

`

4.3.3. DNA isolation and transformation

Isolation of plasmid DNA from *E. coli* was performed using the Zyppy™ Plasmid Miniprep Kit (Zymo, USA). Electroporation was used for all *E. coli* transformations.

4.3.4. Pre-screening of potential n-butanol tolerance-conferring genes

Strains harboring the genes to be pre-screened were cultured in M9 minimal media (5 g/L glucose) and incubated overnight at 37°C to be used as inoculum. On the next day, 100 µL cultures were prepared in 96-well microtiter plates for growth kinetic analysis in the absence and presence of 0.8% (v/v) n-butanol in M9 (corresponding to an inhibition in growth rate of more than 50%) at 37ºC, using an Infinite M200 Microplate reader (TECAN®). Four technical replicates were obtained per sample in the pre-screen. The growth kinetic parameter "*s*" described below was calculated. Statistical significance was assessed using a Student's t-test analysis using a p-value cut-off of 0.05.

4.3.5. Calculation of growth kinetic parameters

The growth kinetics parameters: "percentage of inhibition", "relative fitness coefficient (*s*)" and "relative increase in fitness (*RIF*)" were calculated using Equation 4.1,

Equation 4.2 and Equation 4.3, respectively. These parameters were calculated using the measured maximum specific growth rate (μ_i) of each strain (strain *i*).

Equation 4.1 $$\text{Inhibition (\%)} = \left[1 - \left(\frac{\mu_{\text{clone @ stressful condition}}}{\mu_{\text{clone in absence of stressor}}}\right)\right] \times 100\%$$

Equation 4.2 $$s\ (\%) = \left[\left(\frac{\mu_{\text{clone @ stressful condition}}}{\mu_{\text{reference strain @ stressful condition}}}\right) - 1\right] \times 100\%$$

Equation 4.3 $$RIF\ (\%) = \left[1 - \left(\frac{\text{Inhibition}_{\text{clone @ stressful condition}}}{\text{Inhibition}_{\text{reference strain @ stressful condition}}}\right)\right] \times 100\%$$

The ratio between the specific growth rates of the strain of interest relative to the reference strain under each stress condition was determined using the relative fitness coefficient "*s*" (Equation 4.2). Appropriated reference strains were used for the different calculations. For overexpression studies, the reference strain is the wild-type strain harboring the empty vector. For the phenotypic analysis of the isolated mutants, the ancestral strain expressing the corresponding fluorescent protein was used. The Relative Increase in Fitness, "*RIF*", is a parameter calculated to normalize the relative fitness of the overexpression strain in the presence of the stressor against any fitness defects/advantage exhibited by the strain in the absence of the stressor. Positive values of *RIF* represent a net increase in growth rates in the presence of the stressor. A Student's t-test analysis (p-value < 0.05) was used to assess significance of the aforementioned calculated kinetic parameters.

4.3.6. Detailed phenotypic analysis of selected n-butanol tolerance conferring genes

Clones that showed a statistically significant increase in relative fitness in the presence of n-butanol from the pre-screen were selected for secondary validation in batch cultures. Each strain was cultured in M9 minimal media (5 g/L glucose) and incubated overnight at 37°C to be used as inoculum. The next day, a 5% (v/v) inoculum was used to seed a 30 mL culture in 250 mL closed-cap flasks. At least three biological replicates were used in each experiment. The stressors analyzed in this study were: 0.8% (v/v) n-butanol, 1% (v/v) isobutanol, 4% (v/v) ethanol, 1.75 g/L of acetate (as acetic acid), pH levels of 4.5 and 6.0 (titrated using HCl), and temperatures of 28°C and 42°C. Cultures were incubated at 37°C (except in temperature challenge experiments) with constant shaking at 220 rpm. Growth was monitored using spectrophotometry (OD_{600}) until stationary phase was reached.

4.3.7. Osmotic stress experiments

Cells were grown in M9 minimal media (5 g/L glucose) overnight. 5% (v/v) culture was used to inoculate 5 mL culture in M9 minimal media and incubated at 37°C until an $OD_{600} \sim 0.6$ was reached. Bacterial cultures were normalized to an $OD_{600} = 0.3$ and cells from 1 mL of the normalized culture was recovered by centrifugation. Quick aspiration of the supernatant was followed by resuspension of the cell pellet in 1 mL of 40% (w/v) glucose and incubated at 37°C for 2 hours with constant shaking (220 rpm). After incubation, cells were pelleted and resuspended in 1 mL of M9 minimal media and

then diluted 1:100. 100 μL of the diluted culture was plated on LB agar plates, and incubated overnight at 37°C for colony counting. At least four biological replicas were used in this experiment. A non-stressed culture from the same population was used as control to ensure consistency in dilution. The "*Relative Increase in Survival Rate*" was calculated as the ratio between the number of colony forming units after the osmotic shock between the mutant and the parental strains.

4.4. Results

4.4.1. Phenotypic analyses of the isolated mutants in multiple stressors

In a previous study, we used VERT to isolate *E. coli* mutants with enhanced n-butanol tolerance (21). Two differentially labeled (with GFP or YFP), but otherwise isogenic, strains of *E. coli* were used for the evolutionary experiment in the presence of increasing concentrations of n-butanol, resulting in two different colored subpopulations (independent lineages). Several mutants were isolated and characterized via phenotypic (relative fitness measurements), genotypic (whole genome re-sequencing), and transcriptomic (Gene Expression Microarrays) analyses. We reported differential mechanisms of enhanced tolerance between the two independent lineages. In the GFP-labeled mutants (MG2, MG5, and MG6), cross-tolerance between the cationic antibiotic peptide polymyxin B and n-butanol was identified. The resistance level to polymyxin B in these mutants increased gradually, with the mutants isolated later in the population exhibiting higher antibiotic resistance than earlier GFP-labeled isolates (21). However, this cross-

resistance was not observed in the YFP-labeled mutants (MY1, MY2, MY3, and MY4), none of which exhibited changes in sensitivity to polymyxin B. These prior results led us to hypothesize that additional cross-tolerance and/or antagonistic pleiotropy between n-butanol and other stressors may be present in our isolated mutants.

To identify potential cross-tolerance and antagonistic pleiotropy between n-butanol and other industrially relevant stressors in the isolated mutants, we evaluated their relative fitness under different conditions. The stressors used include organic solvents (n-butanol, isobutanol, and ethanol), organic acid (acetate), acid stress (pH = 4.5 and pH = 6.0), and temperature (42°C) and cold (28°C) stress. Detailed description of each condition is provided in the Materials and Methods section. The results are summarized in Figure 4.1 and Table 4.1.

The two different lineages (GFP- and YFP-labeled mutants) exhibited notable differences in their responses to the applied stressors as shown in Figure 4.1. MY2 showed a 15% increase in relative fitness (*s*) compared with the ancestral strain in the presence of n-butanol, but an increase in *s* of 30% in the absence of n-butanol, indicating that the adaptive mechanisms were not specific for n-butanol tolerance, presumably via an increase in general nutrient utilization. Thus, the specific effects of each challenge on growth kinetics were normalized by calculating the Relative Increase in Fitness (*RIF*) using Equation 4.3, to determine if the increase in specific growth rate was the result of enhanced tolerance or due to a general increased in growth rate (as a result of increased nutrient utilization, energy levels in the cell, etc.). After normalization, the majority of the mutants exhibited negative values in *RIF* under many conditions except in the pres-

ence of n-butanol, isobutanol, and acetate and at low temperatures (see Table 4.1), indicating that the isolated mutants do not have general tolerance to a wide range of stressors. Even among organic solvents, mutant MY4 (YFP-marked) was the only mutant that showed an improvement in tolerance to all three organic solvents tested (n-butanol s = 38 ± 3%, isobutanol s = 34 ± 2% and ethanol s = 41 ± 6%). Interestingly, the GFP-labeled mutants did not exhibit an increase in relative fitness in the presence of ethanol, while the YFP-labeled mutants showed positive values of s under ethanol stress (Table 4.1).

The results presented in Table 4.1 demonstrated a potential antagonist pleiotropy between the molecular mechanisms involved in acid and n-butanol tolerance in the isolated mutants. In all the mutants studied, *RIF* values calculated at pH 4.5 and 6.0 were either negative or non-significant (Table 4.1), suggesting a divergence in the tolerance levels under these two conditions. This antagonistic behavior between the two stresses have been documented previously in *E. coli* (12) and *L. brevis* (15).

Table 4.1. Growth kinetic parameters calculated for the isolated mutants under different stress conditions. The specific growth rates at T = 37ºC were used as references for *RIF* calculations. Statistically significant values are bolded. In case of *s*, the bolded values indicate fitness coefficients significantly different compared to the wild-type strain (Student's t-test with a *p*-value cut-off of 0.05). For *RIF*, the bolded values indicate measurements significantly different than "0".

Stressor	*s*				*RIF*			
	MY2	MG5	MY4	MG6	MY2	MG5	MY4	MG6
n-Butanol	**15 ± 2%**	**85 ± 13%**	**38 ± 3%**	**68 ± 10%**	-48 ± 10%	**216 ± 44%**	**84 ± 15%**	**261 ± 51%**
Isobutanol	**22 ± 2%**	**30 ± 2%**	**34 ± 2%**	**38 ± 3%**	-26 ± 5%	**13 ± 2%**	**64 ± 11%**	**99 ± 15%**
Ethanol	**22 ± 3%**	-6 ± 1%	**41 ± 6%**	6 ± 1%	**-25 ± 6%**	**-122 ± 23%**	**96 ± 20%**	**-68 ± 13%**
Acetate	**37 ± 4%**	**29 ± 3%**	**18 ± 2%**	**29 ± 2%**	**26 ± 5%**	8 ± 1%	-12 ± 2%	**52 ± 8%**
pH = 6.0	**15 ± 1%**	**10 ± 1%**	**9 ± 1%**	**12 ± 1%**	**-48 ± 9%**	**-62 ± 9%**	**-54 ± 9%**	**-34 ± 5%**
pH = 4.5	**30 ± 1%**	**19 ± 1%**	**19 ± 1%**	**20 ± 1%**	0 ± 0%	**-30 ± 4%**	-8 ± 1%	5 ± 1%
T = 28°C	**35 ± 5%**	**33 ± 4%**	**19 ± 3%**	**34 ± 5%**	**19 ± 4%**	**24 ± 4%**	**-9 ± 2%**	**80 ± 15%**
T = 37°C	**30 ± 5%**	**27 ± 4%**	**21 ± 3%**	**19 ± 3%**	0 ± 0%	0 ± 0%	0 ± 0%	0 ± 0%
T = 42°C	**27 ± 8%**	**26 ± 5%**	**22 ± 5%**	**23 ± 4%**	-10 ± 4%	-5 ± 1%	10 ± 3%	**22 ± 5%**

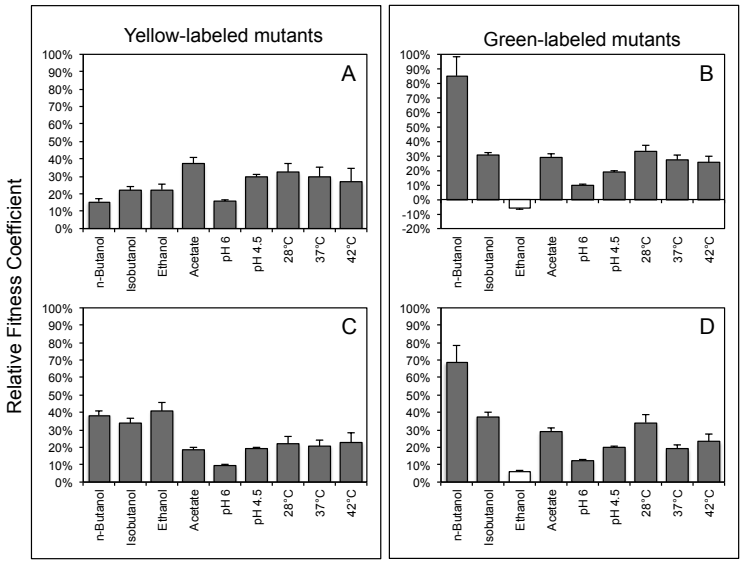


Figure 4.1. Cross-tolerance and antagonistic pleiotropy in the n-butanol evolved mutants: A. MY2, B. MG5, C. MY4, D. MG6. Filled bars indicate statistically significant values compared with the wild type strain (Student's t-test with a *p*-value

4.4.2. Identifying genetic determinants underlying n-butanol tolerance in the YFP-labeled mutants

The phenotypic results demonstrated different molecular mechanisms of n-butanol tolerance between the two lineages from the evolution experiment. To determine the molecular mechanisms behind the differences in phenotypic profiles observed between the GFP-labeled and YFP-labeled mutants, we analyzed the transcriptome profiles in the YFP-labeled mutants to identify potential genetic determinants underlying n-butanol tolerance for further analysis. We focused our analysis on the top ~10% of the upregulated genes involved in membrane-related processes, stress response, or global regulation from each of the YFP-labeled mutants (MY2, MY3 and MY4). A total of 46 genes were selected for further characterization by overexpressing each in the parental strain BW25113.

Initially, a pre-screen of the candidate genes was carried out in a high-throughput manner via growth kinetic analysis in the presence of 0.8% (v/v) n-butanol using a microplate reader. The genes that, when overexpressed, showed potential enhanced tolerance to n-butanol were further verified using batch cultures in close-capped flasks. The two kinetic parameters, *s* and *RIF*, were calculated to assess the improvement in n-butanol tolerance in the overexpression strains. Ten genes were found to have statistically significant effects on increasing n-butanol tolerance when overexpressed and their functions include transporters, membrane components and stress response (Table 4.2). Table 4.3 includes the list of all the screened genes and their different kinetic parameters.

Table 4.2. Summary of the list of genes that showed increased n-butanol tolerance when overexpressed in the ancestral strain BW25113. The *s* and *RIF* values were determined in the presence of 0.8% (v/v) n-butanol.

Overexpressed	*S*		*RIF*		Function
Gene	Average	*p*-value	Average	*p*-value	
sodB	24 ± 2%	1%	10 ± 1%	0%	Iron-superoxide dismutase
gcvH	23 ± 2%	0%	9 ± 1%	0%	Glycine cleavage system
hyfD	25 ± 2%	0%	10 ± 1%	1%	Hydrogenase
nuoI	22 ± 1%	0%	11 ± 1%	0%	NADH:ubiquinone oxidoreductase
pepB	40 ± 3%	0%	7 ± 1%	0%	Proteinase
treF	37 ± 3%	0%	7 ± 1%	0%	Trehalose biosynthesis
ygfO	35 ± 2%	0%	19 ± 1%	0%	Proton motive xanthine transporter
setA	24 ± 3%	4%	8 ± 1%	1%	Sugar efflux pump
mdtA	11 ± 1%	0%	5 ± 1%	0%	Drug resistance efflux pump
pgsA	15 ± 1%	0%	15 ± 1%	0%	Involved in cardiolipin biosynthesis

4.4.3. Membrane-associated genes

Modifications in the outer and cytoplasmic membranes are among the most common mechanisms for solvent tolerance in *E. coli*. We found several genes that encode for membrane-associated proteins to be upregulated in the YFP-labeled mutants.

The roles of several of these genes (*ygfO, setA, mdtA* and *pgsA*) in enhancing n-butanol tolerance were confirmed via overexpression studies in wild-type *E. coli* (see data in Table 4.2). The *mdtABC* (multidrug transporter ABC) operon, encoding the resistance-nodulation-cell division (RND) drug efflux system, is responsible for resistance against different compounds (161). Genes within the *mdt* operon have been previously found to be upregulated in n-butanol stress in *E. coli* (162). Several efflux pumps are known to be involved in solvent tolerance in bacteria, such as *srpABC* from P. putida in the export of octanol, hexane and other hydrocarbons (163), the *acrAB-tolC* pump in *E. coli* in tolerance to hexane, heptane, octane and nonane (164), and several other heterologously expressed pumps in *E. coli* conferred tolerance to different solvents (165). However no efflux pumps has been identified to be effective in exporting short-chain alcohols as n-butanol (165). The other membrane-related genes include the sugar transporter, *setA*, the xanthine transporter *ygfO*, and a biosynthetic gene for the phospholipid cardiolipin, *pgsA*.

Table 4.3. Relative fitness measurements for the top 10% of the most upregulated genes in at least one of the YFP-labeled mutants.

Gene	Microplate reader measurements		Closed-caps flask measurements			
	Fitness (*s*, %)	*p*-Value (%)	Fitness (*s*, %)	*p*-Value (%)	*RIF* (%)	*p*-Value (%)
accD	11	6	---	---	---	---
baeR	31	34	---	---	---	---
ccmC	2	23	---	---	---	---
citF	0	94	---	---	---	---
cusA	4	68	---	---	---	---
gcvH	16	2	23	0	9	0
hyfD	15	1	25	0	10	1
iraP	-3	81	---	---	---	---
iscS	7	22	---	---	---	---

Table 4.3 Continued.

Gene	Microplate reader measurements		Closed-caps flask measurements			
	Fitness (*s*, %)	*p*-Value (%)	Fitness (*s*, %)	*p*-Value (%)	*RIF* (%)	*p*-Value (%)
lpxK	-19	0	---	---	---	---
macB	-19	70	---	---	---	---
mdtA	7	3	11	0	5	0
mreB	12	43	---	---	---	---
nac	34	29	---	---	---	---
nmpC	-29	0	---	---	---	---
nuoI	20	3	22	0	11	0
pepB	14	2	40	0	7	0
pgsA	12	1	15	0	15	0
qseC	20	19	---	---	---	---
rfaJ	2	76	---	---	---	---
rnb	73	3	7	12	---	---
setA	10	1	24	4	8	1
sodB	110	0	24	1	10	1
tatA	9	9	---	---	---	---
tauD	72	0	4	59	---	---
tolC	26	1	25	14	---	---
tolR	81	0	6	51	---	---
treF	28	3	37	0	7	0
trkA	-63	0	---	---	---	---
waaU	-18	36	---	---	---	---
yadC	178	2	7	19	---	---
yagE	-17	2	---	---	---	---
ybaJ	148	0	8	13	---	---
ybhA	147	10	---	---	---	---
yccM	152	6	---	---	---	---
ycdU	-47	0	---	---	---	---
yfcR	-14	3	---	---	---	---
yeaV	0	96	---	---	---	---
yeeS	-39	0	---	---	---	---
ygfO	19	5	35	0	19	0
yqcE	-10	2	---	---	---	---
zntA	0	98	---	---	---	---

In order to determine whether the overexpression of the above-mentioned genes enhance general tolerance in the presence of growth inhibitors or exclusively in enhancing tolerance to n-butanol, their effects under different conditions (0.8% (v/v) n-butanol, 4% (v/v) ethanol, 1% (v/v) isobutanol and 1.75 g/L acetate) were assessed. The results are depicted in Figure 4.2. Interestingly, the overexpression of *ygfO, mdtA* and *pgsA* significantly decreased fitness when the strains were grown in the presence of isobutanol. In the case of *ygfO* overexpression, growth was completely inhibited in the presence of isobutanol, at least within the 12 hours during which the experiment was conducted (Figure 4.2A). No cross-tolerance between n-butanol and ethanol were observed when any of the genes were overexpressed. Overexpression of *ygfO* significantly decreased fitness in all the conditions tested except for n-butanol, suggesting the mechanism of n-butanol tolerance in this strain caused a broad range antagonistic pleiotropy with other inhibitors. Cross-tolerance between acetate and n-butanol stress have been identified previously in *C. acetobutylicum* (166), and thus was included as a test condition here. In the case of acetate, *setA* was the only gene that conferred enhanced tolerance when overexpressed.

Even though these membrane related genes were identified based on the transcriptome data of the YFP-labeled mutants, the phenotypic profiles related with the overexpression of individual genes (Figure 4.2) differ from that observed in mutants MY2 and MY4 (Figure 4.1).

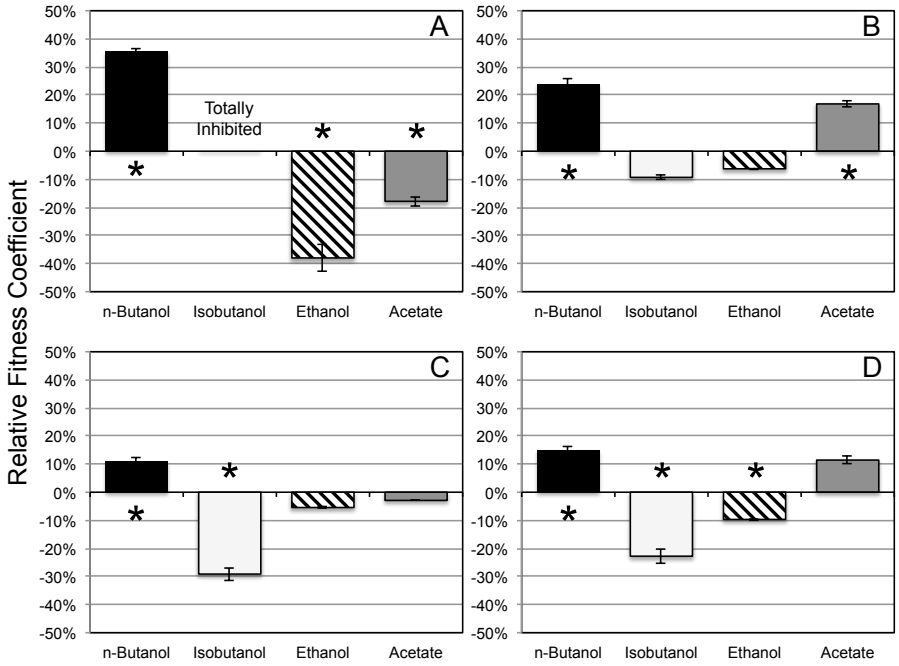


Figure 4.2. The relative fitness coefficient of overexpression of genes related with different transporters and membrane components. A. *ygfO*, B. *setA*, C. *mdtA* and D. *pgsA*. The asterisks indicate statistically significant values using a Student's t-test (p-value < 0.05). Growth was completely inhibited when *ygfO* was overexpressed in the presence of 1% (v/v) isobutanol.

4.4.4. Osmotic stress

The transcriptome analyses revealed several genes known or potentially involved in osmoprotection (*gcvH, treF, setA* and *pgsA*) to be perturbed in the YFP-labeled mutants, but not in the GFP-labeled mutants. Overexpression of these genes in a wild-type strain was found to improve n-butanol tolerance as shown in Table 4.2. Glycine and glycine-betaine (coded by members of the GcvA and BetI regulons) have been identified to contribute to ethanol tolerance in *E. coli*, possibly through higher production of osmolytes (64). The biosynthesis of the disaccharide trehalose (produced by the cytoplasmic TreF and the periplasmic TreA) acts as an osmotic and stress protectant in *E. coli* (167).

This osmoprotectant action was of importance in the presence of different osmotic agents such as inorganic salts and high concentrations of hexose sugars. SetA is an efflux pump capable to transport different sugars, and the expression of *setA* was found to be increased under glucose/phosphate stress (168), potentially alleviating the accumulation of non-metabolized sugar-phosphates. However, SetA also transports, albeit inefficiently, other substrates as the antibiotics kanamycin and neomycin, as well as glucosides and galactosides with alkyl or aryl substituents (169). Since there are no sugar analogs in the media, it is currently unclear whether the overexpression of *setA* contributes to osmoprotection in MY2 and MY4 mutants.

The gene *pgsA*, related to the biosynthesis of the phospholipid cardiolipin, was perturbed in MY2, suggesting a potential increase in cardiolipid biosynthesis in this mutant. Cardiolipin (CL) is a glycerophospholipid, which is important in microbial cell membranes (170). CL plays important roles in bacterial cell division (158, 171), osmotic stress (172), and essential function in the bioenergetics of the cell (173) In *E. coli*, *pgsA* encodes the committed step of CL biosynthesis. Overexpression of *pgsA* has been shown to modify cellular phospholipid composition, by increasing the concentrations of the acidic phospholipids phosphatidylglycerol and CL (174).

Since several potentially osmoprotection-related genes were overexpressed in the evolved YFP-labeled mutants, and their overexpression was verified to be involved in n-butanol tolerance, we hypothesized that the YFP-labeled mutants are also more tolerant to osmotic stress. An osmotic shock experiment using high concentrations of glucose (40% w/v) as described in the Materials and Methods section was used. As shown in

Figure 4.3, statistically significant improvements in relative survival rates were seen in the three YFP-labeled mutants (244% for MY2, 499% for MY3 and 655% for MY4). Cross-tolerance between osmotic tolerance and solvent stress responses has been observed previously: cross-tolerance between ethanol and osmotic stress in *E. coli* (167, 175) and *S. cerevisiae* (176, 177), and between n-butanol and osmotic stress in *E. coli* (178) and *Clostridium* (103). Intriguingly, the GFP-labeled mutants show antagonistic pleiotropy between n-butanol and osmotic stress (relative survival rate of -65% for MG5 and -41% for MG6).

4.4.5. Other potential mechanisms

We found that overexpression of *sodB*, an iron-superoxide dismutase, significantly increased n-butanol tolerance in *E. coli*. Overexpression of *sodB* has also been found in response to different aromatic substrates and phenol-induced stress in *P. putida* (179, 180), as well as organic solvent stress response in the denitrifying bacteria *Aromatoleum aromaticum* (181). Oxidative stress response has been previously associated with n-butanol stress response in *E. coli*, particularly the overexpression of *sodA, sodC* and *yqhD* (12, 17). In fact, the n-butanol stress response has been connected with different well-studied stress responses including oxidative stress response, heat shock, cell envelope stress and perturbations of different respiratory functions (12, 17).

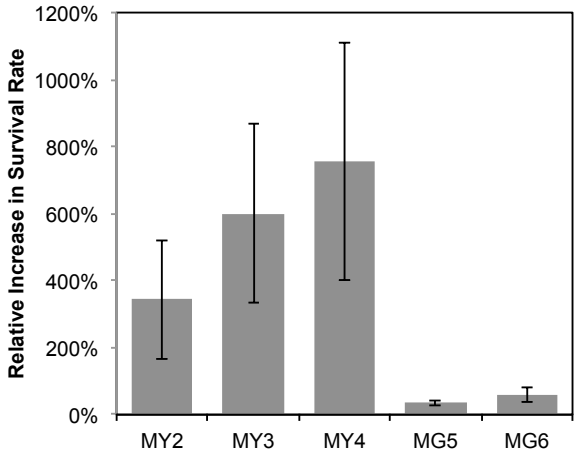


Figure 4.3. Relative changes in survival rates in osmotic stress. E. coli mutants from the YFP-labeled lineage showed increased tolerance to osmotic stress. All the values presented here are statistically significantly different compared with their respective colored parental strain (p-value < 0.05). Error bars represent the standard deviation.

The proteinase PepB, found to increase n-butanol tolerance by more than 40% when overexpressed, has been discovered to be significantly downregulated during acid stress in *Streptococcus* mutants (182). Other mechanisms of protein folding and degradation have been observed in response to n-butanol stress, such as activation of GroESL chaperone system in *E. coli* (159, 183) and *C. acetobutylicum* (160).

Under n-butanol exposure, Rutherford, *et. al*. (162) observed upregulation of the *nuo* and *cyo* operons, indicating either an increased energy requirement or impairment of respiration. Studies have demonstrated that these genes are necessary for the generation of proton gradients across the inner membrane for aerobic energy generation (184). Here we showed an increase of more than 23% in fitness relative to wild type under n-butanol stress when *nuoI* was overexpressed.

4.4.6. Divergence in evolutionary trajectories for enhanced n-butanol tolerance in *E. coli*

The observed differences in the phenotypic profiles between the YFP-labeled and GFP-labeled mutants demonstrated a potential divergence in n-butanol tolerance mechanisms between the two lineages. We next asked the question whether the two different mechanisms could be combined to generate a more n-butanol tolerant strain. The iron-transport-related genes, *feoA* and *entC*, were found to be involved in n-butanol tolerance in the GFP-labeled mutants (21). These two genes were individually overexpressed in the YFP-labeled mutants and the levels of n-butanol tolerance of the resulting strains were assessed. As shown in Figure 4.4, in several cases the relative fitness coefficient decreased significantly with the expression of the iron-related genes in the YFP-labeled mutants, suggesting an antagonistic relationship between the n-butanol tolerance mechanisms in the two differentially colored lineages. Overexpression of *feoA*, involved in the cellular uptake of iron, showed the higher antagonistic effect on the YFP-labeled mutants, where the fitness advantage of MY3 and MY4 in the presence of n-butanol decreased significantly (negative fitness coefficients were observed). In case of *entC*, which is involved in the biosynthesis of enterobactin, decreased the relative fitness in MY3 (fitness was unchanged in MY2 and MY4) in the presence of n-butanol when overexpressed.

