ADAPTIVE LABORATORY EVOLUTION OF *ACINETOBACTER BAYLYI* FOR IMPROVED GROWTH ON GUAIACOL

by

ALAA ASHRAF MAHMOUD AHMED

(Under the Direction of Ellen L. Neidle)

ABSTRACT

During adaptive laboratory evolution, an organism is evolved for several hundred to several thousand generations. Usually, these studies yield multiple mutations in any selected isolate. Mutations that occur in uncharacterized genes, or those with no clear role in the adaptive process may be challenging to investigate yet may provide the most novel information. In this dissertation, I present my analysis of an evolved ADP1-derived strain selected for growth on guaiacol, a lignin-derived aromatic compound. Guaiacol is abundant in many types of lignin streams. Improving the bioconversion of guaiacol represents an important step in the overall goal of producing valuable products from lignin. My analysis revealed an unexpected role of a regulatory protein in the adaptation to growth on guaiacol. Ribosomal silencing factor proteins (RsfA), are known to inhibit protein synthesis during stationary phase and in nutrient-poor conditions. Through modified competition assays, I show that lack of RsfA provides cells with a competitive growth advantage during the transition from stationary phase to fresh medium and only when guaiacol is the carbon source. This advantage is presumably due to derepressed protein synthesis in mutant cells relative to the wild-type cells, enabling the former to remain in a state of continued protein synthesis. RsfA has not been previously characterized in ADP1 and

previous reports in *E. coli* have shown that *rsfA* mutants are always outcompeted by wild-type cells. To our knowledge, this is the first report of conditions under which lack of RsfA is beneficial. RsfA is involved in the conserved process of ribosome hibernation. The results presented here may inform the design of future laboratory evolution experiments. During batch cultures, serial transfers may be done after cells enter stationary phase, and strains that lack RsfA at the onset of evolution may be advantageous. In addition, other previously uncharacterized genes are suggested as targets for future investigative efforts. Finally, preliminary results are presented on development of genetic methods and suggested modified culture conditions for laboratory evolution. Overall, my results have implications that extend beyond the metabolism of lignin-derived aromatic compounds to broader areas such as experimental design and global regulation.

INDEX WORDS: ADP1, *Acinetobacter baylyi*, laboratory evolution, guaiacol, lignin valorization, RsfA, ribosomal silencing factor, ribosome hibernation, adaptive mutations

ADAPTIVE LABORATORY EVOLUTION OF *ACINETOBACTER BAYLYI* FOR IMPROVED GROWTH ON GUAIACOL

by

ALAA ASHRAF MAHMOUD AHMED

Bachelor of Pharmaceutical Sciences, Cairo University, Egypt, 2006

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2019

© 2019

Alaa Ashraf Mahmoud Ahmed

All Rights Reserved

ADAPTIVE LABORATORY EVOLUTION OF *ACINETOBACTER BAYLYI* FOR IMPROVED GROWTH ON GUAIACOL

by

ALAA ASHRAF MAHMOUD AHMED

Major Professor:	Ellen L. Neidle
Committee:	Mark A. Eiteman
	Anna C. Karls
	Jan Mrázek

Electronic Version Approved:

Ron Walcott
Interim Dean of the Graduate School
The University of Georgia
December 2019

DEDICATION

I dedicate my dissertation to my grandma who passed away before my graduation. She was a big inspiration for me in life. She was very supportive, caring, and loving. I miss our lovely conversations and I miss you so much. This is for you, mama.

ACKNOWLEDGEMENTS

It takes a village. While a PhD is a personal journey, I wouldn't have made it without the support and advice of many people.

- Ellen Neidle, my PI for continuous support, guidance, and for making me a better scientist.
- My advisory committee: Dr. Eiteman, Dr. Karls, and Dr. Mrázek for always giving me honest and helpful feedback. And for giving me a "pass" in my oral exam, even when I thought I did a terrible job.
- My wife and daughter who had to travel thousands of miles to allow me to follow my path. I have no words to express how much I love you both and appreciate all your sacrifices.
- My parents who supported me in every possible way. Without you, I wouldn't have accomplished anything.
- My brother who's had a great influence on the person I grew up to be.
- Dr. Melissa Tumen-Velasquez, the former postdoc of the Neidle Lab.
- Members of the Neidle lab, current and former: Stacy Bedore, Emily McIntyre, Chantel Duscent-Maitland, Alyssa Baugh, Walker Whitley, Chris Moxley, Lauren Slarks, and Alicia Schmidt.
- The microbiology office staff, for putting up with my repeated questions and for helping with everything.
- The International Student Life department and Office of Global engagement for being an integral part of international students' experience at UGA, and for recognizing me with multiple awards.

- The Neidle lab collaborators: Dr. John McGeehan (Portsmouth) for being an amazing scientist and for generously hosting me for a three-week visit to his lab. Dr. Gregg Beckham (NREL) for great collaboration and bringing together a lot of amazing scientists.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Purpose of the study

This dissertation project focuses on an accelerated experimental evolution system that utilizes the genetically tractable soil bacterium, *Acinetobacter baylyi* ADP1 [1, 2]. This microbe possesses several traits that make it ideal for metabolic engineering and evolutionary studies. Features that are important for this project include ADP1's natural ability to metabolize a wide array of aromatic compounds, its natural transformation and recombination system, and the relative ease of chromosomal engineering. I sought to expand the aromatic compound degrading capabilities of ADP1 for the purpose of channeling lignin-derived compounds into central metabolism. This goal is motivated by a desire to use lignin as a renewable biomass feedstock to make valuable compounds, a process termed lignin valorization. The complexity of the chemical structure of lignin is currently hindering the economic feasibility of converting lignocellulosic plant material to biofuels and other commercially and industrially desirable chemicals. Initial pretreatment of lignin yields heterogenous mixtures rich in aromatic compounds, such as guaiacol.

Using guaiacol as a model compound for engineering improved lignin biodegradation, my first experiments were conducted to construct and evolve ADP1-derived strains that can use this compound as the sole carbon source (thereby generating a Gua^+ phenotype). To enable such consumption, foreign *gcoAB* genes from *Amycolatopsis* sp. ATCC 39116 were inserted in the ADP1 chromosome. These genes encode a recently characterized two-component cytochrome

P450 enzyme that demethylates guaiacol to produce catechol [3], a natural growth substrate for ADP1. During evolutionary studies, mutants emerged that acquired genetic changes both in the *gcoAB* genes and elsewhere in the chromosome [4]. Analysis of these mutations led to the discovery of novel features of post-transcriptional regulation, protein function, and genetic processes. In addition, new analytical tools were developed to assess the significance of multiple genomic mutations that arise during laboratory evolution. Specifically, I developed methods for testing the effects of combinations of mutations in an easier and more time-efficient way than traditional methods by utilizing the natural transformation and recombination system of ADP1.

As background for this project, this introductory chapter provides information on the genetic system of ADP1, the aromatic compound degrading capabilities of this bacterium, and a recently developed method designated Evolution by Amplification and Synthetic Biology (EASy) [4]. Information is also presented on guaiacol in the context of lignin structure and the importance of demethylation as a key step to funnel more compounds from lignin streams into central metabolism for the ultimate goal of lignin valorization. Additionally, some background information is provided on bacterial stress response, ribosome hibernation, and some other genes and proteins that pertain to key findings from the evolution experiments.

ADP1: An ideal organism for biotechnology applications

Acinetobacter baylyi ADP1 is a Gram-negative, strictly aerobic, non-motile, and non-pathogenic soil bacterium. This organism was previously named *Acinetobacter calcoaceticus* before it received its current species designation [2]. The parent strain of ADP1 was originally isolated in the 1960s as an organism that can grow on butane diol as a carbon source. Genetic transformation was observed when encapsulated cells were grown with unencapsulated mutants, resulting in the latter gaining the ability to form stable capsules [1]. Using UV irradiation, an unencapsulated mutant was obtained and named BD413UE [1]. This strain was later designated ADP1 as the parent strain in physiological studies conducted during decades of research in the laboratory of Prof. L. Nicholas Ornston (Yale University). The discovery of natural transformation in ADP1 allowed its use in studies on genetic transfer [5]. In addition to natural transformation, ADP1 boasts an array of features that make it an ideal and desirable platform organism for genetic engineering and evolutionary studies. These features include the availability of whole genome sequence data [6], a refined genome-scale metabolic model [7], a single knockout collection [8], simple growth requirements, metabolic versatility [9], and the ease of chromosomal manipulation afforded by its exceptionally efficient natural competence and recombination systems [10].

Natural transformation is one of the mechanisms by which bacteria exchange genetic material, which involves the active uptake of free DNA from the environment [11]. For this uptake to occur, a bacterial cell is required to be in a physiological state called “competence”. ADP1 becomes competent for natural transformation upon transfer of cells into fresh medium. The cells remain competent throughout the exponential growth phase and remain competent for a few hours in stationary phase [12]. Early studies showed that in ADP1, transformation is

independent of the carbon source or the type of the growth medium. However, divalent cations, such as Ca^{2+}, Mg^{2+}, and Mn^{2+} are required for DNA uptake in ADP1. In addition, an increase in transformation frequencies is seen with increasing both incubation time and DNA concentrations, until saturation [12]. After competence is induced, three steps comprise the process of natural transformation: DNA binding, DNA translocation, and recombination. Components of the natural transformation system in ADP1 have been characterized and have been reviewed in [13]. The sixteen proteins involved in competence in ADP1 are categorized into either group I, which are DNA translocator-specific proteins, or group II, which are type IV pili related proteins. The genes encoding competence proteins are organized in seven chromosomal loci. Among those components, some proteins appear to be essential for transformation. For example, *comA* and *comP* mutants were found to be defective in competence, while other mutants, such as *comE* and *comF* mutants display reduced transformation frequencies.

Potential benefits of transformation may include the use of DNA as a nutrient source, a role in DNA repair, or an advantage for adaptation and evolution. In a study that looked at the effects of genetic exchange on laboratory evolution, and vice versa, it was found that both competent and non-competent lineages evolved equivalently and all adapted lineages displayed diminished transformability regardless of competence or mutation frequency [14]. In addition, the same study confirmed earlier results that showed that the growth rate of ADP1 was impaired in the presence of free DNA in the growth medium, therefore providing evidence against competence being beneficial for nutritional purposes [15].

The genome of ADP1 is 3.6 Mb in length and has a G+C content of approximately 40% [6]. One characteristic feature of its genome is the presence of five genetic islands of catabolic

diversity. Because ADP1 is a soil bacterium that degrades a wide variety of aromatic compounds of plant source, a large portion of the ADP1 genome is dedicated to degradation pathways. In these genetic islands of catabolic diversity, groups of genes that encode related metabolic functions are clustered together. For example, the catabolic island II harbors the *sal-are-ben-cat* gene cluster which encodes enzymes involved in the conversion of different aromatic substrates to catechol. On the other hand, catabolic island IV harbors the *dca-pca-qui-pob-hca* gene set which encodes enzymes that convert substrates to protocatechuate. Two exceptions exist for this clustering: the *ant* genes for anthranilate metabolism which are distant from the functionally related island II, and the *van* genes which encode enzymes related to vanillate conversion to protocatechuate and are distant from island IV to which they would be expected to cluster, based on the metabolic functions they encode.

Another feature of ADP1's genome relevant to its transformability and evolvability, is the presence of mobile elements. These elements include six copies of an insertion element of the IS*3* family, IS*1236* [16]. It is worth mentioning that one copy, IS*1236_6*, does not seem to be functional due to a mutation affecting its catalytic triad [17]. Among the other IS copies, two are in close proximity to each other, giving rise to the Tn*5163* transposon. These two copies contribute to the instability of this genomic region, which also harbors the *vanAB* genes [18]. In a study on gene amplification events in ADP1, the majority of duplications investigated were found to be insertion sequence mediated, when the genes under study were moved to be in close proximity to insertion sequences. [17]. In contrast, insertion sequences only accounted for 2% of duplication events when the same genes were at their native locus [19]. Another study on genome instability of ADP1 during laboratory evolution showed that IS*1236* accounted for a large fraction of mutations among all evolved populations [20]. In the same study, IS*1236* was

found to mediate loss of competence, by insertions in competence-associated genes. Removal of all six copies of this mobile element allowed the generation of an ADP1-derived strain with reduced mutation rates and improved transformability. The reduced mutations rates might be a desirable trait in biotechnological applications, where the genetic stability of a strain is required for continued performance. However, in a different context, such as some evolutionary studies, the genetic diversity generated via IS-mediated events might be desirable.

In addition to insertion sequences, two prophage regions contribute to genome instability in ADP1. Bacteriophages are viruses that infect bacterial cells and are considered an important factor in the shape and evolution of bacterial communities [21]. Bacteriophages can be classified into two groups, based on their lifecycle. Lytic or virulent phages are those that cause lysis of their bacterial cell host after the phage replication is complete. Temperate phages, while capable of a lytic cycle, can also adopt a non-lytic lifecycle. This group of phages integrate their DNA into the bacterial chromosome of their host and stay dormant, thus becoming a prophage and replicating with the host chromosome. During a long-term laboratory evolution study, a filamentous phage was reactivated in the genome of ADP1, and was shown to be associated with reduced transformability [22]. In this study, ADP1 mutants lacking the genes required for competence were found to be resistant to infection by this phage. Sequencing data showed a 12-fold increase in coverage (number of reads of the region encoding phage proteins) of this region. Similarly, we observed an increase in the coverage of the same region in populations that have undergone long-term laboratory evolution. While the mechanism and significance of this increase in gene dosage has not been investigated, some of the isolates were defective in their competence and ability to be transformed with linear plasmids.

The ease of chromosomal manipulation in ADP1 makes this organism ideal for genetic engineering [23]. The design of strains with deletions, insertions and rearrangements can be achieved with relative ease compared to other model organisms. For example, for gene deletions, overlap extension PCR [24] can be used to generate DNA constructs with sequence identity to regions in the genome in which a selectable/counter-selectable marker, such as *sacB*-Km^R [25] replaces a gene targeted for deletion by allelic replacement (Figure 1.1) [26]. Transforming ADP1 recipients with such a fragment gives rise to transformants that can be selected by resistance to Km and growth inhibition or cell lysis on plates supplemented with sucrose (negative selection). Another splicing PCR reaction can generate a DNA fragment that contains the desired unmarked deletion. Following transformation, the *sacB*-Km^R marker is replaced by the engineered allele. Successful transformants can be selected in this step by the ability to grow in the presence of sucrose. With minor modifications, the same strategy can be used for the design of many different types of chromosomal changes.

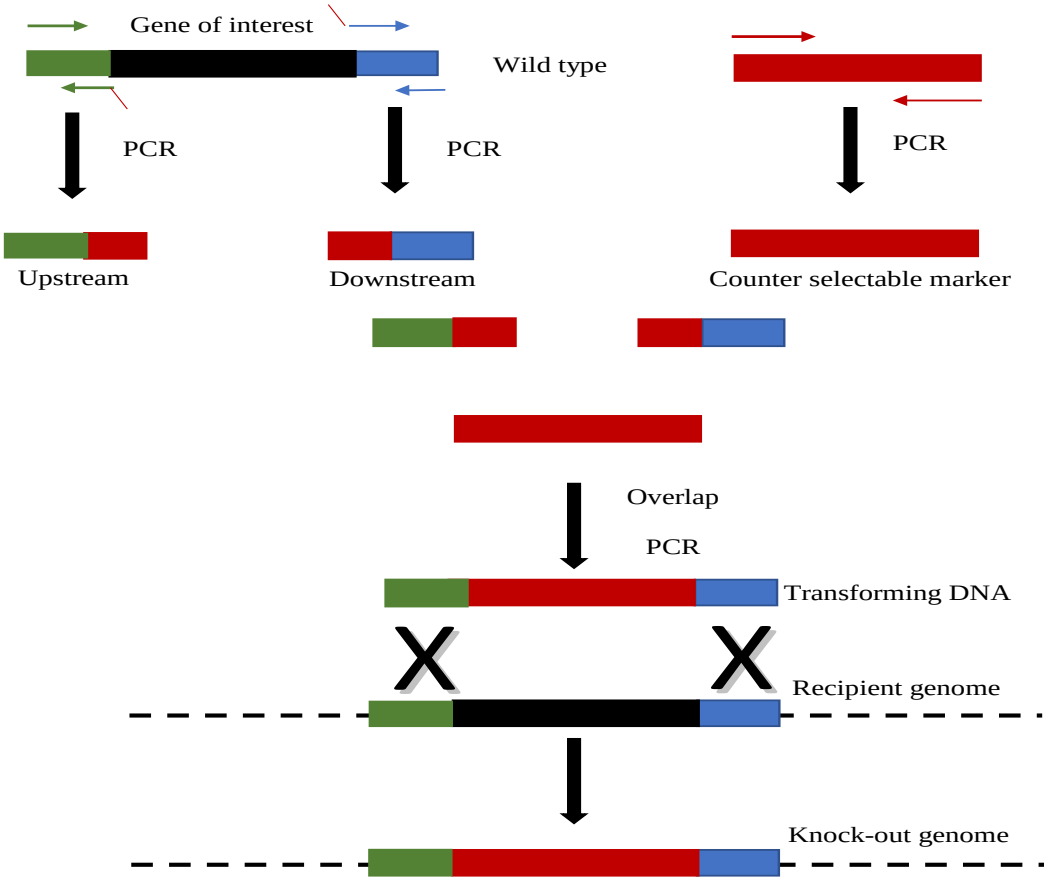


Figure 1.1. Genetic manipulation of the genome of *A. baylyi* ADP1. The chromosomal region to be manipulated (black) can be replaced with a selectable/counter-selectable marker, such as a *sacB*/Km^{R} cassette (red). The marker is amplified by PCR, in addition to the upstream (green) and downstream (blue) regions of the gene to be replaced using primers that generate overlapping sequence with the primers used for cassette amplification. The three fragments are used in a splicing PCR reaction to generate the transforming DNA. This piece of DNA can be used for transforming competent ADP1 recipients where the cassette replaces the gene of interest by allelic replacement (X), generating a knockout mutant that can be selected for by Km resistance. An unmarked deletion can be generated by using primers that generate an overlap between the upstream and downstream regions of the gene of interest. Figure adapted from [26].

The metabolic versatility of ADP1 is manifested in its ability to utilize a variety of carbon sources. Similar to other soil bacteria, ADP1 can funnel a diverse set of aromatic compounds into a small number of substrates that can be subjected to ring cleavage before entering central metabolism. In ADP1, ring cleavage is done through the β-ketoadipate pathway where there are two ring-cleavage substrates, catechol and protocatechuate [27]. The products of ring-cleavage can be further metabolized to give acetyl-CoA and succinyl-CoA, which can enter the tricarboxylic acid cycle. These steps are carried out via the two parallel branches of the β-ketoadipate pathway. The two branches of the pathway converge at β-ketoadipate enol lactone. Seminal studies on the biochemistry of the β-ketoadipate pathway in *Pseudomonas putida* revealed that the pathway is similar in *A. baylyi* [28]. In *A. baylyi*, however, the final three steps of the pathway are carried out by two sets of three isozymes, whereas in *P. putida*, the same three final steps are each catalyzed by a single enzyme. These enzymatic differences reflect variations in transcriptional regulation and the involvement of different metabolites as pathway inducers.

In their natural habitats, bacteria encounter changing environments and nutrient availability. Consistent with the need to govern gene expression under changing conditions, complex regulatory mechanisms have evolved that regulate catabolic genes. The β-ketoadipate pathway in *A. baylyi* is an example of complex regulatory patterns at multiple levels. In addition to mechanisms that allow the expression of certain genes when the relevant metabolite is present, other regulatory mechanisms govern metabolism when multiple carbon sources are available. Such preferential or hierarchal consumption is evident in *A. baylyi* as it utilizes benzoate, which is metabolized through the catechol branch before 4-hydroxy benzoate, which is metabolized via the protocatechuate branch [29]. Cross-regulation between the two branches of the β-ketoadipate

pathway can also be mediated by two transcriptional regulators involved in the catechol branch, BenM and CatM. These two LysR-type transcriptional regulators are involved in repressing the expression of genes involved in the protocatechuate branch [30]. Other mechanisms for regulation of carbon source consumption include carbon catabolite repression mediated by the Crc protein (catabolite repression control) [31].

Several studies highlight the potential of ADP1 in metabolic engineering for lignocellulosic biomass conversion. One study used a consortium of ADP1 mutants for the removal of aromatic growth inhibitors of yeast from lignocellulosic lysates in a two-stage batch process [32]. Another study used an engineered ADP1 strain with a single knockout to remove acetate and formate from lignocellulosic hydrolysates [33]. In the context of lignin valorization, ADP1 has been shown to metabolize various aromatic compounds derived from pretreated lignin. In a study that looked into the potential of different organisms for use in lignin consolidated bioprocessing, ADP1 was found to be one of the top performing strains that can both depolymerize high molecular weight lignin moieties and also catabolize a significant fraction of the low molecular weight aromatic species [34].

EASy: harnessing the power of gene amplification

The role of gene duplication and amplification in prokaryotic evolution and adaptation is well-established [35]. In addition to its significance in medically-relevant phenomena such as antibiotic resistance [36] and cancer [37], this type of genome plasticity provides a means for organisms to respond to changing environments and adapt to adverse conditions [38], by temporarily elevating gene expression and gene dosage. Furthermore, experimental evidence highlights the importance of gene amplification for acquiring novel functions [39]. A model termed “innovation-amplification-divergence” describes the process by which a new function

appears before duplication, and selection results in the evolution of divergent and functionally distinct genes [40]. If selection is removed, or new mutations arise that provide benefit in a single copy, the amplified array tends to contract and copy number decreases. This transient property of gene duplication and amplification makes it a difficult process to study.

The genetic system of ADP1 has been exploited to characterize dynamic changes in the copy number of chromosomal regions [41, 42]. In the original study describing gene amplification in ADP1, spontaneous mutants that can grow on benzoate as the carbon source without the two required activators, BenM and CatM, were readily obtained at a frequency of approximately 10^{-8} within a three-week period [41]. Spontaneous amplification of the *cat* gene cluster was found to be responsible for the ability of these mutants to grow on benzoate as the sole carbon source (Ben^{+}). Investigation of the genetic basis of genomic rearrangements lead to the identification of novel junctions involved in the initial duplication event. This duplication junction fused two genomic regions normally distant on the wild-type chromosome. In addition, a transformation-based assay was developed to identify the precise endpoints of amplified genomic regions. It was shown that the process can encompass two steps: an initial RecA-independent duplication, via illegitimate recombination, and a subsequent amplification mediated by the region of sequence repeats generated after duplication. Analysis of more than 100 spontaneous *Acinetobacter* amplification mutants revealed no sequence similarity at the recombination sites. Despite this lack of identity, several independent isolates harbored the same junction, which suggests site specificity.