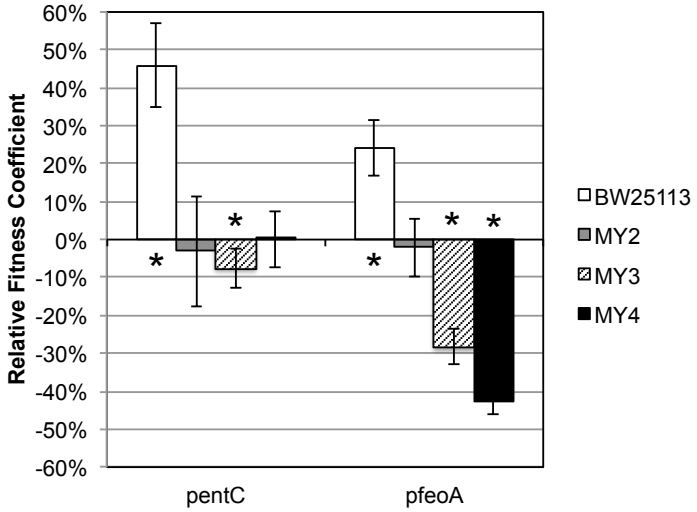


Figure 4.4. Relative fitness coefficients of the YFP-labeled mutants overexpressing *entC* and *feoA* genes. The fitness coefficients were calculated using the following equation $s = (\mu_{strain\_overexpressing\_gene} / \mu_{strain}) - 1$. The asterisks represent statistically significant (p-value < 0.05) values. Error bars represent the standard deviation.

4.5. Discussion

In this study, we demonstrated a case of incompatibility between the mechanisms of n-butanol tolerance between two different lineages from the same evolved population, and identified cross-resistance and antagonistic pleiotropy between different complex phenotypes in the isolated mutants of *E. coli* evolved under n-butanol stress. By using VERT we were able to systematically track the evolutionary dynamics to distinguish parallel mechanisms of n-butanol tolerance in the same population, which would otherwise be difficult to determine. Our results indicate that adaptation (increased n-butanol tolerance) of *E. coli* to increasing concentrations of n-butanol from a single evolving population resulted in two independent and divergent adaptive mechanisms.

MY4 was the only strain that showed a general solvent tolerance to n-butanol, isobutanol, and ethanol (see Table 4.1). Interestingly, MY4 showed a higher fitness advantage in the presence of ethanol compared to n-butanol, suggesting the molecular mechanisms for n-butanol tolerance in this strain is more general towards solvent stress. Whole genome sequencing of MY4 led to the identification of a single nucleotide polymorphism in *nusA* (nusAE212A) as we reported previously (21). NusA interacts with several proteins involved in transcriptional termination including *rho*, a transcriptional terminator affecting expression at whole-cell level. Point mutation in *rho* has been shown to be a major contributor to ethanol tolerance in *E. coli* evolved under ethanol stress (64). Mutations in global regulators have also been shown to enhance ethanol tolerance and n-butanol tolerance (47, 185, 186). Thus, the observed mutation in *nusA*, in part, potentially explains the generalist behavior of this mutant toward different stressors by globally perturbing transcriptional regulation in the cell. On the other hand, the GFP-labeled mutants exhibited much higher relative fitness coefficients in the presence of n-butanol compared to any other stress conditions tested, suggesting that GFP-labeled mutants had evolved a more "specialist" adaptive mechanisms for n-butanol tolerance. However, MG5 and MG6 (the GFP-labeled mutants) both share a same point mutation in *rho* in addition to an IS5 insertion in *feoA* (presented in Table 3.5). As *feoA* has been confirmed as a likely genetic determinant for enhanced n-butanol tolerance in the GFP-labeled mutants, the results currently suggest that the downstream effects of increased expression of iron uptake-related gene(s) is dominant over any potential downstream ef-

fects of the *rho* mutation; further investigation is needed to decipher the respective effects of each mutation in these strains.

The phenotypic response of the alleles under different stress conditions tested indicates divergence in the acquired mechanisms to increase n-butanol tolerance between the isolated mutants from the two lineages. Opposite trends in fitness under several conditions were observed. The most notable one is the differential responses to osmotic stress. Under osmotic stress, all the isolated mutants from the yellow lineage showed improved tolerance, while mutants from the green lineage showed an opposite response (Figure 4.3). In addition, Reyes *et al* (21) documented a cross-tolerance between the cationic antibiotic peptide polymyxin B and n-butanol in only the GFP-labeled mutants. This differential cross-resistance in mutants from the two lineages indicates the possibility of diverse mechanisms of n-butanol tolerance in *E. coli*. Furthermore, the mechanisms of n-butanol tolerance between the two lineages appear to be antagonistic, as the introduction of *feoA* into MY2 or MY4 resulted in a significant decrease in relative fitness in the presence of n-butanol. A similar phenomenon was observed previously between different lineages from a yeast population evolved in glucose-limited conditions (187).

4.6. Conclusion

The use of microbial systems in the production of chemical and fuels from sustainable feedstock requires robust biocatalysts. As the production environment poten-

tially involves multiple inhibitory factors, knowledge regarding the cross tolerance/antagonistic pleiotropy between different conditions of an adaptive allele becomes important. This work further characterized isolated evolved mutants from a single n-butanol challenged population and identified additional genetic determinants involved in n-butanol tolerance in *E. coli*, discovered several cases of cross tolerance/antagonistic pleiotropy between different stressors in the isolated mutants, and a case of negative epistasis between mechanisms of n-butanol tolerance between independent lineages from the same population. Our results suggest that while MG5 and MG6 are our best performers under n-butanol stress, these two mutants may not perform as well as MY2 or MY4 in a production environment; thus, highlighting the importance of characterizing the effects of identified genetic determinants under multiple relevant conditions to better guide strain engineering efforts.

5. EVOLUTIONARY ENGINEERING OF SACCHAROMYCES CEREVISIAE FOR ENHANCED TOLERANCE TO HYDROLYSATES OF LIGNOCELLULOSIC BIOMASS[*]

5.1. Summary

Lignocellulosic biomass has become an important feedstock to mitigate current ethical and economical concerns related to the bio-based production of fuels and chemicals. During the pre-treatment and hydrolysis of the lignocellulosic biomass, a complex mixture of sugars and inhibitors are formed. The inhibitors interfere with microbial growth and product yields. This study uses an adaptive laboratory evolution method called Visualizing Evolution in Real-Time (VERT) to uncover the molecular mechanisms associated with tolerance to hydrolysates of lignocellulosic biomass in *Saccharomyces cerevisiae*. VERT enables a more rational scheme for isolating adaptive mutants for characterization and molecular analyses. Subsequent growth kinetic analyses of the mutants in individual and combinations of common inhibitors present in hydrolysates (acetic acid, furfural, and hydroxymethylfurfural) showed differential levels of resistance to different inhibitors, with enhanced growth rates up to 57%, 12%, 22%, and 24% in hydrolysates, acetic acid, HMF and furfural, respectively. Interestingly, some of the adaptive mutants exhibited reduced fitness in the presence of individual inhibitors, but

[*] Reprinted with permission from Almario, M. P., Reyes, L. H., & Kao, K. C. (2013). Evolutionary engineering of Saccharomyces cerevisiae for enhanced tolerance to hydrolysates of lignocellulosic biomass. Biotechnology and Bioengineering, doi:10.1002/bit.24938. Copyright 2012 John Wiley and Sons.

showed enhanced fitness in the presence of combinations of inhibitors compared to the parental strains. Transcriptomic analysis revealed different mechanisms for resistance to hydrolysates and a potential cross adaptation between oxidative stress and hydrolysates tolerance in several of the mutants.

5.2. Introduction

There is an increased interest in the use of renewable feedstock for the production of second-generation biofuels and chemicals. Lignocellulosic biomass is an abundant renewable resource that is estimated to reach more than 1 billion dry tons annually in the U.S. by year 2030 (188). In general, the lignocellulosic biomass needs to be pretreated and hydrolyzed to convert its cellulose and hemicellulose into simple sugars for fermentation. The breakdown of lignin produces phenolic compounds (189). Hydrolysates of lignocellulosic biomass are complex mixtures of different pentose and hexose sugars, inhibitors (e.g. furfural, 5-hydroxymethyl furfural, phenolic compounds, etc.), and salts (e.g. $MgSO_4$, NaCl) (190, 191). Dilute acid pretreatment is a commonly used method to facilitate enzymatic hydrolysis and improve sugar yield (190, 192). The pretreatment process generates numerous inhibitors that affect cellular growth and fermentation performance (189, 193, 194). The byproducts generated after dilute acid pretreatment that have the most potent inhibitory effects include acetic acid (AA), formic acid, phenolic compounds and the furan aldehydes furfural and 5-hydroxymethylfurfural (HMF) (195, 196). AA is formed by de-acetylation of hemicelluloses (189) and several

mechanisms have been determined to be involved in acetic acid tolerance in yeast (197, 198). The furans are derived from dehydration of hexoses and pentoses (189). The inhibitors present in hydrolysates cause reduction in biological and enzymatic functions in yeast and *Escherichia coli* (199-201). In addition, synergistic effects between different inhibitors have been found to further enhance the negative impact on the growth rate of *Saccharomyces cerevisiae* (199, 202). Several mechanisms to overcome inhibition have been identified, one of which is the *in situ* detoxification of HMF and furfural to less toxic compounds, furan di-methanol and furan methanol respectively, through NADPH-dependent reductions (189, 194, 202). Mechanisms to overcome the toxic effects of weak acids, such as acetic acid, appear to be complex, involving the reduced uptake of extracellular acetate, increased activity of plasma membrane H^+-ATPase and efflux of the acetate through multidrug resistance transporters (MDR) (203, 204). However, the basic mechanisms to overcome inhibition in the presence of the complex mixtures of potential inhibitors present in hydrolysates of lignocellulosic biomass are not well known (189, 199).

The purpose of this study is to understand the mechanisms associated with tolerance of *S. cerevisiae* to hydrolysates of lignocellulosic biomass, using the adaptive laboratory evolution method called visualizing evolution in real time (VERT) (21, 99). During adaptive evolution, mutants with enhanced fitness (adaptive mutants) arise and expand in the population. We refer to these expansions as adaptive events. VERT helps to identify adaptive events in an evolving population via tracking the changes in the relative frequencies of differentially labeled colored-subpopulations using flow cytometry.

Thus, an observed expansion in a colored-subpopulation is indicative of the rise and expansion of an adaptive mutant within the expanding colored-subpopulation. Adaptive mutants can then be isolated from the expanding colored-subpopulation for subsequent analyses. Using VERT, *S. cerevisiae* was evolved in the presence of hydrolysates and individual adaptive mutants were isolated throughout the course of the evolution. The relative resistances of the isolated adaptive mutants to the different inhibitory compounds found in the hydrolysates were assessed. In conjunction with the phenotypic analysis, transcriptome analysis was conducted for the different isolated mutants to elucidate the molecular mechanisms involved in hydrolysates tolerance.

5.3. Materials and Methods

5.3.1. Strains and growth conditions

Fluorescently marked *S. cerevisiae* strains were derived from FY2, a derivative of S288c, as previously described (54). Unless otherwise specified, yeast cells were cultured in YNB supplemented with no carbon source and mixed with different concentrations of hydrolysates (see Table 5.1 and Table 5.2) of lignocellulosic biomass at 30ºC.

5.3.2. Hydrolysates of lignocellulosic biomass preparation method

To perform the enzymatic hydrolysis reaction, the pH of the solution (slurry/ water) was titrated at pH=5.0 using 1M KOH. The following equations were used to determine the required amount of enzyme:

Cellulase: For X ml of corn slurry at pH 5, for m minutes of incubation at 37C.

Equation 5.1 $$\text{Cellulase(mg)} = \frac{\text{w(umol/mL)*X(ml of corn slurry)}}{\text{m(min/U)*1(umol/min)*1.24U/mg}}$$

Cellobiase: For X ml of corn slurry at pH 5, for m minutes of incubation at 37C.

Equation 5.2 $$\text{Cellobiase(mL)} = \frac{\text{w(umol/mL)*X(ml of corn slurry)}}{\text{m(min/U)*2(umol/min)*}\frac{250\text{U}}{\text{g}}\text{*1.2}(\frac{\text{g}}{\text{L}})}(2)$$

Note: W g/L or w ($\frac{G(g)*1000}{180.2(\frac{g}{mol})}$umol/mL) is the desired concentration of glucose in the solution. After the addition of the enzymes, the corn slurry solution was incubated at 37C for m=60 minutes with continuous shake of 220 rpm.

Table 5.1. Concentration of each inhibitor and sugar present in various batches of hydrolysates.

Batch Hydrolysates	Glucose % (w/v)	Xylose %(w/v)	Arabinose %(w/v)	HMF (g/L)	Furfural (g/L)	Acetic Acid (mM)
Batch 1	0.58	1.03	0.16	0.88	0.03	9.27
Batch 2	0.21	0.63	0.08	0.60	0.02	6.68
Batch 3	0.58	1.78	0.26	2.38	0.07	20.34
Batch 4	0.67	2.00	0.29	2.79	0.07	23.47
Batch 5	0.67	2.67	0.4	3.00	0.07	31
Batch 6	0.88	2.57	0.38	3.18	0.08	29.17

5.3.3. Preparation of solutions of hydrolysates of lignocellulosic biomass

Hydrolysates of lignocellulosic biomass were obtained using dilute acid pretreated biomass slurry of corn stover (courtesy of Dr. Dan Schell and Dr. John Ashworth at the National Renewable Energy Laboratory). A determined amount (g) of this slurry

(with solids loading of roughly 35%) was diluted in DI water for enzymatic hydrolysis. The composition of each batch of hydrolysates after enzyme hydrolysis was quantified (details below) and listed in Table 5.1. The resulting hydrolysates were filter sterilized using 0.22 μm filters and subsequently diluted in YNB with no carbon source (5.95 g YNB without amino acids or ammonium sulfate and 17.5 g Ammonium Sulfate). The composition of each batch of hydrolysates/YNB mixture is listed in Table 5.2. Hydrolysates composition was analyzed using High-Performance Liquid Chromatography (HPLC; Agilent Technologies 1260 Infinity) using an Aminex HPX-87H column (Bio-Rad). Glucose, xylose, arabinose, acetic acid and ethanol were quantified using the RI detector at 52℃, using 0.6 ml/min of 5 mM H2SO4 as mobile phase. Furfural and HMF were detected using the UV/vis detector at 254 nm.

Table 5.2. Composition of selective media used for evolution and phenotypic analyses.

Selective media	Generation		Hydrolysates Concentration	
	From	To	Batch[1]	Dilution with YNB[2]
A	0	54	1	17%
B	56	62	2	48%
C	64	70	3	17%
D	73	77	4	15%
E	79	83	4	16%
F	85	91	3	25%
G	93	104	4	25%
H	106	120	5	25%
I	122	160	6	25%

[1]Batch number of hydrolysates from Table 5.1 used for preparation of selective media.
[2]Percent by volume of hydrolysate used in dilution with YNB (without glucose supplementation).

5.3.4. Neutrality test

A neutrality test was performed to ensure that there are no fitness differences between the three fluorescently marked strains used for this work (GFP, YFP, DsRed). For

this test, roughly equal numbers of these labeled strains were fed in a batch culture with YNB and glucose as carbon source (Figure 5.1 shows the results).

Table 5.2. (Cont.) Composition of selective media used for evolution and phenotypic analyses

Selective media	Glucose[3] %(w/v)	Xylose[3] %(w/v)	Arabinose[3] %(w/v)	Acetic Acid[3] mM	HMF[3] g/L	Furfural[3] g/L
A	0.1	0.18	0.03	1.60	0.15	0.01
B	0.1	0.30	0.04	3.18	0.29	0.01
C	0.1	0.31	0.04	3.51	0.41	0.01
D	0.1	0.30	0.04	3.50	0.42	0.01
E	0.1	0.32	0.05	3.98	0.41	0.01
F	0.14	0.445	0.06	5.08	0.59	0.01
G	0.17	0.53	0.07	6.17	0.67	0.01
H	0.17	0.67	0.10	7.75	0.75	0.02
I	0.22	0.642	0.10	7.30	0.8	0.02

[3]Actual concentration in each batch of selective media.

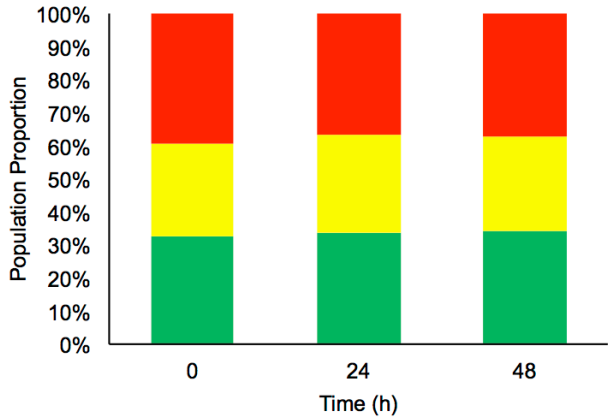


Figure 5.1. Neutrality test. The yellow bars indicate the relative proportion of the yellow subpopulation, green bars indicate the proportion of green subpopulation, and red bars indicate the proportion of the red subpopulation.

5.3.5. Adaptive evolution experiment

The adaptive evolution experiments were performed in 125 ml flasks at 30ºC in 10 mL of hydrolysates/YNB with three parallel populations. Roughly equal numbers of

the three fluorescently marked (GFP, YFP, and DsRed) strains were used to seed each population. A serial transfer of 1% of the total volume into fresh medium was performed every 24-48 hours. Samples were taken prior to each serial transfer to monitor the relative proportions of the three colored subpopulations using a flow cytometer (FACS) (BD FACScan™) and to generate frozen stocks to be stored at -80ºC in 17.5% glycerol for further analysis. The concentration of hydrolysates was increased in a stepwise fashion (see Table 5.2). Table 5.2 lists the concentrations of sugars and inhibitors in each batch of selective media (selective media *A*-*I*) used for the evolution experiments. The concentration of hydrolysates was increased when an adaptive event (increase in more than 2 consecutive measurements in the relative proportion of a colored subpopulation) was observed. This experiment was conducted for approximately 463 generations.

5.3.6. Isolation of adaptive mutants

For the isolation of adaptive mutants responsible for each of the observed adaptive events, five clones from each expanding colored-subpopulation were isolated and their growth kinetics measured in selective media *G* (hydrolysates/YNB medium) (see Table 5.2) in 96-well plates using a microplate reader (TECAN™ Infinite M200) at 30ºC. The parameter "Relative increase in maximum specific growth rate" (21) for each clone was calculated using Equation 5.3, where μ_i is the maximum specific growth rate for strain *i*.

Equation 5.3 Relative increase in $\mu_i = \frac{\left(\mu_{i@Hydro} / \mu_{i@No\ Hydro}\right) - \left(\mu_{wild\ type@Hydro} / \mu_{wild\ type@No\ Hydro}\right)}{\left(\mu_{wild\ type@Hydro} / \mu_{wild\ type@No\ Hydro}\right)}$

The clone with the highest "Relative increase in μ_i" was chosen as the adaptive clone from the expanding subpopulation. Student's t-test using four biological replicates was used to assess significance. The relative fitness coefficient, *s*, for each mutant was determined using Equation 5.4.

Equation 5.4 $$s = \left(\frac{\mu_{Adaptive\_Mutant}}{\mu_{Wild\ type}}\right) - 1$$

Thus a mutant with a relative fitness coefficient greater than "0" has higher fitness compared with the parental strain at that condition. Statistical significance for *s* was assessed by testing the null hypothesis that $\mu_{Adaptive\_Mutant}$ was not higher (or lower, depending on the sign of *s*) than $\mu_{Wild\ type}$ using a one-tailed student t-test (with p-value threshold of 0.05).

5.3.7. Growth kinetics

Three biological replicates of the selected isolated mutants were grown in 125ml flasks in batch cultures to determine the relative fitness coefficient, *s*, of each mutant against its parental strain in different conditions. The growth kinetics for each culture was determined by monitoring the OD_{600} every 2 hours from an initial OD_{600} of ~0.05.

The relative fitness coefficient *s* was determined using Equation 5.4. The analyzed conditions were: a) hydrolysates of biomass with YNB, b) YNB supplemented with 1% (w/v) glucose, c) YNB supplemented with 1% (w/v) glucose and 0.9g/L HMF, d) YNB supplemented with 1% (w/v) glucose and 50mM acetic acid, e) YNB supplemented with 1% (w/v) glucose and 0.9g/L HMF and 0.4g/L furfural, and f) YNB supplemented with 1% (w/v) glucose and 0.4g/L furfural. Cultures grown in condition b) were also analyzed for their rates of glucose consumption using the HPLC.

5.3.8. RNA extraction

Each isolated mutant was inoculated in 30ml of selective media *H* (hydrolysates/YNB) on 125ml flask at an initial OD_{600} of 0.05. The cells were harvested in mid-exponential phase by filtration using analytical test filter funnels (Nalgene), immediately resuspended in 10mL of RNA*later* (Sigma) and stored at -80ºC for future analysis. For the extraction of the total RNA, the ZR Fungal/Bacterial RNA MiniPrep™ (Zymo) kit was used following manufacturers' instructions using 2mL of each stored sample in RNA*later*. The extracted RNA was quantified using the NanoDrop™ 1000® (Thermo Scientific).

5.3.9. Microarray hybridization and data analysis

The SuperScript® Indirect cDNA Labeling System (Invitrogen) with $Oligo(dT)_{20}$ primer was used to synthesized the cDNA with incorporation of aminoallyl dUTP. cDNA was recovered with ice cold ethanol precipitation. Cy3- and Cy5- mono-Reactive

Dye Pack (GE Healthcare) were used to label cDNA samples. The labeled cDNA was hybridized to the *S. cerevisiae* G4813A Gene Expression Microarray (Agilent Technologies). The arrays were scanned using the GenePix 4100A Microarray Scanner and GenePix Pro 6.0 software (Molecular Devices) for image analysis. The Microarray Data Analysis System (MIDAS) software was used to normalize the data using LOWESS based normalization algorithm (118). Differentially expressed genes were identify using the rank product method with a critical p-value <0.05. The MeV (TM4) (205) microarray analysis software was used for clustering and other expression profile analysis. The *Saccharomyces* Genome Database SGD (206) was use to analyze the data using gene ontology. The Network Component Analysis (NCA) (134) was used to analyze the activity of *S. cerevisiae* transcription factors. The number of permutations used to calculate the FDR (false discovery rate) was 1000. The transcription factors reported have a statistical significance with a p-value threshold of <0.05.

5.4. Results and Discussion

5.4.1. Evolutionary dynamics of *S. cerevisiae* during the adaptive evolution for tolerance to hydrolysates of lignocellulosic biomass

Approximate equal numbers of fluorescently (GFP, YFP and DsRed) marked strains were used to seed three hydrolysates-challenged populations (P1, P2 and P3) in batch cultures in selective media *A* (Table 5.2). The concentration of the hydrolysates was increased in a step-wise manner over the course of 463 generations (see Table 5.2

and Figure 5.2). The relative proportions of the three different colored-subpopulations were monitored using FACS. As described in Reyes *et al* (21), the expansions and contractions of different colored subpopulations observed using VERT are the results of rise and expansions of adaptive mutants (fitter mutants) within the different colored subpopulations. Thus, an expansion in a colored subpopulation is indicative of the occurrence and expansion of an adaptive mutant in that colored subpopulation, and fitter mutants can then be isolated from the expanding colored subpopulation for subsequent analyses.

Upon conclusion of the adaptive evolution experiments, population P3, which reached the highest concentration of hydrolysates, was chosen for further analysis. Seven adaptive events were identified in P3, where the expanding subpopulation reached their maximum proportion at generations 132, 164, 231, 259, 338, 353, and 438 (Figure 5.2**)**. To isolate the adaptive mutants responsible for each expansion, five clones were randomly picked from each expanding subpopulation at the above-mentioned generations. The clone with the highest specific growth rate in hydrolysates/YNB was chosen as the adaptive mutant from the expanding subpopulation. These adaptive mutants were named according to their color and generation from which they were isolated: PG-132, PR-164, PG-259, PY-231, PR-338, PG-353, and PR-438.

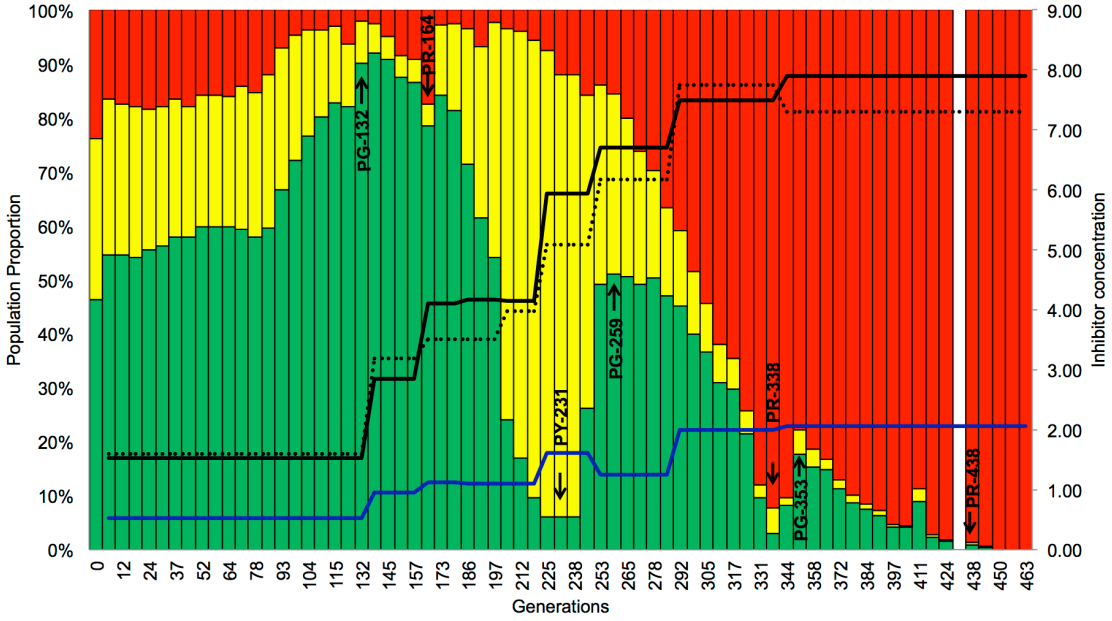


Figure 5.2. The evolutionary dynamics of population P3 during *in vitro* evolution in hydrolysates-challenged condition. The yellow bars indicate the relative proportion of the yellow subpopulation, green bars indicate the proportion of the green subpopulation, and red bars indicate the proportion of the red subpopulation. The black line indicates the approximately concentration of HMF (x10, g/L), dashed black line corresponds to acetic acid (mM), and blue line is furfural (×100, g/L). The arrows indicate the adaptive mutants isolated.

5.4.2. Relative fitness in the presence of hydrolysates and individual inhibitors

The growth kinetics for each adaptive mutant was determined in batch cultures to compare their growth rates against their respective parental strain (of the same color) in several different conditions and their relative fitness coefficients were calculated. First, the specific growth rates of the mutants and parental strains were measured in selective media *H* (Table 5.2). Six out of seven of the isolated adaptive mutants showed a statistically significant positive relative fitness coefficient compared with the parental strain

(see Figure 5.3). The largest improvement was observed in mutant PR-164 with a 57% improvement in specific growth rate over the parental strain in hydrolysates.

Since hydrolysates of biomass contain a complex mixture of inhibitors, adaptation of the evolved mutants may be a result of enhanced resistance to one or more of the inhibitors present. To determine which of the most commonly studied inhibitors (furfural, HMF, and AA) the evolved mutants have enhanced resistance to, the relative fitness coefficients of each strain in the presence of each of the inhibitors were determined in YNB supplemented with 1% (w/v) glucose (see Figure 5.3). All but one mutant (PY-231) showed enhanced fitness in at least one of the conditions tested. The concentration of HMF used in the test was 0.9 g/L (the highest concentration of HMF in the hydrolysate was 0.8 g/L). Since the highest concentration of furfural present in the hydrolysates (0.02 g/L) was not inhibitory by itself, the relative fitness coefficient measurements were carried out in a concentration of 0.4 g/L of furfural.

Adaptive mutants PG-132 and PR-164 showed a statistically significant increase in relative fitness in the presence of furfural, and mutants PR-164 and PR-438 exhibited higher tolerance to HMF compared to their parental strains. All adaptive mutants from the red subpopulation (PR-164, PR-338, and PR-438) and the green mutant PG-353 showed higher tolerance to AA (50 mM). Adaptive mutants PR-164 and PR-438 were the most resistant to AA with relative fitness coefficients of 12%. In addition to individual inhibitors, we also tested the resistances of each mutant in the presence of both HMF (0.9 g/L) and furfural (0.4 g/L). Four out of seven adaptive mutants showed statistically significantly improved fitness in the presence of both HMF and furfural, with PG-353

exhibiting the highest tolerance with a fitness improvement of 46% over the parental strain.

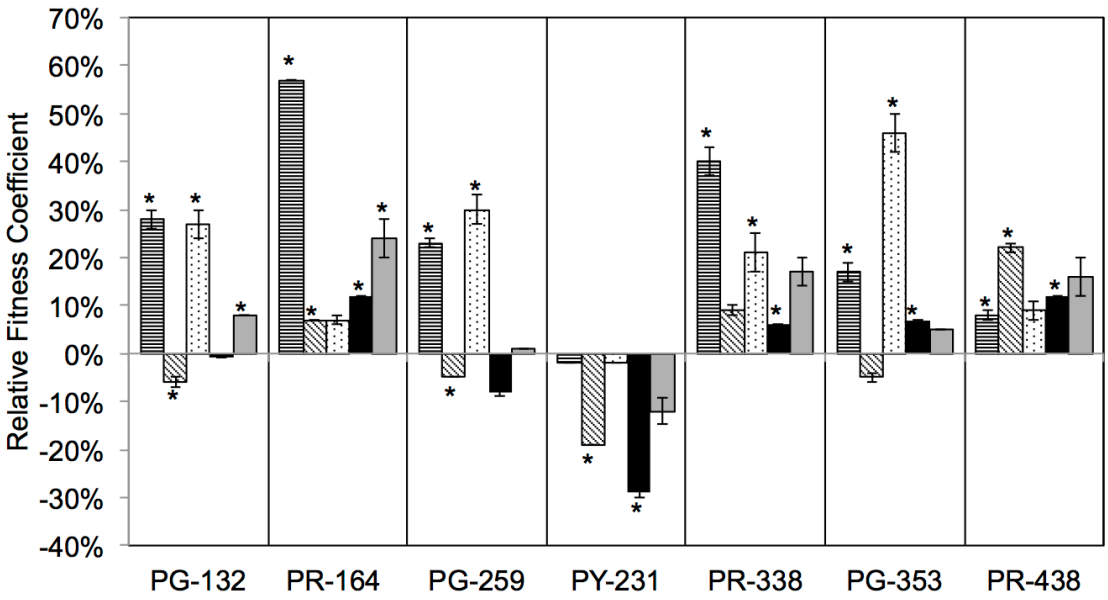


Figure 5.3. The relative fitness coefficients of isolated mutants from population P3 in the presence of each inhibitor. Horizontal lined bars: Selective media *H* (hydrolysates in YNB), diagonal lined bars: HMF (0.9 g/L), dotted bars: HMF (0.9 g/L) and furfural (0.4 g/L), solid black bars: AA (50 mM), and solid gray bars: furfural (0.4 g/L). YNB supplemented with 1% glucose is the base media used for conditions containing individual inhibitors. Asterisks indicate statistical significance between the specific growth rates of isolated mutant and parental strains with a p-value threshold of 0.05 (using a one-tailed student t-test with a minimum of 3 biological replicates).

5.4.3. Transcriptome analysis

Transcriptome analysis using DNA microarrays was used to identify any differences in relative transcript abundances between the isolated mutants and the parental strains in the presence of hydrolysates. The results showed significant perturbations in different metabolic pathways in the mutants, and there exists significant differences in transcriptional regulation between linages. The transcriptome data was analyzed based

on the individual performance of each adaptive mutant to the major inhibitors present in hydrolysates. Some of the key perturbations are described below. Transcription factors activities were assessed using Network Component Analysis (NCA) (133, 134).

5.4.3.1. Acetic acid

The adaptive mutants with a statistically significant positive relative fitness coefficient against their parental strains in the presence of AA were PR-164, PR-338, PG-353 and PR-438. Commonly up-regulated genes between these mutants were *ATP5*, *VMA3 VPH1*, and *SPI1*. *ATP5* is a required gene for ATP synthesis coupled proton transport. *VPH1* and *VMA3* are genes involved with vacuolar ATPase acidification. *SPI1* is a gene involved in weak acid resistance and its expression is controlled by the transcription factors Msn2p/ Msn4p (207). Depletion of ATP (uncoupling theory) and intracellular anion accumulation have been suggested to be responsible for the toxicity effects of weak acids (208). The uncoupling theory states that the decrease in cytoplasmic pH generated by the influx of AA results in the activation of ATP-dependent proton pumps to neutralize the pH, leading to a depletion of ATP (208). The intracellular anion accumulation theory, on the other hand, states that accumulation of dissociated AA inside the cell (since only protonated or associated AA can pass across the cytoplasmic membrane) due to the lower extracellular pH causes a toxicity effect (209). Thus, the overexpression of ATP synthase and H^+-ATPase observed in the adaptive mutants are potential mechanisms for their enhanced tolerance to AA. Indeed, the H^+-ATPase present in the vacuolar membrane has been shown to be important for weak acid tolerance

(203). Table 5.3 shows selected up-regulated or down-regulated genes for the isolated mutants related with acetic acid tolerance.

Detoxification through multidrug resistance transporters (MDR) has also been found to play a role in yeast tolerance to weak acids (203). Using NCA, we identified several transcription factors whose activities were significantly perturbed. The activities of MSN4, RIM101, PDR1, and PDR3 were increased and the activity of WAR1 was decreased (Table 5.4). All aforementioned transcription factors have been found to be related with adaptive response of yeast to drug and weak acids (203, 210).

Table 5.3. Selected up-regulated (bold) or down-regulated (non-bolded) genes in the isolated mutants related with acetic acid tolerance.

Strain	Cell Wall and protein Function	Lipid Metabolism	Plasma Membrane and Vacuolar H+ATpase
PR-164	**HSP82**		**ATP15, ATP16, ATP17, ATP5**
PR-338			**VMA3, ATP14, ATP5, VPH1**
PG-353	SPI1	**YPC1, YDC1, PDR16**	**VPH1, PMP1, ATP5, ATP4**
PR-438	**SPI1, HSP78, HSP104, SSE2**	**YPC1**	**SIA1, ATP14, ATP17, ATP4, ATP5**

Table 5.4. Transcription factors whose activities were increased (bold) or decreased (non-bold) in adaptive mutants associated with resistance to acetic acid.

Strain	Transcription factors
PR-164	**MSN4,** WAR1, **RIM101, PDR1**
PR-338	WAR1, **PDR1**
PG-353	**MSN4,** WAR1, **PDR1, PDR3**
PR-438	**MSN4,** WAR1, **PDR3**

5.4.3.2. HMF and furfural

HMF and furfural affect cell growth by similar mechanisms, including inhibition of alcohol dehydrogenase (*ADH*), pyruvate dehydrogenase (*PDH*), and aldehyde dehydrogenase (*ALDH*) and damages to cell membrane (189). Adaptive mutants PG-132 and

PR-164 were statistically significantly better than the wild type in 0.4 g/L furfural. Additionally, adaptive mutants PR-164 and PR-438 (from the red subpopulation) have positive relative fitness coefficients in presence of 0.9 g/L HMF. Interestingly, adaptive mutants PG-259, PR-338, and PG-353 have statistically significant positive fitness coefficients only in the combination of HMF and furfural. Table 5.5 shows a summary of the differentially regulated genes in the mutants that are potentially associated with HMF and/or furfural tolerance.

S. cerevisiae has the ability to convert furfural to furan methanol and HMF to furan di-methanol using *in situ* detoxification as a response mechanism in the presence of these inhibitors (193, 194). It has previously been proposed that the detoxification is through multiple NADPH-dependent aldehyde reduction, such as the reductases ALD4 and GRE3 (193), which was up-regulated in mutant PR-164, potentially causing NAD(P)H depletion. Serine, lysine, and arginine biosynthesis were down-regulated in all adaptive mutants with enhanced tolerance to HMF (PR-164 and PR-438). The repression of these amino acid biosynthesis pathways have been found to increase regeneration of ATP and NAD(P)H in the TCA cycle (194), possibly as compensatory mechanism for NAD(P)H depletion.

Transcription factors involved in pleiotropic drug resistance (PDR) and oxidative stress tolerance were significant activated in almost all of the adaptive mutants with enhanced tolerance to furfural and/or HMF (see Table 5.6), with the latter suggesting the presence of oxidative stress during growth on hydrolysate. To determine which of the isolated adaptive mutants are also more tolerant to oxidative stress, the growth kinetics

of each of the adaptive mutants were determined in the presence of 2 mM of H_2O_2. All the adaptive mutants from the red subpopulation (PR-164, PR-338, and PR-438) and one mutant from the green subpopulation (PG-353) showed significantly enhanced tolerance to hydrogen peroxide compared to their parental strain (see Table 5.7). In addition, all these adaptive mutants have also enhanced tolerance to AA.

Table 5.5. Selected up-regulated (bold) or down-regulated (non-bolded) genes in the isolated mutants associated with resistance to HMF and/or furfural.

Strain	Aldehyde Reductase	Amino Acid Biosynthesis
PG-132	GRE3, ADH6	ARG7, ARG8, ARG4
PR-164	ALD4	ARG3
PG-259	GRE3	ARG7, ARG2, ARG8, ARG4
PR-338	ADH7	ARG1
PG-353	GRE3, ADH6	CPA1, ARG8, ARG4
PR-438	GRE3, ADH6, ADH7	CPA2, CPA1, ARG3, ARG1, ARG4

All the other adaptive mutants did not show a significant improvement in the presence of hydrogen peroxide. The heat map of expressions from select genes for the adaptive mutants from the red lineage (Figure 5.4) shows that the higher up-regulated genes correspond to drug resistance (*IMD2*), cytochrome c oxidase (*COX13, COX9*), DNA replication stress (*SAP155*), phosphate metabolism (*PHO11*), and mitochondrial inner membrane protein (*FCJ1*). The most down-regulated genes are related with cell wall mannoprotein of the Srp1p/Tip1p family of serine-alanine-rich proteins (*TIR3*), pyruvate decarboxylase (*PDC5*), and Alpha-agglutinin (*SAG1*).

Table 5.6. Transcription factors whose activities were increased (bold) or decreased (non-bold) in adaptive mutants associated with resistance to HMF and/or furfural.

Strain	Transcription Factors Pleiotropic Drug Resistance	Transcription Factors Oxidative Stress
PG-132	-	-
PR-164	**PDR1, YRR1, YAP6**	**YAP1**
PG-259	**PDR1, PDR3, YRR1**	**-**
PR-338	**PDR1, YRR1,**	**YAP1**
PG-353	**PDR1, PDR3, YRR1,**	**YAP1**
PR-438	**PDR3, YRR1, YAP6**	**YAP1**

The adaptive mutants PG-132 and PG-259 exhibited negative relative fitness coefficients in the presence of the individual inhibitor HMF, but positive relative fitness coefficients in the presence of multiple inhibitors (HMF/furfural and hydrolysates). These adaptive mutants were isolated early on during the evolution. It is possible that the mutations acquired for these strains only allow them to grow in presence of more than one inhibitor at a time.

5.4.4. Glucose consumption

To determine if the adaptive mutants evolved not only to be more fit in the presence of hydrolysates but also able to more efficiently utilize and metabolize nutrients in the media, the rate of glucose consumption was measured using HPLC for cultures grown in YNB supplemented with 1% glucose in batch cultures. All the adaptive mutants from the red subpopulation showed significantly higher rate of glucose consumption compared to the ancestral strain. All the other mutants exhibited similar rate of glu-

cose consumption compared to the parental strains, suggesting that the non-red mutants are not more adaptive as a result of enhanced glucose utilization.

Table 5.7. Percentage of improvement in relative fitness in the presence of 2 mM H_2O_2.

Adaptive Mutants	% Improvement in H_2O_2 (2 mM)
PR-164	33%
PR-338	40%
PG-353	70%
PR-438	39%

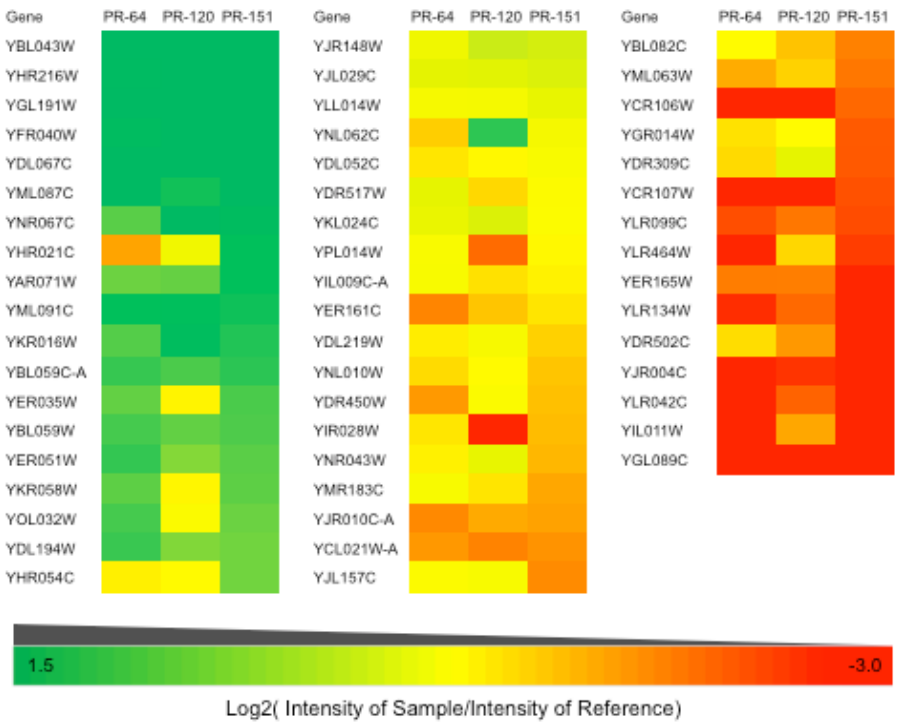


Figure 5.4. Heat map of gene expression changes of isolated mutants from the red subpopulation.

5.5. Conclusions

In conclusion, we isolated adaptive mutants with enhanced fitness in hydrolysates using a previously published adaptive evolution method, VERT. All but one of the isolated mutants showed increased relative fitness compared with the parental strains in the presence of hydrolysates, with the best mutant exhibiting a relative fitness coefficient of 57%. The mutants showed differential resistances to the inhibitors present in the hydrolysates and some had increased glucose uptake rates, indicative of the complex nature of adaptation to this feedstock. Using transcriptome analysis, we elucidated some potential mechanisms for enhanced tolerance in the isolated adaptive mutants.

5.6. Data Availability

The microarray data has been deposited in the Gene Expression Omnibus with accession number GSE44085.

6. IMPROVEMENT OF CAROTENOIDS PRODUCTION IN YEAST VIA ADAPTIVE LABORATORY EVOLUTION

6.1. Summary

The application of adaptive evolution for increased production of secondary metabolites (compounds not directly involved in normal cell growth) generally ends with decreased production due to the metabolic burden imposed for the channeled flux or toxicity of the produced compound. This phenomenon is due to the nature of the technique, where selection must be causally connected to the differential ability to survive and reproduce.

In this study, laboratory adaptive evolution was successfully exploited in order to improve carotenoids production in an engineered *S. cerevisiae* carotenoids producing strain. The effective design of the evolutionary strategy was based in the antioxidant properties of the carotenoids. Scheduled hydrogen peroxide shocking experiments ensured an increase of carotenoids production in more than 200% in a short-term experiment. Transcriptome analysis helped to elucidate the molecular mechanisms for increased carotenoids production, including upregulation of genes related with lipid biosynthesis, especially upregulation of genes part of the melavonate pathway.

6.2. Introduction

Carotenoids are an attractive class of tetraterpenoid pigmented compounds naturally produced by diverse organisms including plants and numerous fungi and bacteria (211). These compounds are known antioxidants (212) that have potential positive impact on human health (playing an important role as pro-vitamin A compounds), and they are currently used in the nutraceutical industry as supplements, fortified foods and cosmetics with a total market value estimated in $1.2 billions in 2010 (BCC research published Sept 2011).

Currently, most industrially produced carotenoids are chemically synthesized through multistep chemical synthesis or by solvent-based chemical extraction from non-microbial sources. Chemical synthesis is not a viable option to produce most carotenoids due to their structural complexity (213). Recently, production of carotenoids has been amply explored using biosynthetic routes, via metabolic engineering and synthetic biology of different microbial systems for a more environmentally friendly and efficient production platform (214-216).

The biological pathways of all carotenoids use isopentenyl diphosphate (IPP) as the precursor. The biosynthesis of IPP falls in two different pathways: The mevalonate (MVA) pathway and the mevalonate independent methyl erythritol 4-phosphate (MEP) pathway. In the MVA pathway, acetyl-coA is the precursor for mevalonate. In the MEP pathway, MEP is produced from 1-deoxy-D-xylulose-5-phosphate (DXP), which is synthesized from pyruvate and glyceraldehyde-3-phosphate (G3P). The MVA pathway is

the one naturally used by yeast for the production of IPP (211). Both bacteria and yeast have been successfully engineered to produce higher titers of carotenoids through metabolic engineering (214-216). These efforts have thus far involved the optimizations of native pathways, introduction of foreign genes to enhance metabolic flux, and through co-factor balancing. One option that has not been explored is the use of laboratory adaptive evolution (or whole genome directed evolution) to improve the production of carotenoids. Laboratory adaptive evolution uses a selective pressure as driving force for the generation of mutants with enhanced properties. Since carotenoids have antioxidant properties (217), we hypothesized that oxidative stress can be used as a driving force for the directed evolution of microbial systems for enhanced carotenoid production.

The main objective of this work is to develop a proper selective pressure in order to exploit adaptive laboratory evolution to improve heterologous carotenoids production in a microbial host. Here, a *Saccharomyces cerevisiae* (*S. cerevisiae*) strain was engineered to produce carotenoids using heterologous genes from the wild yeast *Xanthophyllomyces dendrorhous* (218). Enhancement of carotenoids yield was achieved through adaptive evolution using a regime of shocking experiments with hydrogen peroxide for selected of high producers. The parameters involved in the enhanced production of carotenoids were determined. Upregulation of genes involved in lipid biosynthesis was determined as an adaptive response for increased carotenoids production. The amount of produced β-carotene reached after a short-term evolution experiment was 18 mg/g(dcw), corresponding to an increase of 3-fold in comparison with the ancestral strain (6 mg/g(dcw)).

6.3. Materials and Methods

6.3.1. Strains, plasmids and growth conditions

GFP fluorescently marked *S. cerevisiae* strain GSY1136 (54), derived from FY2, was chosen as the platform strain for the adaptive evolution experiments. All *S. cerevisiae* strains used are listed in Table 6.1. The evolution experiments were carried out in Yeast Extract Peptone Dextrose media (YPD) to ensure high biomass formation at 30°C. The shuttle vector YIplac211YB/I/E* (214) was used to introduce *crtE*, *crtYB* and *crtI* genes into the yeast strain. For the determination of carotenoids yields on glycerol as a carbon source, YPG media (Yeast Extract Peptone Glycerol) supplied with 4% (v/v) glycerol was used.

Table 6.1. List of strains used in this work

Name	Relevant phenotype	Source
GSY1136	Matα, *ura3-52, gal+* in S288c background, *YBR209W::*Act1p-GFP-Act1t-*URA3*	Kao and Sherlock (54)
YLH0	GSY1136 Δ*URA3*	This work
YLH1	GSY1136 YIplac211YB/I/E*	This work
YLH2	GSY1136 YIplac211YB/I/E* Δ*CTT1*	This work
YLH3	GSY1136 Δ*CTT1*	This work

6.3.2. Generation the carotenoids yeast producer

The *URA3* gene was excised out from GSY1136 using 5-fluoroorotic acid (5-FOA) as counter-selection in order to create the uracil auxotrophic strain YLH0. The transformation of the YIplac211YB/I/E* shuttle vector into YLH0 was carried out by lithium acetate procedure (219), using *URA3* as selectable marker for integration, gener-

ating the strain YLH1. The transformed cells were selected on YNB + 2% D-glucose (w/v), incubated at 30°C for 3 days, and verified using PCR amplification.

6.3.3. Deletion of CTT1 catalase

The catalase *CTT1* was knocked out from the genotype of interest using homologous recombination with the *NEO* gene. The primers used for *NEO* amplification, including the homologous regions used to delete *CTT1*, were: forward: 5'–TTA AAA AAA TCC TTC TCT TGT CTC ATG CCA ATA AGA TCA ATC AGC TCA GCT TCA CAA ATG CGG ATC CCC GGG TTA ATT AA–3' and reverse: 5'–TAT AAT TAC GAA TAA TTA TGA ATA AAT AGT GCT GCC TTA ATT GGC ACT TGC AAT GGA CCA GAA TTC GAG CTC GTT TAA AC–3'. The plasmid pFA6a-kanMX6 (220) was used as template for the *NEO* cassette. Transformation of the recombination cassette was carried out using lithium acetate procedure. The recombinants were selected on YPD + geneticin (G418) plates, incubated at 30°C for 2 days, and verified using PCR amplification. The primers used for verification were: forward 5'-ATT CGA CGT AGC CTG GAC AC-3' and reverse 5'-TAA TCG TTG AGT TCA TGC CG-3'.

6.3.4. Carotenoids quantification

For the carotenoids quantification of the evolving populations, 3 mL of YPD were inoculated with cells from frozen stocks, and incubated at 30°C for 72 hours. Cell density was determined using a spectrophotometer at OD_{600}. 250 µL of culture were

transferred to a 2 mL collection tube and the cells were collected by centrifugation at 12,000 rpm for 2 min. Supernatant was vacuum aspirated. The pelleted cells were disrupted using approximately 250 μL of 425-600 μm acid washed glass beads (Sigma) and dodecane (1 mL) to extract the carotenoids. Yeast cells were lysed using an analog Disruptor Genie Cell Disruptor (Scientific Industries). Samples were treated twice for 6 min each to ensure maximum cell disruption. Cell debris and glass beads were separated from the supernatant by centrifugation at 15,000 rpm for 2 min. 200 μL of the supernatant was transferred to a Corning® 96 well black wall clear bottom plate for quantification. Carotenoids were quantified using a survey scan from OD_{350} to OD_{600} to determine any wavelength shifting using a microplate reader TECAN Infinite® M200. The relative total carotenoids production was determined by calculating the area under the curve of the survey scan, using the ancestral strain YLH2 (see Table 6.1 for genotype information) as reference. β-carotene quantification was determined by the absorption at OD_{454} (214).

6.3.5. Isolation of hyper-carotenoids production mutants

Selected populations were plated on YPD agar plates for isolation of single colonies. Plates were incubated for 72 hours and several colonies were chosen based on the apparent red color and normal colony size (compared with the ancestral strain YLH2). The colonies were grown in 3 mL of YPD media and incubated for 72 hours for total carotenoids quantification. The highest producer was selected for further analysis.

6.3.6. Hydrogen peroxide shock experiments

Each isolated mutant was inoculated in 3 mL of YPD media and incubated for 72 hours. Samples were normalized to an OD_{600} ~ 2.0. 500 µL of the normalized culture was transferred to a microcentrifuge tube and shocked with 1.05 M hydrogen peroxide for 30 minutes. Samples were diluted, plated in YPD plates and incubated at 30°C for 48 hours for colony counting. A non-shocked replica was plated on YPD plates to ensure proper normalization of the cell density and dilution.

6.3.7. RNA extraction

The isolated mutants were inoculated in 25 mL of YPD media on 125 mL flask at an initial OD_{600} ~ 0.05. The cells were harvested in late-exponential phase (OD ~ 4.0) by filtration using NALGENE analytical test filter funnels, immediately resuspended in 10 mL of RNA*later* (Sigma) and stored at -80 ºC for future analysis. For the extraction of total RNA, the ZR Fungal/Bacterial RNA MiniPrep™ (Zymo) kit was used following manufacturers' instructions using 3 mL of each stored sample in RNA*later*. The extracted RNA was quantified using the NanoDrop™ 1000® (Thermo Scientific).

6.3.8. Labeled cDNA generation, microarray hybridization and data analysis

The reverse transcription reaction was prepared mixing 10 µg of isolated RNA, 1U SuperScript® III reverse transcriptase (Life Technologies), nucleotides (dATP-5mM, dGTP-5mM, dCTP-5mM, dTTP-2mM and amino-allyl dUTP-3mM) and $Oligo(dT)_{20}$

primers (Life Technologies) to synthesized the cDNA. cDNA was recovered with ice cold ethanol precipitation. Cy3- and Cy5- mono-Reactive Dye Pack (GE Healthcare) were used to label cDNA samples. The labeled cDNA was hybridized to the *S. cerevisiae* G4813A Gene Expression Microarray (Agilent Technologies). The arrays were scanned using the GenePix 4100A Microarray Scanner and GenePix Pro 6.0 software (Molecular Devices) for image analysis. The Microarray Data Analysis System (MIDAS) software was used to normalize the data using LOWESS based normalization algorithm (118). Differentially expressed genes were identified using the rank product method with a critical p-value < 0.05. The MeV (TM4) (205) microarray analysis software was used for clustering and other expression profile analysis. The *Saccharomyces* Genome Database (SGD) (206) was use to analyze the data using gene ontology. Gene ontology analysis was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) (221, 222), in order to identify enriched biological functions.

6.3.9. Real-time PCR

The CFX384 Real-Time PCR Detection System (Bio-Rad Laboratories, CA) was used to carry out the quantitative real-time PCR (qRT-PCR) experiments. RNA extraction and reverse transcription of the studied samples were performed as established in the previous section. GoTaq® qPCR Master Mix (Promega) was used for all the qRT-PCR experiments, using 20ng of cDNA and 0.5 μM of every primer. The primers used for the qRT-PCR experiments are: Control gene (*COQ5*): For: GAC TTC AAT ACA

GTC TTC GAA CCA AA, Rev: TCC TTA TAC AGC TGC TGT TAC AAT T. *crtYB* gene: TGC CAC AAT TGA CAT GGT CT, Rev: AGG CGA AAT GGT ATT GAA CG. *crtI* gene: For: GAA GTC GAG CGT TTT GAA GG, Rev: AGG ATT TGG CCA ATG AAC TG. *crtE* gene: For: GGG ATT CCG CAG ACA ATA AA, Rev: CTT TCG AGA ACG GAA TCT GC. The thermocycler protocol used was: 95 °C for 3 min followed by 39 cycles of 95 °C for 10 s followed by 55 °C for 30 s.

6.3.10. Bioreactor studies

The media used for these studies consisted in YNB media supplemented with 20 g/L glucose. The seed cultures for fermentation were prepared as follows: Samples from single colonies were used to inoculate 3 mL YNB culture (20 g/L glucose) and incubated at 30 °C for 24 hours with constant agitation of 250 rpm. The contents of the culture were used to inoculate 50 mL of fresh YNB media (20 g/L glucose) and incubated at the same conditions as established before for 24 hours. The 50 mL culture was used to inoculate a batch bioreactor with a total volume of 3 L.

The bioreactor studies were carried in a 7 L glass autoclavable bioreactor Applikon®. The pH was maintained at 5.5, controlled automatically by addition of 2 N HCl or 2 N NaOH as necessary. The temperature was set at 30 °C. All the bioreactor experiments were performed in batch until the ethanol produced was consumed.

6.4. Results and Discussion

6.4.1. Establishment of appropriated selective pressure

Selective pressure refers to any environmental factor that influences the direction of natural selection in biological systems. In natural selection, genotypic variations with positive fitness advantage will likely increase in frequency (223). However, the heterologous production of carotenoids by *S. cerevisiae* is not growth-coupled and its production poses a metabolic burden for yeast cells (218). Thus, in order for evolutionary engineering to be applicable to the increased formation of carotenoids, an appropriate selective pressure must be chosen to allow the higher producers a growth advantage. Based on the known antioxidant properties of carotenoids (217), we hypothesized that yeast strain producing carotenoids will have a fitness advantage compared with a non-producer in the presence of oxidative stress, specifically in the presence of hydrogen peroxide. Since yeast produces catalases for hydrogen peroxide detoxification, we first deleted the cytosolic catalase, encoded by *CTT1*, from a heterologous carotenoids producer (see Materials and Methods for construction of the YLH2 strain). The *CTT1* deletion resulted in appreciable decrease in carotenoids production (data not shown). However, carotenoids production led to an increase in cell viability upon short-term exposure to high concentration of hydrogen peroxide in the absence of *CTT1* (See Figure 6.1A). These protective effects of carotenoids on yeast cells have been observed in a prior study (224). In addition, this protective effect of carotenoids towards oxidative stress is more pronounced when the cells reach stationary phase (Figure 6.1B), most likely due to the

fact that the amount of intracellular carotenoids is higher in stationary phase compared with exponential growth phase.