Recently, a method has been developed that allows tandem arrays of chromosomal regions to be precisely engineered in ADP1 [4]. This method, Evolution by Amplification and Synthetic Biology (EASy), utilizes the genetic features of ADP1 to create specific duplication

sites that can undergo further gene amplification and accelerate the evolution of foreign genes (Figure 1.2). EASy builds on the discovery of gene amplification in ADP1 [41]. Where gene amplification occurs in nature in a stochastic and slow manner, EASy allows the targeted amplification of precise chromosomal regions. To generate an amplification of the targeted region, a linear DNA fragment is generated which harbors regions of sequence identity to the recipient strain's chromosome. When this DNA, called the synthetic bridging fragment (SBF) is used to transform the recipient strain, it serves as a platform for homologous recombination. Selective conditions demanding higher copy number (such as resistance to elevated antibiotic concentration) help maintain a multicopy array of the genes. During laboratory evolution, this tandem array will expand and contract through recombination. The selective conditions demanding the function of the amplified genes will establish the initial copy number. EASy was demonstrated to enable *A. baylyi* to consume guaiacol, a key component of lignin. Insertion into the ADP1 chromosome of foreign *Amycolatopsis* genes (*gcoAB*) did not enable growth on guaiacol as a sole carbon sourced despite the predicted ability of the encoded guaiacol-*O*-demethylase to produce catechol, a natural growth substrate. However, when the copy number of *gcoAB* was increased via transformation of an appropriate SBF, gene amplification conferred a Gua^+ phenotype.

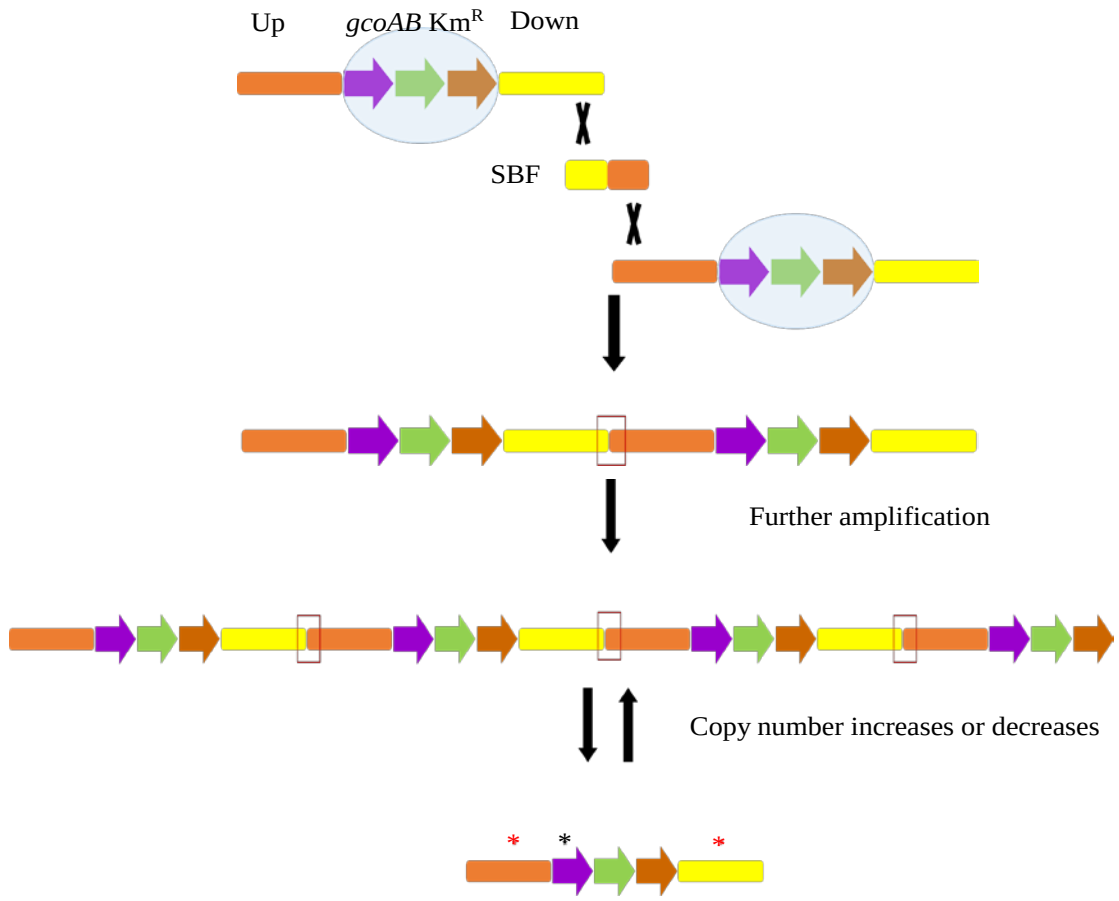


Figure 1.2. Amplification of chromosomal regions by EASy. Foreign genes, *gcoAB* and Km^R (highlighted in blue) are introduced in the chromosome of ADP1. A recipient strain with a single chromosomal copy of these genes is transformed with a linear DNA fragment, the synthetic bridging fragment (SBF), which induces homologous recombination between two normally distant regions upstream (up) and downstream (down) of the region targeted for amplification. Under selective conditions demanding the function encoded by the foreign genes, the copy number will increase or decrease. Emergence of beneficial mutations in the amplified region (black asterisk) or elsewhere in the chromosome (red asterisk) may enable growth on guaiacol with a single chromosomal copy of *gcoAB*. The red box highlights a novel DNA junction introduced by the initial duplication.

Biological conversion of lignin

Lignin is the most abundant aromatic heteropolymer on earth. In plants, it serves as a barrier against pathogen attack, plays a role in water and nutrient transport, and provides rigidity [43]. In addition to its presence in nature, large amounts of lignin are produced from paper and pulp industries, where it is discarded as waste [44]. In biorefineries, cellulosic carbohydrates can be used to generate biofuels while the most productive use of the lignin portion is typically being burned for process heat. Because of its abundance and high energy content, lignin has attracted attention as a desirable source for renewable energy and the production of valuable chemicals, but several challenges prevent the full utilization of this untapped source [45]. Some of these challenges are structural heterogeneity, variability in the ratio of lignin units by plant type, and different pretreatments resulting in a highly variable mixture of aromatic compounds. Lignin is synthesized from three main building blocks (also called monolignols), namely *p*-coumaryl alcohol, coniferyl alcohol, and sinapyl alcohol which give rise to three aromatic components of lignin: *p*-hydroxyphenyl (H), guaiacyl (G), and syringyl (S), respectively. The degree of methoxylation distinguishes the three units [46]. The relative distribution of these units can vary greatly between different plant species. For example, hardwood is composed of both G and S lignin, while softwood is mostly composed of G lignin. Because of the variation in the relative abundance of each, some efforts focus on modifying the structure of lignin to be more amenable to degradation, for example by altering the ratio of the monolignols [47], or by increasing the amount of easily cleavable linkages [48].

Biological conversion approaches capitalize on the ability of some organisms to convert a wide array of lignin-derived aromatic compounds into a few ring-cleavage substrates, in what has been termed “biological funneling” [49]. These substrates, such as catechol and

protocatechuate can be subjected to ring cleavage by dioxygenases. Dioxygenases cleave the aromatic ring in either an intradiol (*ortho*) or extradiol (*meta*) manner, depending on the position of the two hydroxyl groups. The open ring species can then be metabolized through several steps of the β-ketoadipate pathway to produce acetyl-CoA and succinyl-CoA [27] or through the meta-cleavage pathway [50]. However, no single organism can utilize all lignin-derived aromatic compounds and the choice of organism for biological conversion may vary by the type of lignin stream. In this study, I sought to expand the aromatic compound degradation capabilities of ADP1 to allow the utilization of guaiacol, a key component in pretreated lignin that ADP1 does not naturally metabolize. The importance of guaiacol to lignin valorization stems from its presence in almost all lignin, specifically in softwood lignin. In order for guaiacol to be prepared for ring cleavage, it has to be demethylated to yield catechol, which can be further metabolized via the β-ketoadipate pathway.

Motivated by the importance of guaiacol demethylation for lignin valorization, database searches were conducted using a partial N-terminal amino acid sequence of a putative guaiacol *O*-demethylating cytochrome P450 enzyme from *Rhodococcus rhodochrous* [51]. This search lead to the discovery of genes in *Amycolatopsis* sp. ATCC39116 and other organisms that encode proteins homologous to the one identified in *Rhodococcus* [4]. The genes from *Amycolatopsis*, designated *gcoAB* encode a promiscuous two component guaiacol *O*-demethylase that was recently characterized [3]. This enzyme catalyzes the conversion of guaiacol into catechol and formaldehyde (Figure 1.3).

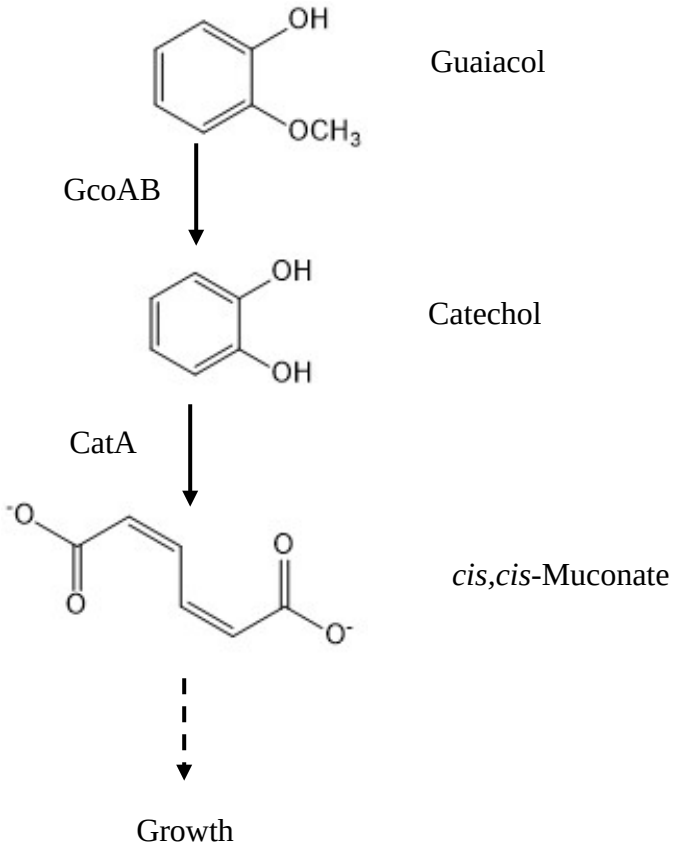


Figure 1.3. Demethylation of guaiacol by GcoAB. Guaiacol can be funneled towards central metabolism through the catechol branch of the β-ketoadipate pathway in ADP1. Guaiacol is first demethylated by GcoAB from *Amycolatopsis* sp. to produce catechol, which is subjected to ring cleavage via CatA (catechol 1,2-dioxygenase) to give *cis,cis*-Muconate, which undergoes several more catalytic steps until finally being converted into acetyl-CoA and succinyl-CoA.

The first component in this system, GcoA is a cytochrome P450 enzyme that belongs to the CYP255A family. Structural characterization showed that GcoA possesses a typical P450 single-domain architecture. Three phenylalanine residues in the active site cavity (Phe75, Phe169, and Phe395) are responsible for positioning the aromatic ring of the substrate, while Val241 and Gly245 coordinate the oxygen atoms of the substrate. In addition, screening trials showed that other analogous substrates can be accommodated in the active site of GcoA, such as guaethol, vanillin and syringol, although the extent of tightly coupled enzymatic conversion to a corresponding demethylated product varied. These efforts revealed mechanistic insights into the conformational changes associated with binding each substrate. In addition, structure-guided engineering enabled a GcoA variant to demethylate syringol, which highlights the importance of engineering P450 systems in the biological conversion of lignin-derived aromatics [52].

The second component, GcoB is a three-domain reductase. Sequence information revealed that GcoB is homologous to three-domain reductases such as benzoate 1,2-dioxygenase reductase (BenC) [53] and anthranilate 1,2-dioxygenase reductase (AntC) from *Acinetobacter baylyi* [54]. GcoB has three domains: an N-terminal 2Fe-2S ferredoxin-like domain, a central flavin-binding domain, and a C-terminal domain that interacts with NAD(P)H. In typical bacterial cytochrome P450 systems, two distinct redox partners are required for transfer of electrons from NAD(P)H and therefore, GcoAB represents a new two-component class of P450 systems.

Laboratory evolution as a tool for improving phenotypes

Adaptive laboratory evolution allows the study of populations under well-defined conditions [55]. The use of microorganisms provides certain advantages for use in such experiments. For example, microbial cells have simple nutritional requirements and fast growth rates, reaching many generations in a relatively short time. In addition, the ability to freeze cells almost indefinitely, provides the opportunity to "replay" an experiment and compare evolutionary trajectories. Laboratory evolution has proven important in addressing multiple questions, from the basic understanding of evolutionary processes, to improving performance of industrial strains, and in cancer research [56].

Typically, a microorganism is grown under controlled settings for extended periods of time, which can range from several weeks to years. Standard culture methods for such experiments include chemostats and batch cultures, with each method providing different advantages and drawbacks. Chemostats provide constant conditions, by adding fresh medium at a fixed rate and removing a fraction of the culture at the same rate. The cells are grown in a steady state at a fixed growth rate dictated by the dilution rate. In addition, chemostats provide the ability to tightly control physiological conditions and population sizes. However, chemostats usually require high operation costs [57]. Batch cultures, employing serial transfers remain an attractive option for running parallel experiments at a lower cost compared to chemostats. Yet, batch cultures subject the cells to alternating cycles of high and low densities, as well as periods of nutrient availability and deprivation. These factors result in non-uniform selective conditions. Some trials attempted to overcome some of these drawbacks by automating the process of maintaining the cultures [58].

Serial transfers have been used for long-term evolution experiments that have been in progress since 1988 [59]. In these studies, 12 *E. coli* populations were serially passaged in minimal medium supplemented with low levels of glucose as the carbon source. Changes in phenotype and fitness have been, and continue to be, monitored and reported periodically. Results of this project provide valuable insights into understanding evolutionary theory and they continue to show the complexity of dynamic long-term adaptation, even in constant environments [60]. Observations have included the evolution of high mutation rates [61], the rapid evolution of core genes relative to non-core genes [62], and the emergence of a variant with the ability to utilize citrate as a carbon source after more than 30,000 generations [63]. Some of the processes underlying these observations are described in the next section.

Typically, selective pressure involves improved growth on a certain substrate or higher tolerance to stress. Once a desired phenotype is achieved, whole-genome or whole population DNA sequencing can identify mutations that may be responsible for the phenotype. The relatively affordable cost of whole-genome sequencing, and continuous improvements in library preparation allow researchers to identify genetic changes in evolved microbial populations [64]. However, the identification of causal mutations under certain conditions requires establishing a link between the mutations and phenotype, a process that is not always straightforward. While evolution experiments are typically started from a clonal strain, evolved populations are typically more heterogenous. The dynamic nature of evolution requires investigating multiple evolutionary scenarios such as clonal interference, hitchhiking mutations, and epistasis [65]. In clonal interference, more than one beneficial mutation may arise in different clones of the same population and compete with each other. For example, a beneficial mutation that appears initially in the population, may decline before it completely sweeps the population, when another

mutation with a larger fitness benefit appears. Hitchhiking mutations are those deleterious or neutral mutations that may increase in frequency in a given population due to being linked in the genome to a region conferring a beneficial mutation. Finally, epistasis describes the scenario where the effect of one mutation is influenced by another mutation. Epistasis can be further classified depending on the combined effect of the mutations. For example, in diminishing-returns epistasis [66] the outcome of a combination of mutations is less than expected, while in all-or-none-epistasis [67], initial, seemingly neutral mutations are needed before a driver mutation appears later that creates the new or improved phenotype.

To assess phenotypic changes, the fitness of an evolved isolate is compared relative to its ancestral strain [59]. This is typically done in a head-to-head competition under conditions that mimic those under which the adapted organism evolved (Figure 1.4). The population size of the competing strains at the onset and the end of the competition is used to calculate the relative fitness of the evolved isolate as follows [68]:

$$w = \frac{\ln\left(\frac{A_f}{A_i}\right)}{\ln(\frac{B_f}{B_i})}$$

Where *w* is the relative fitness, *A* and *B* represent the population sizes of the mutant and ancestral strain, respectively. Subscripts *i* and *f* denote initial and final time points, respectively. Mutants with improved fitness relative to the ancestor will have a relative fitness of more than 1.

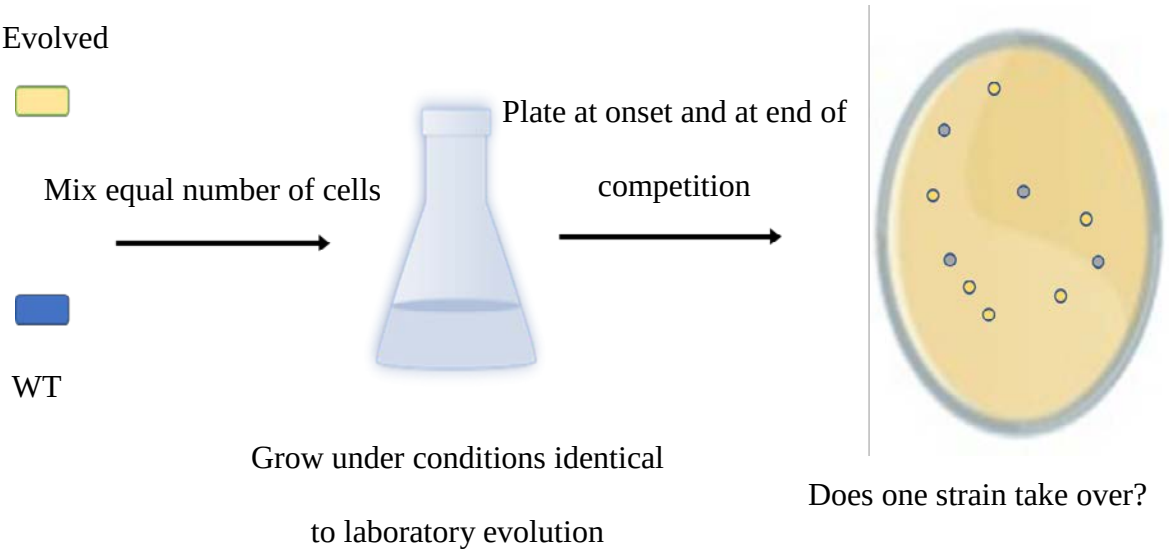


Figure 1.4. Competition assays. A head-to-head competition can show whether an adapted strain has a fitness improvement relative to its ancestor. Ideally, the two competing strains should be easily distinguished, for example by different drug resistance profiles or colony morphology. The two strains are mixed in a 1:1 ratio and grown under the same conditions under which the adapted strain was evolved. The frequency of each strain at the beginning and at the end of the competition period is used to calculate the relative fitness, *w*.

In an ideal adaptive scenario, an evolved isolate with improved fitness or phenotype will display mutation(s) in the gene of interest. However, recent evidence suggests that mutations elsewhere in the chromosome occur frequently [69]. In this study, eight *E. coli* populations were evolved from a parent strain that lacks a gene coding for a protein essential for arginine synthesis, ArgC. This experimental condition required the promiscuous activity of another protein, ProA. After evolution for up to 1000 generations, the growth rates of the populations increased by approximately three-fold. Genetic analysis showed that amplification of regions encompassing *proA* was observed in all populations. The copy number of the *proA* amplicon and the size of the amplified region varied, but none of the populations reached a single copy at the end of adaptive evolution. In only one population, mutations were identified in *proA*. Whole-genome sequencing revealed some mutations elsewhere in the chromosome that were common to multiple populations, suggesting that these mutations may provide a fitness advantage. One of these mutations was shown to be a general adaptation to growth in minimal medium, while other mutations were specific adaptations to the conditions investigated. One mechanism that improved arginine synthesis without affecting the promiscuous activity of ProA involved an increased abundance of ArgB that catalyzes the second step in arginine synthesis by improving translation efficiency. Another mechanism involved CarB, the large subunit of carbamoyl phosphate synthetase and increased flux towards the lower initiation of arginine synthesis.

As it pertains to results in this dissertation project, similar observations have been made during studies on gene duplication and amplification. According to evolutionary models, divergence of amplified genes may occur when a secondary activity of a protein becomes relevant under certain environmental changes [40]. Because initially this secondary activity is weak, duplication and further amplification may provide a rapid mechanism for adaptation to the

selective conditions, by increasing the amount of the weak enzyme. In attempts to study this model, one investigation followed the evolution of amplified genes in real-time as two specialized alleles emerged [40]. As the model predicts, after improved alleles became fixed in the population, the pressure to keep an extra "less fit" copy is relaxed and therefore copy number was reduced. In their study, the authors followed the fate of the diverging genes only and did not report on mutations elsewhere in the genome.

In Chapter 2, information is presented on results similar to those discussed above. In our initial EASy trial, seven populations were evolved to grow on guaiacol as the carbon source, a phenotype observed only when the copy number of the genes encoding this function is increased [4]. However, unlike the study of Morgenthaler *et al* where spontaneous gene amplification was observed, in our system gene amplification was induced at the onset of laboratory evolution and the precise endpoints of the amplicon were defined. After approximately 1000 generations, whole-genome sequencing revealed that most of the populations harbored genetic rearrangements in the *gcoAB* region, encoding fusions between CatA and GcoA. However, in only one population point mutations in *gcoAB* were found. Chapter 2 discusses the analysis of multiple mutations identified in one isolate from that population. The rest of this introductory chapter will present background information about some of the genes found to carry mutations that could have affected evolutionary trajectories. These genes include one encoding a ribosomal silencing factor, a gene encoding a penicillin-binding protein predicted to play a role in peptidoglycan synthesis, and a third gene encoding a transport protein. These results support the important role of genetic changes throughout the chromosome as the evolved populations adapt to grow on guaiacol. My studies demonstrate the highly dynamic nature of adaptive evolution

and highlight the need to develop new tools for addressing the significance of how mutations may function in combination.

Stationary phase and the role of ribosome hibernation

Laboratory conditions for bacterial growth typically provide abundant nutrients. However, in natural habitats bacteria rarely encounter such luxury and constantly face changing physiological conditions and fluctuating nutrient availability, in what has been compared to a "feast or famine" situation [70]. To survive, bacteria must rapidly adapt and respond to this harsh lifestyle. It has been estimated that more than half of Earth's biomass consists of microorganisms in a non-growing or "quiescent" state [71]. Such conditions resemble the stationary phase of batch cultures. Thus, stationary phase and conditions of nutrient scarcity appear to constitute a significant portion of the lifecycle of bacteria in nature.

Regulatory mechanisms govern cellular responses to different stressors. For instance, during nutrient starvation or upon entering stationary phase, a general bacterial stress response can be activated that alters the expression profile of a large number of genes via the alternate sigma factor RpoS [72, 73]. RpoS is responsible for orchestrating entrance into stationary phase and producing a resistant cell that can tolerate stress until favorable conditions are available. The regulation of RpoS itself is a complex process that requires coordination between different regulators, non-coding RNAs, and changes in macromolecule stability. RpoS belongs to group 2 of the σ70 family and displays an increased concentration at the onset of stationary phase and under stress conditions [74]. This increase in concentration is attributed to improved transcription of *rpoS*, higher efficiency of translation, and increased protein stability.