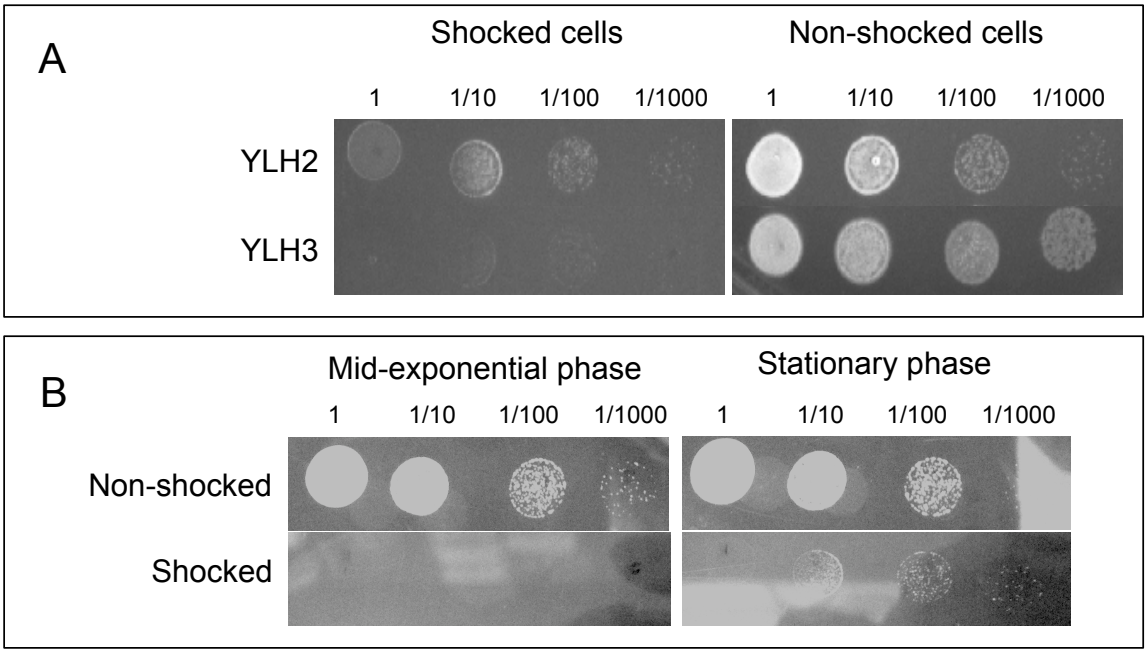


Figure 6.1. Shock experiments using hydrogen peroxide (690 mM). A. Results indicate that carotenoids in the cells increased cell viability in the presence of hydrogen peroxide in the carotenoids producer Δ*CTT1* strain YLH2 compared with the wild-type Δ*CTT1* strain YLH3. B. Cells were shocked when cell density was high to ensure high amount of intracellular carotenoids. At this point the carotenoids antioxidant properties start playing an important role.

6.4.2. Laboratory adaptive evolution experiments

Prior to the start of the adaptive evolution experiments, the strain YLH2 (YLH1 Δ*CTT1*) was serially transferred in the absence of hydrogen peroxide to determine if higher carotenoids production is selected for in the absence of hydrogen peroxide. As demonstrated in Figure 6.2C, the amount of carotenoids produced in a short-term experiment does not increase in comparison to the ancestral strain. Moreover, the level of ca-

rotenoids produced decreased by approximately 30%, possibly due to the presence of jackpot mutant(s) in the ancestral population or due to the metabolic burden of carotenoids production.

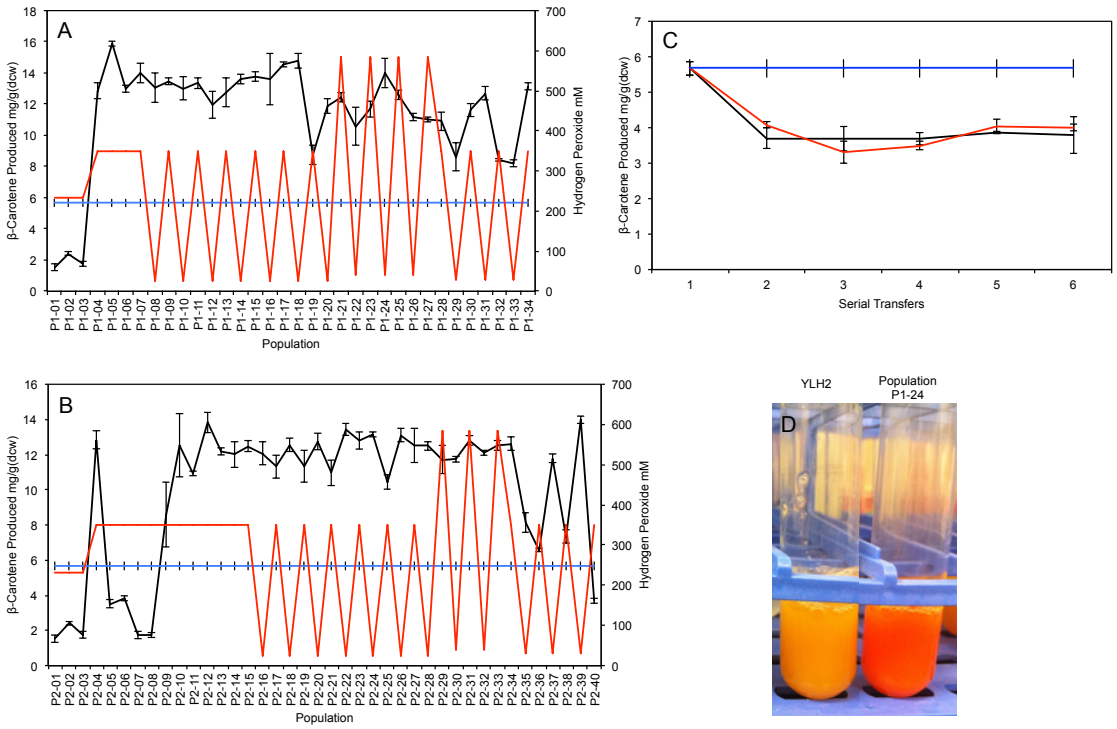


Figure 6.2. The observed improvements in carotenoids production during hydrogen peroxide challenged adaptive evolution. Average β-carotene produced and the hydrogen peroxide concentration used in two independent populations, P1 (A) and P2 (B). C. Changes in β-carotene concentration in two independent cultures (black and red lines) of YLH2 serially propagated in the absence of hydrogen peroxide. D. Comparison of the amount of carotenoids produced between the ancestral strain and the population P1-24. Blue line: the level of β-carotene produced for the ancestor strain. Black line: the average amount of carotenoids produced for the population. Red line: the amount of hydrogen peroxide used to shock the cultures.

For the adaptive laboratory evolution experiments, serial batch transfers were performed at 30 °C in YPD media. Two independent colonies of YLH2 (ancestral

strain) were initially inoculated into two 3 mL cultures to form the two populations (P1 and P2) for the evolution experiment. Periodically, approximately 7% of the total culture volume was serially transferred into fresh medium under different hydrogen peroxide shock schemes (described below). This procedure was performed every 24 hours. Two different phases of hydrogen peroxide shocks were used as selective pressure to increase carotenoids production (details shown in Table 6.2**)**. Prior to each "shock treatment", cells were grown in YPD for 24 hours in order to achieve high biomass yield in the absence of hydrogen peroxide. 500 µL of culture were then transferred to a sterile microcentrifuge tube to be shocked with hydrogen peroxide for 30 minutes. Concentrations of hydrogen peroxide used for the shocking experiments are specified in Figure 6.2A and Figure 6.2B (red line). 200 µL of the shocked culture were inoculated into 3 mL of fresh YPD media and incubated at 30°C for 24 hours for recovery (recovery phase). This shocking phase was performed on a daily basis. In population P1, this hydrogen peroxide challenge scheme was performed for 7 serial transfers, and in population P2, 14 serial transfers were performed (Table 6.2). As aforementioned, the initial inoculum at the start of the evolution experiment did not contain H_2O_2, however, upon initiation of the "shock treatment," the concentration of hydrogen peroxide in the growing medium during each recovery phase was not zero since approximately 7% of the H_2O_2 present in the shocked culture was transferred to the new culture. Therefore, H_2O_2 likely accumulated quickly in the medium through each serial transfer, negatively impacting the growth rate, decreasing the cell yield and in some cases delaying the transfer by more than 24 hours (due to low cell density). As consequence, a second phase of

evolution, with increased recovery phase by performing "shock treatments" every other serial transfer (see Table 6.2 for details), was initiated. This increased recovery phase helped to ameliorate the increased toxicity due H_2O_2 accumulation.

Table 6.2. Hydrogen peroxide shocks schemes used to increase carotenoids production.

Selective pressure scheme	Duration	First 24 h cycle	Second 24 h cycle
Phase 1:	P1-01 to P1-07	Shocked	Shocked
Shock Treatment	P2-01 to P2-15		
Phase 2:	P1-08 to P1-34	Shocked	Non Shocked
Alternated shocking/non-shocking periods	P2-16 to P2-40		(Recovery phase)

Using adaptive evolution under periodic hydrogen peroxide shocks, the total amount of β-carotene produced (and the total carotenoids content in general) increased shortly after the start of the evolution experiment, reaching more than 12 mg/g(dry cell weight (dcw)) and generating a boost in the observed coloration of the cultures as shown in Figure 6.2D. This improvement corresponds to more than 100% increase in the amount of carotenoids produced compared to the ancestral population, originally producing 6 ± 1 mg/g(dcw). Unfortunately, prolonged propagation under the second hydrogen peroxide shock scheme did not result in additional increases in the average amount of carotenoids produced by the evolving populations. Although there is indication that the carotenoids are transported outside the cell, the rate seems to be very low, and it is likely that we have reached a physical limit in the amount of carotenoids that can be accumulated inside the cell.

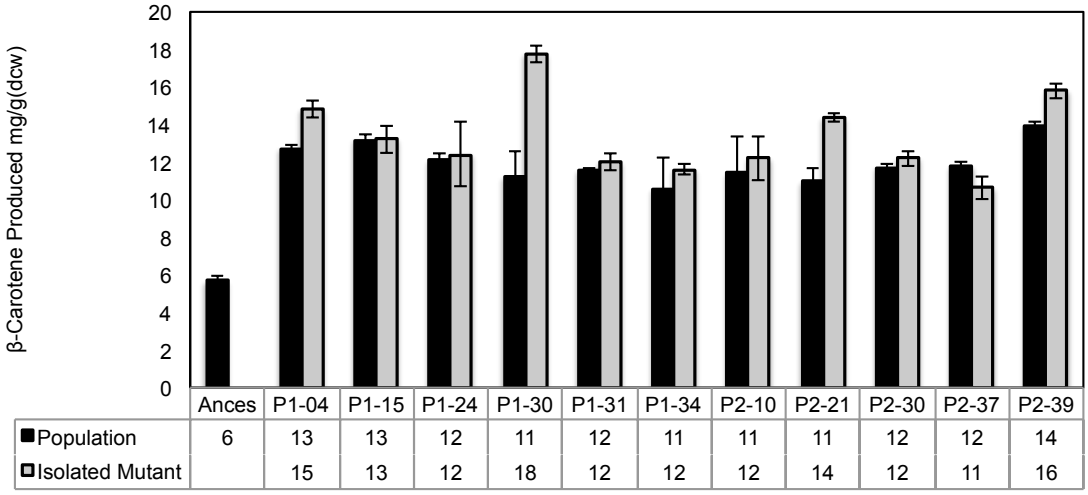
6.4.3. Isolation of adaptive mutants

Upon the conclusion of the adaptive evolution experiment, several mutants were isolated from different populations throughout the evolution. The populations to be analyzed further were chosen based on the observed peaks in the average amount of carotenoids produced (see Figure 2A and 2B). Mutants were isolated from each of the chosen populations as described in the Materials and Methods section. Details about each isolated mutant can be found in Table 6.3.

Table 6.3. Best β-carotene producers isolated from the evolved populations.

Isolated mutant	Progenitor population
SM11	P1-04
SM12	P1-15
SM13	P1-24
SM14	P1-30
SM15	P1-31
SM16	P1-34
SM21	P2-10
SM22	P2-30
SM23	P2-37
SM24	P2-39

The amounts of β-carotene produced in the isolated mutants were measured and compared with the population where they were isolated (the results are summarized in Figure 6.3). In general, the amount of β-carotene produced by the isolated mutants was significantly higher than the average amount produced by the population they were isolated from. The isolated mutant SM14 was amongst the highest producers, producing greater than 200% more β-carotene (18 ± 1 mg/g(dcw)) compared to the parental YLH2 strain (6 ± 1 mg/g(dcw)).



	Ances	P1-04	P1-15	P1-24	P1-30	P1-31	P1-34	P2-10	P2-21	P2-30	P2-37	P2-39
Population	6	13	13	12	11	12	11	11	11	12	12	14
Isolated Mutant		15	13	12	18	12	12	12	14	12	11	16

Figure 6.3. Comparison of the produced β-carotene between the evolved populations (black) and selected mutants isolated from each respective population (gray).

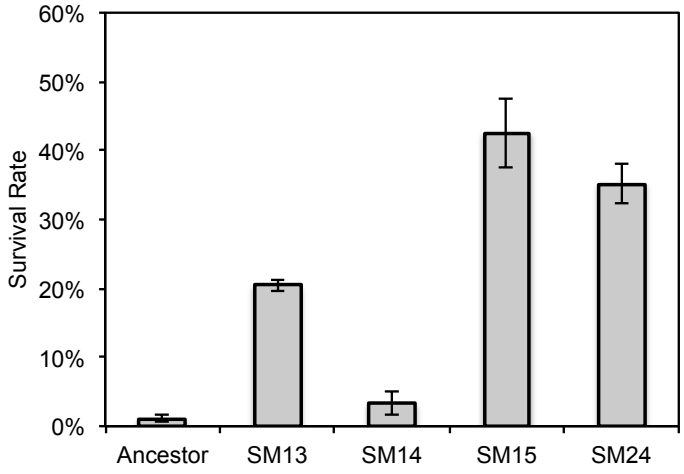


Figure 6.4. Hydrogen peroxide shock experiment with selected isolated mutants. In general, the survival rate of the isolated mutants after exposure to 1.05M hydrogen peroxide increased in comparison with the ancestral strain (all 4 mutants shown here are statistically significantly different from the ancestral strain (p-value < 0.05)).

A hydrogen peroxide shock experiment using 1.05M H_2O_2 was performed in order to determine whether the carotenoids hyper-producing mutants have increased survival rates in the presence of high concentrations of hydrogen peroxide. Increment in

the cell survival rate was verified in all of the isolated mutants tested, and the results are shown in Figure 6.4.

6.4.4. Transcriptome analysis of adaptive mutants

The four highest carotenoids producers (SM12, SM13, SM14, and SM22) were chosen for further molecular analysis to identify potential mechanisms of increased carotenoids production in these strains. Since increases in the gene expressions of the carotenoids biosynthetic genes (either due to copy number amplification or mutations in the promoter regions) is a likely cause for increased carotenoids production, we used qRT-PCR to compare the expression of the genes *crtI*, *crtE* and *crtYB* (heterologous carotenogenic genes) and the *COQ5* gene, whose expression was found to be unaltered in the mutants based on our transcriptome data (see below), as control in a subset of the hyper-producing strains and the YLH2 strain. The results demonstrated that there are no expression differences in the carotenogenic genes between the hyper-producing strains and YLH2. In addition, no increase in gene copy number was found between the hyper-producing strains and YLH2 from whole genome re-sequencing. These results suggest a different route for enhanced carotenoids production in these strains.

Using DNA microarrays, the transcriptome of every isolated mutant was compared to YLH2 as reference. All the strains were grown in YNB media supplemented with 20 g/L glucose at 30 °C until late-exponential phase. Three biological replicas were performed per strain. The genes significantly (p-value < 0.05) differentially expressed

were analyzed further to identify potential mechanisms for increased carotenoids production (on average, 107 genes were upregulated and 97 downregulated).

A high similarity in gene expression profiles was observed between the 4 mutants analyzed (SM12, SM13, SM14 and SM22), with approximately 50% of the perturbed genes being commonly perturbed between the mutants. The data suggests that the enhanced carotenoids production in these strains is due to channeled metabolic flux towards lipid biosynthesis, specifically to the mevalonate pathway, ultimately increasing carotenoids production due to IPP accumulation. Some of the key perturbations are described below.

6.4.4.1. Genes involved in lipid biosynthesis

Analysis of the transcriptome in the isolated mutants allowed us to identify several genes involved in the production of lipids. *CAB1*, a gene expressing the pantothenate kinase, catalyzing the first step in the synthesis of coenzyme A, regulated via sterol response element (225), and *NSG1* and *CYB5*, involved in the regulation of the sterol and lipid biosynthesis (226, 227) were found to be upregulated. Upregulation of *NSG1* has been identified to play a key role in reducing Hmg2p degradation (226). Hmg2p is one of the two HMG-CoA coenzymes, and increased production of HMG-CoA has been reported to improve production of β-carotene in *S. cerevisiae* (228). The increase in the level of transcription of *NSG1* suggests reduced levels of sterols in the cell, possibly because part of the metabolic flux is being channeled towards carotenoids production (Figure 6.5). The cytochrome b5, encoded for the gene *CYB5*, acts as an electron donor

to support the sterol C-5(6) desaturation (229). This process is carried out through the C-5 desaturase, encoded by *ERG3* (Figure 6.5). These observations are coherent with the transcriptional changes in the melavonate pathway that were also observed in all the isolated mutants. The genes *ERG13* and *ERG27* were found to be upregulated in all the mutants studied here. *ERG13* is a 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) synthase, catalyzing the formation of HMG-CoA from acetyl-CoA and acetoacetyl-CoA and is involved in the second step in mevalonate biosynthesis (230). Upregulation of *ERG13* suggests an increase in the metabolic flux towards melavonate pathway, which likely led to an eventual increase in carotenoids production (Figure 6.5). On the other hand, the upregulation of *ERG27*, a 3-keto sterol reductase, catalyzing the last of three steps in ergosterol biosynthesis (230), is a competing branch to the carotenoids pathway. We hypothesized that the channeling of the metabolic flux towards the carotenoids production depleted the cell of available sterols. Increases in expression observed in the genes *ERG27* and *CYB5* possibly act as a compensating effect. On the other hand, overexpression of *ERG13* and *NSG1* directly increase carotenoids production. Downregulation of the ergosterol biosynthetic pathway via deletion of the *ERG24* gene, as well as downregulation of lipid and amino acids biosynthesis, seemed to affect bisabolene and carotenoids production (231) in prior study.

Other lipid biosynthesis-related genes that were differentially expressed are summarized in Figure 6.6.

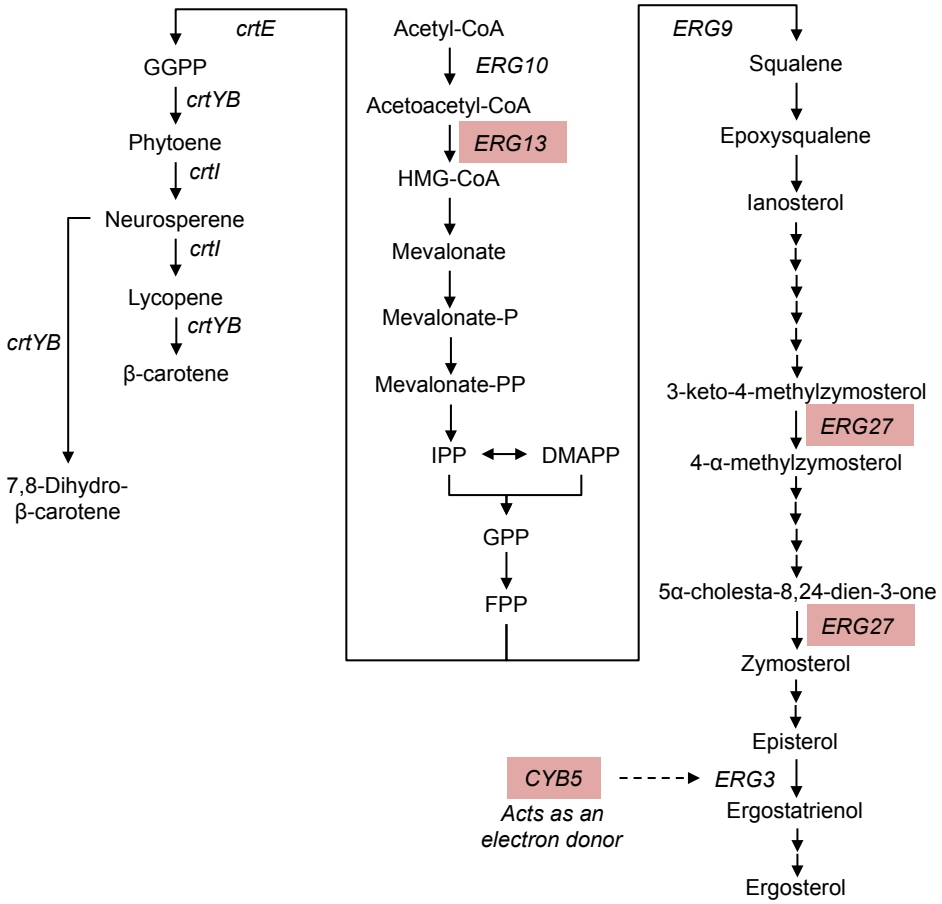


Figure 6.5. Superpathway of ergosterol biosynthesis, including the carotenoids synthesis pathway from *X. dendrorhous*. Upregulated genes are highlighted.

6.4.4.2. Change in expression of genes involved in stress response

We found significant expression changes in genes involved in stress response. The genes *AAD6, MXR2, UGA1* and *ROX1*, involved in oxidative stress response, were downregulated in the isolated mutants. On the other hand, the genes *GND1*, *SKO1* and *HOR2*, whose expressions have been shown to decrease under oxidative stress (232-234), were upregulated in our mutants under the experimental conditions. This result suggest that the evolved mutants, despite being exposed on a daily basis to extreme oxidative stress conditions, did not increase their tolerance to hydrogen peroxide through

the known mechanisms of oxidative stress response, instead likely increased carotenoids production as an antioxidant response.

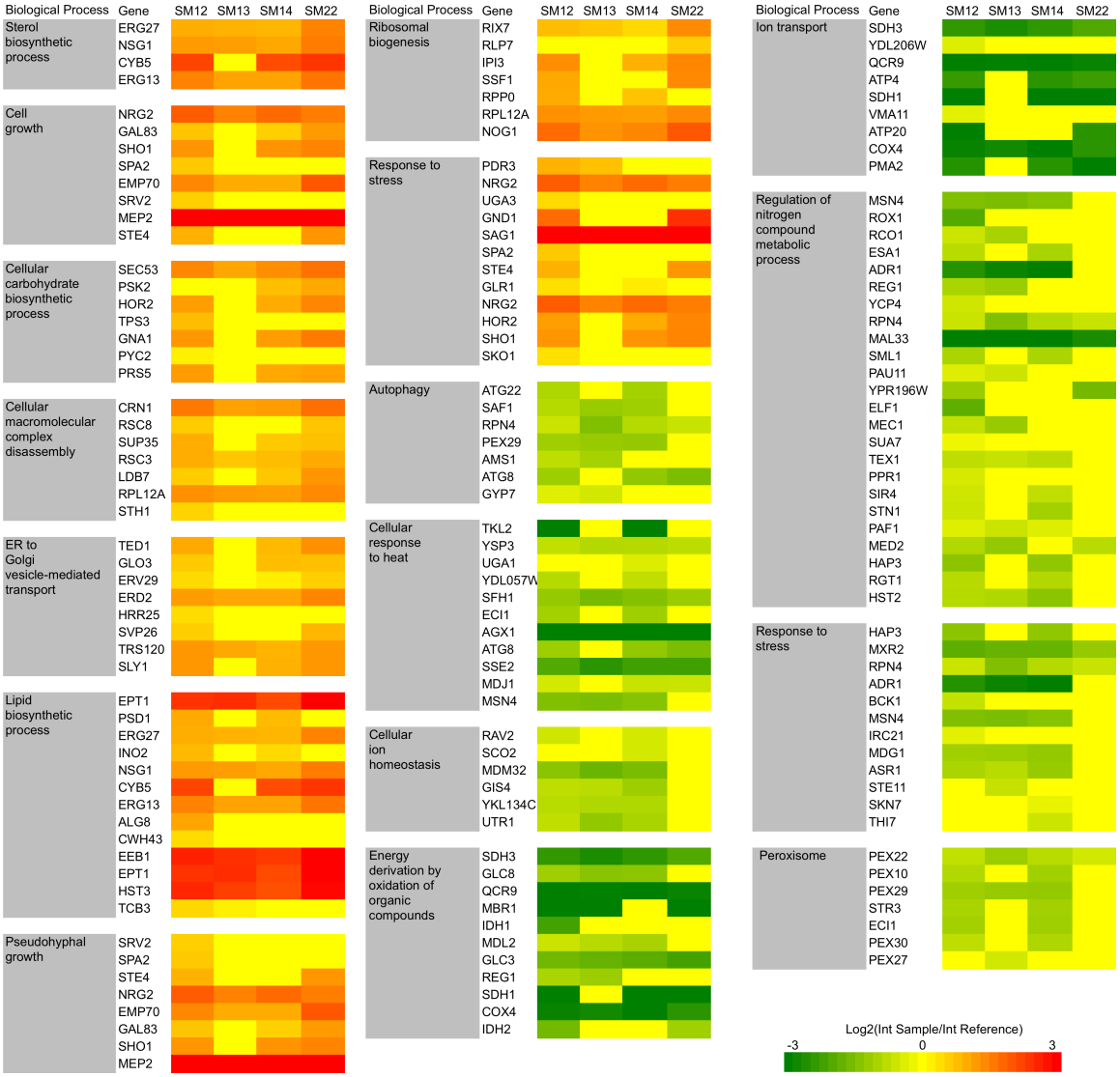


Figure 6.6. Selected genes perturbed in the evolved carotenoids producer strains indicating the biological function based in gene ontology terms.

The gene encoding a transcriptional activator of the pleiotropic drug resistance network, *PDR3*, was upregulated in the mutants SM12 and SM13. Upregulation of PDR

genes, encoding ABC-type and major facilitator transporters, reported to be involved in secretion of toxic compounds out of cells, has previously been observed in transcriptome analyses in carotenoids producing cells, possibly due to an inhibitory effect of the carotenoids (218). Two experiments trying to quantify the amount of carotenoids exported to the growth media using two-phases cultures using vegetable oil as organic phase were carried out. First, incubation of the evolved mutants for 72 hours in two-phases YPD/dodecane culture produced an orange coloration in the organic phase due to dissolved carotenoids as reported by Verwaal *et al* (218). However, we failed to detect carotenoids after cultivation in YPD only media for 72 hours and then using vegetable oil to recover carotenoids by vortexing the two-phases for 2 minutes (data not shown), suggesting that the carotenoids are continuously exported to the organic phase at a slow rate.

Numerous genes involved in response to hyperosmotic and chemical stimulus were differentially expressed in the mutants (see Figure 6.6). The accumulation of the hydrophobic intracellular droplets of carotenoids could trigger different mechanisms of stress response, particularly response to osmotic stress. Several genes involved in the biogenesis and regulation of peroxisomes were downregulated in the different isolated mutants as shown in Figure 6.6. Since different hydrogen peroxide oxidases and catalases are compartmentalized in peroxisomes in yeast (235), it is currently not clear how downregulation of peroxisome biogenesis genes contribute to enhanced carotenoids production.

6.4.4.3. Other transcriptional changes

When all differentially expressed genes were categorized according to their cellular component ontology, we found numerous genes involved in mitochondrial proteins (~11% of the perturbed genes) to be perturbed. Most of the perturbed genes were downregulated, and their functions are involved in electron transport and cellular respiration related functions (*MIC14*, *COX4*, *QCR9*, *MBR1, SDH3, SDH1, IDH1* and *IDH2*).