In addition to this response to starvation, other factors help cells adapt to stressful conditions. Interplay between these different processes further adds to the complexity of cellular

responses to adverse conditions. For example, the stringent response is triggered against amino acid starvation by downregulating the synthesis of ribosomal components and upregulating levels of RpoS and other stress-related proteins [75]. The stringent response is mediated by the alarmone (p)ppGpp, or guanosine 3', 5'-bispyrophosphate. Among other functions, ppGpp binds the β subunit of the RNA polymerase, altering the expression of several genes. In addition, ppGpp works with other cellular factors to help RpoS outcompete the housekeeping sigma factor σ70 for RNA polymerase binding. Cells entering stationary phase also exhibit morphological changes in addition to these metabolic changes and altered gene expression. The nucleoid condenses in a process that is mediated by a DNA-binding protein called Dps. This nucleoprotein, which is induced during stationary phase, forms a stable complex with DNA that serves to compact the nucleoid and protect the DNA [76].

Because protein synthesis is one of the most energy-consuming cellular processes [77], bacterial regulatory mechanisms also adjust protein expression to match the physiological state of the cell. One such mechanism is ribosome hibernation [78], whereby several ribosome-associated proteins convert translationally active ribosomes into an inactive state. These factors include ribosome modulation factor, which acts to form 90S dimers [79], a hibernation promoting factor, which forms a 100S complex from 90S dimers [80], and ribosome-associated inhibitor A, which does not form dimers but maintains 70S ribosomes in an inactive state [81]. In addition, alternate factors can down regulate protein synthesis, such as a phase-induced ribosome-associated protein, an energy-dependent translational throttle A, and ribosomal silencing factor A (RsfA) [82].

The role of RsfA (also known as RsfS) in ADP1 and other bacteria is highlighted in this dissertation because one of our evolved Gua^+ isolates was found to have a deletion of the

corresponding gene. RsfA proteins are highly conserved in almost all bacterial genomes and eukaryotic organelles. Early studies in *Zea mays* showed that Iojap, the RsfA homolog in plastids is involved in ribosome biogenesis [83]. Similarly, the mitochondrial RsfA homolog, C7orf30 was found to be involved in the biogenesis of the large ribosomal subunit [84]. These proteins are known to inhibit protein synthesis during stationary phase and media shifts from nutrient rich to nutrient poor, and they help cells adapt to these conditions [82]. Structural studies showed that RsfA from *Mycobacterium tuberculosis* is a dimer in solution, but a monomer binds the L14 ribosomal protein on the large ribosomal subunit. Such binding at the 50S-30S interface prevents the two ribosomal subunits from joining [85].

In addition to its role as a posttranscriptional regulator, several studies suggest its involvement in other niche-specific processes. For example, a study in *Acinetobacter baumannii* showed that RsfA may be involved in this pathogen's resistance to colistin [86]. Another study found that a RsfA homolog in *Exiguobacterium antarcticum* could play a role in this psychrophile's adaptation to extreme environments [87]. These results, and those presented in chapter 2 of this dissertation, highlight the potential role of stationary phase regulation beyond merely orchestrating entry into this phase of bacterial growth. Since data presented later pertain to the metabolism of aromatic compounds, a tightly regulated process, mutations in regulatory proteins may optimize regulation and enable improved growth on aromatic substrates. RsfA has not been experimentally characterized in ADP1 and has not been studied in the context of degrading aromatic compounds. In addition, our analysis warrants further investigation of genes that we would not have predicted to play a role under conditions of growth on guaiacol. Among these genes, one encodes a penicillin-binding protein, which is discussed in more detail in the next section. Collectively, these results underscore the importance of discovering new targets for

genetic engineering that may enable improved strain design for lignin valorization and other industrial applications.

The role of penicillin-binding proteins in cell wall synthesis

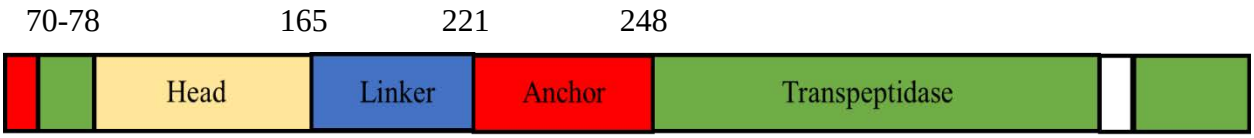
Analysis of mutations in a Gua^+ mutant (ACN1850) suggested the importance of unexpected genes in the adaptive process. One gene is predicted to encode a penicillin-binding protein. These proteins are known to be involved in the synthesis of peptidoglycan (PG). PG is an essential mesh-like layer surrounding the cell membrane of bacteria and provides protection and preserves cell morphology [88]. PG is a polymer of alternating sugar units called *N*-acetylglucosamine and *N*-acetylmuramic acid, with a pentapeptide attached to the *N*-acetylmuramic acid unit. In Gram-negative bacteria, PG is typically 3-6 nm thick and comprised of a single layer, while in Gram-positive bacteria, PG is thicker (10 – 40 nm). Penicillin-binding proteins (PBPs) are involved in the synthesis and crosslinking of PG. PBPs are targeted by penicillin due to the structural similarity between the latter and the natural substrate of PBPs, the D-Ala-D-Ala motif of the stem pentapeptide that connects the glycan strands [89]. PBPs are generally classified according to their molecular weights. High molecular mass PBPs include two classes, A and B, while low molecular mass PBPs are known as class C [90]. High molecular mass PBPs vary in the structure and catalytic function of their N-terminal domain. Class A PBPs have an N-terminal domain with a glycosyltransferase activity, which catalyzes the elongation of uncrosslinked glycan strands, while the N-terminal domain of class B PBPs may be involved in cell morphology, in cooperation with other proteins. The C-terminal domains of both classes possess a transpeptidase activity that catalyzes crosslinking of adjacent glycan strands.

Among the three class A PBPs in *E. coli*, two are most important, PBP1a and PBP1b, as mutations removing both of these PBPs were found to be lethal [91, 92]. The two class B PBPs in *E. coli*, PBP2 and PBP3, are involved in cell elongation and cell division, respectively. The crystal structure of PBP2 from *E. coli* was recently solved during a study on developing new compounds active against pathogenic bacteria [93]. Structural data showed that this protein displays an N-terminus with three subdomains: anchor, linker, and head domains. The anchor and head domains are involved in protein-protein interactions with other PBPs [94]. The C-terminus contains the catalytic transpeptidase domain and displays the conserved motifs common to this class of PBPs. Particularly, the catalytic serine of the SXXK motif is at position 330. The PBP2 structure from *E. coli* was used, in addition to the structure from *Helicobacter pylori*, for homology modeling of PBP2 proteins of other pathogenic bacteria. This modeling showed that the overall active site structure is well conserved, while the anchor and head domains were more variable [93].

In ADP1, PBP2 is encoded by the gene ACIAD1101 (*pbpA*). According to classification schemes, this protein belongs to the high molecular mass group of PBPs, specifically class B, and it is predicted to function as a DD-transpeptidase. This catalytic function involves an attack of the PBP2's catalytic serine on the carbonyl group of the D-Ala-D-Ala bond of one PG strand, called the donor strand, followed by crosslinking with a second PG strand, called the acceptor strand [95]. The active site serine is part of the conserved SXXK motif in PBPs. The catalytic serine in ADP1's PBP2 maps to position 328. The wild-type protein is 675 amino acids long in ADP1. Figure 1.5 shows a partial alignment of the amino acid sequence from ADP1 and *E. coli* highlighting the SXXK motif of which the catalytic serine is part. Results presented in chapter 2 of this dissertation discuss an EASy-derived isolate which was found to harbor a mutation in the

gene encoding PBP2. This mutation encodes an early termination at position 365, which removes a large portion of the catalytic domain. It is not known if the remainder of the protein is stably synthesized in the corresponding mutant.

A



B

Figure 1.5. Structural features and important residues in PBP2 proteins. A) Representation of the primary structure of PBP2 from *E. coli* (PDB: 6G9P), showing the main domains. Numbers mark the first residue of each secondary structure. B) Sequence alignment of a C-terminal region of PBP2 proteins from ADP1, *A. baumannii*, *E. coli* and *Ps. putida* showing the highly conserved motif SXXK of class B PBPs which is highlighted by the red box. Alignment generated with BoxShade. Panel A adapted from [93].

Transport proteins

Another mutation identified in the isolate in which a point mutation in *gcoA* was found (ACN1850), may prove important in the context of aromatic compound metabolism. A gene encoding a predicted transporter, ACIAD_RS15610 (ACIAD3450) had a one bp deletion that results in a frameshift insertion of 11 residues and an early termination of the gene product. Sequence similarity indicates that the protein encoded by ACIAD_RS15610 belongs to the Major Facilitator Superfamily (MFS) of transporters, a large and diverse family of proteins involved in the import and export of small solutes across the cell membrane in response to chemiosmotic ion gradients [96]. In ADP1, several MFS transporters have been previously characterized and are involved in the uptake of aromatic compounds. For example, BenK and MucK function in the uptake of benzoate and *cis,cis*-muconate, respectively [97, 98]. In addition, VanK, is involved in the uptake of vanillate, and PcaK in the uptake of both 4-hydroxybenzoate and protocatechuate [99]. VanK and PcaK display overlapping specificities which is evident by the impaired growth of double mutants when protocatechuate is the sole carbon source [100]. In the context of lignin valorization, transport of lignin depolymerization products is one of the least well understood steps. Therefore, it may prove critical for enhanced utilization of lignin-derived aromatics to improve our knowledge of transport processes and the proteins involved.

While initially thought to be involved in the transport of sugars only, studies have shown that MFS transporters are more diverse than that and can transport different solutes. Information from the crystal structures of lactose permease (LacY) and glycerol-3-phospate transporter (GlpT), revealed that MFS transporters display a characteristic structure with 12 transmembrane helices that exhibit a two-fold symmetry [101, 102]. Early reports proposed that the characteristic 12-helix fold of MFS proteins arose through an intragenic duplication event from

an ancestral 6-transmembrane helix unit [103]. Transport of substrates occurs through an alternating access mechanism, often called a “rocker switch” mechanism.

According to the TransportDB database, 51 MFS transporters are encoded in the ADP1 genome, out of 366 total transporters [104]. Given its metabolic versatility and ability to grow on a wide variety of growth substrates, transport is an important function for ADP1. This fact is reflected in the dedication of almost 10% of its genes to transport functions [6]. The protein encoded by ACIAD_RS15610 (ACIAD3450) is predicted to display a typical 12 transmembrane helices topology (Figure 1.6). Mapping the ACN1850 mutation in ACIAD_RS15610 to the predicted amino acid sequence of the transporter shows that the mutation results in a frame shift after position 345. The wild-type protein is 454 amino acids long and this frame shift occurs in the predicted 11th transmembrane helix. It is unknown whether a shortened protein is stably produced in the corresponding mutant. Based on homology and according to standard classification [96], the transporter encoded by ACIAD_RS15610 belongs to family 3/drug efflux family of MFS proteins. While ADP1 is not pathogenic, studies in this organism showed that single-step exposure to chloramphenicol gave rise to antibiotic resistant mutants, in which a MFS protein was upregulated, highlighting the role of multidrug efflux systems in developing antibiotic resistance [105]. Therefore, further analysis of this mutation may provide insights into transport relevant to diverse cellular processes.

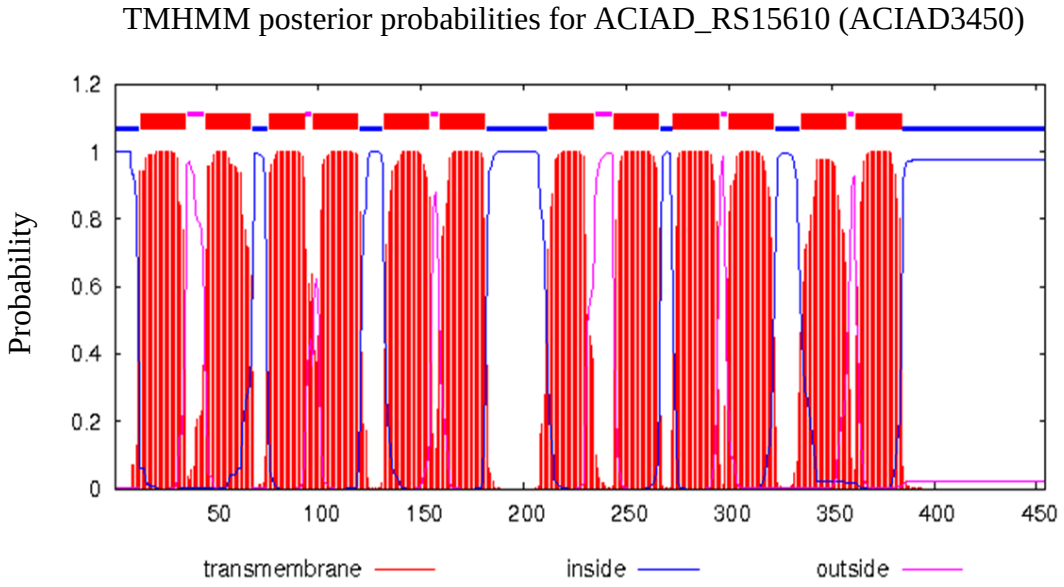


Figure 1.6. Predicted transmembrane helices of an ADP1 MFS transport protein. A typical 12 transmembrane helices topology of the transporter encoded by ACIAD_RS15610 (ACIAD3450). Helices are in red, while intracellular and extracellular loops are depicted in blue and pink, respectively. Prediction generated by TMHMM 2.0 [106]. The server uses a method based on hidden Markov model to predict the most probable location and orientation of transmembrane helices in the input sequence. The probabilities plot is generated by calculating the total probability that a residue sits in a helix, inside, or outside summed over all possible paths through the model.

Lon protease

In addition to the genes encoding RsfA, PBP2, and an MFS transporter, other mutations were identified in the same guaiacol-degrading mutant (ACN1850). One of those mutations occurs in ACIAD_RS05110, which is predicted to encode an DNA-binding ATP-dependent (Lon) protease, also known as protease La [107]. In ACN1850, this gene has a 39 bp deletion, which results in an in-frame removal of 13 residues between residues 237 and 251. The wild-type protein in ADP1 is 808 amino acids long. Lon protease homologs are found in all kingdoms of life. These proteases belong to the serine protease family, so called because of the conserved serine residue in their active site. In *E. coli*, Lon protease has a Ser-Lys dyad involved in the catalytic activity [108], and early studies showed that Lon binds DNA in a sequence-specific manner [109]. Lon proteases can be generally classified into two subfamilies: LonA, which includes bacterial and eukaryotic Lon proteases, and LonB, which includes the archaeal Lon proteases [110]. Sequence comparisons show that proteins belonging to the LonA subfamily consist of three domains: A variable N-terminal domain, an ATPase domain, and a proteolytic C-terminal domain, while those belonging to LonB lack the N-terminal domain. Lon proteases have been shown to mediate selective degradation of abnormal and misfolded proteins, as well as short-lived regulatory proteins. For example, in *E. coli*, Lon protease has been shown to target the transcriptional activator SoxS, which is involved in defense mechanisms against superoxide stress, resulting in reduced levels of SoxS once the stress cause is removed [111, 112].

During the long-term evolution experiment, evolved populations harbored mutations in, or in the vicinity of, the *lon* gene after 40,000 generations [113]. While in our study the mutation in *lon* may have been selected in the context of aromatic compounds metabolism, recent evidence suggests the involvement of Lon protease in important processes such as pathogenicity

and virulence in other organisms. For example, in a study in *Brucella* spp, RNA-seq data showed that the expression of genes involved in stress response and quorum sensing was altered in a *lon* mutant [114]. Another study in *Acinetobacter baumannii* suggested a role for Lon protease in biofilm formation and motility [115]. Therefore, several lines of evidence suggest the involvement of Lon protease in diverse processes. Further investigation of the role of mutations in the *lon* gene may provide insights into diverse and important phenomena.

Dissertation outline

The rest of this dissertation focuses on key findings from laboratory evolution experiments and on the development of methods and tools relevant to this type of experimentation. Chapter 2 presents the results of analyzing specific mutations discovered in a guaiacol-degrading evolved isolate. The findings include the role of a previously uncharacterized gene encoding a post-transcriptional regulator (RsfA) in ADP1 in adaptation during growth on guaiacol. Those results highlight the importance of experimental design.

In chapter 3, a summary is presented on methods and tools developed for laboratory evolution. One section discusses the effect of adding external DNA on evolutionary trajectories of the evolving populations. Another section presents preliminary results on using natural transformation for the analysis of adaptive mutations. Altogether, the results presented in chapter 3 provide a platform for improving experimental design and for development of analytical tools for laboratory evolution.

The fourth and final chapter provides a “broader impacts” point of view regarding the lessons learned during my project. The chapter highlights some avenues for future exploration that build on the results presented here.

Abbreviations

Ap	Ampicillin
EASy	Evolution by Amplification and Synthetic biology
epPCR	Error-prone PCR
gDNA	Genomic DNA
HPLC	High-performance liquid chromatography
Km	Kanamycin
LB	Lysogeny Broth
MM	Minimal Medium
MFS	Major facilitator superfamily
NGS	Next-generation sequencing
OD	Optical Density
P450	Cytochrome P450
PBP2	Penicillin-binding protein 2
PBPs	Penicillin-binding proteins
PG	Peptidoglycan
qPCR	Quantitative PCR
RsfA (RsfS)	Ribosomal Silencing Factor A
SBF	Synthetic bridging fragment
Sm	Streptomycin
Sp	Spectinomycin
WGS	Whole-genome sequencing
WT	Wild type

References

1. Juni, E. and A. Janik, Transformation of *Acinetobacter calco-aceticus* (*Bacterium anitratum*). J Bacteriol, 1969. **98**(1): p. 281-8.
2. Vaneechoutte, M., et al., Naturally transformable *Acinetobacter* sp. strain ADP1 belongs to the newly described species *Acinetobacter baylyi*. Appl Environ Microbiol, 2006. **72**(1): p. 932-6.
3. Mallinson, S.J.B., et al., A promiscuous cytochrome P450 aromatic O-demethylase for lignin bioconversion. Nature Communications, 2018. **9**(1): p. 2487.
4. Tumen-Velasquez, M., et al., Accelerating pathway evolution by increasing the gene dosage of chromosomal segments. Proceedings of the National Academy of Sciences, 2018. **115**(27): p. 7105-7110.
5. Juni, E., Interspecies transformation of *Acinetobacter*: genetic evidence for a ubiquitous genus. Journal of bacteriology, 1972. **112**(2): p. 917-931.
6. Barbe, V., et al., Unique features revealed by the genome sequence of *Acinetobacter* sp. ADP1, a versatile and naturally transformation competent bacterium. Nucleic Acids Res, 2004. **32**(19): p. 5766-79.
7. Durot, M., et al., Iterative reconstruction of a global metabolic model of *Acinetobacter baylyi* ADP1 using high-throughput growth phenotype and gene essentiality data. BMC Syst Biol, 2008. **2**: p. 85.
8. de Berardinis, V., et al., A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1. Molecular systems biology, 2008. **4**: p. 174-174.
9. Juni, E., Genetics and physiology of *Acinetobacter*. Annu Rev Microbiol, 1978. **32**: p. 349-71.

10. Young, D.M., D. Parke, and L.N. Ornston, Opportunities for genetic investigation afforded by *Acinetobacter baylyi*, a nutritionally versatile bacterial species that is highly competent for natural transformation. Annu Rev Microbiol, 2005. **59**: p. 519-51.
11. Lorenz, M.G. and W. Wackernagel, Bacterial gene transfer by natural genetic transformation in the environment. Microbiological reviews, 1994. **58**(3): p. 563-602.
12. Palmen, R., et al., Physiological characterization of natural transformation in *Acinetobacter calcoaceticus.* Microbiology, 1993. **139**(2): p. 295-305.
13. Averhoff, B. and I. Graf, The natural transformation system of *Acinetobacter baylyi* ADP1: a unique DNA transport machinery, p 119–139. Acinetobacter: molecular biology. Caister AP, Norfolk, United Kingdom, 2008.
14. Bacher, J.M., D. Metzgar, and V. de Crécy-Lagard, Rapid Evolution of Diminished Transformability in *Acinetobacter baylyi.* Journal of Bacteriology, 2006. **188**(24): p. 8534-8542.
15. Palmen, R., P. Buijsman, and K.J. Hellingwerf, Physiological regulation of competence induction for natural transformation in *Acinetobacter calcoaceticus.* Archives of Microbiology, 1994. **162**(5): p. 344-351.
16. Gerischer, U., D.A. D'Argenio, and L.N. Ornston, IS 1236, a newly discovered member of the IS3 family, exhibits varied patterns of insertion into the *Acinetobacter calcoaceticus* chromosome. Microbiology, 1996. **142**(7): p. 1825-1831.
17. Cuff, L.E., et al., Analysis of IS*1236*-Mediated Gene Amplification Events in *Acinetobacter baylyi* ADP1. Journal of Bacteriology, 2012. **194**(16): p. 4395-4405.

18. Segura, A., et al., Genetic analysis of a chromosomal region containing *vanA* and *vanB*, genes required for conversion of either ferulate or vanillate to protocatechuate in *Acinetobacter*. J Bacteriol, 1999. **181**(11): p. 3494-504.

19. Reams, A.B. and E.L. Neidle, Gene amplification involves site-specific short homology-independent illegitimate recombination in Acinetobacter sp. strain ADP1. J Mol Biol, 2004. **338**(4): p. 643-56.

20. Suarez, G.A., et al., Reduced Mutation Rate and Increased Transformability of Transposon-Free *Acinetobacter baylyi* ADP1-ISx. Appl Environ Microbiol, 2017. **83**(17).

21. Fortier, L.-C. and O. Sekulovic, Importance of prophages to evolution and virulence of bacterial pathogens. Virulence, 2013. **4**(5): p. 354-365.

22. Renda, B.A., et al., Emergence of a Competence-Reducing Filamentous Phage from the Genome of *Acinetobacter baylyi* ADP1. Journal of Bacteriology, 2016. **198**(23): p. 3209-3219.

23. Elliott, K.T. and E.L. Neidle, *Acinetobacter baylyi* ADP1: transforming the choice of model organism. IUBMB Life, 2011. **63**(12): p. 1075-80.

24. Higuchi, R., B. Krummel, and R.K. Saiki, A general method of in vitro preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. Nucleic acids research, 1988. **16**(15): p. 7351-7367.

25. Jones, R.M. and P.A. Williams, Mutational analysis of the critical bases involved in activation of the AreR-regulated sigma54-dependent promoter in Acinetobacter sp. strain ADP1. Applied and environmental microbiology, 2003. **69**(9): p. 5627-5635.

26. Metzgar, D., et al., *Acinetobacter* sp. ADP1: an ideal model organism for genetic analysis and genome engineering. Nucleic Acids Research, 2004. **32**(19): p. 5780-5790.

27. Harwood, C.S. and R.E. Parales, The beta-ketoadipate pathway and the biology of self-identity. Annu Rev Microbiol, 1996. **50**: p. 553-90.

28. Ornston, L.N. and R.Y. Stanier, The conversion of catechol and protocatechuate to beta-ketoadipate by *Pseudomonas putida.* J Biol Chem, 1966. **241**(16): p. 3776-86.

29. Gaines, G.L., 3rd, L. Smith, and E.L. Neidle, Novel nuclear magnetic resonance spectroscopy methods demonstrate preferential carbon source utilization by *Acinetobacter calcoaceticus.* J Bacteriol, 1996. **178**(23): p. 6833-41.

30. Brzostowicz, P.C., et al., Transcriptional cross-regulation of the catechol and protocatechuate branches of the β-ketoadipate pathway contributes to carbon source-dependent expression of the *Acinetobacter* sp. strain ADP1 *pobA* Gene. Applied and Environmental Microbiology, 2003. **69**(3): p. 1598-1606.

31. Zimmermann, T., et al., Role of *Acinetobacter baylyi* Crc in catabolite repression of enzymes for aromatic compound catabolism. Journal of bacteriology, 2009. **191**(8): p. 2834-2842.

32. Singh, A., et al., Removal of aromatic inhibitors produced from lignocellulosic hydrolysates by *Acinetobacter baylyi* ADP1 with formation of ethanol by *Kluyveromyces marxianus.* Biotechnol Biofuels, 2019. **12**: p. 91.

33. Kannisto, M.S., et al., Metabolic engineering of *Acinetobacter baylyi* ADP1 for removal of *Clostridium butyricum* growth inhibitors produced from lignocellulosic hydrolysates. Biotechnol Biofuels, 2015. **8**: p. 198.

34. Salvachúa, D., et al., Towards lignin consolidated bioprocessing: simultaneous lignin depolymerization and product generation by bacteria. Green Chemistry, 2015. **17**(11): p. 4951-4967.

35. Elliott, K.T., L.E. Cuff, and E.L. Neidle, Copy number change: evolving views on gene amplification. Future Microbiol, 2013. **8**(7): p. 887-99.

36. Sandegren, L. and D.I. Andersson, Bacterial gene amplification: implications for the evolution of antibiotic resistance. Nat Rev Microbiol, 2009. **7**(8): p. 578-88.

37. Albertson, D.G., Gene amplification in cancer. Trends Genet, 2006. **22**(8): p. 447-55.

38. Andersson, D.I. and D. Hughes, Gene amplification and adaptive evolution in bacteria. Annu Rev Genet, 2009. **43**: p. 167-95.

39. Bergthorsson, U., D.I. Andersson, and J.R. Roth, Ohno's dilemma: Evolution of new genes under continuous selection. Proceedings of the National Academy of Sciences, 2007. **104**(43): p. 17004-17009.

40. Näsvall, J., et al., Real-Time Evolution of New Genes by Innovation, Amplification, and Divergence. Science, 2012. **338**(6105): p. 384-387.

41. Reams, A.B. and E.L. Neidle, Genome plasticity in *Acinetobacter:* new degradative capabilities acquired by the spontaneous amplification of large chromosomal segments. Mol Microbiol, 2003. **47**(5): p. 1291-304.

42. Seaton, S.C., et al., Genome-wide selection for increased copy number in *Acinetobacter baylyi* ADP1: locus and context-dependent variation in gene amplification. Molecular Microbiology, 2012. **83**(3): p. 520-535.

43. Boerjan, W., J. Ralph, and M. Baucher, Lignin Biosynthesis. Annual Review of Plant Biology, 2003. **54**(1): p. 519-546.

44. Ragauskas, A.J., et al., Lignin Valorization: Improving Lignin Processing in the Biorefinery. Science, 2014. **344**(6185).

45. Beckham, G.T., et al., Opportunities and challenges in biological lignin valorization. Curr Opin Biotechnol, 2016. **42**: p. 40-53.

46. Freudenberg, K. and A.C. Neish, Constitution and biosynthesis of lignin. 1968, Berlin: Springer-Verlag. 129 pp.

47. Bonawitz, N.D., et al., Disruption of Mediator rescues the stunted growth of a lignin-deficient *Arabidopsis* mutant. Nature, 2014. **509**: p. 376.

48. Wilkerson, C.G., et al., Monolignol Ferulate Transferase Introduces Chemically Labile Linkages into the Lignin Backbone. Science, 2014. **344**(6179): p. 90-93.

49. Linger, J.G., et al., Lignin valorization through integrated biological funneling and chemical catalysis. Proceedings of the National Academy of Sciences, 2014. **111**(33): p. 12013-12018.

50. Seaton, S.C. and E.L. Neidle, Chapter 10 Using Aerobic Pathways for Aromatic Compound Degradation to Engineer Lignin Metabolism, in Lignin Valorization: Emerging Approaches. 2018, The Royal Society of Chemistry. p. 252-289.

51. Eltis, L.D., U. Karlson, and K.N. Timmis, Purification and characterization of cytochrome P450RR1 from *Rhodococcus rhodochrous.* Eur J Biochem, 1993. **213**(1): p. 211-6.

52. Machovina, M.M., et al., Enabling microbial syringol conversion through structure-guided protein engineering. Proceedings of the National Academy of Sciences, 2019. **116**(28): p. 13970-13976.

53. Karlsson, A., et al., X-ray crystal structure of benzoate 1,2-dioxygenase reductase from *Acinetobacter* sp. strain ADP1. J Mol Biol, 2002. **318**(2): p. 261-72.

54. Eby, D.M., et al., Characterization and evolution of anthranilate 1,2-dioxygenase from *Acinetobacter* sp. strain ADP1. J Bacteriol, 2001. **183**(1): p. 109-18.

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

56. Taylor, T.B., et al., First steps in experimental cancer evolution. Evolutionary Applications, 2013. **6**(3): p. 535-548.

57. Gresham, D. and J. Hong, The functional basis of adaptive evolution in chemostats. FEMS Microbiology Reviews, 2014. **39**(1): p. 2-16.

58. Horinouchi, T., et al., Development of an Automated Culture System for Laboratory Evolution. Journal of Laboratory Automation, 2014. **19**(5): p. 478-482.

59. Lenski, R.E., et al., Long-Term Experimental Evolution in *Escherichia coli*. I. Adaptation and Divergence During 2,000 Generations. The American Naturalist, 1991. **138**(6): p. 1315-1341.

60. Good, B.H., et al., The dynamics of molecular evolution over 60,000 generations. Nature, 2017. **551**(7678): p. 45-50.

61. Sniegowski, P.D., P.J. Gerrish, and R.E. Lenski, Evolution of high mutation rates in experimental populations of *E. coli.* Nature, 1997. **387**(6634): p. 703-5.

62. Maddamsetti, R., et al., Core Genes Evolve Rapidly in the Long-term Evolution Experiment with *Escherichia coli.* Genome Biol Evol, 2017.

63. Blount, Z.D., C.Z. Borland, and R.E. Lenski, Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli.* Proceedings of the National Academy of Sciences, 2008. **105**(23): p. 7899-7906.

64. van Dijk, E.L., et al., Ten years of next-generation sequencing technology. Trends Genet, 2014. **30**(9): p. 418-26.

65. Barrick, J.E. and R.E. Lenski, Genome dynamics during experimental evolution. Nature Reviews Genetics, 2013. **14**: p. 827.

66. Chou, H.H., et al., Diminishing returns epistasis among beneficial mutations decelerates adaptation. Science, 2011. **332**(6034): p. 1190-2.

67. Meyer, J.R., et al., Repeatability and contingency in the evolution of a key innovation in phage lambda. Science, 2012. **335**(6067): p. 428-32.

68. Wiser, M.J. and R.E. Lenski, A Comparison of Methods to Measure Fitness in *Escherichia coli.* PLOS ONE, 2015. **10**(5): p. e0126210.

69. Morgenthaler, A.B., et al., Mutations that improve the efficiency of a weak-link enzyme are rare compared to adaptive mutations elsewhere in the genome. bioRxiv, 2019: p. 624205.

70. Navarro Llorens, J.M., A. Tormo, and E. Martínez-García, Stationary phase in gram-negative bacteria. FEMS Microbiology Reviews, 2010. **34**(4): p. 476-495.

71. Gray, J.V., et al., "Sleeping beauty": quiescence in *Saccharomyces cerevisiae.* Microbiology and molecular biology reviews : MMBR, 2004. **68**(2): p. 187-206.

72. Lange, R. and R. Hengge-Aronis, Identification of a central regulator of stationary-phase gene expression in *Escherichia coli.* Mol Microbiol, 1991. **5**(1): p. 49-59.

73. Weber, H., et al., Genome-wide analysis of the general stress response network in *Escherichia coli*: sigmaS-dependent genes, promoters, and sigma factor selectivity. J Bacteriol, 2005. **187**(5): p. 1591-603.

74. Hengge-Aronis, R., Signal transduction and regulatory mechanisms involved in control of the sigma(S) (RpoS) subunit of RNA polymerase. Microbiol Mol Biol Rev, 2002. **66**(3): p. 373-95, table of contents.

75. Traxler, M.F., et al., The global, ppGpp-mediated stringent response to amino acid starvation in *Escherichia coli.* Molecular microbiology, 2008. **68**(5): p. 1128-1148.

76. Wolf, S.G., et al., DNA protection by stress-induced biocrystallization. Nature, 1999. **400**(6739): p. 83-5.

77. Russell, J.B. and G.M. Cook, Energetics of bacterial growth: balance of anabolic and catabolic reactions. Microbiological reviews, 1995. **59**(1): p. 48-62.

78. Yoshida, H., et al., The ribosome modulation factor (RMF) binding site on the 100S ribosome of *Escherichia coli.* J Biochem, 2002. **132**(6): p. 983-9.

79. Wada, A., et al., Ribosome modulation factor: stationary growth phase-specific inhibitor of ribosome functions from *Escherichia coli.* Biochem Biophys Res Commun, 1995. **214**(2): p. 410-7.

80. Ueta, M., et al., Role of HPF (hibernation promoting factor) in translational activity in *Escherichia coli.* J Biochem, 2008. **143**(3): p. 425-33.

81. Maki, Y., H. Yoshida, and A. Wada, Two proteins, YfiA and YhbH, associated with resting ribosomes in stationary phase *Escherichia coli.* Genes Cells, 2000. **5**(12): p. 965-74.

82. Hauser, R., et al., RsfA (YbeB) proteins are conserved ribosomal silencing factors. PLoS Genet, 2012. **8**(7): p. e1002815.

83. Walbot, V. and E.H. Coe, Nuclear gene iojap conditions a programmed change to ribosome-less plastids in *Zea mays.* Proc Natl Acad Sci U S A, 1979. **76**(6): p. 2760-4.

84. Rorbach, J., P.A. Gammage, and M. Minczuk, C7orf30 is necessary for biogenesis of the large subunit of the mitochondrial ribosome. Nucleic Acids Res, 2012. **40**(9): p. 4097-109.

85. Li, X., et al., Structure of Ribosomal Silencing Factor Bound to *Mycobacterium tuberculosis* Ribosome. Structure, 2015. **23**(12): p. 2387.

86. Mu, X., et al., The Effect of Colistin Resistance-Associated Mutations on the Fitness of *Acinetobacter baumannii*. Frontiers in Microbiology, 2016. **7**(1715).

87. da Costa, W.L.O., et al., Functional annotation of hypothetical proteins from the *Exiguobacterium antarcticum* strain B7 reveals proteins involved in adaptation to extreme environments, including high arsenic resistance. PLOS ONE, 2018. **13**(6): p. e0198965.

88. Sauvage, E., et al., The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. FEMS Microbiol Rev, 2008. **32**(2): p. 234-58.

89. Tipper, D.J. and J.L. Strominger, Mechanism of action of penicillins: a proposal based on their structural similarity to acyl-D-alanyl-D-alanine. Proceedings of the National Academy of Sciences of the United States of America, 1965. **54**(4): p. 1133-1141.

90. Goffin, C. and J.M. Ghuysen, Multimodular penicillin-binding proteins: an enigmatic family of orthologs and paralogs. Microbiol Mol Biol Rev, 1998. **62**(4): p. 1079-93.

91. Suzuki, H., Y. Nishimura, and Y. Hirota, On the process of cellular division in *Escherichia coli*: a series of mutants of *E. coli* altered in the penicillin-binding proteins. Proc Natl Acad Sci U S A, 1978. **75**(2): p. 664-8.

92. Denome, S.A., et al., *Escherichia coli* mutants lacking all possible combinations of eight penicillin binding proteins: viability, characteristics, and implications for peptidoglycan synthesis. J Bacteriol, 1999. **181**(13): p. 3981-93.

93. Levy, N., et al., Structural basis for *E. coli* penicillin binding protein (PBP) 2 inhibition, a platform for drug design. Journal of Medicinal Chemistry, 2019. **62**(9): p. 4742-4754.

94. Banzhaf, M., et al., Cooperativity of peptidoglycan synthases active in bacterial cell elongation. Mol Microbiol, 2012. **85**(1): p. 179-94.

95. Rhazi, N., et al., Catalytic mechanism of the *Streptomyces* K15 DD-transpeptidase/penicillin-binding protein probed by site-directed mutagenesis and structural analysis. Biochemistry, 2003. **42**(10): p. 2895-906.

96. Pao, S.S., I.T. Paulsen, and M.H. Saier, Jr., Major facilitator superfamily. Microbiology and molecular biology reviews : MMBR, 1998. **62**(1): p. 1-34.

97. Collier, L.S., N.N. Nichols, and E.L. Neidle, benK encodes a hydrophobic permease-like protein involved in benzoate degradation by Acinetobacter sp. strain ADP1. Journal of Bacteriology, 1997. **179**(18): p. 5943-5946.

98. Williams, P.A. and L.E. Shaw, *mucK*, a gene in *Acinetobacter calcoaceticus* ADP1 (BD413), encodes the ability to grow on exogenous cis,cis-muconate as the sole carbon source. J Bacteriol, 1997. **179**(18): p. 5935-42.

99. Nichols, N.N. and C.S. Harwood, PcaK, a high-affinity permease for the aromatic compounds 4-hydroxybenzoate and protocatechuate from *Pseudomonas putida*. J Bacteriol, 1997. **179**(16): p. 5056-61.

100. D'Argenio, D.A., et al., The physiological contribution of *Acinetobacter* PcaK, a transport system that acts upon protocatechuate, can be masked by the overlapping specificity of VanK. J Bacteriol, 1999. **181**(11): p. 3505-15.

101. Abramson, J., et al., Structure and mechanism of the lactose permease of *Escherichia coli.* Science, 2003. **301**(5633): p. 610-5.

102. Huang, Y., et al., Structure and mechanism of the glycerol-3-phosphate transporter from *Escherichia coli.* Science, 2003. **301**(5633): p. 616-20.

103. Rubin, R.A., et al., Gene duplication in the evolution of the two complementing domains of gram-negative bacterial tetracycline efflux proteins. Gene, 1990. **87**(1): p. 7-13.

104. Elbourne, L.D.H., et al., TransportDB 2.0: a database for exploring membrane transporters in sequenced genomes from all domains of life. Nucleic acids research, 2017. **45**(D1): p. D320-D324.

105. Brzoska, A.J., et al., Single-step selection of drug resistant *Acinetobacter baylyi* ADP1 mutants reveals a functional redundancy in the recruitment of multidrug efflux systems. PLOS ONE, 2013. **8**(2): p. e56090.

106. Krogh, A., et al., Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol, 2001. **305**(3): p. 567-80.

107. Chung, C.H. and A.L. Goldberg, The product of the *lon* (*capR*) gene in *Escherichia coli* is the ATP-dependent protease, protease La. Proceedings of the National Academy of Sciences of the United States of America, 1981. **78**(8): p. 4931-4935.

108. Botos, I., et al., The catalytic domain of *Escherichia coli* Lon protease has a unique fold and a ser-lys dyad in the active site. Journal of Biological Chemistry, 2004. **279**(9): p. 8140-8148.

109. Fu, G.K., M.J. Smith, and D.M. Markovitz, Bacterial protease Lon is a site-specific DNA-binding protein. Journal of Biological Chemistry, 1997. **272**(1): p. 534-538.

110. Rotanova, T.V., et al., Classification of ATP-dependent proteases Lon and comparison of the active sites of their proteolytic domains. European Journal of Biochemistry, 2004. **271**(23 -24): p. 4865

111. Griffith, K.L., I.M. Shah, and R.E. Wolf, Jr., Proteolytic degradation of *Escherichia coli* transcription activators SoxS and MarA as the mechanism for reversing the induction of the superoxide (SoxRS) and multiple antibiotic resistance (Mar) regulons. Mol Microbiol, 2004. **51**(6): p. 1801-16.

112. Shah, I.M. and R.E. Wolf, Jr., Sequence requirements for Lon-dependent degradation of the *Escherichia coli* transcription activator SoxS: identification of the SoxS residues critical to proteolysis and specific inhibition of in vitro degradation by a peptide comprised of the N-terminal 21 amino acid residues. J Mol Biol, 2006. **357**(3): p. 718-31.

113. Barrick, J.E., et al., Genome evolution and adaptation in a long-term experiment with *Escherichia coli.* Nature, 2009. **461**(7268): p. 1243-7.

114. Liu, Y., et al., RNA-seq reveals the critical role of Lon protease in stress response and *Brucella* virulence. Microbial Pathogenesis, 2019. **130**: p. 112-119.

115. Ching, C., et al., Lon protease has multifaceted biological functions in *Acinetobacter baumannii.* J Bacteriol, 2019. **201**(2).

CHAPTER 2

MODULATING STATIONARY PHASE REGULATION IMPROVES GROWTH ON A LIGNIN-DERIVED SUBSTRATE[1]

[1]Ahmed[*], A., Tumen-Velasquez, M., Schmidt, A.L., Duscent-Maitland, C.D., and Neidle, E.L.

To be submitted to *Molecular Microbiology*.

Introduction

Adaptive laboratory evolution has become an important tool for understanding evolutionary processes and adaptation of microorganisms under well-defined conditions [1]. Typically, mutants evolve during selection for improved growth or stress resistance for several hundred to several thousand generations. Microbial cells provide a good choice for such studies due to their relatively short generation times, fast growth rates, and large population sizes. In adaptive laboratory evolution, several mutations may be identified in a single selected mutant, some of which may have an obvious role in the new phenotype. However, those mutations with no clear role in conferring selective benefits may provide the most novel information. Evaluation of mutations is further complicated when considering interactions among different evolved alleles, such as synergistic effects, diminishing-returns epistasis [2], and all-or-none epistasis [3].

We previously reported the development of a new method for accelerated laboratory evolution, called Evolution by Amplification and Synthetic Biology (EASy) [4]. This method utilizes the genetic features of *Acinetobacter baylyi* ADP1 to induce gene amplification of precise chromosomal regions for accelerating laboratory evolution. The utility of EASy was demonstrated on the catabolism of guaiacol, a common component in pretreated lignin. Two genes from *Amycolatopsis* sp ATCC 39116 were introduced into the chromosome of *A. baylyi*. The two genes, *gcoAB* encode a two-component cytochrome P450 enzyme that catalyzes the demethylation of guaiacol to catechol [5]. A single chromosomal copy of *gcoAB* was insufficient to allow growth of *A. baylyi* on guaiacol as the sole carbon source, despite catechol being a natural growth substrate. Using EASy, a multicopy array of *gcoAB* was generated that allowed growth on guaiacol (Gua^+).

The first component in this system, GcoA is a cytochrome P450 enzyme that belongs to the CYP255A family. GcoA is 407 amino acids long and possesses a typical P450 single-domain architecture, with three phenylalanine residues in the active site cavity (Phe75, Phe169, and Phe395) responsible for positioning the aromatic ring of the substrate. In addition, structural studies showed that GcoA displays a broad substrate specificity, as evidenced by the accommodation of various substrates in its active site [5]. Recently, structure-guided design led to engineering a GcoA variant that can demethylate syringol, another important aromatic compound derived from lignin [6]. The second component of this system is GcoB, a three-domain reductase that is 334 amino acids long. Sequence information revealed that GcoB is homologous to three-domain reductases such as benzoate 1,2-dioxygenase reductase (BenC) [7] and anthranilate 1,2-dioxygenase reductase (AntC) from *A. baylyi* [8]. GcoB is composed of an N-terminal 2Fe-2S ferredoxin-like domain, a central flavin-binding domain, and a C-terminal domain that interacts with NAD(P)H. In typical bacterial cytochrome P450 systems, two distinct redox partners are required for transfer of electrons from NAD(P)H and therefore, GcoAB represents a new two-component class of P450 systems.

During laboratory evolution of Gua^+ populations for approximately 1,000 generations, improved *gcoAB* alleles emerged that allowed growth on guaiacol in a single chromosomal copy, after initially requiring multiple copies of *gcoAB.* In one variant, a fusion between GcoA and CatA (catechol 1,2-dioxygenase) was isolated. This chimeric protein represents an evolutionary outcome that would not have been rationally designed. This CatA-GcoA chimeric protein seems to provide a selective advantage via metabolite channeling. In addition, when chromosomally expressed in a different bacterium, *Pseudomonas putida*, this chimeric protein allowed growth on

guaiacol as sole carbon source, demonstrating the ability to use EASy-derived mutants for the generation of enzymes that can be successfully expressed in other organisms [4].

In a Gua^+ isolate, ACN1850, derived from a parent strain with multiple *gcoAB* copies, whole genome sequencing (WGS) identified two mutations in *gcoAB*. The first mutation causes an amino acid replacement from glycine to aspartate in GcoA at position 72. This position is not highly conserved in GcoA homologs and according to structural data, this residue does not seem to play a role in the catalytic activity. However, this residue lies in a helix that may interact with a loop that is involved in substrate access and release. The other mutation causes an amino acid replacement from alanine to threonine in GcoB at position 4. This replacement occurs at the 5' end of *gcoB* and does appear to affect the encoded ferredoxin-like N-terminal domain of GcoB.

To investigate whether either mutation is solely sufficient for growth on guaiacol, the two amino acid replacements were reconstructed in engineered mutants individually and in combination. We generated the following reconstructed strains, each with a single chromosomal copy of *gcoAB* encoding the following mutations: ACN1881 (*gcoA* mutation), ACN1863 (*gcoB* mutation), and ACN1886 (both *gcoAB* mutations) (Table 2.2). Each of these strains were otherwise analogous to the parent strain ACN1667 which has a single chromosomal copy of WT *gcoAB*. The three reconstructed strains were tested for growth on guaiacol, but despite frequent trials on plates and in liquid media, none of the strains utilized guaiacol as sole carbon source (Gua^-). This result led us to investigate other mutations in ACN1850 that might contribute to its Gua^+ phenotype.

WGS revealed seven mutations in ACN1850 outside the *gcoAB* region (Table 2.1). For all the genes listed, bioinformatic analysis and sequence information may provide a glimpse into the predicted functions of the gene products and their potential role in adaptation to growth on

guaiacol. For example, one gene encodes a putative transport protein. Transport is a critical function for the uptake of aromatic compounds, and mutations in this gene may affect guaiacol metabolism. Another gene encodes a putative posttranscriptional regulator. Metabolism of aromatic compounds is a tightly regulated process in order to prevent accumulation of toxic intermediates and ensure efficient consumption of carbon sources. Mutations in regulatory proteins may optimize regulation and enable improved growth on aromatic substrates. However, none of the products of the genes listed in Table 2.1 have been experimentally characterized in ADP1.