The genes *GPT2* (glycerol-3-phosphate/dihydroxyacetone phosphate, involved in the lipid biosynthesis via acylation of glycerol-3-phosphate and dihydroxyacetone) and *ADR1* (transcription factor required in the expression of several genes involved in glycerol, ethanol and fatty acid utilization), involved in glycerol and fatty acid utilization, were found downregulated. Indeed, we observed a decreased growth rate (Figure 6.7) when the evolved cells were grown in glycerol as a carbon source, compared to the YLH2 strain.

6.4.5. Scale-up studies of carotenoids production

In order to maximize the carotenoids production, scale-up experiments were carried out in a 7 liters bioreactor. The strain SM14 was chosen for scale-up test since it was the highest producer. The seed train was initialized by inoculating 3 ml of YNB media (20 g/L glucose) from a single colony and incubated for 24 hours at 30°C. The culture was transferred to 50 mL of fresh YNB media and incubated for 24 hours at 30°C at constant shaking. The bioreactor was inoculated with the entire contents of the

50 mL culture. The bioreactor was maintained at pH = 5, temperature at 30°C and constantly agitated at 400 rpm.

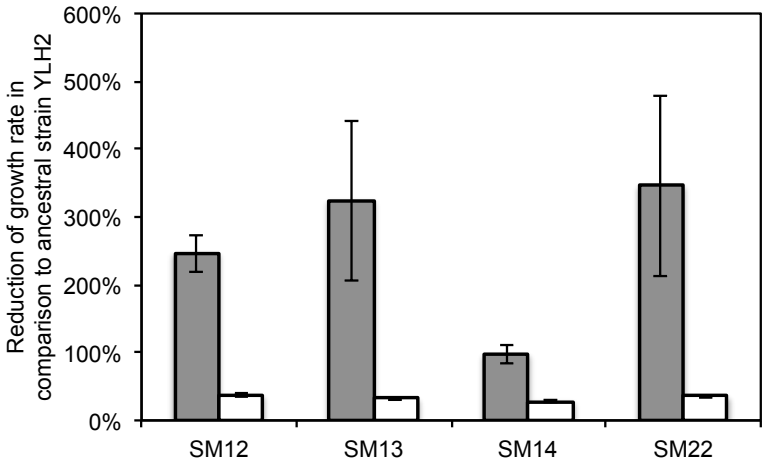


Figure 6.7. Reduction in the maximum specific growth rate of the evolved strains when grown in glucose and glycerol as a carbon source, in comparison with the ancestral strain YLH2. Open bars: the reduction of growth rate in glucose as carbon source. Gray bars: glycerol as carbon source.

In an initial study, the bioreactor was supplied with constant airflow at 6.0 L/min in order to maintain at least 50% dissolved oxygen throughout the experiment. Within 24 hours, the supplied glucose was consumed (verified using HPLC) and the β-carotene productivity at that point was 6 mg/g(dcw). After 24 hours, the accumulated ethanol started being consumed as carbon source, and the net production of β-carotene increased noticeably (reaching 15 mg/g(dcw)) as show in Figure 6.8. The β-carotene productivity increased more than 100% when the carbon source was switched from glucose to ethanol (from 6 mg/g(dcw) to 15 mg/g(dcw). To verify that using ethanol as a carbon source increases β-carotene productivity compared with glucose as a carbon source, two batch

culture experiments were carried out with SM14 grown in YNB media supplemented with either 2% glucose (w/v) or 4% ethanol (v/v) until late exponential phase in order to avoid consumption of produced ethanol in case of cultivation using glucose as carbon source. The results showed an increase of 73% ± 6% in β-carotene productivity when ethanol, relative to glucose, was used as the carbon source. Carotenoids accumulation in most yeast strains starts in the late exponential phase and continues into the stationary phase, which is typically observed in production of secondary metabolites (236).

However, the obtained productivity in the bioreactor was lower than the productivity obtained in test tubes (18 ± 1 mg/g(dcw)), although the growth conditions used in both cases were different (growth media, pH, aeration).

A second attempt to improve the β-carotene yield in the bioreactor was carried out by growing the cells in the absence of airflow. The cells went from a micro-aerobic to an anaerobic environment in less than 24 hours, when the dissolved oxygen was completely spent, as well as the glucose in the medium. However the productivity at this point was lower than the productivity reached in aerobic conditions (3 mg/g(dcw)) (Figure 6.8). The growth stopped after the culture reached anaerobic conditions. The reduction in productivity and the absence of growth in an anaerobic environment could be due to the lack of sterol biosynthesis under this condition. The sterol biosynthesis in *S. cerevisiae* is an exclusive aerobic process since molecular oxygen is required in several steps of the ergosterol biosynthetic pathway (229). It is likely that the low carotenoids productivity in the absence of airflow was due to the low biomass achieved under this condition.

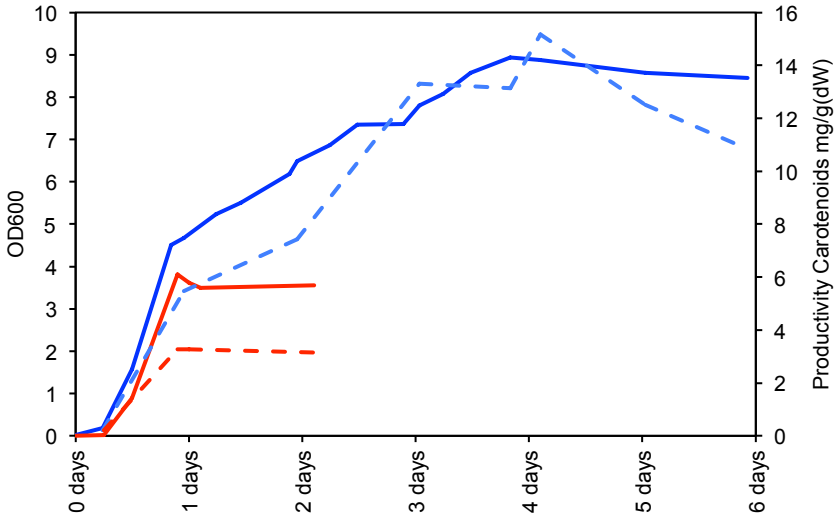


Figure 6.8. Scale-up studied of carotenoids production in SM14. Blue lines: under oxygen saturation conditions. Red lines: in the absence of airflow (initially micro-aerobic culture, then anaerobic conditions). Solid lines: OD_{600}. Dashed lines: β-carotene productivity.

6.5. Conclusions

We successfully applied adaptive evolution for increased carotenoids production in *Saccharomyces cerevisiae.* An effective design of the evolutionary pressure was based in the antioxidant properties of the carotenoids. The producer strain was shocked using hydrogen peroxide as oxidative agent, increasing the production of carotenoids in more than 200% in a short-term experiment.

Using transcriptome analysis, different molecular mechanisms for increased carotenoids production were elucidated. We found upregulation of genes related with lipid biosynthesis, especially upregulation of genes part of the melavonate pathway, in all the isolated mutants. Upregulation in mechanisms of stress response, including pleiotropic drug resistance genes, suggested ameliorating response for carotenoids toxicity.

7. BONUS CHAPTER: GENOMIC LIBRARY SCREENS FOR GENES INVOLVED IN N-BUTANOL TOLERANCE IN *ESCHERICHIA COLI*[*]

7.1. Summary

n-Butanol is a promising emerging biofuel, and recent metabolic engineering efforts have demonstrated the use of several microbial hosts for its production. However, most organisms have very low tolerance to n-butanol (up to 2% (v/v)), limiting the economic viability of this biofuel. The rational engineering of more robust n-butanol production hosts relies upon understanding the mechanisms involved in tolerance. However, the existing knowledge of genes involved in n-butanol tolerance is limited. The goal of this study is therefore to identify *E. coli* genes that are involved in n-butanol tolerance. Using a genomic library enrichment strategy, we identified approximately 270 genes that were enriched or depleted in n-butanol challenge. The effects of these candidate genes on n-butanol tolerance were experimentally determined using overexpression or deletion libraries. Among the 55 enriched genes tested, 11 were experimentally shown to confer enhanced tolerance to n-butanol when overexpressed compared to the wild type. Among the 84 depleted genes tested, three conferred increased n-butanol resistance when deleted. The overexpressed genes that conferred the largest increase in n-butanol tolerance were related to iron transport and metabolism, *entC* and *feoA*, which

[*] Reprinted with permission from Luis H. Reyes, Maria P. Almario, and Katy C. Kao, (2011). "Genomic library screens for genes involved in n-butanol tolerance in Escherichia coli". *PloS one*, *6*(3), e17678. DOI:10.1371/journal.pone.0017678.

increased the n-butanol tolerance by 32.8 ± 4.0% and 49.1 ± 3.3%, respectively. The deleted gene that resulted in the largest increase in resistance to n-butanol was *astE*, which enhanced n-butanol tolerance by 48.7 ± 6.3%. We identified and experimentally verified 14 genes that decreased the inhibitory effect of n-butanol tolerance on *E. coli*. From the data, we were able to expand the current knowledge on the genes involved in n-butanol tolerance; the results suggest that an increased iron transport and metabolism and decreased acid resistance may enhance n-butanol tolerance. The genes and mechanisms identified in this study will be helpful in the rational engineering of more robust biofuel producers.

7.2. Introduction

There has been renewed interest in the four-carbon alcohol, n-butanol, within the scientific and industrial fields due to its potential as an alternative liquid fuel. n-Butanol has physiochemical properties comparable to gasoline, allowing its use as a fuel replacement in internal combustion engines without any modification (237). Currently, members of the *Clostridia* genus are the only native n-butanol producers known (238, 239). The solvent production in *Clostridia* is coupled to its complex growth phases, which creates difficulties in the engineering of the organism for improved n-butanol production. The complex growth and production phases and the strict anaerobic nature of the native producers have prompted researchers to pursue heterologous hosts for bio-butanol production. In the last few years, with the advances in metabolic engineering,

non-native producers of n-butanol such as *Escherichia coli* (7, 12, 23), *Saccharomyces cerevisiae* (157), *Lactobacillus brevis* (155), *Pseudomonas putida* (156) and *Bacillus subtilis* (156), have been demonstrated as potential hosts for use in n-butanol production. However, n-butanol is highly toxic to microorganisms (240), with most organisms able to tolerate up to 2% (v/v). An exceptional example corresponds to several adapted *P. putida* strains reported to be able to tolerate concentrations of n-butanol higher than 3% (v/v) in rich medium supplemented with glucose; however the tolerance level of the strains without glucose supplementation or in minimum medium were still 1% - 2% (v/v) (241). Understanding the mechanisms involved in n-butanol response can help to facilitate the engineering of production hosts for improved tolerance.

The toxic effects of n-butanol are believed to result from increased membrane fluidity in the presence of the solvent, disrupting the functions of membrane components (242). Solvents affect the membrane by disrupting their fatty acid and protein structure. These disruptions alter membrane fluidity (33), impair internal pH regulation (240), disrupt protein-lipid interactions (33) and negatively impact energy generation by inhibiting nutrient transport (240). Bacteria and other microorganisms can adopt diverse mechanisms to overcome the action of organic solvents. Examples of those mechanisms include: *i.* changes in the hydrophobicity of the outer envelope (243), *ii.* alterations of the cytoplasmic membrane, modifying its structure by changing the saturation of the fatty acids in the phospholipid layer (243), *iii.* changes in the permeability of the membrane by modifications of the lipopolysaccharides and porins (244-246) of the outer mem-

brane, and *iv.* enhanced efflux pump activity to excrete the solvent present in the cytoplasm (163).

Transcriptional analyses and genomic libraries have been used to investigate the molecular mechanisms involved in n-butanol tolerance in *C. acetobutylicum*. Tomas *et al* (247), using transcriptional analysis, determined that genes involved in general stress response and solvent formation in *C. acetobutylicum*, were upregulated under n-butanol stress. In a study using a *C. acetobutylicum* genomic library enrichment, overexpression of genes encoding for transcriptional regulators, specifically the genes CAC0003 and CAC1869 were identified to increase n-butanol tolerance by 13% and 81% respectively (33). The response of *E. coli* to isobutanol via transcriptional analysis has elucidated that quinone malfunction and the action of ArcA are some of the key perturbations during solvent stress (248). Rutherford *et al* (17) showed that n-butanol stress response in *E. coli* share components with other common stress responses. These commonalities include changes in respiratory functions (*nuo* and *cyo* operons), responses to heat shock, oxidative, and cell envelope stress (*rpoE*, *clpB*, *htpG*, *cpxR*, *cpxP*, *sodA*, *sodC*, and *yqhD*), and changes in metabolite transport and biosynthesis (*malE* and *opp* operon). These studies demonstrated that the response to n-butanol is a complex phenotype, involving multiple mechanisms.

Thus far, few genes have been directly identified to be involved in enhanced tolerance to n-butanol. Using an *E. coli* genomic library enrichment strategy, we identified several candidate genes that are involved in n-butanol tolerance. Candidate genes that are enriched or depleted from the genomic enrichment were tested using overexpression

and knockout libraries, respectively. Several of the candidate genes tested were confirmed to reduce the growth inhibitory effects of n-butanol on *E. coli*.

7.3. Materials and Methods

7.3.1. Bacterial strains, plasmid constructs and genomic library construction

The *E. coli* K-12 strain, BW25113 (F^-, Δ(*araD-araB*)*567*, Δ*lacZ4787*(::*rrnB*-3), *lambda*$^-$, *rph*-1, Δ(*rhaD-rhaB*)568, *hsdR514*), was used in this study. Overnight cultures from frozen stocks were grown in 5 ml of Luria-Bertani broth or on solid LB agar plates supplemented with kanamycin (30 μg/ml) and incubated at 37°C.

Genomic DNA was extracted using DNeasy Blood & Tissue Kit (QIAGEN). The genomic DNA was fragmented to pieces between 2000 and 3000 base pairs using sonication (Ultrasonic Liquid Processor S-4000, Misonix, Inc). The ends of the fragmented DNA were repaired using T4 DNA polymerase (New England Biolabs). The library of repaired DNA fragments were ligated to the pSMART-LC Kan vector (Lucigen Corporation), following the manufacture's instructions and transformed into *E. coli* by electroporation using the Gene PulserMXcell Electroporation System (Bio-rad). Cells (approximately 14,000 colonies) were recovered from the plates and frozen stocks of the genomic library were made and saved at -80°C.

7.3.2. n-Butanol challenge

The genomic library was inoculated in 25 ml of LB and incubated at 37°C until OD_{600} of approximately 0.6 was reached. A sample was collected to be used as the reference. The enrichment strategy involves the serial transfers of batch cultures in increasing n-butanol concentrations (0%, 0.9%, 1.3% and 1.7% n-butanol v/v) along with the respective controls (enrichment scheme shown in Figure 7.1). For each serial transfer, when the cultures reached the desired OD_{600} (approximately 0.7), a sample was taken, and the plasmids from the enriched libraries were recovered using alkaline lysis procedure. The constructs were verified via PCR, using the primers SL1 5'-CAG TCC AGT TAC GCT GGA GTC-3' and SR2 5'-GGT CAG GTA TGA TTT AAA TGG TCA GT-3'.

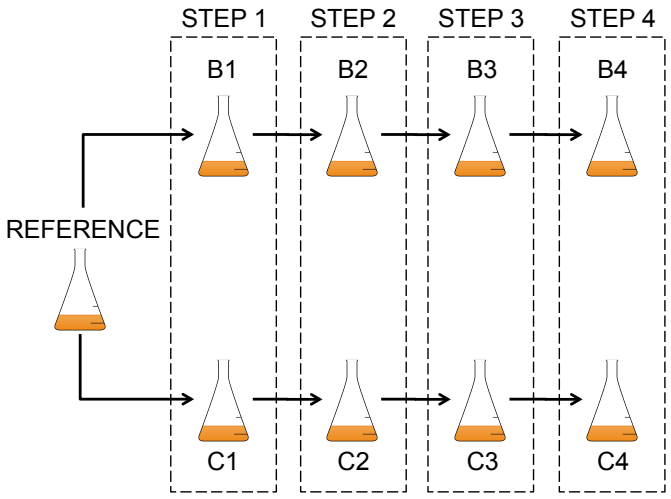


Figure 7.1. n-Butanol challenge strategy. The library was serially transferred in batch cultures with increasing n-butanol concentration. Control serial transfers in the absence of n-butanol was included.

7.3.3. Comparative genome hybridization microarray (array-CGH)

The plasmid DNA (5 μg) isolated from each step of the enrichment, was digested at 37°C for two hours with 10 units each of AluI and RsaI (Invitrogen Corporation) in a reaction containing 10mM $MgCl_2$ and 50mM Tris-HCl (pH=8.0). Samples were cleaned using Zymo Clean & Concentrate-5 columns (Zymo Research), and eluted in TE (pH=8.0). The fragmented plasmid DNA was labeled and hybridized using the BioPrime® Total kit (Invitrogen Corporation) for Agilent aCGH, following manufacture's protocols.

Each labeled sample along with the differentially labeled reference were hybridized to Agilent *E. coli* catalog arrays (*E. coli* gene expression microarray, Agilent Technologies) according to the manufacture's instructions. The arrays were scanned using the GenePix 4100A Microarray Scanner and image analysis performed using GenePix Pro 6.0 Software (Molecular Devices). The Microarray Data Analysis System software was used to normalize the data using LOWESS based normalization algorithm (118). Subsequently, a Student's t-test was used to identify the genes that are statistically significantly enriched or depleted (p-value below 5%) in the n-butanol challenge. The selected genes were clustered via Cluster Affinity Search Technique (249), using the software MeV (Multiexperiment viewer) from the TM4 Microarray Software Suite (119), to group genes with similar enrichment profiles.

7.3.4. Growth kinetic parameters calculated for the genes enriched (via ASKA collection) and depleted (via Keio collection)

The parameters "Improvement in the Inhibitory Effect" (IIE) and "Reduction of Specific Growth Rate in absence of n-butanol" (RSGR) were calculated using Equation 7.1 and Equation 7.2 respectively. Those parameters were determined by measuring the maximum specific growth rate (μ_{max}) of the wild type and the clone (carrying the over-expression plasmid or the deletion clone) in M9 minimal medium (supplied with 5g/L glucose) at two different concentrations of n-butanol, 0% and 0.5% (v/v). The growth kinetics for each strain was measured using a TECAN Infinite M200 Microplate reader (TECAN). Four biological replicas were obtained per sample. A Student's t-test was carried out on the four biological replicates to determine if there was a significant improvement in the n-butanol tolerance when the gene was overexpressed or deleted from the genome.

Equation 7.1

$$IIE = \frac{\left(\frac{\mu_{\text{ASKA or Keio @ 0.5\% n-Butanol}}}{\mu_{\text{ASKA or Keio @ 0\% n-Butanol}}\right) - \left(\frac{\mu_{\text{WT @ 0.5\% n-Butanol}}}{\mu_{\text{WT @ 0\% n-Butanol}}\right)}{\left(\frac{\mu_{\text{WT @ 0.5\% n-Butanol}}}{\mu_{\text{WT @ 0\% n-Butanol}}\right)}$$

Equation 7.2

$$RSGR = 1 - \left(\frac{\mu_{\text{ASKA or Keio @ 0\% n-Butanol}}}{\mu_{\text{WT @ 0\% n-Butanol}}\right)$$

Where $\mu_{\text{ASKA or Keio @ 0.5\% n-Butanol}}$ and $\mu_{\text{WT @ 0.5\% n-Butanol}}$ are the specific growth rates of the overexpression/deletion strain or wild type strain in 0.5% (v/v) n-butanol,

respectively, and $\mu_{\text{ASKA or Keio @ 0\% n-Butanol}}$ and $\mu_{\text{WT @ 0\% n-Butanol}}$ are the specific growth rates of the overexpression/deletion strain or wild type strain in the absence of n-butanol, respectively.

7.4. Results and Discussion

7.4.1. Genomic library construction and description of n-butanol challenge

An *E. coli* genomic library with an approximately seven-fold-coverage of the *E. coli* genome was generated (details are described in the *Materials and Methods* section). The genomic library was exposed to increasing concentrations of n-butanol (0.5%, 0.9%, 1.3%, and 1.7% (v/v)) via batch serial transfers. To reduce false positives, control enrichments in the absence of n-butanol were included. Samples were collected after each step in the n-butanol challenge for subsequent analysis to identify the genes that are enriched or depleted in the presence of n-butanol.

7.4.2. Identifying enriched genes via array-CGH

The plasmids from the genomic library after each step of the serial n-butanol challenge were extracted and hybridized to Comparative Genome Hybridization microarrays (array-CGH), using the unchallenged (original) *E. coli* genomic library as reference. The data obtained from the array-CGH were analyzed as described in the *Materials and Methods* section. Some of the enriched genes identified from the n-butanol challenge may indeed confer enhancements in n-butanol tolerance. However, certain genes

may be enriched as a result of metabolic enhancement (*e.g.* more efficient nutrient uptake and utilization) rather than solvent tolerance. Since, the enriched genes from the controls likely confer general growth advantage through metabolic enhancements, any gene enriched in the n-butanol-challenged libraries that was also enriched in the control experiments was removed from further analysis. In the end, a total of 193 candidate genes were identified to be enriched from the n-butanol challenge. Their enrichment profiles are shown in Figure 7.2.

Among the enriched set of genes shown in Figure 7.2, approximately 30% have membrane-related functions based on Gene Ontology (GO) terms (whereas around 17% of the currently annotated *E. coli* genes are membrane-related), which corresponds with the main cellular response to the presence of other organic solvents (241, 250-252). The main groups of enriched membrane-related genes are those constituting efflux pumps and anti-porters, amino acid and sugar transporter systems, membrane lipoproteins, multidrug resistance and stress response genes. Table 7.1 shows the list of enriched genes with membrane-related functions.

The genes *acrB, argO, mdtB, emrA* were enriched in the n-butanol challenge. Studies in *E. coli* have shown that the AcrAB efflux system is important in multidrug, cyclohexane, n-hexane and n-pentane resistance (146). Our result suggests that AcrB plays a role in n-butanol tolerance as well, possibly by alleviating the cytoplasmic concentration of solvent. Similar conclusions can be drawn for the arginine effluxer (ArgO), the MdtABC multidrug export system (161) and the EmrAB transport system (253).

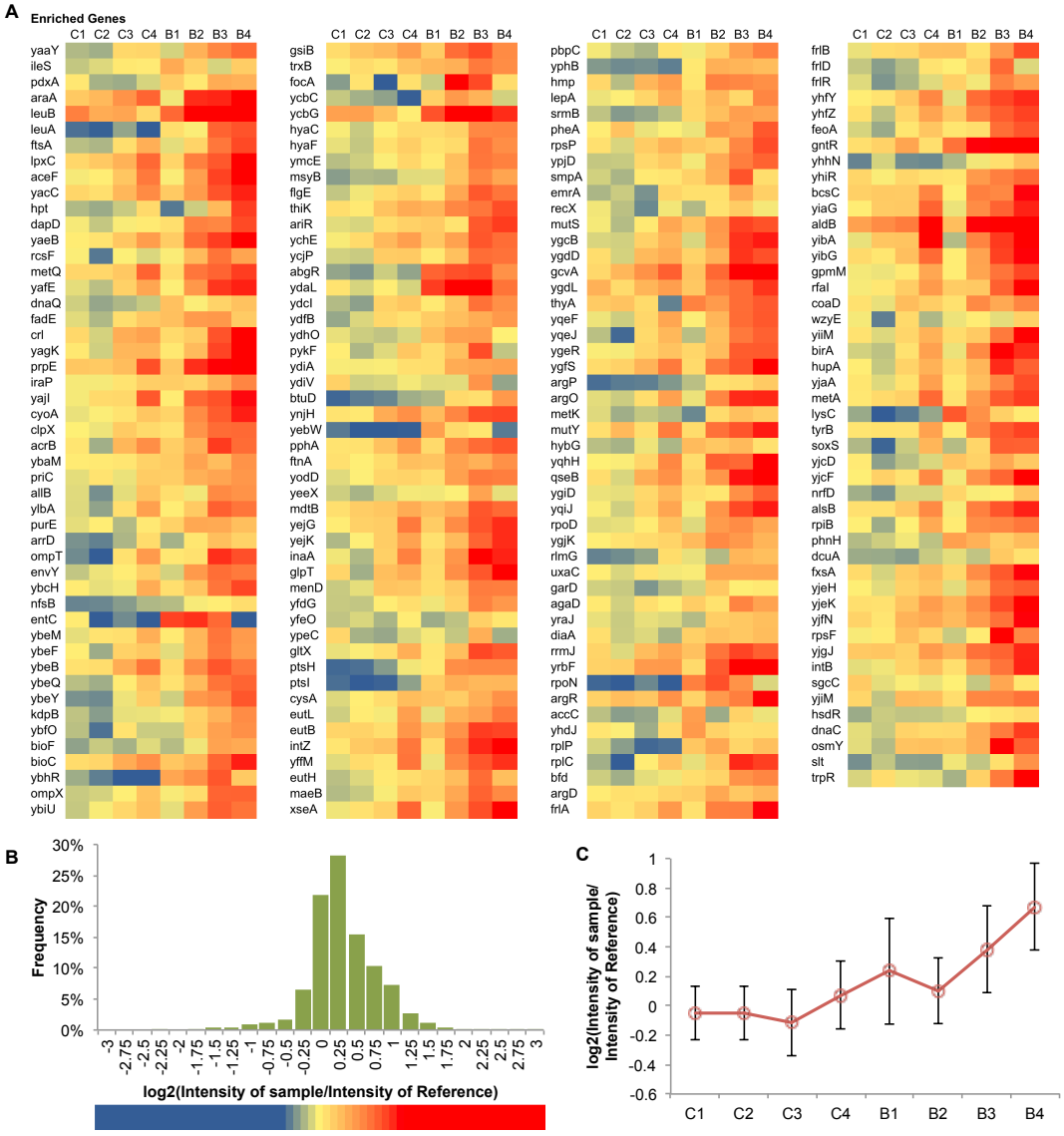


Figure 7.2. Profiles of genes significantly enriched in the n-butanol challenge. A. Heat map of all genes enriched. B. Histogram of the range of normalized *log_2(Intensity of sample/Intensity of reference)*. The colored bar at the bottom part of the figure is the legend for A. C. The averaged profile.

Enrichment of genes involved in amino acid and sugar transport, such as *argD, argR, dapD, lysC, leuA* and *leuB*, suggest that higher energy requirements may be needed to overcome the solvent challenge. The enrichment of genes such as *ompX*, which is

a part of a complex regulatory network involved in the control of outer membrane adaptability and permeability (254), and *smpA*, encoding for the small outer-membrane lipoprotein regulated by σ^E (255), potentially suggest that one mechanism for n-butanol resistance is by preventing n-butanol influx to the cytosol and the disruption of the cell envelope. The xanthine/uracil permease (YjcD), enriched in our experiment, has been predicted to belong to the *purR* regulon (256), which has been identified to be involved in organic solvent tolerance (257). YjaA and YodD are proteins involved in stress response of *E. coli* to hydrogen peroxide, cadmium and acid (50), and our data suggests a potential link of those genes with tolerance to n-butanol. SoxS, a transcriptional activator, has been found as an important transcription factor in the nitric acid, hydrogen peroxide and oxidative stress (258, 259), and tolerance to multiple drugs (145) and cyclohexane (146), possibly via lipopolysaccharide modification.

Table 7.1. Membrane related genes enriched in the n-butanol challenge.

Function	Genes enriched
Efflux pump and anti-porters	*acrB, argO, emrA, focA* and *ybhR*
Amino acid and sugar transporter systems	*agaD, alsB, btuD, dcuA, frlA, glpT, gsiB, kdpB, metQ, sgcC, ycjP* and *yjeH*
Membrane lipoproteins	*cyoA, eutH, eutL, hyaC, ompT, ompX, rfaI, smpA, yajI, yfdG, ygdD, yjcD* and *ypjD*
Multidrug resistance	*acrB, emrA, mdtB* and *ychE*
Stress response	*ompT, yjaA* and *yodD*

A gene ontology analysis of the enriched set of genes, using the toolkit GOEAST (Gene Ontology Enrichment Analysis Software Toolkit) (260), was carried out to identify significantly enriched Gene Ontology (GO) groups in our dataset. The enriched GO terms from the list of enriched genes are summarized in Table 7.2.