Here, we describe the analysis of some of the mutations identified in ACN1850. One mutation prevents the translation of a post-transcriptional regulator involved in stationary phase. Using competition assays, we provide evidence that lack of this regulator is beneficial under specific growth conditions. Another mutation occurs in an unexpected gene that encodes a protein involved in cell wall synthesis. Our analysis provides insights into new genetic targets for altering regulation for the purpose of improving the metabolism of lignin derived aromatic compounds. In many adaptive laboratory evolution experiments, focus is given to changes in the genes required for the selective conditions or in genes with clear roles in the adaptive process. However, those mutations with no obvious role or in previously uncharacterized genes may provide the most novel information. In addition, recent evidence suggests that beneficial mutations are more common in genes not directly involved in a selective condition, compensating for the need to improve a “weak-link” enzyme [9]. We present evidence that modulating protein synthesis regulation in stationary phase provides an advantage to cells growing on a lignin derived aromatic compound. While focus may be given to improvements in growth during exponential phase, it has been shown that stationary phase is a dynamic part of

bacterial growth [10]. Moreover, the growth advantage in stationary phase phenotype has n well documented, where aged cells outcompete young cells in a mixed culture [11, 12]. Our results highlight the important link between experimental conditions and the type of adaptive mutations that emerge. These results may be extended to other growth conditions, under which stress resistance and bypassing inhibition of translation may be beneficial.

Table 2.1. Mutations outside *gcoAB* identified in ACN1850 by whole genome sequencing[a,b]

Gene identifier (Old locus tag)[c]	Position of mutation[d]	Nucleotide change (Codon Change)	Amino acid change Position of change WT protein length	Predicted function
ACIAD_RS04905 (ACIAD1064)	1,061,077	C→T GGC→GAC Reverse strand	G→D 141 312	Alpha/beta hydrolase
ACIAD_RS05060 (ACIAD1101)	1,091,644	C→T UGG→UAG Reverse strand	W→Stop 366 675	Penicillin-binding protein 2
ACIAD_RS05110 (ACIAD1115)	1,103,385-1,103,423	39 bp deletion (reverse strand)	In frame deletion R237-E251 (13 aa) 808	DNA-binding ATP-dependent protease
ACIAD_RS06240 (ACIAD1354)	1,350,431	G→A UGG→UGA	W→Stop 385 386	phospholipase A
ACIAD_RS08335 (ACIAD1807)	1,814,453-1,814,499	47 bp deletion	Adds 5 residues and stop after Y238 386	prolyl oligopeptidase family serine peptidase
ACIAD_RS13910 (ACIAD3076)	3,003,892-3,004,098	207 bp deletion	Prevents translation 133	Ribosome silencing factor
ACIAD_RS15610 (ACIAD3450)	3,372,250	1 bp deletion	Adds 11 residues and stop after G345 454	MFS transporter

[a]Mutations were identified as differences between the sequence of the parent strain and the evolved strain

[b]Sequence changes were observed in 80% or more of the reads

[c]Old locus tags refer to GenBank accession CR543861.1

[d]Numbers correspond to positions in the ADP1 genome deposited in NCBI (NC_005966)

Results

A mutation in *gcoA* provides a weak benefit during selection for growth on guaiacol

Although the mutation in *gcoA* was insufficient to confer growth on guaiacol, other observations suggested this mutation could provide a benefit. When screening for spontaneous Gua^+ mutants, two Gua^- reconstructed strains, ACN1881 and ACN1886, that harbor the mutation in *gcoA*, gave rise to Gua^+ mutants within a two weeks incubation period at 30° C. In contrast, no spontaneous Gua^+ mutants were obtained from strains that have the wild-type sequence of *gcoAB* (ACN1667 and ACN2142), despite extended incubation. Similarly, no spontaneous Gua^+ mutants were obtained from a strain which harbors only the mutation in *gcoB* (ACN1863). The spontaneous Gua^+ mutants have either undergone spontaneous amplification of the *gcoAB* genes or harbored a deletion of the DNA upstream of *gcoA*, resulting in a genetic fusion of *catA* and *gcoA*. The latter result is similar to what we frequently observed in previously reported mutants [4]. In addition to allowing the isolation of spontaneous Gua^+ mutants, *in vitro* data suggest that the GcoA variant (GcoA-G72D) has more than two-fold increase in its catalytic efficiency relative to the wild-type protein. This result is primarily due to a reduction in the G72D variant's K_M [13]. Taken together, these observations suggest that while the *gcoA* mutation encoding the GcoA(G72D) variant is insufficient to enable growth on guaiacol, it provides some weak advantage that may enable the cells to survive on guaiacol plates until secondary beneficial mutations arise.

A mutation in a stationary phase related gene provides additional benefit

Because the *gcoA* mutation was insufficient for growth *in vivo*, the mutations involved in the ability of ACN1850 to grow on guaiacol are likely elsewhere in the chromosome. Therefore, we sought to investigate whether any of those other mutations may provide an additional benefit

that, in combination with the *gcoA* mutation, yields a Gua^+ phenotype or improves the ability to grow on guaiacol. Initial attempts to reconstruct strains with each of the mutations did not result in a Gua^+ phenotype. However, in only one case, we observed an effect of one mutation on the ability to grow on guaiacol. Figure 2.1 describes the construction of some of the strains used in this study and their relevant characteristics. More details on the genotype of strains can be found in Table 2.1. Two strains were constructed with an unmarked deletion of ACIAD_RS13910 (old locus tag ACIAD3076) and a single copy of either the WT *gcoAB* sequence (ACN2472), or a single copy of the ACN1850 mutated *gcoAB* (ACN2482). Neither of those two strains were able to grow on guaiacol. However, a reconstructed Gua^+ strain (ACN2484) that had a deletion of ACIAD_RS13910 was able to grow on a higher concentration of guaiacol (13 mM) compared to a similar Gua^+ reconstructed strain (ACN1739) with a wild-type allele of *rsfA*, which could only grow on up to 11 mM of guaiacol. Both ACN1739 and ACN2484 carry a *catA-gcoA* fusion that enables growth on guaiacol with a single chromosomal copy [4]. Sequence data predict the product of ACIAD_RS13910 to function as a ribosomal silencing factor (RsfA or RsfS). RsfA proteins are known to block protein synthesis during stationary phase and under nutrient-poor conditions [14]. In ACN1850, a 207-bp deletion in *rsfA* prevents the translation of the corresponding protein. Next, we measured the frequency of isolating spontaneous Gua^+ mutants from a reconstructed strain with a deletion of *rsfA* and one copy of the mutated *gcoA*, encoding the G72D variant (ACN2482). We compared this frequency to that of a comparable strain with a WT allele of *rsfA* (ACN1886). Both ACN1886 and ACN2484 yielded spontaneous Gua^+ mutants at similar frequencies, 1.4×10^{-9} and 1.8×10^{-9}, respectively. However, ACN2482 needed a shorter incubation period (6 – 8 days) than ACN1886 (12 – 17 days) for the Gua^+ mutants to arise. Taken together, it appears that while each mutation alone does not allow growth on

guaiacol, each provides a weak benefit to cells, but additional mutations might be needed to yield a Gua^+ phenotype.

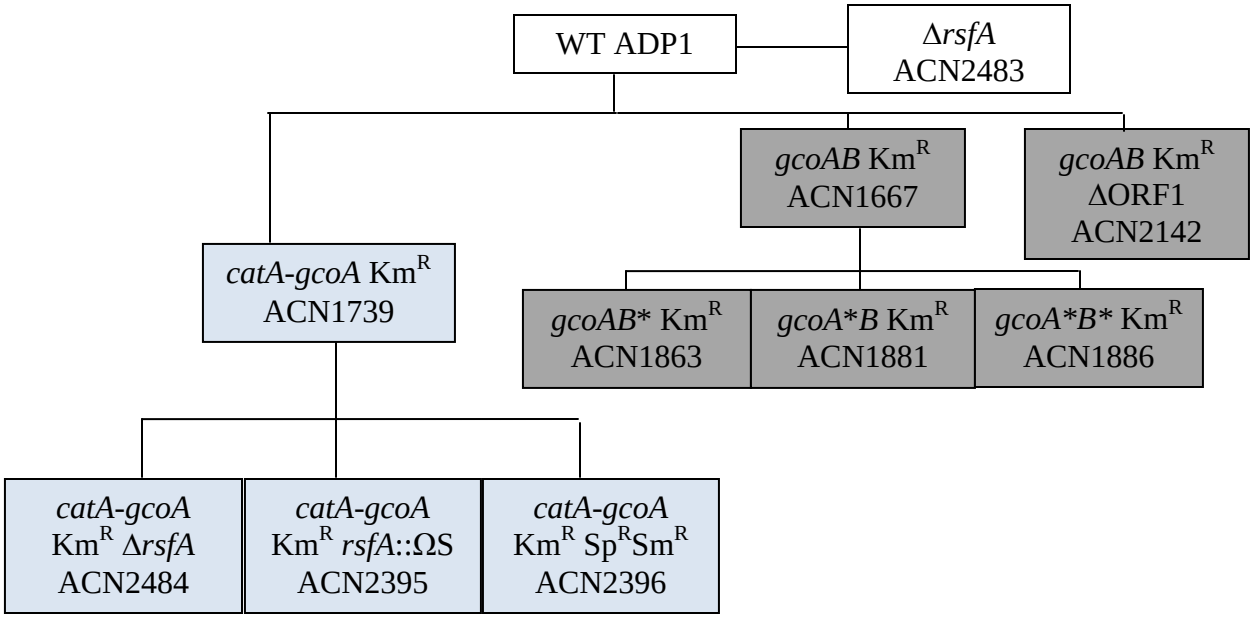


Figure 2.1. Construction and relevant characteristics of some mutants used in this study. The ability of strains to grow on guaiacol as the sole carbon source (Gua^+) is indicated by blue color, while the inability to utilize guaiacol is indicated by grey color. Strains either harbored a single chromosomal copy of *gcoAB* or a *catA-gcoA* fusion encoding a chimeric protein that enables growth on guaiacol. *gcoA** encodes GcoA(G72D) and *gcoB** encodes GcoB(A4T). ACN2142 differs from ACN1667 in the DNA region between *catA* and *gcoA*. ΩS indicates the omega drug-resistance cassette for Sp and Sm resistance. ACN2396 has the drug marker inserted in a different gene. See Table 2.2 for more information about the genotypes of the strains.

Lack of RsfA does not impair growth under standard lab conditions

In every report to date, *rsfA* mutants in *E. coli* have been shown either to suffer a competitive disadvantage relative to the wild type [14], or to show no impairment in growth [15], depending on the culture conditions. In the ADP1 single gene deletion collection, *rsfA* was found to be dispensable [16], a result confirmed by more recent evidence [17] and by our ability to generate *rsfA* null mutants. Therefore, the implications of our results that cells lacking RsfA may possess an advantage demanded further investigation. Because the mutant in which we identified a deletion of *rsfA*, was evolved during growth on guaiacol, we were interested in the potential benefit of this deletion under the same conditions. However, because RsfA has not been previously characterized in ADP1, we first sought to identify any phenotypes related to lack of RsfA under normal laboratory conditions. To this end, we compared the growth of ADP1 to strain ACN2483 which has an unmarked deletion of *rsfA* (Figure 2.2A). The two strains grew similarly when utilizing pyruvate as the carbon source. In addition, the two strains reached similar maximum optical densities after overnight growth on different carbon sources (Figure 2.2B). We concluded that, under standard laboratory growth conditions, lack of RsfA does not impair growth.

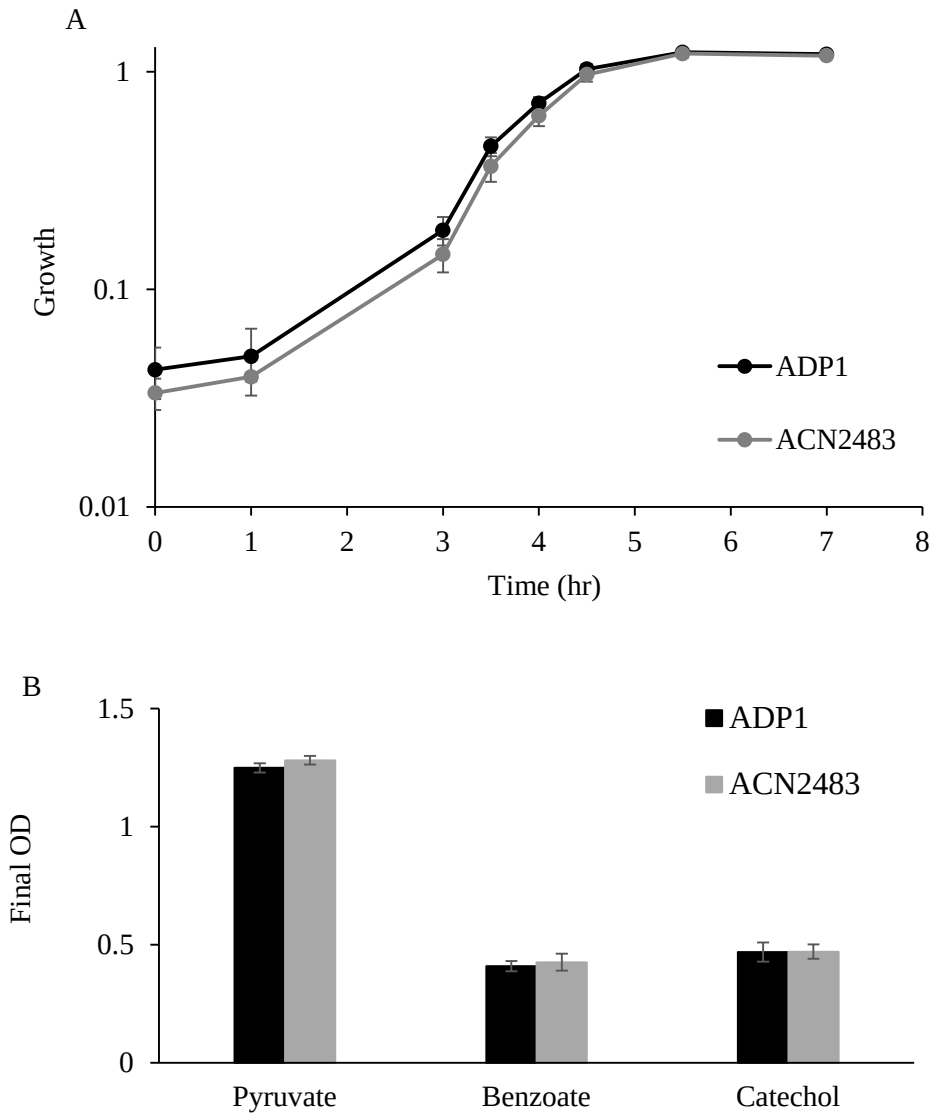


Figure 2.2. Lack of RsfA does not affect general growth characteristics under normal lab conditions. A) WT ADP1 and ACN2483 (Δ*rsfA*) show similar growth parameters when grown on 20 mM pyruvate as the carbon source. B) Final optical densities of the two strains on various carbon sources (20 mM pyruvate, 2 mM benzoate, and 2 mM catechol). OD measurements were taken after overnight growth. Results shown are averages of three replicates. Error bars represent one standard deviation.

***rsfA* mutants persist when grown on guaiacol**

We sought to identify differences in the ability to grow on guaiacol between Gua$^+$ strains that either had a wild-type *rsfA* or a deletion of that gene. When grown on guaiacol, the doubling time of a Gua$^+$ mutant strain with unmarked deletion of *rsfA* (ACN2484) was similar to that of a comparable strain (ACN1739) with a WT allele of *rsfA*. (Figure 2.3). In addition, no significant differences were observed between the growth curves of both strains (Figure 2.4).

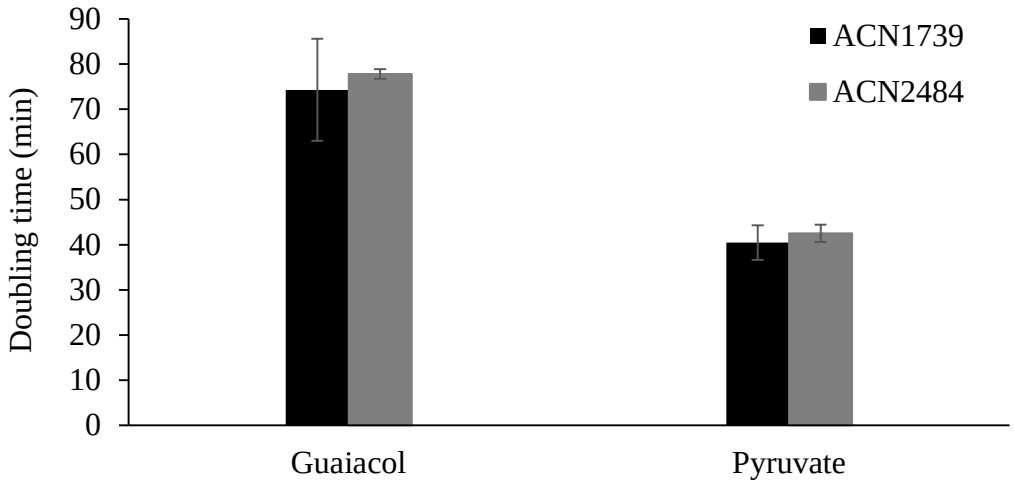


Figure 2.3. Cells lacking RsfA grow similarly to wild-type cells on guaiacol. Doubling times of ACN1739 (WT *rsfA*) and ACN2484 (Δ*rsfA*) on either 1 mM guaiacol or 20 mM pyruvate as the carbon source. The doubling times on pyruvate are comparable to WT ADP1 (37 +/- 3.4 minutes). Doubling times are averages of at least three replicates. Error bars represent one standard deviation.

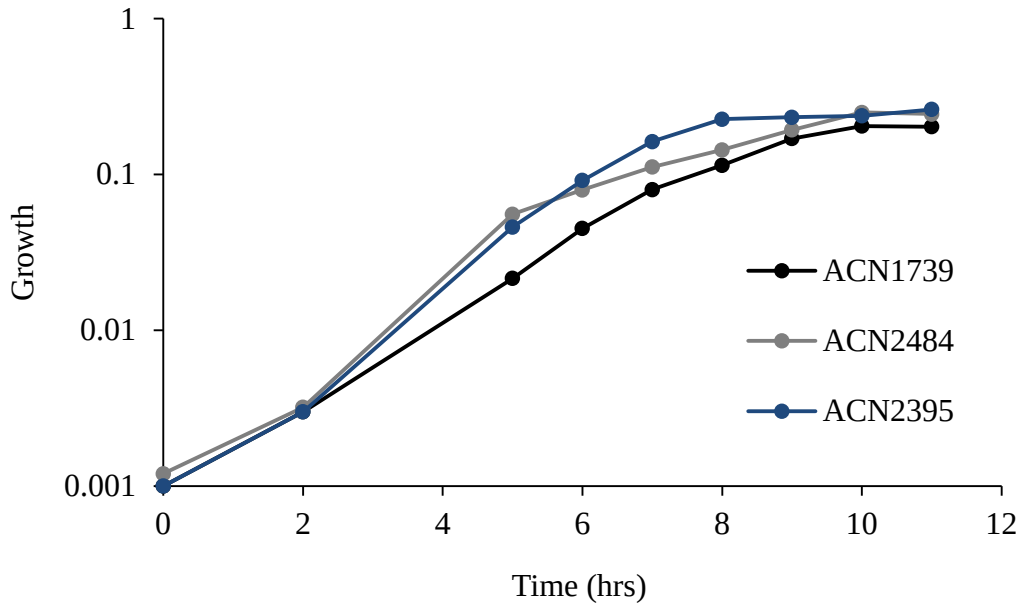


Figure 2.4. Insertion of a drug marker does not impact growth on guaiacol. Growth curves of ACN1739 (WT *rsfA*), ACN2484 (Δ*rsfA*), and ACN2395 (*rsfA*::ΩS) on 1 mM guaiacol as the carbon source. The three strains show comparable growth with both ACN2484 and ACN2395 showing overall slightly higher ODs. Consumption of guaiacol (not shown for clarity of figure) was similar for the three strains. The growth curve shown is representative of trends seen upon replication.

Because Gua$^+$ strains with a WT *rsfA* allele or with a deletion of *rsfA* grew similarly on guaiacol, we conducted a head-to-head competition assay between the two strains to identify any subtle differences in phenotype. To distinguish between the two competing strains, we used a strain with a drug marker disrupting *rsfA* (ACN2395) as our Gua$^+$ *rsfA* mutant. The other Gua$^+$ strain (ACN1739) had a WT *rsfA* allele. Prior to conducting the competition assays, we compared the final optical density (OD) of the two competing strains and a third mutant (ACN2396) in which the drug marker was inserted elsewhere in the chromosome (Figure 2.5). The final ODs of all strains were comparable.

The competing strains were independently grown in minimal medium with 1 mM guaiacol as the sole carbon source. After overnight growth, equal numbers of cells from each strain were mixed in a 1:1 ratio in fresh medium. At designated timepoints, we plated the competition mixture on plates with and without antibiotic. The fraction of viable cells of ACN2395 was determined as the ratio of colonies formed on plates containing antibiotic to the total number of colonies formed on antibiotic-free plates. As a control, we conducted another competition between the two strains using pyruvate as the carbon source. We accounted for the possibility that the drug marker insertion may affect the outcomes of the competition assay by conducting the same competitions as described below between ACN1739 and ACN2396. The latter strain has a WT *rsfA* gene and the same drug marker inserted elsewhere in the chromosome. The drug marker had no effect on the results we obtained. See Experimental Procedures and Figures 2.5 and 2.6 for more details.

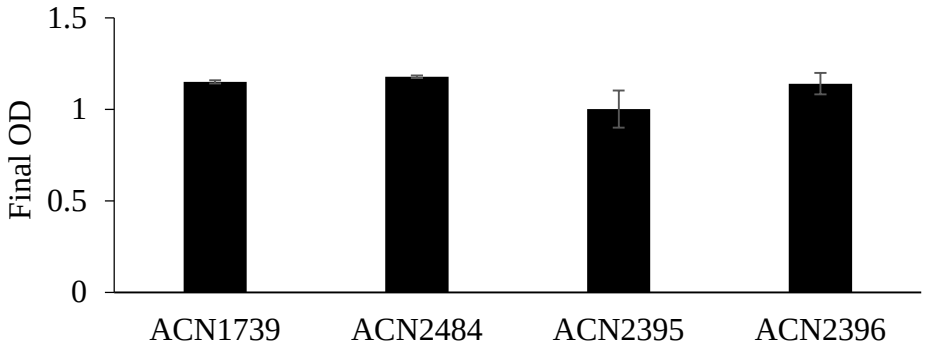


Figure 2.5. Insertion of a drug marker in *rsfA* does not affect final OD. Strains with wild-type *rsfA* (ACN1739), an unmarked deletion of *rsfA* (ACN2484), or *rsfA* disrupted with a drug marker (ACN2395) show similar final OD after overnight growth. An additional strain (ACN2396) an insertion of the same drug marker as ACN2395 but in a different gene. Doubling times are averages of at least three replicates. Error bars represent one standard deviation.

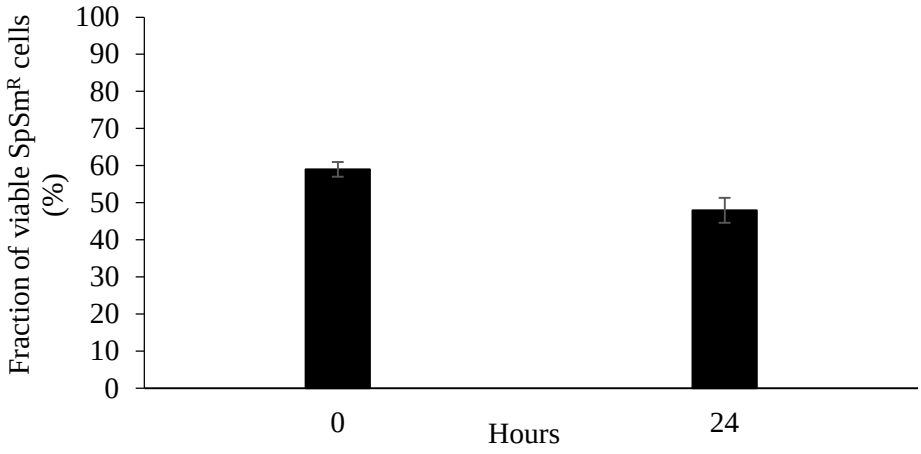


Figure 2.6. Insertion of a drug marker does not affect competition assay outcomes. Competition assay between ACN1739 (WT *rsfA*) and ACN2396 (WT *rsfA* and same drug marker as ACN2395 inserted in a different gene) on 20 mM pyruvate. The fraction of viable ACN2396 cells slightly decreases but remains at almost 50% of the population after 24 hours, unlike ACN2395 which drops to less than 20% (see Figure. 2.7).

When pyruvate was used as the carbon source, ACN2395 is outcompeted by ACN1739, decreasing to less than 20% of the population at the end of the competition. On the other hand, when guaiacol is used as the carbon source, ACN2395 remains around 50% of the population after 24 hours of the competition. In other words, on guaiacol ACN2395 does not suffer the same growth disadvantage it displays when grown on pyruvate. While the results of the competition assay on pyruvate agree with reports on *rsfA* mutants in other organisms [14], the ability of the *rsfA* mutant strain to survive during competition on guaiacol and to display no apparent disadvantage presents new questions about the role of this mutation during the adaptation of ACN1850 to growth on guaiacol. In addition, this observation opens the door to investigate the possibility of targeting *rsfA* and other ribosome hibernation genes for genetic engineering to improve fitness of evolving strains.

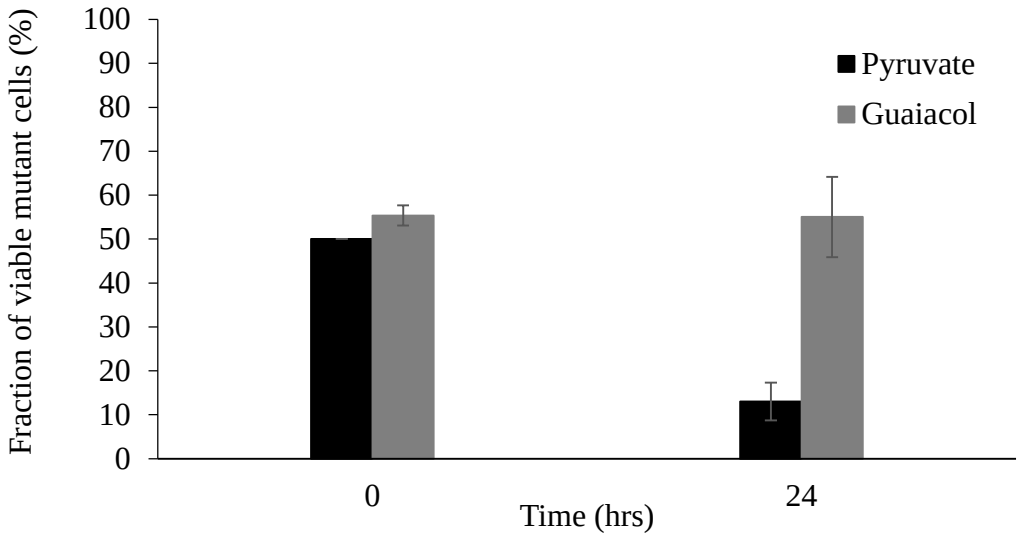


Figure 2.7. Cells lacking RsfA do not suffer a fitness disadvantage during growth on guaiacol. The fraction of viable ACN2395 (*rsfA*::ΩS) cells during growth on either 20 mM pyruvate (black) or 1 mM guaiacol (grey). Equal numbers of cells from ACN1739 and ACN2395 were mixed after overnight growth and plated immediately and after 24 hours of growth. On guaiacol, ACN2395 cells remain approximately 50% of the population. On pyruvate, mutant cells are outcompeted and constitute less than 20% of the population after 24 hours of growth. Results are averages of at least three replicates. Error bars represent one standard deviation.

Growth conditions affect adaptive mutations during evolution

Because RsfA is known to downregulate protein synthesis during stationary phase and under stressful conditions, we propose a model for the specific conditions under which a deletion of *rsfA* may have been selected (Figure 2.8). During the evolution of the population from which ACN1850 was derived, the cells were allowed to grow overnight and enter stationary phase before the subsequent serial transfer occurred. Assuming the existence of two subpopulations, where the first consists of cells harboring the wild-type allele of *rsfA* and the other subpopulation carries a deletion of *rsfA*, the two subpopulations are expected to have different protein synthesis profiles immediately before and after serial transfers. Cells with a WT *rsfA* allele would have an active RsfA that blocks protein synthesis upon entering stationary phase, while in cells lacking RsfA, protein synthesis should be less inhibited. Upon transfer to fresh medium, the cells lacking RsfA would have a competitive advantage as their protein synthesis machinery is active and can immediately utilize the nutrients available in the medium, while the WT cells should require a time period to transition from the physiological cellular state of stationary phase (where protein synthesis is repressed) to reactivate protein synthesis.

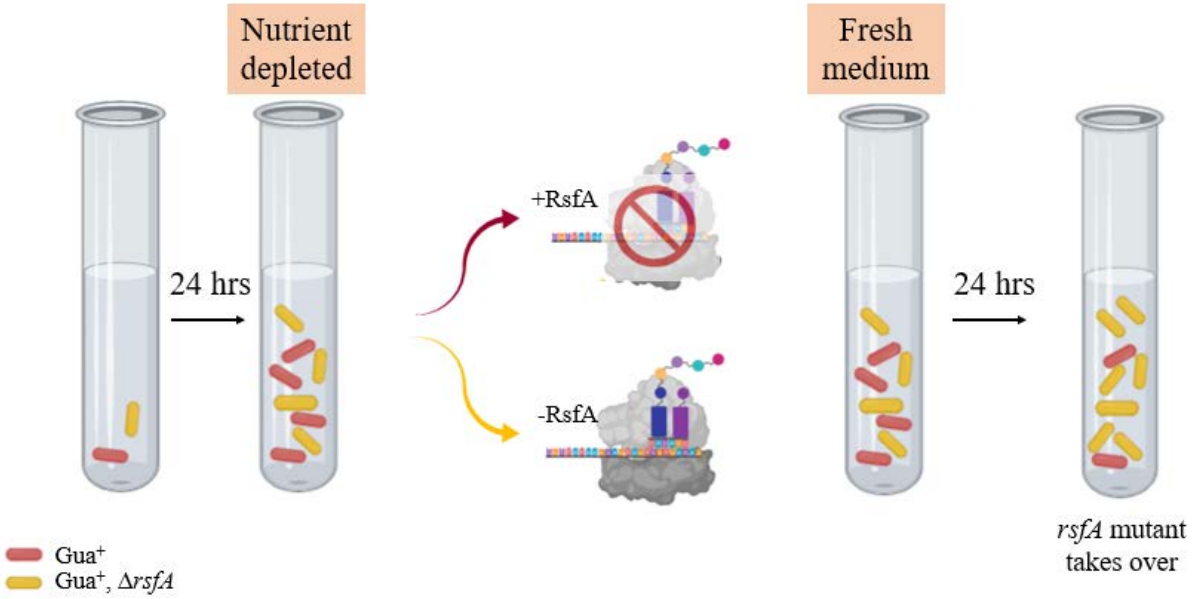


Figure 2.8. **Competitive advantage conferred by lack of RsfA selected during laboratory evolution**. In a previous EASy trial, a population was evolved to grow on guaiacol as sole carbon source. Serial transfers to fresh medium were done after cells reached stationary phase. The evolving population is assumed to be heterogenous, with cells belonging to either a subpopulation that carries a wild-type *rsfA* allele (red), or another subpopulation that has a deletion of *rsfA* (yellow). After 24 hours of growth (or in competition), cells with a functional RsfA should display repressed protein synthesis due to the blocking of 70S ribosome formation by RsfA. In contrast, cells lacking RsfA, protein synthesis should be de-repressed. As a result, when the population is transferred to fresh medium, cells lacking RsfA may be able to immediately utilize the newly available carbon source, while cells with a functional RsfA may require a transition period to reverse the action of ribosome silencing and form functional 70S ribosomes, before being able to utilize the carbon source. This advantage may allow *rsfA* mutants to take over the population after serial transfers. Image created using BioRender website.

Lack or RsfA is beneficial during transition from nutrient-depleted to nutrient-replete conditions

A typical competition assay for 24 hours did not lead to Δ*rsfA* cells taking over the population. However, unlike competition under other conditions, the mutant cells persisted after 24 hours in competition and remained at 50% of the population. To test the model proposed in Figure 2.8, the competition assay was repeated with serial transfers conducted either before the cells reached stationary phase (nutrient-replete), or after the cells entered stationary phase (nutrient-depleted). The results of both conditions were compared. We carried out the same competition between ACN1739 and ACN2395 but performed serial transfers after the initial 24 hours under culture conditions similar to those employed during the evolution of our EASy-derived populations [4]. We followed the fraction of viable ACN2395 cells in the mixed culture after serial transfers and observed a gradual increase in that fraction until ACN2395 reached more than 90% of the population within three transfers (Figure 2.9A). When serial transfers were performed during log phase, we observed the opposite; ACN1739 cells outgrew those of ACN2395. The latter cells constituted less than 20% of the population after two transfers. These results support the model proposed in Figure 2.8: the lack of RsfA is beneficial for cells upon transitioning from a nutrient poor medium (stationary phase, where carbon source is depleted) to a nutrient rich medium (fresh medium after serial transfers), but not the opposite. These results have important implications for linking the culture conditions used during laboratory evolution to mutations identified in evolved isolates, as well as implications for the design of batch cultures and when to transfer cells into fresh medium. In addition, these results show that, unlike previously reported in *E. coli* [14], *rsfA* mutants may display a fitness advantage upon poor to rich medium transitions.

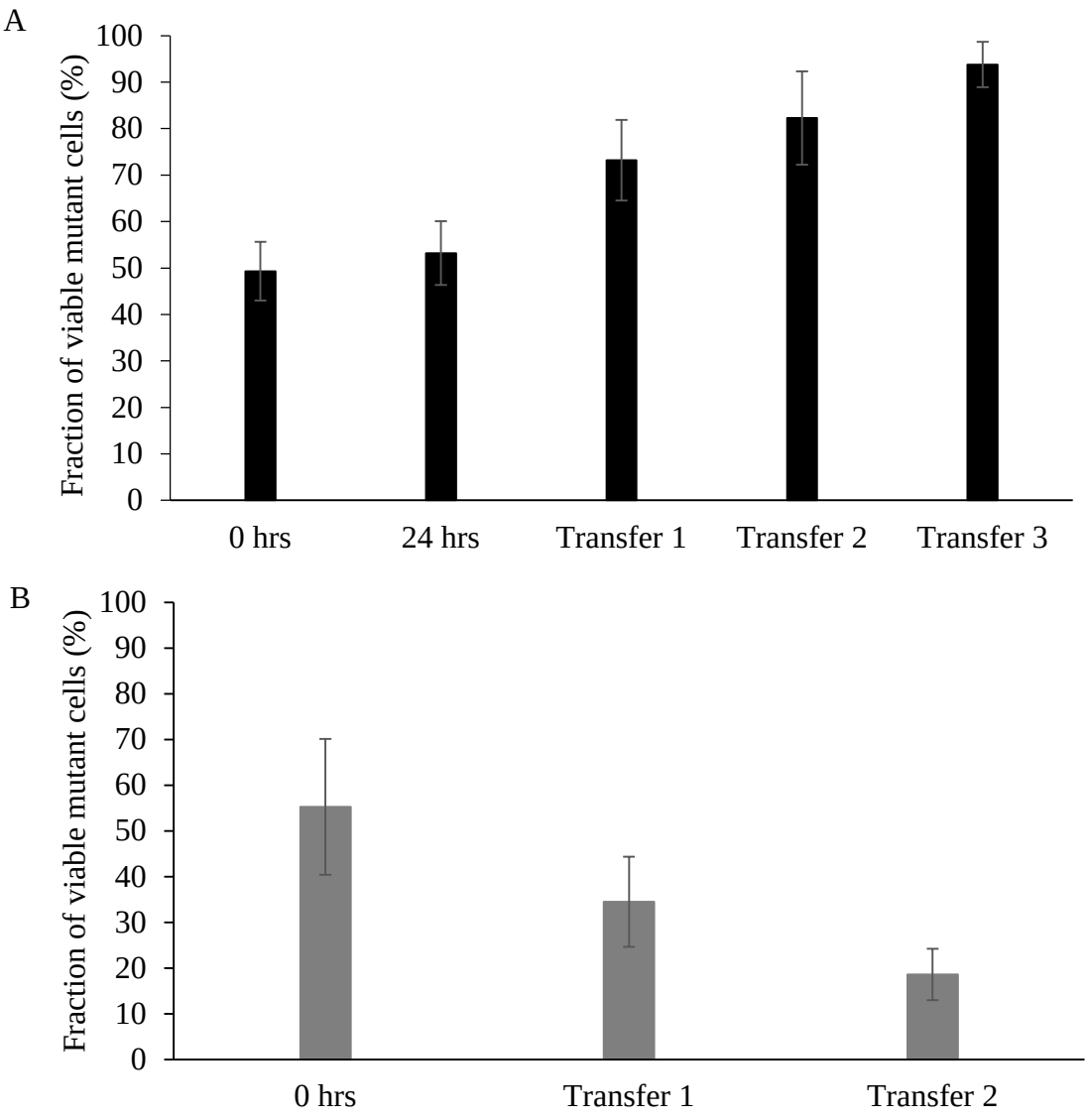


Figure 2.9. *rsfA* mutant takes over population during the transition from poor to rich medium. A) Competition assay between ACN1739 and ACN2395 conducted as described in Figure 2.6 with 1 mM guaiacol as the carbons source but with serial transfers done every 24 hours. ACN2395 reached more than 90% of the population within three transfers. Results are from five replicate experiments. B) The same competition assay as panel A with serial transfers done every 4 to 5 hours (log phase). In contrast to panel A, ACN2395 was outcompeted by ACN1739 within two transfers. Results are from three replicate experiments. For both panels, error bars represent one standard deviation.

Discussion

Adaptive laboratory evolution provides a tool to develop strains with desirable and improved phenotypes. During selection for a certain function, the fate of the gene encoding that function is followed. In an ideal scenario, mutations within the gene of interest may be identified and directly yield the improved phenotype upon reconstitution in a wild-type background. Another fortuitous scenario may yield mutations elsewhere in the chromosome in genes that encode well-characterized proteins. In both scenarios, investigative efforts are focused on those mutations that may display a clear role in the adaptive process. However, those mutations in uncharacterized genes, or that may not show an immediate and obvious role in the adaptive process may provide significant information. In our attempt to analyze the genetic basis of a mutant's ability to grow on guaiacol, we were presented with a less-than-ideal scenario. In ACN1850, mutations were identified within the genes under study, but those mutations were insufficient to confer a Gua^+ phenotype. WGS identified mutations in seven genes elsewhere in the chromosome. None of the mutated genes or their products have been previously characterized in ADP1, which presented both a challenging yet exciting task to distinguish the beneficial mutations from the neutral ones and to elucidate the mechanisms by which beneficial mutations provide an advantage under conditions of growth on guaiacol.

Stationary phase regulation plays an important role during laboratory evolution

In their natural habitat, bacteria rarely encounter an abundance of nutrients for long periods of time. It has been estimated that more than half of Earth's biomass consists of microorganisms in non-growing or "quiescent" state [18]. Since these conditions resemble the stationary phase, regulatory mechanisms enable the cells to appropriately respond to such conditions. For example, during nutrient starvation or upon entering stationary phase, a general

stress response can be activated that alters the expression profile of a large number of genes via the alternate sigma factor RpoS [19, 20]. In addition, because protein synthesis is one of the most energy-consuming cellular processes [21], several mechanisms govern the regulation of protein synthesis during different growth phases. In stationary phase, one such mechanism is ribosome hibernation [22]. This widespread phenomenon involves multiple proteins that stabilize ribosomes into inactive states. Among those proteins, ribosomal silencing factor A (RsfA), prevents the formation of 70S ribosomes by binding to the L14 ribosomal protein at the 50S-30S interface [14, 23]. To our knowledge, every report on *rsfA* mutants demonstrated that lack of a functional RsfA comes with a fitness cost. Here, we present evidence that during growth on guaiacol as the carbon source, *rsfA* mutants do not suffer from a fitness disadvantage relative to the wild-type cells (Figure 2.7). Furthermore, our results show that during laboratory evolution, when serial transfers are made after cells reach stationary phase the loss of RsfA is advantageous (Figure 2.9). We propose a model in Figure 2.8 that this advantage is derived from derepression of protein synthesis in cells lacking RsfA, which allows those cells to utilize the newly available carbon source, while wild-type cells require a small transitional period to reactivate protein synthesis. These results have implications that extend to the design of laboratory evolution experiments. Standard culture methods for laboratory evolution include chemostats and batch cultures. The latter typically employ serial transfers. While in some experiments serial transfers may be conducted during the exponential growth phase, these conditions may not necessarily reflect all selective pressures that cells may face during adaptation [24]. In our previous study, we allowed the evolving populations to reach stationary phase before the cells were transferred to fresh medium, which may have contributed to the selection of a mutation in *rsfA* [4].

The fact that the advantage of lacking RsfA was only evident when cells were grown on guaiacol can be attributed to the specific transition from a nutrient-depleted medium to fresh medium with abundant nutrients. Another possible explanation may stem from the differences in the pathways required for utilizing guaiacol or pyruvate as the carbon source. Guaiacol requires the function of *gcoAB*, foreign genes that require an elevated copy number to confer the ability to grow on guaiacol [4]. It is possible that a tradeoff exists between the cost and benefit of this mutation. Under certain conditions, the benefit of derepressing protein synthesis via loss of RsfA may outweigh the fitness cost. This may not be exclusive to growth on guaiacol and could be extended to other carbon sources and/or experimental conditions.

New engineering targets for lignin valorization

Lignin is the most abundant aromatic heteropolymer on earth. Because of its abundance and high energy content, lignin has attracted attention as a desirable source for renewable energy and the production of valuable chemicals, a process called lignin valorization. Currently, lignin is mostly discarded as waste in the paper and pulp industries or burned for process heat. This presents a roadblock towards the economic feasibility of lignin valorization [25]. Approaches to biological conversion of lignin capitalize on the ability of some organisms to convert a wide array of lignin-derived aromatic compounds into a few ring-cleavage substrates, in what has been termed “biological funneling” [26]. Because a single organism is unlikely to be capable of complete utilization of lignin, it is of great importance to engineer microbial strains with improved metabolism of lignin-derived aromatic compounds. We previously evolved ADP1-derived populations under conditions of growth on guaiacol, a key substrate derived from lignin. While we identified mutations in the genes encoding the function under study, they were insufficient for growth on guaiacol. Our whole-genome sequencing analysis revealed mutations

in genes that did not provide initial clues to their potential benefit in the adaptive process. Our main findings involved RsfA, a protein involved in the widely spread phenomenon of ribosome hibernation [27]. The advantageous phenotype we observed in *rsfA* mutants may not be specific to guaiacol metabolism. Similar observations may be observed under different culture conditions and may provide microbial cells with the ability to grow better and/or for longer periods on lignin derived aromatic compounds. Ribosome hibernation involves multiple proteins that may provide additional targets to overcome inhibition of protein synthesis during stationary phase. Overall, our results highlight the importance of identifying new engineering targets that modulate global regulation for lignin valorization. RsfA proteins have not been previously characterized in ADP1 or in the context of aromatic compound metabolism, and thus our results provide an exciting new avenue for identifying genetic targets for altering regulation in strains used for lignin valorization.

RsfA: Beyond ribosome silencing

In addition to their role in regulating protein synthesis in stationary phase, several studies suggest the involvement of RsfA proteins in other important and niche-specific processes. For example, a study looked at the resistance of *Acinetobacter baumannii* to colistin [28]. Colistin-resistant mutants were found to harbor mutations in genes known to play a role in altering the lipid A biosynthesis pathway, the target of colistin. However, colistin resistance resulted in a fitness cost in the mutants, which exhibited reduced growth rates and increased sensitivity to other drugs. In one isolate, this fitness deficit was restored due to a mutation in gene A1S_0570, which encodes *A. baumannii*'s RsfA ortholog. The authors used serial passaging and either increasing or constant concentrations of colistin during their experiment, which further highlights the importance of growth conditions in selecting for adaptive mutations.

Another study that aimed at improving functional annotation of hypothetical proteins in the psychrophile *Exiguobacterium antarcticum* found that a RsfA homolog could play a role in this organism's adaptation to extreme environments [29]. In addition, protein-protein interaction data from the same study predict that RsfA has a wide interaction network among the proteins examined. Most of those interactions involved protein synthesis processes. Interestingly, in *E. coli* the gene encoding RsfA is in the same operon as another gene orthologous to one of the mutated genes in ACN1850 (Table 2.1). This gene, ACIAD_RS05060 (old locus tag ACIAD1101) encodes a putative penicillin-binding protein and is predicted to function in the final stages of peptidoglycan synthesis. While the two genes are not in close proximity in the ADP1 genome, this observation provides further investigation routes to identify interactions between the two genes or their products in ADP1.