Biotin (*birA*, *bioC*, and *bioF*) and amino acid biosynthesis (arginine, lysine, and leucine) were among the functions enriched from the GO-term analysis. Enzymes requiring biotin include acetyl-CoA carboxylase, pyruvate carboxylase, propionyl-CoA carboxylase, methylcrotonyl-CoA carboxylase, geranoyl-CoA carboxylase, oxaloacetate decarboxylase, methylmalonyl-CoA decarboxylase, transcarboxylase and urea amidolyase, which are involved in a variety of different processes such as fatty acid biosynthesis, amino acid metabolism and the citric acid cycle. In fatty acid biosynthesis, biotin has been demonstrated to affect the lipid composition of the cell wall and membrane of *E. coli* (124); cells deficient in biotin showed a decrease in unsaturated fatty acids, the presence of unsaponifiable lipid material and the lack of a lipopolysaccharide fraction in the cell wall and membrane (124).

Table 7.2. Gene Ontology terms enriched in the enriched set of genes.

GO ID	Term	Log odd-ratio	Corrected p-value
GO:0003700	Sequence-specific DNA binding transcription factor activity	0.62	0.07
GO:0016564	Transcription repressor activity	0.90	0.07
GO:0050897	Cobalt ion binding	1.71	0.06
GO:0030145	Manganese ion binding	1.03	0.09
GO:0006525	Arginine metabolic process	1.71	0.06
GO:0009085	Lysine biosynthetic process	2.64	0.01
GO:0019867	Outer membrane	0.97	0.07
GO:0009102	Biotin biosynthetic process	2.93	0.01
GO:0030955	Potassium ion binding	2.20	0.02
GO:0046912	Transferase activity, transferring acyl groups, acyl groups converted into alkyl on transfer	2.93	0.02
GO:0009098	Leucine biosynthetic process	3.20	0.02
GO:0006352	Transcription initiation	2.93	0.02
GO:0016987	Sigma factor activity	2.71	0.03
GO:0044011	Single-species biofilm formation on inanimate substrate	3.52	0.02
GO:0070301	Cellular response to hydrogen peroxide	3.10	0.02

One of the microbial defense mechanisms against organic solvents involves alterations of the cytoplasmic membrane structure, either by modifying the degree of saturation of the fatty acids, isomerization of unsaturated fatty acids, or altering the dynamics of the phospholipid turnover, thereby reestablishing the fluidity and stability of the membrane (261). Modifications of the lipopolysaccharides in the presence of organic solvents has also been identified (244). Thus, the enrichment in biotin biosynthesis genes suggests that increased biosynthesis of biotin may help to enhance cell wall and/or membrane integrity. However, the enrichment of *birA*, which is a repressor of the biotin biosynthesis genes, runs counter to this argument. Since BirA also serves the role of the biotin-ligase in the activation of the enzyme acetyl-CoA carboxylase (ACC) (262), which is the first committed step in fatty acid biosynthesis, the enrichment of *birA* seems to suggest that the activation of ACC may have a larger effect on n-butanol tolerance than reduction in biotin biosynthesis. Several ion-binding proteins were enriched in our studies (*allB, metK, pdxA, araA, leuB, menD, pphA and pykF*). Enrichment in the potassium transporter, *kdpB*, suggests that ion transport may be involved in n-butanol tolerance, possibly by increasing the motive force of many efflux pumps systems (263). In addition, several genes with transcriptional regulation-related functions, such as *srmB*, *rpoD*, *rpoN*, *rplP*, *rplC*, *rpiB* and *rpsF*, were also enriched. Borden and Papoutsakis also found that 4 out of 16 loci that were enriched in a *C. acetobutylicum* genomic library under n-butanol stress were transcriptional regulators (33). This suggests that global transcriptional perturbations may be involved in n-butanol tolerance.

7.4.3. Analysis of genes enriched during n-butanol challenge through the use of an overexpression library

To validate whether the genes enriched from the n-butanol-challenged libraries were indeed involved in enhanced n-butanol tolerance, we used clones from the ASKA collection (121), which is an ORFeome library collection for *E. coli* K-12. Two parameters were calculated to determine the enhancement in n-butanol tolerance due to overexpression of a gene, the Improvement in the Inhibitory Effect (IIE) and the Reduction of Specific Growth Rate in absence of n-butanol (RSGR), as described in the *Materials and Methods* section. IIE measure the increase (in percentage) in the n-butanol tolerance (defined as the improvement of the specific growth rate in presence of n-butanol in comparison with the specific growth rate in absence of the solvent) of the overexpression strain in comparison with the wild type strain. Positive values of IIE signify improvements in n-butanol tolerance in the overexpression strain compared to the wild type. RSGR measures the change of the specific growth rate due to the overexpression of the gene. Under the hypothesis that an increase in the maximum specific growth rate (μ_{max}) is an indication of enhanced tolerance to the solvent, we calculated the parameters IIE and RSGR for each of the strains overexpressing the candidate genes tested. Another alternative measurement to determine the enhancement in n-butanol tolerance is the growth yield. However, based on our data, the specific growth rate seems to be a more sensitive measurement of such improvement (overexpression of some genes decrease the specific growth rate without a significant effect on the growth yield).

We screened 55 out of the 194 genes that were enriched in the n-butanol-challenged library, and identified 11 genes that conferred significant increase in n-butanol tolerance when overexpressed (Table 7.3). Two genes involved in iron metabolism (*entC* and *feoA*) were found to confer a significant increase in n-butanol resistance. Iron metabolism has not been previously associated with enhanced n-butanol tolerance. However, several genes related to iron metabolism were downregulated in *E. coli* under isobutanol stress (248), suggesting a disruption in iron metabolism. Thus, the enhanced n-butanol resistance in *entC* and *feoA* overexpressing strains may be due to the compensatory effects of such a disruption in n-butanol stress. Interestingly, three of the 11 genes (*yibA*, *metA* and *ymcE*) are heat shock related genes (264, 265). These genes are under the control of σ^{32}, which is a sigma factor that is active under several stress conditions. Overexpression of the outer membrane protease, OmpT, which is active under extreme denaturing conditions (266), was found to increase n-butanol tolerance. The formate transporter, encoded by the gene *focA*, which can also act as an efflux pump that regulates the intracellular formate pool (267), also enhanced n-butanol tolerance when overexpressed.

7.4.4. Depleted genes

Along with the enriched genes, depleted genes from the n-butanol-challenged libraries identified in the array-CGH were also analyzed, as some of these genes may help to enhance n-butanol tolerance when their expression is decreased. Similar selection criteria as those used for the enriched gene set were applied to identify and analyze the

genes that are significantly depleted. A total of 84 significantly depleted genes were identified (see Figure 7.3 for the list of genes).

Table 7.3. Genes that significantly increase n-butanol tolerance when they are overexpressed using ASKA collection.

Clone	IIE	RSGR	p-Value
ompT	10.8 ± 0.9%	-13.3 ± 0.7%	0.01
entC	32.8 ± 4.0%	-0.8 ± 0.1%	0.05
yibA	12.7 ± 0.8%	-8.4 ± 0.3%	0.02
metA	14.9 ± 0.9%	-7.2 ± 0.2%	0.01
alsB	13.9 ± 1.0%	-12.2 ± 0.6%	0.02
phnH	42.4± 3.0%	18.4 ± 0.4%	0.01
feoA	49.1 ± 3.3%	3.6 ± 0.1%	0.00
focA	4.3 ± 0.2%	-15.4 ± 0.3%	0.02
hyaF	20.8 ± 2.1%	15.4 ± 0.6%	0.03
ymcE	13.2 ± 0.4%	-11.1 ± 0.2%	0.02
yfdG	20.3 ± 1.4%	4.9 ± 0.2%	0.00

Analysis of the depleted genes may reveal the possible negative effects of higher expression of these genes under n-butanol stress. Those effects can be grouped in two main categories. The first group are genes that when overexpressed possibly increase the metabolic burden to the cell. Genes like *purP*, which is involved in energized high-affinity adenine uptake (268, 269), and *luxS*, which synthesizes the quorum sensing molecule autoinducer-2 (AI-2) (270), are likely not directly involved in increase n-butanol susceptibility. Their depletion from the library is likely due to the increased metabolic burden. The second group constitutes genes that may increase the concentration of n-butanol in the cell. OmpG, which is a nonspecific and efficient channel for sugar and large solutes (271), may also allow the diffusion of n-butanol into the cell. Table 7.4 shows the results of the gene ontology analysis of the set of depleted genes.

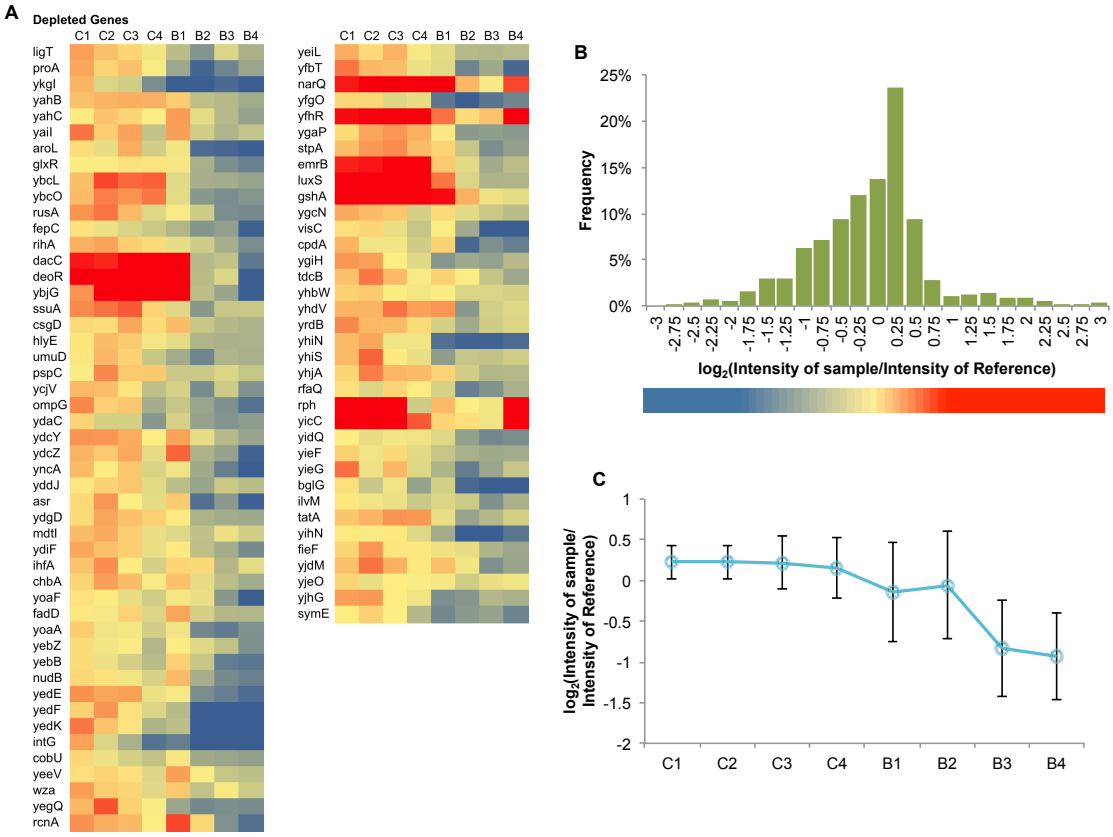


Figure 7.3. Profiles of genes significantly depleted in the n-butanol challenge. A. Heat map of all genes depleted. B. Histogram of the range of normalized *log$_2$(Intensity of sample/Intensity of Reference*. The colored bar at the bottom part of the figure is the legend for A. C. The averaged profile.

7.4.5. Analysis of genes depleted during the n-butanol challenge using the *E. coli* knockout collection

Strains from the Keio knockout collection (122, 272) were used to examine if deletion of the depleted genes could increase the n-butanol tolerance of *E. coli*. The IIE and RSGR parameters were calculated from the wild type strain and the deletion mutant in M9 minimal medium at 0% and 0.5% (v/v) n-butanol.

Table 7.4. Gene Ontology terms enriched in the depleted gene set

GO ID	Term	Log odd-ratio	Corrected p-value
GO:0006508	Proteolysis	2.54	0.00
GO:0008360	Regulation of cell shape	1.94	0.09
GO:0008658	Penicillin binding	3.42	0.09
GO:0008236	Serine-type peptidase activity	3.57	0.00
GO:0009081	Branched chain family amino acid metabolic process	2.42	0.05
GO:0009405	Pathogenesis	3.42	0.10
GO:0003984	Acetolactate synthase activity	3.42	0.10
GO:0046654	Tetrahydrofolate biosynthetic process	3.42	0.04
GO:0046930	Pore complex	2.94	0.02
GO:0043190	ATP-binding cassette (ABC) transporter complex	1.89	0.09
GO:0009432	SOS response	2.42	0.05
GO:0015774	Polysaccharide transport	2.57	0.09

Out of 84 genes tested, three genes were found to significantly reduce the inhibitory effect of n-butanol when they were deleted: *astE, ygiH* and *rph.* The calculated parameters are shown in Table 7.5. Improvements in the relative specific growth rates were observed in all three deletions strains in the presence of n-butanol compared with the wild type (see Figure 7.4).

Table 7.5. Genes that significantly enhance n-butanol tolerance when deleted from the *E. coli* genome.

Mutant	IIE	RSGR	p-value
astE	48.7± 6.3%	-3.3± 0.3%	0.00
ygiH	14.8± 1.2%	12.3 ± 0.6%	0.02
rph	48.4± 4.1%	-10.2 ± 0.6%	0.01

AstE hydrolyzes N^2-succinylglutamate into succinate and L-glutamate. L-glutamate has been identified to be involved in acid stress response in *E. coli* (273, 274). Recent studies have demonstrated that n-butanol response in *Lactobacillus brevis* (15) downregulated the acid stress response significantly. Thus, deletion of *astE* may lead to

decreased L-glutamate pool, resulting in increased n-butanol tolerance. Deletion of *ygiH*, the gene encoding an inner membrane protein, increased resistance to n-butanol by 14.8±1.2%. Studies have found that PlsY proteins in *Bacillus subtilis* and *Streptococcus pneumoniae* exhibit similarities with YgiH, as they both function as the glycerol-3-phosphate acyltransferases for phospholipid biosynthesis (275). However, in *E. coli*, the function of PlsY is replaced by PlsB, and PlsX and YgiH play important roles in regulating the intracellular levels of acyl-ACP, an important precursor in the fatty acid biosynthesis. Studies demonstrated that single deletions of the PlsX or YgiH do not strongly affect cell growth, however double deletion is synthetically lethal (275).

The depletion of YgiH suggests that phospholipid biosynthesis may be optimized to the requirements needed to overcome the solvent stress. Deletion of the RNase PH gene, Rph, resulted in an increase in n-butanol tolerance by 48.4 ± 4.1%. However, the *E. coli* strain BW25113, used in this study, has a *rph*- background, with a frameshift mutation inactivating *rph* function. Complete deletion of this gene may ameliorate transcriptional polarity on the *pyrE* gene, increasing pyrimidine biosynthesis (276). Thus, *rph* most likely is not directly involved in n-butanol tolerance in *E. coli*.

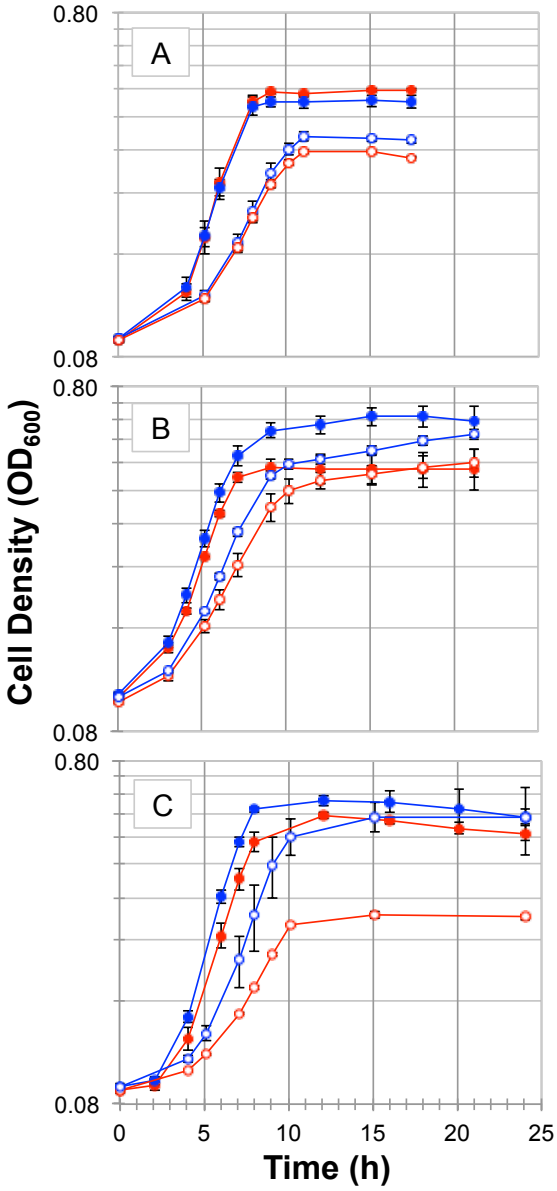


Figure 7.4. The growth kinetics of A. *ΔygiH*, B. *ΔastE*, and C. *Δrph* vs. wild type. Red lines represent the growth kinetics of wild type in absence (open circle) and presence (solid circles) of 0.5% (v/v) n-butanol. Blue lines represent the growth curves of the deletion strains in absence (open circle) and presence (solid circle) of 0.5% (v/v) n-butanol.

7.5. Conclusions

Using a genomic library enrichment strategy, we identified genes involved in n-butanol tolerance in *E. coli.* We identified two groups of genes from the n-butanol challenge: genes that were enriched and depleted during the exposure to n-butanol. From the data, we were able to expand the current knowledge on the genes involved in n-butanol tolerance; we observed enrichment of genes involved in membrane functions, transport systems (encoded by *acrB, argO, mdtB* and *emrA*), amino acid transport, sugar transport and stress response proteins. We also found enrichment in genes involved in biotin synthesis (*bioC* and *bioF*), indicating that an increase in this cofactor may help to enhance membrane integrity. Among the depleted genes, we identified genes that when overexpressed may cause undesirable increase in n-butanol inside the cell. We experimentally verified 14 genes that decreased the growth-inhibitory effects of n-butanol on *E. coli*. The overexpression of the iron transport and metabolism related genes, *entC* and *feoA*, increased n-butanol tolerance by 32.8 ± 4.0% and 49.1 ± 3.3%, respectively. Deletion of *astE*, which may lead to decreased L-glutamate (potentially decreasing acid resistance), enhanced n-butanol tolerance by 48.7 ± 6.3%. The genes and mechanisms identified in this study will be useful in the rational engineering of more robust biofuel producers. In addition, since organic solvent tolerance is known as a complex phenotype, there may be potential synergistic effects between different combinations of deletions and overexpressions of genes identified in this work; we will be investigating such effects in subsequent works.

7.6. Data Availability

All raw data is MIAME compliant and have been deposited in the GEO database with accession number GSE26223.

8. CONCLUSIONS AND RECOMMENDATIONS

This dissertation examined the use of laboratory adaptive evolution as a useful technique to study complex phenotypes in microbial systems.

8.1. Visualizing Evolution in Real-Time

The method based on *in vitro* adaptive evolution, Visualizing Evolution in Real-Time (VERT) was successfully developed and applied in different studies of complex phenotypes in microbial systems as the n-butanol tolerance in *Escherichia coli* and tolerance to hydrolysates of lignocellulosic biomass in *Saccharomyces cerevisiae*. VERT facilitates the identification of fitter mutants throughout the course of evolution and greatly enhances the mapping of adaptive landscapes of industrially relevant phenotypes.

VERT uses isogenic, but differentially labeled (typically with fluorescent proteins) strains to seed the initial evolving population. When a beneficial mutant arises and expands in the population, the labeled subpopulation that it belongs is expected to increase in proportion. These changes can be measured by tracking the relative proportions of each of the labeled subpopulations at each point in time. Each continued expansion in the proportion of a colored subpopulation is called an "adaptive event". Thus, tracking the different colored subpopulations permit the researcher to determine when a fitter mutant arises in the population, facilitating the map-out of fitness landscapes and a more rational scheme for the isolation of adaptive mutants for further characterization.

8.2. Effect of Clonal Interference in E. coli Evolved Under n-Butanol Stress

The gram-negative bacteria *E. coli* was evolved under n-butanol stress using continuous cultures (chemostats) as a platform to carry out the evolution experiment. Using a two-colored VERT system (GFP- and YFP-labeled subpopulations), the evolutionary dynamics of *E. coli* under increasing concentrations of n-butanol was tracked. The results showed that while consecutive adaptive mutants have fitness advantages over their predecessors, they might not always be better than the ancestral population, likely due to the heterogeneities in the evolving population and potential epistatic interactions between clones. Thus, this result further highlights the benefit of tracking the adaptive events in a population to isolate each adaptive mutant, as the last one isolated from the population may not necessarily be the best one.

Using a combination of phenotypic studies, whole genome transcriptome profiling and resequencing analyses, underlying n-butanol tolerance mechanisms were identified in mutants isolated from the two independent subpopulations. Divergent evolutionary trajectories for increased n-butanol tolerance were identified between the GFP-labeled and YFP-labeled mutants. In case of the GFP-labeled mutants, a reduced activity of the ferric uptake regulator Fur, leading to increased siderophore biosynthesis and transport, which ultimately led to membrane alterations, was identified to be a likely mechanism of enhanced n-butanol tolerance. The deactivation of Fur also led to cross-tolerance between n-butanol and the cationic antimicrobial peptide polymyxin B. However, this tolerance mechanism was only observed in mutants from the GFP-labeled sub-

population and not in the YFP-labeled subpopulation. Molecular mechanisms of n-butanol tolerance of the YFP-labeled mutants indicated additional membrane-related and osmotic stress related genes conferring n-butanol tolerance in *E. coli*. Interestingly, the YFP-labeled mutants showed cross-tolerance between n-butanol and osmotic stress while the GFP-labeled mutants showed antagonistic pleiotropy between n-butanol and osmotic stress tolerance, contrary to the response observed under polymyxin B stress. Further efforts trying to combine the two independent mechanisms of tolerance demonstrated divergence in the evolutionary trajectories, since they resulted in significant decrease in relative fitness in presence of n-butanol.

8.3. Effect of Clonal Interference in S. cerevisiae Evolved Under Hydrolysates of Lignocellulosic Biomass

In vitro evolution using serial batch transfers was used to uncover molecular mechanisms of tolerance to hydrolysates of lignocellulosic biomass of *S. cerevisiae*. The hydrolysates are composed of a mixture of sugars and inhibitors formed during the pre-treatment processes. A three-color VERT system was used in order to enable a rational scheme for isolating adaptive mutants and map-out the adaptive landscape.

Phenotypic analysis of several isolated adaptive mutants with enhanced fitness in hydrolysates demonstrated different evolutionary routes. Even though the relative fitness of the isolated mutants compared with the parental strains in the presence of hydrolysates was positive (increase growth advantage), analysis of the mutants in individu-

al and combination of common inhibitors present in the hydrolysates (acetic acid, furfural and hydroxymethylfurfural (HMF)) showed differential levels of resistance to the different growth conditions. Using transcriptome analysis, some potential mechanisms for enhanced tolerance in the isolated adaptive mutants were elucidated.

8.4. Effective Design of Evolutionary Pressure for the Increased Production of Carotenoids in S. cerevisiae Using Adaptive Evolution

The application of adaptive evolution for increased production of secondary metabolites has been constantly avoided due to the nature of the technique. Adaptive evolution is a natural process where the Darwinian principle of natural selection applies, and it establishes that genetic variation is causally connected to the differential ability to survive and reproduce, and positive differential reproduction will probably succeed; however, secondary metabolites are compounds not directly involved in normal cell growth. Application of adaptive evolution for increased production of such metabolites generally ends with decreased production due to metabolic burden imposed for the channeled flux towards the bioproduct.

In this dissertation, laboratory adaptive evolution was successfully exploited in order to improve secondary metabolite production, specifically the improvement of carotenoids production in an engineered *S. cerevisiae* carotenoids producing strain. The effective design of the evolutionary strategy was based in the antioxidant properties of the carotenoids. Scheduled hydrogen peroxide shocking experiments ensured an in-

crease of carotenoids production in more than 200% in a short-term experiment. Transcriptome analysis helped to elucidate the molecular mechanisms for increased carotenoids production.

8.5. Recommendations

- Redesign VERT using DNA barcoding in order to increase the number of studied independent subpopulations (not limited for the available fluorescent proteins or the detection limits imposed for the instrument), increasing the resolution of the mapped evolutionary dynamics and avoiding extra metabolic burden to the microbial system (evolution can lead to fluorescence loss). Subpopulations can be tracked in real-time using qPCR, extracting gDNA from evolved populations and quantifying relative amounts of barcodes present. Selection of mutants can be easily achieved by integration of selective markers under inducible promoters, using such constructs as DNA barcodes for the evolution tracking.
- Implementation of VERT in n-butanol tolerant mutants, described in this dissertation, under acid stress, in order to demonstrate divergence in the evolutionary trajectories (antagonistic pleiotropy) observed not only in *E. coli*, but also in *Lactobacillus brevis*. This will allow to determine if increased acid tolerance implies decreased n-butanol tolerance, or alternative adaptive routes in the fitness landscape allows cross-tolerance between the two stressors.

- Demonstrate that the design of effective selective pressure is not an isolated successful event, but the aforementioned design can be applied to increased carotenoids production in other microbial systems as *E. coli*.