Developing tools for combinatorial analysis of adaptive mutations

A systematic approach to the analysis of adaptive mutations entails reconstructing each mutation in a wild-type genetic background and screening for the desired phenotype. This approach is indeed an obstacle to identifying the significance of a mutation, or a combination of mutations, where our ability to identify mutations by whole-genome sequencing exceeds our ability to address their relative significance in an evolutionary scenario. Moreover, the lack of experimental evidence on many genes and their products, presents another challenge when prioritizing the investigation of specific mutations. Regardless of their location in the chromosome, beneficial mutations may yield an improved phenotype. in our case, all mutations occurred in genes that have not been previously characterized in ADP1. In addition, sequence information provided no clue as to which mutations(s) might be beneficial and which are neutral. While we initially combined two of the mutations via established ADP1 genome engineering

methods (ACN2482), this combination did not yield a Gua^+ phenotype. Reconstituting seven mutations in all possible combinations, while theoretically possible, is a resource and time intensive task that would require the generation of a large number of mutants, especially in model organisms. Therefore, a need is evident for developing tools that allow rapid analysis of the significance of combinations of mutations, without requiring any *a priori* information on the functions encoded by uncharacterized genes. The highly efficient natural transformation and recombination system of ADP1 provides a platform for developing such analytical tools. With the ability to use linearized DNA to transform ADP1 recipients, a method can be envisioned that employs combining multiple DNA fragments, generated by PCR or linearized plasmids to transform a recipient and directly screen for the desired phenotype. Similar approaches have been previously used in studies on gene amplification in ADP1 [30, 31]. A transformation-based assay may provide a tool that could allow the rapid deciphering of the significance of combinations of mutations.

Experimental Procedures

Bacterial strains, plasmids, and growth Conditions

All strains and plasmids used in this study and their relevant characteristics are listed in tables 2.2 and 2.3, respectively. *Escherichia coli* XL1-Blue competent cells were used as plasmid hosts. *E. coli* strains were grown in Lysogeny Broth (LB) medium at 37°C [32]. *A. baylyi* strains were grown in LB or minimal medium (MM) at 37°C with aeration. The composition of MM is as follows: 2.5% 0.5 M KH_2PO_4, 25% 0.5 M Na_2HPO_4, 1% of a 10% w/v $(NH_4)_2SO_4$ and 0.1% concentrated base solution. For solid media, 1.5% agar was added before autoclaving. Carbon sources were filter sterilized and added aseptically to liquid media after autoclavation. Cultures of *A. baylyi* strains growing on guaiacol were incubated at 30°C. Carbon sources were provided at the following final concentrations, unless otherwise noted: 1 mM guaiacol, 20 mM pyruvate, 2 mM benzoate, and 2 mM catechol. Antibiotics were added as needed to the following final concentrations: Ampicillin (Ap), 150 μg/ml, kanamycin (Km), 25 μg/ml, spectinomycin (Sp), 12.5 μg/ml, and streptomycin (Sm), 12.5 μg/ml. For selection of amplification mutants, 1 mg/ml of Km was used. For growth curves, single colonies were used to start liquid cultures in 5 ml total volume of minimal medium and the respective carbon source. After overnight growth, cells were diluted 1:100 (1:50 for guaiacol) by volume in fresh medium. Growth was monitored as optical density at 600 nm. For final OD measurement, strains were grown overnight, and OD was recorded after 18 – 20 hours of growth.

Genome modification of *A. baylyi*

Recipient strains were transformed with linearized plasmid DNA. Transformed colonies were selected based on drug resistance or the loss of a *sacB* marker in the presence of 10% sucrose. Successful transformants were confirmed by PCR analysis and/or DNA sequencing.

Competition assays

Cultures of ACN1739 (WT *rsfA*) and either ACN2395 (Δ*rsfA*) or ACN2396 (WT *rsfA*, Sp^R, Sm^R) were grown separately overnight in minimal medium and either 20 mM pyruvate or 1 mM guaiacol as the carbon source. At an optical density of approximately 1, equal numbers of cells from both cultures were mixed in a 5 ml total volume of MM with the respective carbon source. Immediately after mixing and at after 24 hours of incubation, the competition mixture was serially diluted and 100 µl of a dilution that yields approximately 20 to 200 colonies were plated in triplicate on medium containing the respective carbon source with and without 12.5 µg/ml Sm and Sp. The two strains were distinguishable by the antibiotic marker. The Δ*rsfA* mutant is Sm^RSp^R, the WT strain is not. The number of viable cells of the Δ*rsfA* mutant was determined at each time point by counting colonies on the plates containing the antibiotics. To estimate the fraction of viable Δ*rsfA* cells in the population, we used the ratio of the number of colonies on plates with antibiotic to the total number of colonies on plates without the antibiotic. Counts from at least 3 plates per time point were averaged. To ensure plating efficiency did not affect the results, approximately 150 colonies from each time point were patched from antibiotic-free plates to plates with and without antibiotics and the results conformed to the numbers obtained from direct colony counts. To incorporate serial dilutions, the procedure mentioned above was followed but after 24 hours of incubation, the competition mixture was transferred to fresh medium and allowed to grow for another 24 hours before plating again. For serial transfers during log phase, the competition mixture was transferred to fresh medium every 4 to 5 hours which corresponds to early log phase.

Whole Genome Sequencing

Genomic DNA was isolated from overnight cultures using the NucleoSpin Tissue kit (Macherey-Nagel). Approximately 1 µg of gDNA was fragmented by sonication to yield a fragment size of 300-500 bp. End repair, A-tailing, and adapter ligation reactions were performed on fragmented DNA using the NEBNext Ultra II kit (New England Biolabs), according to the manufacturer's protocols. Illumina sequencing was performed on a NextSeq500 device at the Georgia Genomics Facility (University of Georgia). Analysis of the raw sequence data and variant calling was performed using Geneious software with default settings [33]. The minimum variant frequency was set to 0.8 for identification of single nucleotide polymorphisms. Duplications or larger amplifications were visually identified by comparing the average coverage to the coverage of specific regions of interest. Reference genomes were constructed by mapping whole genome sequencing reads of single-copy parent strain ACN1667 to ADP1 genome (NCBI accession no. NC_005966).

Frequency of spontaneous Gua^+ mutants

A single colony from each Gua^- parent strain (ACN1667, ACN1886, ACN2143, ACN2472, or ACN2482) was used to start a liquid culture in minimal medium and 20 mM pyruvate. At mid log phase, one ml of the liquid culture was centrifuged and the supernatant discarded. The cell pellet was washed twice in minimal medium to remove any remaining carbon source. The pellet was resuspended in 100 ml of minimal medium. After serial dilution of the cell suspension, approximately $10^7 – 10^8$ cells were plated on guaiacol plates in triplicate and incubated at 30°C. In addition, cells were plated on pyruvate plates in triplicate to determine the total number of cells. Guaiacol plates were incubated up to three weeks. The frequency of appearance of spontaneous Gua^+ mutants from each strain was determined as the ratio of the number of Gua^+ colonies to the total number of cells.

Table 2.2. Strains used in this study

Strain	Relevant characteristics Donor DNA/Restriction enzyme, transforms (X) recipient strain	Source
ADP1	Wild type (BD413)	[34, 35]
ACN1667	ACIAD1443::*gcoAB51661;* Km^R*51667* (ΩK^a cassette inserted downstream of *gcoB*), Gua^- parent strain single copy of *gcoAB* region pBAC1261/EcoRI X ACN1661 selected by Km^R	[4]
ACN1676	ACIAD1443::*gcoAB51661;* Km^R*51667;* multiple chromosomal copies of a 9.7 kbp amplicon including *gcoAB*; SBF*51676,* pBAC1262/EcoRI X ACN1667 selected by high-level Km^R, Gua^-	[4]
ACN1686	ACIAD1443::*gcoAB51661;* Km^R*51667;* SBF*51676* multiple copies of 9.7 kbp amplicon Derived from ACN1676 by direct selection on a guaiacol plate, Gua^+	[4]
ACN1739	*catA-gcoA51738*; *gcoB51661*; Km^R*51667;* Reconstructed strain, Gua^+ pBAC1314/NdeI X ADP1 selected by Km^R	[4]
ACN1850	ACIAD1443::*gcoA51850* [encodes GcoA(G72D)], *gcoB51850* [encodes GcoB(A4T)]; Km^R*51667;* EASy-derived Gua^+ isolate from evolving population of ACN1686 Mutations identified in genome outside the *gcoAB* region	[4]
ACN1863	ACIAD1443::*gcoA51661*,*gcoB51850* [encodes GcoB(A4T)]; Km^R*51667;* Reconstructed strain, Gua^- pBAC1446/NdeI X ADP1 selected by Km^R	[4]
ACN1881	ACIAD1443::*gcoA51850* [encodes GcoA(G72D)], *gcoB51661*; Km^R*51667;* Reconstructed strain, Gua^- pBAC1456/NdeI X ADP1 selected by Km^R	[4]
ACN1886	ACIAD1443::*gcoA51850* [encodes GcoA(G72D)], *gcoB51850* [encodes GcoB(A4T)]; Km^R*51661;* Reconstructed strain, Gua^- pBAC1459/NdeI X ADP1 selected by Km^R	[4]
ACN2142	*gcoAB51667;* Km^R*51667;*(*gcoAB* – ΩK^a cassette replaces ACIAD1443), Gua^- pBAC1618/NdeI X ADP1 selected by Km^R	This study
ACN2143	*gcoA51850* [encodes GcoA(G72D)], *gcoB51661*; Km^R*51667;*(*gcoAB* - ΩK^a cassette replace ACIAD1443) pBAC1619/NdeI X ADP1 selected by Km^R	This study
ACN2395	*catA-gcoA51738*; *gcoB51661*; Km^R*51667;* ACIAD3076::ΩS^b, Gua^+ pBAC1633/NdeI X ACN1739[b] selected by Sm^R	This study
ACN2396	*catA-gcoA51738*; *gcoB51661*; Km^R*51667;* ACIAD1542::ΩS^b,	This study

	Gua^+ pBAC1175/NdeI X ACN1739 selected by Sm^R	
ACN2452	ACIAD3076::*sacB*-$Sm^R Sp^R$*52452;* counter-selectable marker inserted downstream of *hyi* to facilitate subsequent strain construction pBAC1769/NdeI X ADP1 selected by Sm^R	This study
ACN2453	ACIAD1443::*gcoA51850* [encodes GcoA(G72D)], *gcoB51850* [encodes GcoB(A4T)]; Km^R*51661,* ACIAD3076::*sacB*- $SpSm^R$ *52452 ;* counter-selectable marker inserted downstream of *hyi* to facilitate subsequent strain construction*;* Gua^- pBAC1769/NdeI X ACN1886 selected by Sm^R	This study
ACN2467	*catA-gcoA51738*; *gcoB51661*; Km^R*51667;* ACIAD3076::*sacB*- $Sp^R Sm^R$ *52452 ;* counter-selectable marker inserted downstream of *hyi* to facilitate subsequent strain construction, Gua^+ pBAC1769/NdeI X ACN1739 selected by Sm^R	This study
ACN2482	ACIAD1443::*gcoA51850* [encodes GcoA(G72D)], *gcoB51850* [encodes GcoB(A4T)]; Km^R*51661,*ΔACIAD3076*;* Gua^- pBAC1768/NdeI X ACN2453 selected by growth in the presence of 10% sucrose	This study
ACN2483	ΔACIAD3076 pBAC1768/NdeI X ACN2452 selected by growth in the presence of 10% sucrose	This study
ACN2484	*catA-gcoA51738*; *gcoB51661*; Km^R*51667;* ΔACIAD3076; Gua^+ pBAC1768/NdeI X ACN2467 selected by growth in the presence of 10% sucrose	This study
ACN2484	*catA-gcoA51738*; *gcoB51661*; Km^R*51667;* ΔACIAD3076; Gua^+ pBAC1768/NdeI X ACN2467 selected by growth in the presence of 10% sucrose	This study

[a]ΩK indicates the omega drug-resistance cassette for Km^R from pUI1637 [36]

[b]ΩS indicates the omega drug-resistance cassette for Sp and Sm resistance from pHP45 [37]

Table 2.3. Plasmids used in this study

Plasmid	Relevant characteristics	Source
pUC18, pUC19	Ap^R, cloning vectors	[38]
pUI1637	Source of omega Km^R cassette	[36]
pUI1638	Source of omega $SpSm^R$ cassette	[36]
pBAC1282	Ap^R, Km^R; This plasmid, linearized with XbaI and EcoRV, allows capture of the *A. baylyi* chromosomal region containing *gcoAB* using the gap-repair method.	[4]
pBAC1314	Ap^R, Km^R; *catA-gcoA51738*; *gcoB51661*; Km^R*51667;* DNA recovered from ACN1738 by the gap repair method using linearized pBAC1282	[4]
pBAC1446	Ap^R, Km^R; *gcoB51850*; [encodes GcoB(A4T)] made by replacing the SbfI-PspOMI fragment of pBAC1261 with comparable DNA carrying the *gcoB* mutation discovered in ACN1850	[4]
pBAC1456	Ap^R, Km^R; ACIAD1443::*gcoA51850* ; [encodes GcoB(A4T)] made by replacing the SbfI-BsiWI fragment of pBAC1261 with comparable DNA carrying the *gcoA* mutation discovered in ACN1850	[4]
pBAC1459	Ap^R, Km^R; ACIAD1443::*gcoA51850; gcoB51850* [encodes GcoA(G72D) and GcoB(A4T)] constructed by replacing the BbvCI-BsiWI fragment of pBAC1261 with comparable DNA carrying both mutations discovered in ACN1850	[4]
pBAC1580	Ap^R, Km^R; (ACIAD3076) amplified from ADP1 gDNA with primers ALS157 and ALS158 and cloned in pCR4-TOPO vector	This study
pBAC1618	Ap^R, Km^R; *gcoAB51667;* Km^R*51667* (*gcoAB* – ΩK^a cassette replaces ACIAD1443) constructed by replacing the SwaI-SbfI fragment of pBAC1284 with a PCR generated *gcoAB* DNA fragment	This study
pBAC1619	Ap^R, Km^R; *gcoA51850* [encodes GcoA(G72D)], *gcoB51661;* Km^R*51667* (*gcoAB* – ΩK^a cassette replaces ACIAD1443) constructed by replacing the SwaI-SbfI fragment of pBAC1284 with a PCR generated *gcoAB* DNA fragment	This study
pBAC1633	Ap^R, Km^R, Sp^R, Sm^R; ΩS cassette from pHP45 [37] inserted as a BamHI fragment in pBAC1580	This study
pBAC1768	Ap^R, ΔACIAD_RS13910 (ΔACIAD3076), *in vivo* assembly of two ADP1 DNA fragments and a pUC18 fragment to generate an unmarked deletion of ACIAD3076.	This study
pBAC1769	Ap^R, Sp^RSm^R, *sacB*-Sp^RSm^R, linearized fragment containing the counter selectable marker ligated to pBAC1768 linearized with PsiI	This study

References

1. Dragosits, M. and D. Mattanovich, Adaptive laboratory evolution – principles and applications for biotechnology. Microbial Cell Factories, 2013. **12**(1): p. 64.
2. Chou, H.H., et al., Diminishing returns epistasis among beneficial mutations decelerates adaptation. Science, 2011. **332**(6034): p. 1190-2.
3. Meyer, J.R., et al., Repeatability and contingency in the evolution of a key innovation in phage lambda. Science, 2012. **335**(6067): p. 428-32.
4. Tumen-Velasquez, M., et al., Accelerating pathway evolution by increasing the gene dosage of chromosomal segments. Proceedings of the National Academy of Sciences, 2018. **115**(27): p. 7105-7110.
5. Mallinson, S.J.B., et al., A promiscuous cytochrome P450 aromatic O-demethylase for lignin bioconversion. Nature Communications, 2018. **9**(1): p. 2487.
6. Machovina, M.M., et al., Enabling microbial syringol conversion through structure-guided protein engineering. Proceedings of the National Academy of Sciences, 2019. **116**(28): p. 13970-13976.
7. Karlsson, A., et al., X-ray crystal structure of benzoate 1,2-dioxygenase reductase from *Acinetobacter* sp. strain ADP1. J Mol Biol, 2002. **318**(2): p. 261-72.
8. Eby, D.M., et al., Characterization and evolution of anthranilate 1,2-dioxygenase from *Acinetobacter* sp. strain ADP1. J Bacteriol, 2001. **183**(1): p. 109-18.
9. Morgenthaler, A.B., et al., Mutations that improve the efficiency of a weak-link enzyme are rare compared to adaptive mutations elsewhere in the genome. bioRxiv, 2019: p. 624205.

10. Navarro Llorens, J.M., A. Tormo, and E. Martinez-Garcia, Stationary phase in gram-negative bacteria. FEMS Microbiol Rev, 2010. **34**(4): p. 476-95.
11. Zambrano, M., et al., Microbial competition: *Escherichia coli* mutants that take over stationary phase cultures. Science, 1993. **259**(5102): p. 1757-1760.
12. Finkel, S.E., Long-term survival during stationary phase: evolution and the GASP phenotype. Nature Reviews Microbiology, 2006. **4**(2): p. 113-120.
13. Machovina, M.M., Enzymatic strategies for controlling and harnessing the oxidative power or O2 (Doctoral Dissertation). 2018.
14. Hauser, R., et al., RsfA (YbeB) proteins are conserved ribosomal silencing factors. PLoS Genet, 2012. **8**(7): p. e1002815.
15. Baba, T., et al., Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Molecular systems biology, 2006. **2**: p. 2006.0008-2006.0008.
16. de Berardinis, V., et al., A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1. Molecular systems biology, 2008. **4**: p. 174-174.
17. Suárez, G.A., et al., Rapid and assured genetic engineering methods applied to *Acinetobacter baylyi* ADP1 genome streamlining. bioRxiv, 2019: p. 754242.
18. Gray, J.V., et al., "Sleeping beauty": quiescence in *Saccharomyces cerevisiae*. Microbiology and molecular biology reviews : MMBR, 2004. **68**(2): p. 187-206.
19. Lange, R. and R. Hengge-Aronis, Identification of a central regulator of stationary-phase gene expression in *Escherichia coli*. Mol Microbiol, 1991. **5**(1): p. 49-59.

20. Weber, H., et al., Genome-wide analysis of the general stress response network in *Escherichia coli*: sigmaS-dependent genes, promoters, and sigma factor selectivity. J Bacteriol, 2005. **187**(5): p. 1591-603.

21. Russell, J.B. and G.M. Cook, Energetics of bacterial growth: balance of anabolic and catabolic reactions. Microbiological reviews, 1995. **59**(1): p. 48-62.

22. Yoshida, H., et al., The ribosome modulation factor (RMF) binding site on the 100S ribosome of *Escherichia coli.* J Biochem, 2002. **132**(6): p. 983-9.

23. Li, X., et al., Structure of Ribosomal Silencing Factor Bound to *Mycobacterium tuberculosis* Ribosome. Structure, 2015. **23**(12): p. 2387.

24. Kram, K.E., et al., Adaptation of *Escherichia coli* to Long-Term Serial Passage in Complex Medium: Evidence of Parallel Evolution. mSystems, 2017. **2**(2): p. e00192-16.

25. Ragauskas, A.J., et al., Lignin Valorization: Improving Lignin Processing in the Biorefinery. Science, 2014. **344**(6185).

26. Linger, J.G., et al., Lignin valorization through integrated biological funneling and chemical catalysis. Proceedings of the National Academy of Sciences, 2014. **111**(33): p. 12013-12018.

27. Prossliner, T., et al., Ribosome Hibernation. Annu Rev Genet, 2018. **52**: p. 321-348.

28. Mu, X., et al., The Effect of Colistin Resistance-Associated Mutations on the Fitness of *Acinetobacter baumannii*. Frontiers in Microbiology, 2016. **7**(1715).

29. da Costa, W.L.O., et al., Functional annotation of hypothetical proteins from the *Exiguobacterium antarcticum* strain B7 reveals proteins involved in adaptation to extreme environments, including high arsenic resistance. PLOS ONE, 2018. **13**(6): p. e0198965.

30. Reams, A.B. and E.L. Neidle, Genome plasticity in *Acinetobacter:* new degradative capabilities acquired by the spontaneous amplification of large chromosomal segments. Mol Microbiol, 2003. **47**(5): p. 1291-304.

31. Seaton, S.C., et al., Genome-wide selection for increased copy number in *Acinetobacter baylyi* ADP1: locus and context-dependent variation in gene amplification. Molecular Microbiology, 2012. **83**(3): p. 520-535.

32. Sambrook, J. and D.W. Russell, Molecular cloning: a laboratory manual. Vol. 3. 2001: Cold Spring Harbor Laboratory Press.

33. Kearse, M., et al., Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics (Oxford, England), 2012. **28**(12): p. 1647-1649.

34. Juni, E. and A. Janik, Transformation of *Acinetobacter calco-aceticus* (*Bacterium anitratum*). J Bacteriol, 1969. **98**(1): p. 281-8.

35. Vaneechoutte, M., et al., Naturally transformable *Acinetobacter* sp. strain ADP1 belongs to the newly described species *Acinetobacter baylyi.* Appl Environ Microbiol, 2006. **72**(1): p. 932-6.

36. Eraso, J.M. and S. Kaplan, *prrA*, a putative response regulator involved in oxygen regulation of photosynthesis gene expression in Rhodobacter sphaeroides. Journal of bacteriology, 1994. **176**(1): p. 32-43.

37. Prentki, P. and H.M. Krisch, In vitro insertional mutagenesis with a selectable DNA fragment. Gene, 1984. **29**(3): p. 303-313.

38. Yanisch-Perron, C., J. Vieira, and J. Messing, Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene, 1985. **33**(1): p. 103-19.

CHAPTER 3

METHOD DEVELOPMENT FOR ACCELERATED LABORATORY EVOLUTION

Introduction

In chapter 2, my analysis showed that the lack of RsfA provides a competitive advantage during growth on guaiacol and under specific conditions. However, no individual mutation in the evolved strain that was studied (ACN1850) was found to be solely sufficient for growth on guaiacol as the carbon source. These results highlight the complexity of understanding the roles of multiple mutations that tend to arise during adaptive evolution experiments. Furthermore, the results of our efforts to improve the function of GcoAB for Gua^+ growth revealed two potential problems with targeted protein evolution. First, one type of change may predominate in multiple lineages and therefore reduce the breadth of possible outcomes. In my studies, fusions between GcoA and CatA were common, and this was the only type of protein alteration obtained that was sufficient to confer the selectable phenotype. Second, mutations outside the targeted region may enable growth and thereby obviate the need for additional changes in the targeted protein. The goal of the current study was to exploit the powerful genetic system of ADP1 to develop methods to improve the use of EASy for targeted protein engineering and to assess the significance of multiple mutations. The natural transformation system of ADP1 served as the foundation for method development. We previously showed that EASy allowed the evolution of Gua^+ strains with beneficial *gcoAB* alleles in approximately 1,000 generations [1]. One allele encodes a fusion between GcoA and CatA (catechol dioxygenase). This fusion was sufficient to

confer a Gua$^+$ phenotype in both *A. baylyi* and *P. putida* in a single chromosomal copy. In addition, this fusion represents an evolutionary outcome that would not have been attempted during rational engineering efforts. We identified other fusions that vary in the length of the peptide connection between GcoA and CatA. Another *gcoA* allele was identified and encodes a single amino acid change (G72D) in GcoA. This GcoA variant was shown to be two-fold more catalytically efficient, *in vitro*, than the wild-type protein [2]. However, a single chromosomal copy of the evolved *gcoA* allele encoding the G72D variant was insufficient to confer growth on guaiacol. Mutations in other chromosomal regions were also identified in the same mutant, ACN1850, from which the G72D variant was isolated. One mutation prevents the translation of a predicted ribosomal silencing factor (RsfA).