REFERENCES

1. Demirbas, A. (2009) Political, economic and environmental impacts of biofuels: a review. *Applied Energy* **86**, S108–S117

2. Bioethanol needs biotech now [Editorial] (2006). *Nat. Biotechnol.* **24**, 725

3. Bringezu, S., Schütz, H., O Brien, M., Kauppi, L., and Howarth, R. W. (2009) Toward sustainable production and use of resources: Assessing biofuels. *United Nations Environmental Program*. ISBN: 978-92-807-3052-4, Paris, France, 120 pages

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

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

6. Connor, M. R., and Liao, J. C. (2009) Microbial production of advanced transportation fuels in non-natural hosts. *Current Opinion in Biotechnology* **20**, 307–315

7. Shen, C. R., and Liao, J. C. (2008) Metabolic engineering of Escherichia coli for 1-butanol and 1-propanol production via the keto-acid pathways. *Metab. Eng.* **10**, 312–320

8. Sillers, R., Al-Hinai, M. A., and Papoutsakis, E. T. (2009) Aldehyde-alcohol dehydrogenase and/or thiolase overexpression coupled with CoA transferase downregulation lead to higher alcohol titers and selectivity in Clostridium acetobutylicum fermentations. *Biotechnol. Bioeng.* **102**, 38–49

9. Smith, K. M., and Liao, J. C. (2011) An evolutionary strategy for isobutanol production strain development in Escherichia coli. *Metab. Eng.* **13**, 674–681

10. Trinh, C. T., Li, J., Blanch, H. W., and Clark, D. S. (2011) Redesigning Escherichia coli metabolism for anaerobic production of isobutanol. *Appl. Environ. Microbiol.* **77**, 4894–4904

11. Fischer, C. R., Klein-Marcuschamer, D., and Stephanopoulos, G. (2008) Selec-

tion and optimization of microbial hosts for biofuels production. *Metab. Eng.* **10**, 295–304

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

13. Borden, J. R., Jones, S. W., Indurthi, D., Chen, Y., and Terry Papoutsakis, E. (2010) A genomic-library based discovery of a novel, possibly synthetic, acid-tolerance mechanism in Clostridium acetobutylicum involving non-coding rnas and ribosomal RNA processing. *Metab. Eng.* **12**, 268–281

14. Lynch, M. D., Warnecke, T., and Gill, R. T. (2006) SCALEs: multiscale analysis of library enrichment. *Nat. Methods* **4**, 87–93

15. Winkler, J., and Kao, K. C. (2011) Transcriptional analysis of Lactobacillus brevis to n-butanol and ferulic acid stress responses. *PLoS ONE* **6**, e21438

16. Oh, M.-K., Rohlin, L., Kao, K. C., and Liao, J. C. (2002) Global expression profiling of acetate-grown Escherichia coli. *J. Biol. Chem.* **277**, 13175–13183

17. Rutherford, B. J., Dahl, R. H., Price, R. E., Szmidt, H. L., Benke, P. I., Mukhopadhyay, A., and Keasling, J. D. (2010) Functional genomic study of exogenous n-butanol stress in Escherichia coli. *Appl. Environ. Microbiol.* **76**, 1935–1945

18. Warner, J. R., Reeder, P. J., Karimpour-Fard, A., Woodruff, L. B. A., and Gill, R. T. (2010) Rapid profiling of a microbial genome using mixtures of barcoded oligonucleotides. *Nat. Biotechnol.* **28**, 856–862

19. Liu, H., Yan, M., Lai, C., Xu, L., and Ouyang, P. (2010) gTME for improved xylose fermentation of Saccharomyces cerevisiae. *Appl Biochem Biotechnol* **160**, 574–582

20. Conrad, T. M., Lewis, N. E., and Palsson, B. Ø. (2011) Microbial laboratory evolution in the era of genome-scale science. *Mol Syst Biol* **7**, 509

21. Reyes, L. H., Almario, M. P., Winkler, J., Orozco, M. M., and Kao, K. C. (2012) Visualizing evolution in real time to determine the molecular mechanisms of n-butanol tolerance in Escherichia coli. *Metab. Eng.* **14**, 579–590

22. Minty, J. J., Lesnefsky, A. A., Lin, F., Chen, Y., Zaroff, T. A., Veloso, A. B., Xie, B., McConnell, C. A., Ward, R. J., Schwartz, D. R., Rouillard, J.-M., Gao, Y., Gulari, E., and Lin, X. N. (2011) Evolution combined with genomic study elucidates genetic bases of isobutanol tolerance in Escherichia coli. *Microb. Cell Fact.* **10**, 18

23. Atsumi, S., and Liao, J. C. (2008) Metabolic engineering for advanced biofuels production from Escherichia coli. *Current Opinion in Biotechnology* **19**, 414–419

24. Patnaik, R. (2008) Engineering complex phenotypes in industrial strains. *Biotechnol. Prog.* **24**, 38–47

25. Singer, B., and Kuśmierek, J. T. (1982) Chemical mutagenesis. *Annu. Rev. Biochem.* **51**, 655–693

26. Bailey, J. (1991) Toward a science of metabolic engineering. *Science* **252**, 1668–1675

27. Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., Khandurina, J., Trawick, J. D., Osterhout, R. E., Stephen, R., Estadilla, J., Teisan, S., Schreyer, H. B., Andrae, S., Yang, T. H., Lee, S. Y., Burk, M. J., and Van Dien, S. (2011) Metabolic engineering of Escherichia coli for direct production of 1,4-butanediol. *Nat. Chem. Biol.* **7**, 445–452

28. Bongaerts, J. (2001) Metabolic engineering for microbial production of aromatic amino acids and derived compounds. *Metab. Eng.* **3**, 289–300

29. He, M.-X., Wu, B., Shui, Z.-X., Hu, Q.-C., Wang, W.-G., Tan, F.-R., Tang, X.-Y., Zhu, Q.-L., Pan, K., Li, Q., and Su, X.-H. (2012) Transcriptome profiling of Zymomonas mobilis under ethanol stress. *Biotechnol Biofuels* **5**, 75

30. Carlisle, P. L., and Kadosh, D. (2013) A genome-wide transcriptional analysis of morphology determination in Candida albicans. *Mol. Biol. Cell* **24**, 246–260

31. Chatterjee, A., Saranath, D., Bhatter, P., and Mistry, N. (2013) Global transcriptional profiling of longitudinal clinical isolates of Mycobacterium tuberculosis exhibiting rapid accumulation of drug resistance. *PLoS ONE* **8**, e54717

32. Wang, Z., Gerstein, M., and Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57–63

33. Borden, J. R., and Papoutsakis, E. T. (2007) Dynamics of genomic-library enrichment and identification of solvent tolerance genes for Clostridium acetobutylicum. *Appl. Environ. Microbiol.* **73**, 3061–3068

34. Woodruff, L. B. A., Pandhal, J., Ow, S. Y., Karimpour-Fard, A., Weiss, S. J., Wright, P. C., and Gill, R. T. (2012) Genome-scale identification and characterization of ethanol tolerance genes in Escherichia coli. *Metab. Eng.* **15**, 124–133

35. Sandoval, N. R., Mills, T. Y., Zhang, M., and Gill, R. T. (2011) Elucidating ace-

tate tolerance in E. coli using a genome-wide approach. *Metab. Eng.* **13**, 214–224

36. Knietsch, A., Waschkowitz, T., Bowien, S., Henne, A., and Daniel, R. (2003) Metagenomes of complex microbial consortia derived from different soils as sources for novel genes conferring formation of carbonyls from short-chain polyols on Escherichia coli. *J. Mol. Microbiol. Biotechnol.* **5**, 46–56

37. Suenaga, H., Ohnuki, T., and Miyazaki, K. (2007) Functional screening of a metagenomic library for genes involved in microbial degradation of aromatic compounds. *Environ. Microbiol.* **9**, 2289–2297

38. Amann, R. I., Ludwig, W., and Schleifer, K. H. (1995) Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol. Rev.* **59**, 143–169

39. Cottrell, M. T., Moore, J. A., and Kirchman, D. L. (1999) Chitinases from uncultured marine microorganisms. *Appl. Environ. Microbiol.* **65**, 2553–2557

40. Henne, A., Schmitz, R. A., Bömeke, M., Gottschalk, G., and Daniel, R. (2000) Screening of environmental DNA libraries for the presence of genes conferring lipolytic activity on Escherichia coli. *Appl. Environ. Microbiol.* **66**, 3113–3116

41. Yun, J., Kang, S., Park, S., Yoon, H., Kim, M.-J., Heu, S., and Ryu, S. (2004) Characterization of a novel amylolytic enzyme encoded by a gene from a soil-derived metagenomic library. *Appl. Environ. Microbiol.* **70**, 7229–7235

42. Ferrer, M., Golyshina, O. V., Chernikova, T. N., Khachane, A. N., Reyes-Duarte, D., Santos, V. A. P. M. D., Strompl, C., Elborough, K., Jarvis, G., Neef, A., Yakimov, M. M., Timmis, K. N., and Golyshin, P. N. (2005) Novel hydrolase diversity retrieved from a metagenome library of bovine rumen microflora. *Environ. Microbiol.* **7**, 1996–2010

43. Winkler, J., Rehmann, M., and Kao, K. C. (2010) Novel Escherichia coli hybrids with enhanced butanol tolerance. *Biotechnol Lett* **32**, 915–920

44. Patnaik, R., Louie, S., Gavrilovic, V., Perry, K., Stemmer, W. P. C., Ryan, C. M., and del Cardayré, S. (2002) Genome shuffling of Lactobacillus for improved acid tolerance. *Nat. Biotechnol.* **20**, 707–712

45. Dai, M., Ziesman, S., Ratcliffe, T., Gill, R. T., and Copley, S. D. (2005) Visualization of protoplast fusion and quantitation of recombination in fused protoplasts of auxotrophic strains of Escherichia coli. *Metab. Eng.* **7**, 45–52

46. Yu, L., Pei, X., Lei, T., Wang, Y., and Feng, Y. (2008) Genome shuffling en-

hanced l-lactic acid production by improving glucose tolerance of Lactobacillus rhamnosus. *Journal of Biotechnology* **134**, 154–159

47. Alper, H., and Stephanopoulos, G. (2007) Global transcription machinery engineering: A new approach for improving cellular phenotype. *Metab. Eng.* **9**, 258–267

48. Lanza, A. M., and Alper, H. S. (2012) Using transcription machinery engineering to elicit complex cellular phenotypes. *Methods Mol. Biol.* **813**, 229–248

49. Portnoy, V. A., Bezdan, D., and Zengler, K. (2011) Adaptive laboratory evolution--harnessing the power of biology for metabolic engineering. *Current Opinion in Biotechnology* **22**, 590–594

50. Lee, D.-H., and Palsson, B. Ø. (2010) Adaptive evolution of Escherichia coli K-12 MG1655 during growth on a nonnative carbon source, L-1,2-propanediol. *Appl. Environ. Microbiol.* **76**, 4158–4168

51. Atsumi, S., Wu, T.-Y., Machado, I. M. P., Huang, W.-C., Chen, P.-Y., Pellegrini, M., and Liao, J. C. (2010) Evolution, genomic analysis, and reconstruction of isobutanol tolerance in Escherichia coli. *Mol Syst Biol* **6**, 449

52. Hu, H., and Wood, T. K. (2010) An evolved Escherichia coli strain for producing hydrogen and ethanol from glycerol. *Biochem. Biophys. Res. Commun.* **391**, 1033–1038

53. Wisselink, H. W., Toirkens, M. J., del Rosario Franco Berriel, M., Winkler, A. A., van Dijken, J. P., Pronk, J. T., and van Maris, A. J. A. (2007) Engineering of Saccharomyces cerevisiae for efficient anaerobic alcoholic fermentation of L-arabinose. *Appl. Environ. Microbiol.* **73**, 4881–4891

54. Kao, K. C., and Sherlock, G. (2008) Molecular characterization of clonal interference during adaptive evolution in asexual populations of Saccharomyces cerevisiae. *Nat. Genet.* **40**, 1499–1504

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

56. Sauer, U. (2001) Evolutionary engineering of industrially important microbial phenotypes. *Adv. Biochem. Eng. Biotechnol.* **73**, 129–169

57. Portnoy, V. A., Herrgård, M. J., and Palsson, B. Ø. (2008) Aerobic fermentation of d-glucose by an evolved cytochrome oxidase-deficient Escherichia coli strain. *Appl. Environ. Microbiol.* **74**, 7561–7569

58. Comas, I., Borrell, S., Roetzer, A., Rose, G., Malla, B., Kato-Maeda, M., Galagan, J., Niemann, S., and Gagneux, S. (2012) Whole-genome sequencing of rifampicin-resistant Mycobacterium tuberculosis strains identifies compensatory mutations in RNA polymerase genes. *Nat. Genet.* **44**, 106–110

59. Toprak, E., Veres, A., Michel, J.-B., Chait, R., Hartl, D. L., and Kishony, R. (2011) Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. *Nat. Genet.* **44**, 101–105

60. Fitzgerald, J. R., and Musser, J. M. (2001) Evolutionary genomics of pathogenic bacteria. *Trends in Microbiology* **9**, 547–553

61. Paulsen, I. T., Chen, J., Nelson, K. E., and Saier, M. H. (2001) Comparative genomics of microbial drug efflux systems. *J. Mol. Microbiol. Biotechnol.* **3**, 145–150

62. Callister, S. J., McCue, L. A., Turse, J. E., Monroe, M. E., Auberry, K. J., Smith, R. D., Adkins, J. N., and Lipton, M. S. (2008) Comparative bacterial proteomics: analysis of the core genome concept. *PLoS ONE* **3**, e1542

63. Ding, M.-Z., Zhou, X., and Yuan, Y.-J. (2009) Metabolome profiling reveals adaptive evolution of Saccharomyces cerevisiae during repeated vacuum fermentations. *Metabolomics* **6**, 42–55

64. Goodarzi, H., Bennett, B. D., Amini, S., Reaves, M. L., Hottes, A. K., Rabinowitz, J. D., and Tavazoie, S. (2010) Regulatory and metabolic rewiring during laboratory evolution of ethanol tolerance in E. coli. *Mol Syst Biol* **6**, 378

65. Shaver, A. C., Dombrowski, P. G., Sweeney, J. Y., Treis, T., Zappala, R. M., and Sniegowski, P. D. (2002) Fitness evolution and the rise of mutator alleles in experimental Escherichia coli populations. *Genetics* **162**, 557–566

66. Wright, S. (1931) Evolution in mendelian populations. *Genetics* **16**, 97

67. Wright, S. (1988) Surfaces of selective value revisited. *Am. Nat.* **131**, 115–123

68. Wright, S. (1982) Character change, speciation, and the higher taxa. *Evolution* **36**, 427–443

69. Orr, H. A. (2005) The genetic theory of adaptation: a brief history. *Nat Rev Genet* **6**, 119–127

70. Weinreich, D. M., Delaney, N. F., Depristo, M. A., and Hartl, D. L. (2006) Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* **312**, 111–114

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

72. Desai, M. M., Fisher, D. S., and Murray, A. W. (2007) The speed of evolution and maintenance of variation in asexual populations. *Curr. Biol.* **17**, 385–394

73. Gerrish, P. (2001) The rhythm of microbial adaptation. *Nature* **413**, 299–302

74. Gerrish, P. J., and Lenski, R. E. (1998) The fate of competing beneficial mutations in an asexual population. *Genetica* **102-103**, 127–144

75. Orr, H. A. (2000) The rate of adaptation in asexuals. *Genetics* **155**, 961–968

76. Kim, Y., and Stephan, W. (2003) Selective sweeps in the presence of interference among partially linked loci. *Genetics* **164**, 389–398

77. Campos, P. R. A., and de Oliveira, V. M. (2004) Mutational effects on the clonal interference phenomenon. *Evolution* **58**, 932–937

78. Wilke, C. O. (2004) The speed of adaptation in large asexual populations. *Genetics* **167**, 2045–2053

79. Desai, M. M., and Fisher, D. S. (2007) Beneficial mutation selection balance and the effect of linkage on positive selection. *Genetics* **176**, 1759–1798

80. Yedid, G., and Bell, G. (2001) Microevolution in an electronic microcosm. *Am. Nat.* **157**, 465–487

81. Bachtrog, D., and Gordo, I. (2004) Adaptive evolution of asexual populations under Muller's ratchet. *Evolution* **58**, 1403–1413

82. Arjan, J. A., Visser, M., Zeyl, C. W., Gerrish, P. J., Blanchard, J. L., and Lenski, R. E. (1999) Diminishing returns from mutation supply rate in asexual populations. *Science* **283**, 404–406

83. Elena, S. F., and Lenski, R. E. (2003) Microbial genetics: evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat Rev Genet* **4**, 457–469

84. Gresham, D., Desai, M. M., Tucker, C. M., Jenq, H. T., Pai, D. A., Ward, A., DeSevo, C. G., Botstein, D., and Dunham, M. J. (2008) The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS Genet.* **4**, e1000303

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

86. Elena, S. F., and Lenski, R. E. (1997) Long-term experimental evolution in Escherichia coli. VII. Mechanisms maintaining genetic variability within populations. *Evolution* **51**, 1058–1067

87. Sniegowski, P. D., Gerrish, P. J., and Lenski, R. E. (1997) Evolution of high mutation rates in experimental populations of E. coli. *Nature* **387**, 703–705

88. Vulić, M., Lenski, R. E., and Radman, M. (1999) Mutation, recombination, and incipient speciation of bacteria in the laboratory. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 7348–7351

89. Lenski, R. E., Winkworth, C. L., and Riley, M. A. (2003) Rates of DNA sequence evolution in experimental populations of Escherichia coli during 20,000 generations. *J. Mol. Evol.* **56**, 498–508

90. Selifonova, O., Valle, F., and Schellenberger, V. (2001) Rapid evolution of novel traits in microorganisms. *Appl. Environ. Microbiol.* **67**, 3645–3649

91. Lenski, R. E., Rose, M. R., Simpson, S. C., and Tadler, S. C. (1991) Long-term experimental evolution in Escherichia coli. 1. Adaptation and divergence during 2,000 generations. *American Naturalist* **138**, 1315–1341

92. Masel, J. (2011) Genetic drift. *Current Biology* **21**, R837–R838

93. Kawecki, T. J., Barton, N. H., and Fry, J. D. (1997) Mutational collapse of fitness in marginal habitats and the evolution of ecological specialisation. *J Evolution Biol* **10**, 407–429

94. Chao, L., Levin, B. R., and Stewart, F. M. (1977). A complex community in a simple habitat: an experimental study with bacteria and phage. *Ecology* **58**, 369–378

95. Lenski, R. E., and Levin, B. R. (1985) Constraints on the coevolution of bacteria and virulent phage: a model, some experiments, and predictions for natural communities. *American Naturalist* **125**, 585–602

96. Lenski, R. E. (1988) Experimental studies of pleiotropy and epistasis in Escherichia coli. I. Variation in competitive fitness among mutants resistant to virus T4. *Evolution* **42**, 425–432

97. Zhang, E., and Ferenci, T. (1999) OmpF changes and the complexity of Esche-

richia coliadaptation to prolonged lactose limitation. *FEMS Microbiol Lett* **176**, 395–401

98. Bull, J. J., Badgett, M. R., and Wichman, H. A. (2000) Big-benefit mutations in a bacteriophage inhibited with heat. *Mol. Biol. Evol.* **17**, 942–950

99. Reyes, L. H., Winkler, J., and Kao, K. C. (2012) Visualizing evolution in real-time method for strain engineering. *Front Microbiol* **3**, 198

100. Almario, M. P., Reyes, L. H., and Kao, K. C. (2013) Evolutionary engineering of Saccharomyces cerevisiae for enhanced tolerance to hydrolysates of lignocellulosic biomass. *Biotechnol. Bioeng.*

101. Elena, S. F., and Lenski, R. E. (2003) Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat Rev Genet* **4**, 457–469

102. Winkler, J., and Kao, K. C. (2012) Computational identification of adaptive mutants using the VERT system. *J Biol Eng* **6**, 3

103. Nicolaou, S. A., Gaida, S. M., and Papoutsakis, E. T. (2010) A comparative view of metabolite and substrate stress and tolerance in microbial bioprocessing: From biofuels and chemicals, to biocatalysis and bioremediation. *Metab. Eng.* **12**, 307–331

104. Agrawal, M., Mao, Z., and Chen, R. R. (2011) Adaptation yields a highly efficient xylose-fermenting Zymomonas mobilis strain. *Biotechnol. Bioeng.* **108**, 777–785

105. Cooper, T. F., and Lenski, R. E. (2010) Experimental evolution with E. coli in diverse resource environments. I. Fluctuating environments promote divergence of replicate populations. *BMC Evol. Biol.* **10**, 11

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

107. Paquin, C., and Adams, J. (1983) Frequency of fixation of adaptive mutations is higher in evolving diploid than haploid yeast populations. *Nature* **302**, 495–500

108. Paquin, C. E., and Adams, J. (1983) Relative fitness can decrease in evolving asexual populations of S. cerevisiae. *Nature* **306**, 368–371

109. Desai, M. M., and Fisher, D. S. (2007) Beneficial mutation selection balance and the effect of linkage on positive selection. *Genetics* **176**, 1759–1798

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

111. Gerrish, P. J., and Lenski, R. E. (1998) The fate of competing beneficial mutations in an asexual population. *Genetica* **102-103**, 127–144

112. Kim, Y. (2005) Adaptation in sexuals vs. asexuals: clonal interference and the Fisher-Muller model. *Genetics* **171**, 1377–1386

113. Notley-McRobb, L., and Ferenci, T. (2000) Experimental analysis of molecular events during mutational periodic selections in bacterial evolution. *Genetics* **156**, 1493–1501

114. Haldimann, A., and Wanner, B. L. (2001) Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. *J. Bacteriol.* **183**, 6384–6393

115. Kao, K. C. (2005) A global regulatory role of gluconeogenic genes in Escherichia coli revealed by transcriptome network analysis. *Journal of Biological Chemistry* **280**, 36079–36087

116. Bartlett, M. S., and Gourse, R. L. (1994) Growth rate-dependent control of the rrnB P1 core promoter in Escherichia coli. *J. Bacteriol.* **176**, 5560–5564

117. Rao, L., Ross, W., Appleman, J. A., Gaal, T., Leirmo, S., Schlax, P. J., Record, M. T., and Gourse, R. L. (1994). Factor independent activation of rrnB P1. An "extended" promoter with an upstream element that dramatically increases promoter strength. *Journal of Molecular Biology* **235**, 1421–1435

118. Quackenbush, J. (2002) Microarray data normalization and transformation. *Nat. Genet.* **32**, 496–501

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

120. Saeed, A. I., Sharov, V., White, J., Li, J., and Liang, W. (2003) TM4: a free, open-source system for microarray data management and analysis. *Biotechniques* **34**, 374–378

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

122. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* **2**, 2006.0008

123. Ferea, T. L., Botstein, D., Brown, P. O., and Rosenzweig, R. F. (1999) Systematic changes in gene expression patterns following adaptive evolution in yeast. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 9721–9726

124. Gavin, J. J., and Umbreit, W. W. (1965) Effect of biotin on fatty acid distribution in Escherichia coli. *J. Bacteriol.* **89**, 437–443

125. Yao, Z., and Valvano, M. A. (1994) Genetic analysis of the O-specific lipopolysaccharide biosynthesis region (rfb) of Escherichia coli K-12 W3110: identification of genes that confer group 6 specificity to Shigella flexneri serotypes Y and 4a. *J. Bacteriol.* **176**, 4133–4143

126. Aono, R., and Kobayashi, H. (1997) Cell surface properties of organic solvent-tolerant mutants of Escherichia coli K-12. *Appl. Environ. Microbiol.* **63**, 3637–3642

127. Taylor, F. R., and Cronan, J. E. (1979) Cyclopropane fatty acid synthase of Escherichia coli. Stabilization, purification, and interaction with phospholipid vesicles. *Biochemistry* **18**, 3292–3300

128. Grisham, C. M., and Barnett, R. E. (1973) The effects of long-chain alcohols on membrane lipids and the (Na++K+)-ATPase. *Biochim. Biophys. Acta* **311**, 417–422

129. Hui, F. K., and Barton, P. G. (1973) Mesomorphic behaviour of some phospholip1ds with aliphatic alcohols and other non-ionic substances. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism* **296**, 510–517

130. Sikkema, J., de Bont, J. A., and Poolman, B. (1995) Mechanisms of membrane toxicity of hydrocarbons. *Microbiol. Rev.* **59**, 201–222

131. Neilands, J. B. (1995) Siderophores: structure and function of microbial iron transport compounds. *J. Biol. Chem.* **270**, 26723–26726

132. Bagg, A., and Neilands, J. B. (1987) Ferric uptake regulation protein acts as a repressor, employing iron(II) as a cofactor to bind the operator of an iron transport operon in Escherichia coli. *Biochemistry* **26**, 5471–5477

133. Liao, J. C., Boscolo, R., Yang, Y.-L., Tran, L. M., Sabatti, C., and Roychowdhury, V. P. (2003) Network component analysis: reconstruction of regulatory

signals in biological systems. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 15522–15527

134. Tran, L. M., Brynildsen, M. P., Kao, K. C., Suen, J. K., and Liao, J. C. (2005) gNCA: A framework for determining transcription factor activity based on transcriptome: identifiability and numerical implementation. *Metab. Eng.* **7**, 128–141

135. Hagiwara, D., Yamashino, T., and Mizuno, T. (2004) A Genome-wide view of the Escherichia coli BasS-BasR two-component system implicated in iron-responses. *Biosci. Biotechnol. Biochem.* **68**, 1758–1767

136. Gatzeva-Topalova, P. Z., May, A. P., and Sousa, M. C. (2005) Crystal structure and mechanism of the Escherichia coli ArnA (PmrI) transformylase domain. An enzyme for lipid A modification with 4-amino-4-deoxy-L-arabinose and polymyxin resistance. *Biochemistry* **44**, 5328–5338

137. Gulletta, E., Das, A., and Adhya, S. (1983) The pleiotropic ts15 mutation of E. coli is an IS1 insertion in the rho structural gene. *Genetics* **105**, 265–280

138. Sawers, R. G. (2005) Transcript analysis of Escherichia coli K-12 insertion element IS5. *FEMS Microbiol Lett* **244**, 397–401

139. Schnetz, K., and Rak, B. (1992) IS5: a mobile enhancer of transcription in Escherichia coli. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 1244–1248

140. Schoner, B., and Kahn, M. (1981) The nucleotide sequence of IS5 from Escherichia coli. *Gene* **14**, 165–174

141. Yang, X., and Ishiguro, E. E. (2003) Temperature-sensitive growth and decreased thermotolerance associated with relA mutations in Escherichia coli. *J. Bacteriol.* **185**, 5765–5771

142. Saedler, H., Reif, H. J., Hu, S., and Davidson, N. (1974) IS2, a genetic element for turn-off and turn-on of gene activity in E. coli. *Molec. Gen. Genet.* **132**, 265–289

143. Craven, M. G., and Friedman, D. I. (1991) Analysis of the Escherichia coli nusA10(Cs) allele: relating nucleotide changes to phenotypes. *J. Bacteriol.* **173**, 1485–1491

144. Platt, T. (1986) Transcription termination and the regulation of gene expression. *Annu. Rev. Biochem.* **55**, 339–372

145. Lee, J.-H., Lee, K.-L., Yeo, W.-S., Park, S.-J., and Roe, J.-H. (2009) SoxRS-mediated lipopolysaccharide modification enhances resistance against multiple

drugs in Escherichia coli. *J. Bacteriol.* **191**, 4441–4450

146. White, D. G., Goldman, J. D., Demple, B., and Levy, S. B. (1997) Role of the acrAB locus in organic solvent tolerance mediated by expression of marA, soxS, or robA in Escherichia coli. *J. Bacteriol.* **179**, 6122–6126

147. Hopkin, K. A., Papazian, M. A., and Steinman, H. M. (1992) Functional differences between manganese and iron superoxide dismutases in Escherichia coli K-12. *J. Biol. Chem.* **267**, 24253–24258

148. Dorel, C., Lejeune, P., and Rodrigue, A. (2006) The Cpx system of Escherichia coli, a strategic signaling pathway for confronting adverse conditions and for settling biofilm communities? *Res. Microbiol.* **157**, 306–314

149. Pogliano, J., Lynch, A. S., Belin, D., Lin, E. C., and Beckwith, J. (1997) Regulation of Escherichia coli cell envelope proteins involved in protein folding and degradation by the Cpx two-component system. *Genes & Development* **11**, 1169–1182

150. Giuliodori, A. M., Brandi, A., Giangrossi, M., Gualerzi, C. O., and Pon, C. L. (2007) Cold-stress-induced de novo expression of infC and role of IF3 in cold-shock translational bias. *RNA* **13**, 1355–1365

151. Ko, J.-H., Lee, S.-J., Cho, B., and Lee, Y. (2006) Differential promoter usage of infA in response to cold shock in Escherichia coli. *FEBS Lett.* **580**, 539–544

152. Hirakawa, H., Takumi-Kobayashi, A., Theisen, U., Hirata, T., Nishino, K., and Yamaguchi, A. (2008) AcrS/EnvR represses expression of the acrAB multidrug efflux genes in Escherichia coli. *J. Bacteriol.* **190**, 6276–6279

153. Prasad Maharjan, R., Yu, P.-L., Seeto, S., and Ferenci, T. (2005) The role of isocitrate lyase and the glyoxylate cycle in Escherichia coli growing under glucose limitation. *Res. Microbiol.* **156**, 178–183

154. Pomposiello, P. J., Bennik, M. H., and Demple, B. (2001) Genome-wide transcriptional profiling of the Escherichia coli responses to superoxide stress and sodium salicylate. *J. Bacteriol.* **183**, 3890–3902

155. Berezina, O. V., Zakharova, N. V., Brandt, A., Yarotsky, S. V., Schwarz, W. H., and Zverlov, V. V. (2010) Reconstructing the clostridial n-butanol metabolic pathway in Lactobacillus brevis. *Appl Microbiol Biotechnol* **87**, 635–646

156. Nielsen, D. R., Leonard, E., Yoon, S.-H., Tseng, H.-C., Yuan, C., and Prather, K. L. J. (2009) Engineering alternative butanol production platforms in heterologous bacteria. *Metab. Eng.* **11**, 262–273

157. Steen, E. J., Chan, R., Prasad, N., Myers, S., Petzold, C. J., Redding, A., Ouellet, M., and Keasling, J. D. (2008) Metabolic engineering of Saccharomyces cerevisiae for the production of n-butanol. *Microb. Cell Fact.* **7**, 36

158. Dürre, P. (1998) New insights and novel developments in clostridial acetone/butanol/isopropanol fermentation. *Appl Microbiol Biotechnol* **49**, 639–648

159. Zingaro, K. A., and Terry Papoutsakis, E. (2013) GroESL overexpression imparts Escherichia coli tolerance to i-, n-, and 2-butanol, 1,2,4-butanetriol and ethanol with complex and unpredictable patterns. *Metab. Eng.* **15**, 196–205

160. Tomas, C. A., Welker, N. E., and Papoutsakis, E. T. (2003) Overexpression of groESL in Clostridium acetobutylicum results in increased solvent production and tolerance, prolonged metabolism, and changes in the cell's transcriptional program. *Appl. Environ. Microbiol.* **69**, 4951–4965

161. Nagakubo, S., Nishino, K., Hirata, T., and Yamaguchi, A. (2002) The putative response regulator BaeR stimulates multidrug resistance of Escherichia coli via a novel multidrug exporter system, MdtABC. *J. Bacteriol.* **184**, 4161–4167

162. Rutherford, B. J., Dahl, R. H., Price, R. E., Szmidt, H. L., Benke, P. I., Mukhopadhyay, A., and Keasling, J. D. (2010) Functional genomic study of exogenous n-butanol stress in Escherichia coli. *Appl. Environ. Microbiol.* **76**, 1935–1945

163. Kieboom, J., and de Bont, J. (2001) Identification and molecular characterization of an efflux system involved in Pseudomonas putida S12 multidrug resistance. *Microbiology (Reading, Engl.)* **147**, 43–51

164. Takatsuka, Y., Chen, C., and Nikaido, H. (2010) Mechanism of recognition of compounds of diverse structures by the multidrug efflux pump AcrB of Escherichia coli. *Proceedings of the National Academy of Sciences* **107**, 6559–6565

165. Dunlop, M. J., Dossani, Z. Y., Szmidt, H. L., Chu, H. C., Lee, T. S., Keasling, J. D., Hadi, M. Z., and Mukhopadhyay, A. (2011) Engineering microbial biofuel tolerance and export using efflux pumps. *Mol Syst Biol* **7**, 487

166. Alsaker, K. V., Paredes, C., and Papoutsakis, E. T. (2010) Metabolite stress and tolerance in the production of biofuels and chemicals: gene-expression-based systems analysis of butanol, butyrate, and acetate stresses in the anaerobe Clostridium acetobutylicum. *Biotechnol. Bioeng.* **105**, 1131–1147

167. Purvis, J. E., Yomano, L. P., and Ingram, L. O. (2005) Enhanced trehalose production improves growth of Escherichia coli under osmotic stress. *Appl. Environ. Microbiol.* **71**, 3761–3769

168. Sun, Y., and Vanderpool, C. K. (2011) Regulation and function of Escherichia coli sugar efflux transporter A (SetA) during glucose-phosphate stress. *J. Bacteriol.* **193**, 143–153

169. Liu, J. Y., Miller, P. F., Willard, J., and Olson, E. R. (1999) Functional and biochemical characterization of Escherichia coli sugar efflux transporters. *J. Biol. Chem.* **274**, 22977–22984

170. Zhang, M., Mileykovskaya, E., and Dowhan, W. (2002) Gluing the respiratory chain together. Cardiolipin is required for supercomplex formation in the inner mitochondrial membrane. *J. Biol. Chem.* **277**, 43553–43556

171. Mileykovskaya, E., and Dowhan, W. (2009) Cardiolipin membrane domains in prokaryotes and eukaryotes. *Biochim. Biophys. Acta* **1788**, 2084–2091

172. Romantsov, T., Stalker, L., Culham, D. E., and Wood, J. M. (2008) Cardiolipin controls the osmotic stress response and the subcellular location of transporter ProP in Escherichia coli. *J. Biol. Chem.* **283**, 12314–12323

173. Haines, T. H. (2009) A new look at Cardiolipin. *Biochim. Biophys. Acta* **1788**, 1997–2002

174. Erez, E., Stjepanovic, G., Zelazny, A. M., Brugger, B., Sinning, I., and Bibi, E. (2010) Genetic evidence for functional interaction of the Escherichia coli signal recognition particle receptor with acidic lipids in vivo. *J. Biol. Chem.* **285**, 40508–40514

175. Ma, R., Zhang, Y., Hong, H., Lu, W., Lin, M., Chen, M., and Zhang, W. (2011) Improved osmotic tolerance and ethanol production of ethanologenic Escherichia coli by IrrE, a global regulator of radiation-resistance of Deinococcus radiodurans. *Curr Microbiol* **62**, 659–664

176. Balakumar, S., and Arasaratnam, V. (2012) Osmo-, thermo- and ethanol- tolerances of Saccharomyces cerevisiae S1. *Braz. J. Microbiol.* **43**, 157–166

177. Okolo, B. N., Moneke, A. N., Anyanwu, C. U., Ezeogu, L. I., and Aligwekwe, G. N. (2004) On the pH and Osmotic Stress Tolerance of High Ethanol Tolerant Palm Wine Saccharomyces Yeast Isolates. *Bio-Research* **2**, 1–7

178. Dragosits, M., Mozhayskiy, V., Quinones-Soto, S., Park, J., and Tagkopoulos, I. (2013) Evolutionary potential, cross-stress behavior and the genetic basis of acquired stress resistance in Escherichia coli. *Mol Syst Biol* **9**, 643

179. Kim, Y. H., Cho, K., Yun, S.-H., Kim, J. Y., Kwon, K.-H., Yoo, J. S., and Kim, S. I. (2006) Analysis of aromatic catabolic pathways in Pseudomonas putida KT

2440 using a combined proteomic approach: 2-DE/MS and cleavable isotope-coded affinity tag analysis. *Proteomics* **6**, 1301–1318

180. Santos, P. M., Benndorf, D., and Sá-Correia, I. (2004) Insights into Pseudomonas putida KT2440 response to phenol-induced stress by quantitative proteomics. *Proteomics* **4**, 2640–2652

181. Trautwein, K., Kühner, S., Wöhlbrand, L., Halder, T., Kuchta, K., Steinbüchel, A., and Rabus, R. (2008) Solvent stress response of the denitrifying bacterium "Aromatoleum aromaticum" strain EbN1. *Appl. Environ. Microbiol.* **74**, 2267–2274

182. Len, A. C. L., Harty, D. W. S., and Jacques, N. A. (2004) Stress-responsive proteins are upregulated in Streptococcus mutans during acid tolerance. *Microbiology (Reading, Engl.)* **150**, 1339–1351

183. Zingaro, K. A., and Papoutsakis, E. T. (2012) Toward a semisynthetic stress response system to engineer microbial solvent tolerance. *MBio* **3**

184. Lennen, R. M., Kruziki, M. A., Kumar, K., Zinkel, R. A., Burnum, K. E., Lipton, M. S., Hoover, S. W., Ranatunga, D. R., Wittkopp, T. M., Marner, W. D., and Pfleger, B. F. (2011) Membrane stresses induced by overproduction of free fatty acids in Escherichia coli. *Appl. Environ. Microbiol.* **77**, 8114–8128

185. Klein-Marcuschamer, D., Santos, C. N. S., Yu, H., and Stephanopoulos, G. (2009) Mutagenesis of the bacterial RNA polymerase alpha subunit for improvement of complex phenotypes. *Appl. Environ. Microbiol.* **75**, 2705–2711

186. Santos, C. N. S., and Stephanopoulos, G. (2008) Combinatorial engineering of microbes for optimizing cellular phenotype. *Current Opinion in Chemical Biology* **12**, 168–176

187. Kvitek, D. J., and Sherlock, G. (2011) Reciprocal sign epistasis between frequently experimentally evolved adaptive mutations causes a rugged fitness landscape. *PLoS Genet.* **7**, e1002056

188. Downing, M., Eaton, L. M., Graham, R. L., Langholtz, M. H., Perlack, R. D., Turhollow, A. F., Jr, Stokes, B., and Brandt, C. C. (2011) *U.S. billion-ton update: biomass supply for a bioenergy and bioproducts industry*, U.S. Department of Energy

189. Almeida, J., Modig, T., and Petersson, A. (2007) Increased tolerance and conversion of inhibitors in lignocellulosic hydrolysates by Saccharomyces cerevisiae. *J Chem Technol Biotechnol* **82**, 340–349

190. Mosier, N., Wyman, C., Dale, B., Elander, R., Lee, Y. Y., Holtzapple, M., and Ladisch, M. (2005) Features of promising technologies for pretreatment of lignocellulosic biomass. *Bioresour. Technol.* **96**, 673–686

191. Sun, Y., and Cheng, J. (2002) Hydrolysis of lignocellulosic materials for ethanol production: a review. *Bioresour. Technol.* **83**, 1–11

192. Chen, X., Shekiro, J., Franden, M. A., Wang, W., Zhang, M., Kuhn, E., Johnson, D. K., and Tucker, M. P. (2012) The impacts of deacetylation prior to dilute acid pretreatment on the bioethanol process. *Biotechnol Biofuels* **5**, 8

193. Liu, H., and Naismith, J. H. (2008) An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol.* **8**, 91

194. Ma, M., and Liu, Z. L. (2010) Comparative transcriptome profiling analyses during the lag phase uncover YAP1, PDR1, PDR3, RPN4, and HSF1 as key regulatory genes in genomic adaptation to the lignocellulose derived inhibitor HMF for Saccharomyces cerevisiae. *BMC Genomics* **11**, 660

195. Larsson, S., Cassland, P., and Jönsson, L. J. (2001) Development of a Saccharomyces cerevisiae strain with enhanced resistance to phenolic fermentation inhibitors in lignocellulose hydrolysates by heterologous expression of laccase. *Appl. Environ. Microbiol.* **67**, 1163–1170

196. Martin, C., Alriksson, B., Sjöde, A., Nilvebrant, N.-O., and Jönsson, L. J. (2007) Dilute sulfuric acid pretreatment of agricultural and agro-industrial residues for ethanol production. *Appl Biochem Biotechnol* **137-140**, 339–352

197. Casal, M., Cardoso, H., and Leao, C. (1996) Mechanisms regulating the transport of acetic acid in Saccharomyces cerevisiae. *Microbiology* **142**, 1385–1390

198. Ludovico, P., Rodrigues, F., Almeida, A., Silva, M. T., Barrientos, A., and Côrte-Real, M. (2002) Cytochrome c release and mitochondria involvement in programmed cell death induced by acetic acid in Saccharomyces cerevisiae. *Mol. Biol. Cell* **13**, 2598–2606

199. Liu, Z. L., Slininger, P. J., Dien, B. S., Berhow, M. A., Kurtzman, C. P., and Gorsich, S. W. (2004) Adaptive response of yeasts to furfural and 5-hydroxymethylfurfural and new chemical evidence for HMF conversion to 2,5-bis-hydroxymethylfuran. *J Ind Microbiol Biotechnol* **31**, 345–352

200. Tobias Modig, G. L. M. J. T. (2002) Inhibition effects of furfural on alcohol dehydrogenase, aldehyde dehydrogenase and pyruvate dehydrogenase. *Biochemi-*

cal Journal **363**, 769

201. Sanchez, B., and Bautista, J. (1988) Effects of furfural and 5-hydroxymethylfurfural on the fermentation of Saccharomyces cerevisiae and biomass production from Candida guilliermondii. *Enzyme Microb. Technol.* **10**, 315–318

202. Liu, Z. L., Moon, J., Andersh, B. J., Slininger, P. J., and Weber, S. (2008) Multiple gene-mediated NAD(P)H-dependent aldehyde reduction is a mechanism of in situ detoxification of furfural and 5-hydroxymethylfurfural by Saccharomyces cerevisiae. *Appl Microbiol Biotechnol* **81**, 743–753

203. Mira, N. P., Teixeira, M. C., and Sá-Correia, I. (2010) Adaptive response and tolerance to weak acids in Saccharomyces cerevisiae: a genome-wide view. *OMICS: A Journal of Integrative Biology* **14**, 525–540

204. Zhang, J.-G., Liu, X.-Y., He, X.-P., Guo, X.-N., Lu, Y., and Zhang, B.-R. (2011) Improvement of acetic acid tolerance and fermentation performance of Saccharomyces cerevisiae by disruption of the FPS1 aquaglyceroporin gene. *Biotechnol Lett* **33**, 277–284

205. Saeed, A. I., Bhagabati, N. K., Braisted, J. C., Liang, W., Sharov, V., Howe, E. A., Li, J., Thiagarajan, M., White, J. A., and Quackenbush, J. (2006) TM4 microarray software suite. *Meth. Enzymol.* **411**, 134–193

206. Cherry, J. M., Hong, E. L., Amundsen, C., Balakrishnan, R., Binkley, G., Chan, E. T., Christie, K. R., Costanzo, M. C., Dwight, S. S., Engel, S. R., Fisk, D. G., Hirschman, J. E., Hitz, B. C., Karra, K., Krieger, C. J., Miyasato, S. R., Nash, R. S., Park, J., Skrzypek, M. S., Simison, M., Weng, S., and Wong, E. D. (2011) Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Research* **40**, D700–D705

207. Teixeira, M. C., Monteiro, P., Jain, P., Tenreiro, S., Fernandes, A. R., Mira, N. P., Alenquer, M., Freitas, A. T., Oliveira, A. L., and Sá-Correia, I. (2006) The YEASTRACT database: a tool for the analysis of transcription regulatory associations in Saccharomyces cerevisiae. *Nucleic Acids Research* **34**, D446–51

208. Russell, J. B. (1992) Another explanation for the toxicity of fermentation acids at low pH: anion accumulation versus uncoupling. *J Appl Microbiol* **73**, 363–370

209. Palmqvist, E., and Hahn-Hägerdal, B. (2000) Fermentation of lignocellulosic hydrolysates. II: inhibitors and mechanisms of inhibition. *Bioresour. Technol.* **74**, 25–33

210. Kolaczkowska, A., and Goffeau, A. (1999) Regulation of pleiotropic drug resistance in yeast. *Drug Resistance Updates* **2**, 403–414

211. Armstrong, G. A., and Hearst, J. E. (1996) Carotenoids 2: Genetics and molecular biology of carotenoid pigment biosynthesis. *FASEB J.* **10**, 228–237

212. Edge, R., McGarvey, D. J., and Truscott, T. G. (1997) The carotenoids as antioxidants--a review. *J. Photochem. Photobiol. B, Biol.* **41**, 189–200

213. Vachali, P., Bhosale, P., and Bernstein, P. S. (2012) Microbial carotenoids. *Methods Mol. Biol.* **898**, 41–59

214. Verwaal, R., Wang, J., Meijnen, J.-P., Visser, H., Sandmann, G., van den Berg, J. A., and van Ooyen, A. J. J. (2007) High-level production of beta-carotene in Saccharomyces cerevisiae by successive transformation with carotenogenic genes from Xanthophyllomyces dendrorhous. *Appl. Environ. Microbiol.* **73**, 4342–4350

215. Nishizaki, T., Tsuge, K., Itaya, M., Doi, N., and Yanagawa, H. (2007) Metabolic engineering of carotenoid biosynthesis in Escherichia coli by ordered gene assembly in Bacillus subtilis. *Appl. Environ. Microbiol.* **73**, 1355–1361

216. Verdoes, J. C., Sandmann, G., Visser, H., Diaz, M., van Mossel, M., and van Ooyen, A. J. J. (2003) Metabolic engineering of the carotenoid biosynthetic pathway in the yeast Xanthophyllomyces dendrorhous (Phaffia rhodozyma). *Appl. Environ. Microbiol.* **69**, 3728–3738

217. Paiva, S. A., and Russell, R. M. (1999) Beta-carotene and other carotenoids as antioxidants. *J Am Coll Nutr* **18**, 426–433

218. Verwaal, R., Jiang, Y., Wang, J., Daran, J.-M., Sandmann, G., van den Berg, J. A., and van Ooyen, A. J. J. (2010) Heterologous carotenoid production in Saccharomyces cerevisiae induces the pleiotropic drug resistance stress response. *Yeast* **27**, 983–998

219. D Gietz, A. S. J. R. A. W. R. H. S. (1992) Improved method for high efficiency transformation of intact yeast cells. *Nucleic Acids Research* **20**, 1425

220. Bähler, J., Wu, J. Q., Longtine, M. S., Shah, N. G., McKenzie, A., Steever, A. B., Wach, A., Philippsen, P., and Pringle, J. R. (1998) Heterologous modules for efficient and versatile PCR-based gene targeting in Schizosaccharomyces pombe. *Yeast* **14**, 943–951

221. Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat*

Protoc **4**, 44–57

222. Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* **37**, 1–13

223. Antonovics, J., Ellstrand, N. C., and Brandon, R. N. (1988) in *Plant evolutionary biology* pp. 275–303, Springer Netherlands, Dordrecht

224. Yan, G.-L., Liang, H.-Y., Wang, Z.-Q., Yang, X.-F., Liu, D., Liu, J.-F., and Duan, C.-Q. (2011) Important role of catalase in the production of β-carotene by recombinant Saccharomyces cerevisiae under H2O2 stress. *Curr Microbiol* **62**, 1056–1061

225. Olzhausen, J., Schübbe, S., and Schüller, H.-J. (2009) Genetic analysis of coenzyme A biosynthesis in the yeast Saccharomyces cerevisiae: identification of a conditional mutation in the pantothenate kinase gene CAB1. *Curr. Genet.* **55**, 163–173

226. Flury, I., Garza, R., Shearer, A., Rosen, J., Cronin, S., and Hampton, R. Y. (2005) INSIG: a broadly conserved transmembrane chaperone for sterol-sensing domain proteins. *EMBO J.* **24**, 3917–3926

227. Lamb, D. C., Kelly, D. E., Manning, N. J., Kaderbhai, M. A., and Kelly, S. L. (1999) Biodiversity of the P450 catalytic cycle: yeast cytochrome b5/NADH cytochrome b5 reductase complex efficiently drives the entire sterol 14-demethylation (CYP51) reaction. *FEBS Lett.* **462**, 283–288

228. Yan, G.-L., Wen, K.-R., and Duan, C.-Q. (2012) Enhancement of β-carotene production by over-expression of HMG-CoA reductase coupled with addition of ergosterol biosynthesis inhibitors in recombinant Saccharomyces cerevisiae. *Curr Microbiol* **64**, 159–163

229. Rogers, K. M., Pierson, C. A., Culbertson, N. T., Mo, C., Sturm, A. M., Eckstein, J., Barbuch, R., Lees, N. D., and Bard, M. (2004) Disruption of the Candida albicans CYB5 gene results in increased azole sensitivity. *Antimicrob. Agents Chemother.* **48**, 3425–3435

230. Parks, L. W., Smith, S. J., and Crowley, J. H. (1995) Biochemical and physiological effects of sterol alterations in yeast--a review. *Lipids* **30**, 227–230

231. Ozaydın, B., Burd, H., Lee, T. S., and Keasling, J. D. (2013) Carotenoid-based phenotypic screen of the yeast deletion collection reveals new genes with roles in isoprenoid production. *Metab. Eng.* **15**, 174–183

232. Izawa, S., Maeda, K., Miki, T., Mano, J., Inoue, Y., and Kimura, A. (1998) Importance of glucose-6-phosphate dehydrogenase in the adaptive response to hydrogen peroxide in Saccharomyces cerevisiae. *Biochem. J.* **330 (Pt 2)**, 811–817

233. Rep, M., Proft, M., Remize, F., Tamás, M., Serrano, R., Thevelein, J. M., and Hohmann, S. (2001) The Saccharomyces cerevisiae Sko1p transcription factor mediates HOG pathway-dependent osmotic regulation of a set of genes encoding enzymes implicated in protection from oxidative damage. *Molecular Microbiology* **40**, 1067–1083

234. Pahlman, A. K., Granath, K., Ansell, R., Hohmann, S., and Adler, L. (2001) The yeast glycerol 3-phosphatases Gpp1p and Gpp2p are required for glycerol biosynthesis and differentially involved in the cellular responses to osmotic, anaerobic, and oxidative stress. *J. Biol. Chem.* **276**, 3555–3563

235. Purdue, P. E., and Lazarow, P. B. (2001) Peroxisome biogenesis. *Annu. Rev. Cell Dev. Biol.* **17**, 701–752

236. Marova, I., Certik, M., and Breierova, E. (2011) Production of enriched biomass by carotenogenic yeasts–Application of whole-cell yeast biomass to production of pigments and other lipid compounds. *Remote Sensing of Biomass: Principles and Applications/Book* **4**, 345–384

237. Lin, Y. L., and Blaschek, H. P. (1983) Butanol production by a butanol-tolerant strain of Clostridium acetobutylicum in extruded corn broth. *Appl. Environ. Microbiol.* **45**, 966–973

238. Beesch, S. C. (1953) Acetone-butanol fermentation of starches. *Appl Microbiol* **1**, 85–95

239. Beesch, S. C. (1952) Acetone-butanol fermentation of sugars. *Ind. Eng. Chem.* **44**, 1677–1682

240. Bowles, L. K., and Ellefson, W. L. (1985) Effects of butanol on Clostridium acetobutylicum. *Appl. Environ. Microbiol.* **50**, 1165–1170

241. Rühl, J., Schmid, A., and Blank, L. M. (2009) Selected Pseudomonas putida strains able to grow in the presence of high butanol concentrations. *Appl. Environ. Microbiol.* **75**, 4653–4656

242. Sardessai, Y., and Bhosle, S. (2002) Tolerance of bacteria to organic solvents. *Res. Microbiol.* **153**, 263–268

243. Weber, F. J., and de Bont, J. A. (1996) Adaptation mechanisms of microorganisms to the toxic effects of organic solvents on membranes. *Biochim. Biophys.*

Acta **1286**, 225–245

244. Pinkart, H. C., Wolfram, J. W., Rogers, R., and White, D. C. (1996) Cell envelope changes in solvent-tolerant and solvent-sensitive Pseudomonas putida Strains following exposure to o-xylene. *Appl. Environ. Microbiol.* **62**, 1129–1132

245. Duque, E. (1997) Mechanisms for solvent tolerance in bacteria. *Journal of Biological Chemistry* **272**, 3887–3890

246. Ramos, J. L., Duque, E., Gallegos, M.-T., Godoy, P., Ramos-Gonzalez, M. I., Rojas, A., Teran, W., and Segura, A. (2002) Mechanisms of solvent tolerance in gram-negative bacteria. *Annu. Rev. Microbiol.* **56**, 743–768

247. Tomas, C. A., Beamish, J., and Papoutsakis, E. T. (2004) Transcriptional analysis of butanol stress and tolerance in Clostridium acetobutylicum. *J. Bacteriol.* **186**, 2006–2018

248. Brynildsen, M. P., and Liao, J. C. (2009) An integrated network approach identifies the isobutanol response network of Escherichia coli. *Mol Syst Biol* **5**, 277

249. Ben-Dor, A., Shamir, R., and Yakhini, Z. (1999) Clustering gene expression patterns. *Journal of Computational Biology* **6**, 281–297

250. Heipieper, H. J., Weber, F. J., Sikkema, J., Keweloh, H., and de Bont, J. A. (1994) Mechanisms of resistance of whole cells to toxic organic solvents. *Trends in Biotechnology* **12**, 409–415

251. Isken, S., and de Bont, J. A. (1998) Bacteria tolerant to organic solvents. *Extremophiles* **2**, 229–238

252. Nielsen, L. E., Kadavy, D. R., Rajagopal, S., Drijber, R., and Nickerson, K. W. (2005) Survey of extreme solvent tolerance in gram-positive cocci: membrane fatty acid changes in Staphylococcus haemolyticus grown in toluene. *Appl. Environ. Microbiol.* **71**, 5171–5176

253. Tsukagoshi, N., and Aono, R. (2000) Entry into and release of solvents by Escherichia coli in an organic-aqueous two-liquid-phase system and substrate specificity of the AcrAB-TolC solvent-extruding pump. *J. Bacteriol.* **182**, 4803–4810

254. Dupont, M., James, C. E., Chevalier, J., and Pages, J. M. (2007) An early response to environmental stress involves regulation of OmpX and OmpF, two enterobacterial outer membrane pore-forming proteins. *Antimicrob. Agents Chemother.* **51**, 3190–3198

255. Lewis, C., Skovierova, H., Rowley, G., Rezuchova, B., Homerova, D., Stevenson, A., Sherry, A., Kormanec, J., and Roberts, M. (2008) Small outer-membrane lipoprotein, SmpA, is regulated by sigmaE and has a role in cell envelope integrity and virulence of Salmonella enterica serovar Typhimurium. *Microbiology (Reading, Engl.)* **154**, 979–988

256. Qin, Z. S., McCue, L. A., Thompson, W., Mayerhofer, L., Lawrence, C. E., and Liu, J. S. (2003) Identification of co-regulated genes through Bayesian clustering of predicted regulatory binding sites. *Nat. Biotechnol.* **21**, 435–439

257. Shimizu, K., Hayashi, S., Doukyu, N., Kobayashi, T., and Honda, H. (2005) Time-course data analysis of gene expression profiles reveals purR regulon concerns in organic solvent tolerance in Escherichia coli. *Journal of Bioscience and Bioengineering* **99**, 72–74

258. Demple, B. (1996) Redox signaling and gene control in the Escherichia coli soxRS oxidative stress regulon--a review. *Gene* **179**, 53–57

259. Semchyshyn, H., Bagnyukova, T., and Lushchak, V. (2005) Involvement of soxRS regulon in response of Escherichia coli to oxidative stress induced by hydrogen peroxide. *Biochemistry Mosc.* **70**, 1238–1244

260. Zheng, Q., and Wang, X.-J. (2008) GOEAST: a web-based software toolkit for gene ontology enrichment analysis. *Nucleic Acids Research* **36**, W358–63

261. Weber, F. J., and de Bont, J. A. (1996) Adaptation mechanisms of microorganisms to the toxic effects of organic solvents on membranes. *Biochim. Biophys. Acta* **1286**, 225–245

262. Inui, M., Suda, M., Kimura, S., Yasuda, K., Suzuki, H., Toda, H., Yamamoto, S., Okino, S., Suzuki, N., and Yukawa, H. (2008) Expression of Clostridium acetobutylicum butanol synthetic genes in Escherichia coli. *Appl Microbiol Biotechnol* **77**, 1305–1316

263. Borges-Walmsley, M. I., McKeegan, K. S., and Walmsley, A. R. (2003) Structure and function of efflux pumps that confer resistance to drugs. *Biochem. J.* **376**, 313–338

264. Ron, E. Z., Alajem, S., Biran, D., and Grossman, N. (1990) Adaptation of Escherichia coli to elevated temperatures: The metA gene product is a heat shock protein. *Antonie van Leeuwenhoek* **58**, 169–174

265. Sugai, R., Shimizu, H., Nishiyama, K., and Tokuda, H. (2001) Overexpression of yccL (gnsA) and ydfY (gnsB) increases levels of unsaturated fatty acids and suppresses both the temperature-sensitive fabA6 mutation and cold-sensitive

secG null mutation of Escherichia coli. *J. Bacteriol.* **183**, 5523–5528

266. White, C. B., Chen, Q., Kenyon, G. L., and Babbitt, P. C. (1995) A novel activity of OmpT. *Journal of Biological Chemistry* **270**, 12990–12994

267. Suppmann, B., and Sawers, G. (1994) Isolation and characterization of hypophosphite--resistant mutants of Escherichia coli: identification of the FocA protein, encoded by the pfl operon, as a putative formate transporter. *Molecular Microbiology* **11**, 965–982

268. Burton, K. (1994) Adenine transport in Escherichia coli. *Proceedings of the Royal Society B: Biological Sciences* **255**, 153–157

269. Burton, K. (1977) Transport of adenine, hypoxanthine and uracil into Escherichia coli. *Biochemical Journal* **168**, 195

270. Surette, M. G., Miller, M. B., and Bassler, B. L. (1999) Quorum sensing in Escherichia coli, Salmonella typhimurium, and Vibrio harveyi: a new family of genes responsible for autoinducer production. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 1639–1644

271. Fajardo, D. A., Cheung, J., Ito, C., Sugawara, E., Nikaido, H., and Misra, R. (1998) Biochemistry and regulation of a novel Escherichia coli K-12 porin protein, OmpG, which produces unusually large channels. *J. Bacteriol.* **180**, 4452–4459

272. Datsenko, K. A., and Wanner, B. L. (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 6640–6645

273. Rowbury, R. J., Humphrey, T. J., and Goodson, M. (1999) Properties of an L-glutamate-induced acid tolerance response which involves the functioning of extracellular induction components. *J Appl Microbiol* **86**, 325–330

274. Ma, Z., Richard, H., Tucker, D. L., Conway, T., and Foster, J. W. (2002) Collaborative regulation of Escherichia coli glutamate-dependent acid resistance by two AraC-like regulators, GadX and GadW (YhiW). *J. Bacteriol.* **184**, 7001–7012

275. Yoshimura, M., Oshima, T., and Ogasawara, N. (2007) Involvement of the YneS/YgiH and PlsX proteins in phospholipid biosynthesis in both Bacillus subtilis and Escherichia coli. *BMC Microbiol.* **7**, 69

276. Conrad, T. M., Joyce, A. R., Applebee, M. K., Barrett, C. L., Xie, B., Gao, Y., and Palsson, B. Ø. (2009) Whole-genome resequencing of Escherichia coli K-12

MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations. *Genome Biol* **10**, R118