Based on the information presented previously, two important observations can be made. First, the majority of mutations occurring in *gcoAB* in our evolved populations involved some genetic rearrangement through deletion of DNA fragments. This almost always resulted in a genetic fusion between *gcoA* and *catA*. Point mutations in *gcoAB* were identified in one population only. This observation motivated efforts to broaden the range of beneficial mutations isolated from EASy. One possible approach is to add mutated DNA to evolving cultures, to increase the genetic diversity available to evolutionary forces. Mutated DNA to be added may include randomly mutated DNA generated by error-prone PCR [3] or DNA encoding beneficial mutations. Another approach is to modify the *gcoAB* sequence at the onset of evolution and investigating the effect of altered *gcoAB* sequences on the evolutionary trajectories. For example, evolving cells that acquire the mutation encoding the GcoA-G72D variant may follow a different evolutionary path than those cells with a wild-type *gcoAB* sequence.

The second observation involves the investigation of the significance of multiple mutations in an adaptive scenario. In model organisms, reconstituting multiple mutations is a resource and time intensive task. Additionally, investigating the significance of multiple mutations is further complicated by the dynamic nature of the evolutionary process [4]: how many mutations are required to confer the selected phenotype? Which mutations are "hitchhikers"? Are certain mutations needed first to enable subsequent fitness improvement or innovation? In our case, little information was available on the potential role of the mutations under study. This highlights the need to develop a genetic tool that enables the rapid and easy investigation of the role of different mutations, especially how they may function in combinations.

This chapter presents information on my attempts to address the two observations mentioned above. First, a modified EASy experiment was conducted that incorporates the addition of different types of mutated DNA to evolving cultures. In addition, comparisons are made between the evolutionary trajectories of populations that carry either a wild-type sequence of *gcoAB* or a modified *gcoAB* allele. Similar to our previous EASy experiment, the copy number of *gcoAB* was followed during laboratory evolution. In addition, genetic changes in the evolved populations are currently under investigation. Next, information is presented on attempts to use the natural transformability of ADP1 to elucidate which mutations in a previously evolved mutant, ACN1850, may be contributing to the Gua^+ phenotype of this mutant. These results provide insights into the potential role of previously uncharacterized genes and provide a foundation for developing a combinatorial analytical tool for investigating the evolutionary significance of multiple mutations.

Results

Addition of mutated DNA to EASy cultures

Providing evolving cultures with mutated DNA may increase the genetic diversity available for evolving cells. This could result in the emergence of a broader range of beneficial mutations, beyond DNA remodeling by deletions. Moreover, the addition of mutated DNA may allow evolving populations to select a beneficial *gcoAB* allele at a faster rate. To this end, I sought to investigate the effects of both adding different types of mutated DNA to EASy cultures and starting with populations that encode the G72D variant in their amplified *gcoAB* arrays. As discussed earlier, while the mutation encoding the G72D variant does not confer the ability to grow on guaiacol, this mutation allowed the emergence of spontaneous Gua^{+} mutants, suggesting a weak or minor benefit conferred by this mutation. Therefore, I sought to test whether the evolving populations would have a different trajectory if this mutation were present at the onset of evolution. Amplification mutants were constructed from two parent strains, ACN2142 and ACN2143 (Table 3.1). The latter differs only in a single point mutation in *gcoA* that encodes the G72D change in GcoA previously identified in ACN1850. While both ACN2142 and ACN2143 have a single chromosomal copy of *gcoAB,* their chromosomal configuration differs slightly from ACN1667, the parent strain in the previously reported EASy experiment. This difference lies in the genetic region between *catA* and *gcoA*. In ACN1667, the distance between *catA* and *gcoA* is 255 bp long and a synthetic ribosome binding site is inserted. In both ACN2142 and ACN2143, 84 bp separate *catA* and *gcoA*, where the *gcoAB* and Km^{R} genes replace ACIAD_RS06665 (ACIAD1443 or ORF1) and no artificial ribosome binding site is present (Figure 3.1). This genetic configuration would keep the spacing between *catA* and *gcoA* as would normally exist in the wild-type ADP1 chromosome between *catA* and ACIAD_RS06665. This

modification was intended to allow the emergence of different mutations and decrease the chance of formation of genetic fusions between *catA* and *gcoA*, which was a frequent outcome in the previous EASy trial.

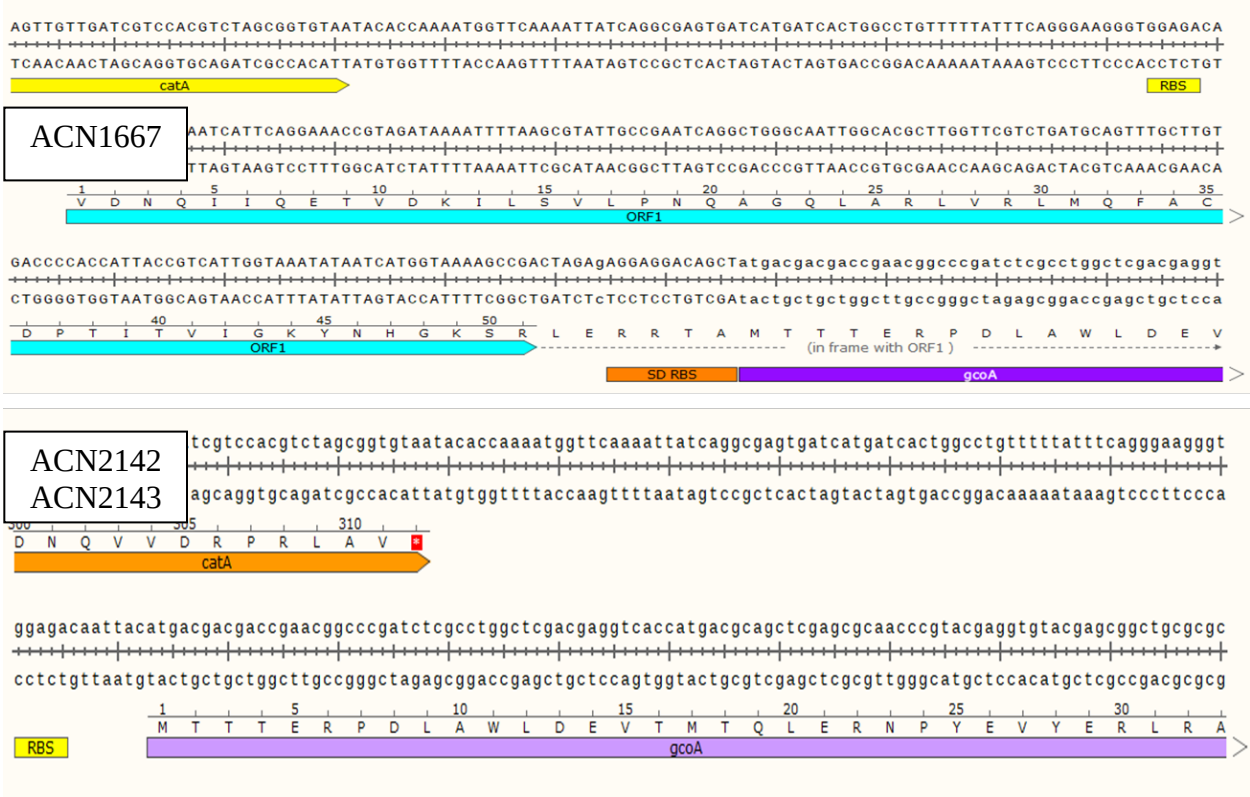


Figure 3.1. Design of parent strains for a new EASy experiment. The two parent strains ACN2142 (WT *gcoAB*) and ACN2143 (*gcoA* encoding G72D) differ from ACN1667 in the genetic region upstream of *gcoA*. In ACN1667, 255 bp separate the stop codon of *catA* and the start codon of *gcoA*, including a synthetic ribosome binding site (SD RBS, orange box). In ACN2142 and ACN2143, *catA* and *gcoA* are separated by 84 bp and no synthetic RBS is present. Instead, the native RBS of ORF1 (ACIAD_RS06665) is present (RBS, yellow box). Not drawn to scale.

Batch cultures were started by growing ACN2148 and ACN2150 on 1 mM guaiacol as the sole carbon source. Serial dilutions were performed daily after approximately 24 hours of growth. For each EASy modification, three replicate cultures were employed. To separate replicate cultures derived from ACN2148, DNA was added as follows: randomly mutated *gcoAB* DNA generated by error-prone PCR (epPCR), *gcoAB* DNA encoding the G72D amino acid change, and a DNA mixture encoding all mutations identified in a previously characterized Gua^+ isolate (ACN1850, Table 2.1). As a control, DNA encoding the wild-type *gcoAB* sequence was added to one set of replicates. The experimental setup and the different conditions are summarized in Figure 3.2.

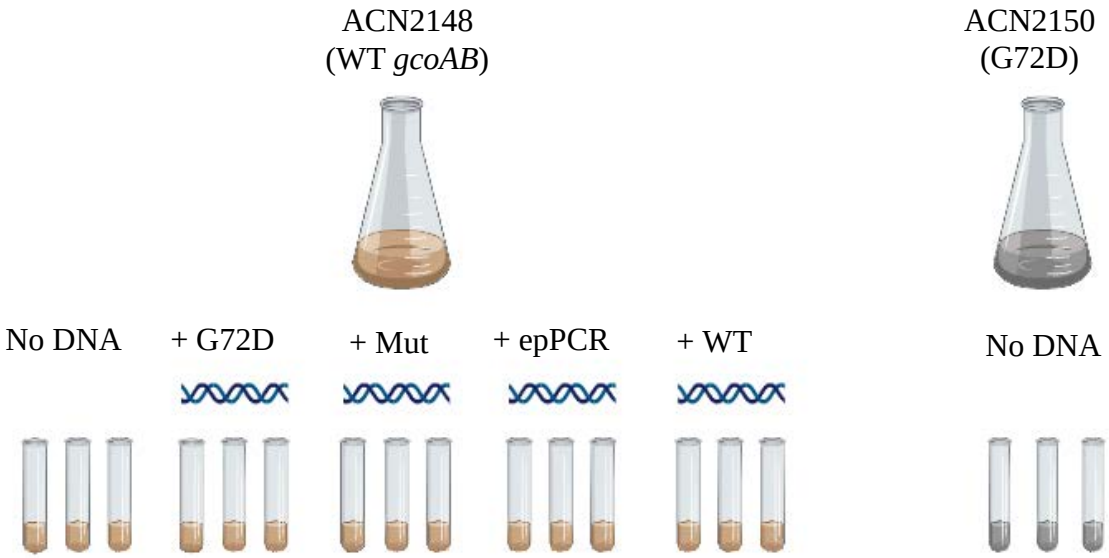


Figure 3.2. Experimental design for EASy cultures. ACN2148 and ACN2150 are the starting amplification strains, with either WT *gcoA* sequence or *gcoA* encoding G72D, respectively. To populations derived from ACN2148, mutated DNA was added: + G72D, *gcoA* encoding G72D; + Mut, all mutations from ACN1850 generated by PCR and pooled; + epPCR, pool of randomly mutated *gocAB* DNA by error-prone PCR. As a control, WT *gcoA* DNA was added to one set (+ WT). ACN2150 derived populations were evolved without the addition of DNA.

During continuous selection, beneficial mutations that improve the ability to grow on guaiacol should relieve the pressure to maintain extra copies of *gcoAB* and the average copy number is expected to decrease. We followed the change in copy number of the Km^R gene inserted downstream of *gcoB* by qPCR for the evolving populations as indicated in Figure 3.3.

As expected, the copy number decreased over time for all populations. However, unlike the previous EASy trial, none of the populations reached an average copy number of one. Instead, for all conditions the average copy number seems to remain stable around 3-4 copies after 120 days in culture. It is possible that some of the populations may require longer to reach an average copy number of one. However, we were interested in comparing the timeline of these modifications to our original study, during which 120 days was the first point when a population reached an average copy number of one. The data presented in Figure 3.3 are from one representative culture per condition for easy visualization, but the trends were observed across replicates. The presence of the variant GcoA(G72D) at the beginning of evolution seems to be slightly advantageous. Populations derived from ACN2150 (carrying multiple copies of the variant) had an overall lower average copy number than populations carrying multiple copies of WT *gcoAB* (Figure 3.3, grey line). This slight advantage is consistent with previous observations that the G72D amino acid change in GcoA provides a weak benefit. Interestingly, PCR analysis revealed that all ACN2148 populations had a genetic rearrangement that involves a deletion of DNA between *catA* and *gcoA*, likely encoding a CatA-GcoA fusion (Figure 3.5). Further investigation by whole-genome sequencing will enable the identification of the exact configuration of this region. On the contrary, ACN2150 populations did not show any signs of DNA deletions in the same region.

The effect of adding mutated DNA varied by the type of DNA. At the end of the experiment, the populations where epPCR DNA was added had the lowest average copy number (Figure 3.3, green line), indicating that the addition of this type of DNA might be the most useful. This benefit may be due to the higher genetic diversity generated in a pool of randomly mutated DNA. The addition of mutated DNA encoding the G72D variant (Figure 3.3, red line) did not result in a lower copy number than the addition of WT *gcoA* DNA (Figure 3.3, yellow line).

Overall, the addition of mutated DNA did not yield populations with an average copy number of one in the time frame of the experiment. Based on the data presented in Figure 3.3, the addition of randomly mutated *gcoAB* DNA seems to have the most beneficial effect, while populations starting with a multicopy array of *gcoAB* DNA encoding the G72D variant seems slightly advantageous. The other conditions tested do not seem to result in a significant benefit or reduction in the time required for isolating beneficial mutations. However, further analysis may provide more conclusive insights (see Discussion).

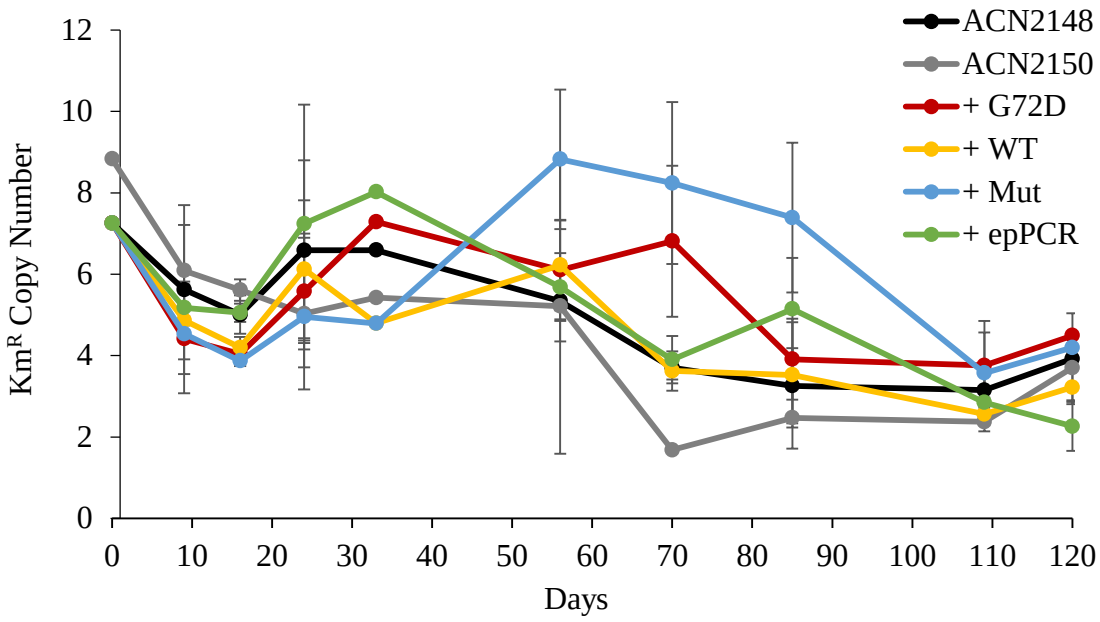


Figure 3.3. Copy number changes during modified EASy culturing. Chromosomal copy number of Km^R was followed by qPCR during the evolution of populations derived from ACN2148 and ACN2150. The DNA additions are indicated with the same designations as in the legend of Figure 3.2. For each culture condition, three replicate populations were evolved and noted A, B, or C. For simplicity, copy number data presented here are from one representative culture per condition. Similar trends have been observed across replicates. Copy number data from all populations are presented in Figure 3.4. Error bars represent one standard deviation.

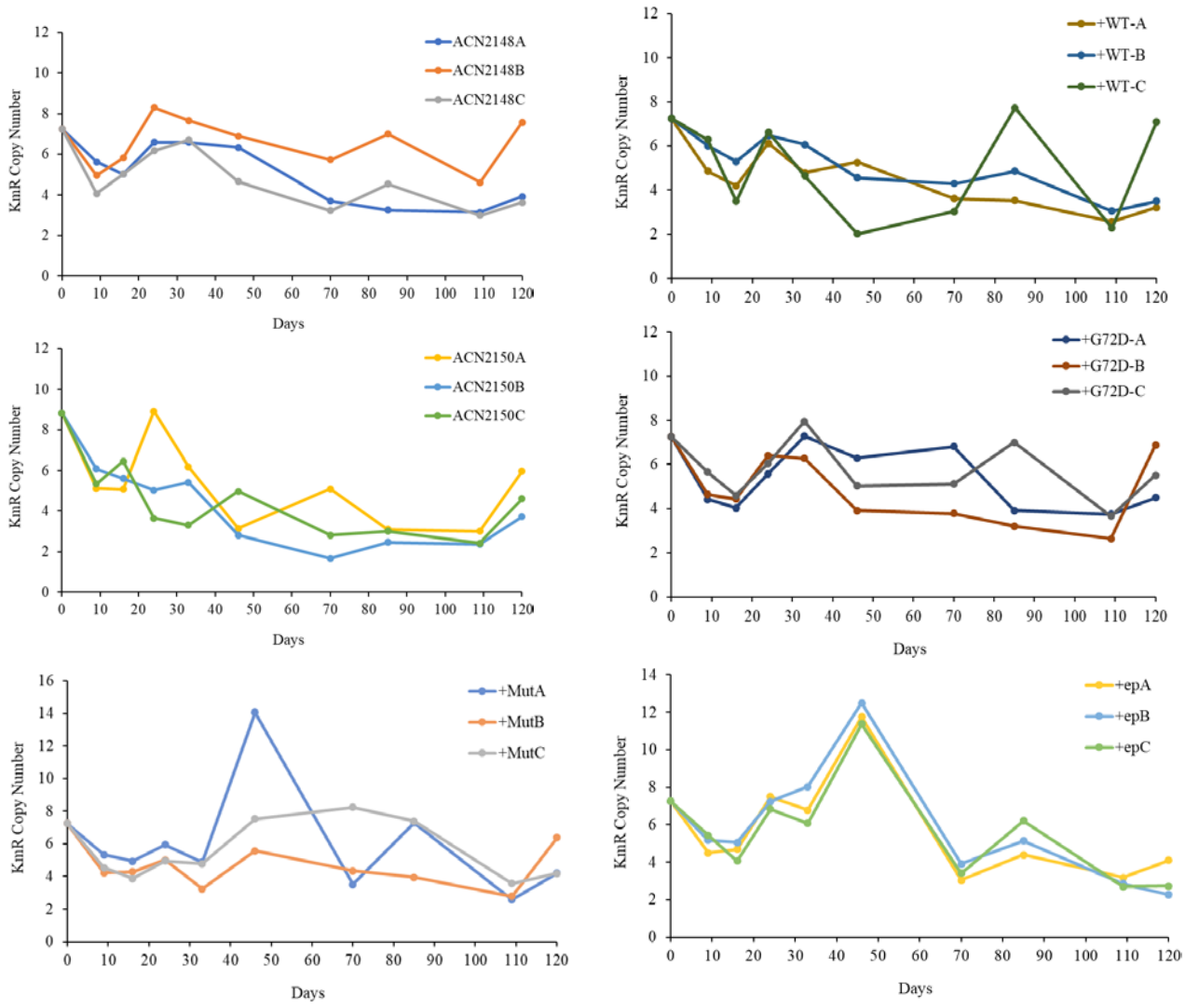


Figure 3.4. Complete copy number data across all modified EASy conditions. Copy number for all evolved populations categorized by condition. Conditions are indicated as described in the legend of Figure 3.3, with populations labelled A, B, or C. Error bars are omitted for clarity.

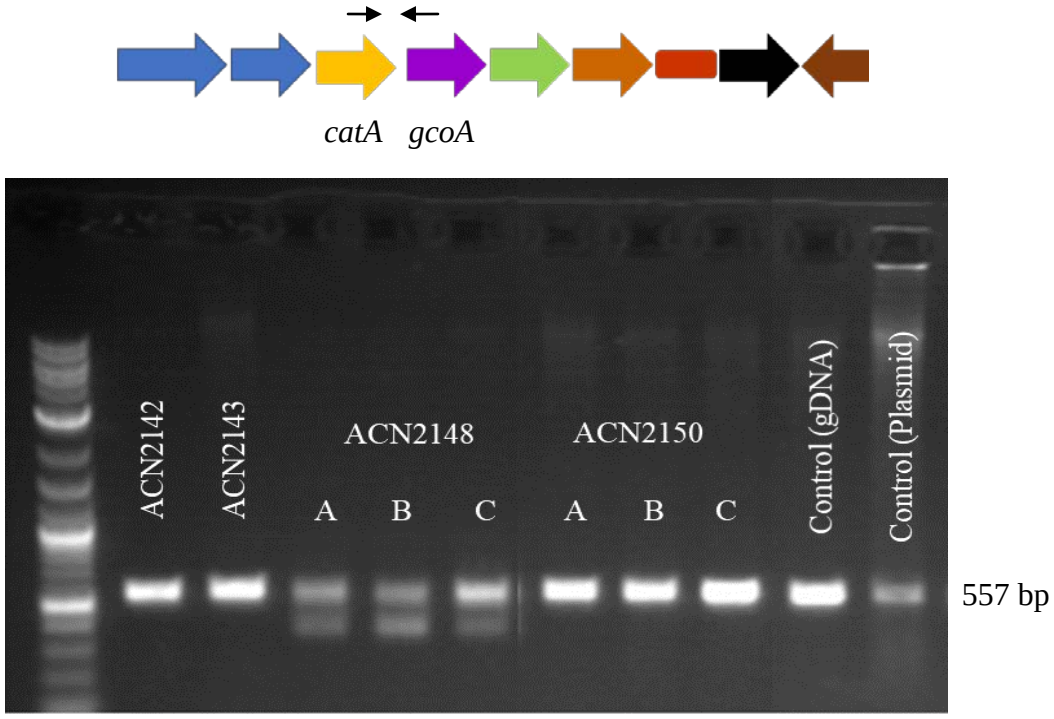


Figure 3.5. Detection of deletions in the *catA-gcoA* region in evolving populations. Upper panel: Primers AA57 and MTV112 (black arrows) amplify a region that encompasses the 3′ end of *catA* and the 5′ end of *gcoA* (not drawn to scale). Bottom panel: Populations derived from ACN2150 (amplified array of *gcoA* encoding G72D) do not harbor deletions in the *catA-gcoA* region, unlike ACN2148-derived populations. Genomic DNA from each population was isolated and used as a template for PCR. In ACN2142 and ACN2143, the parent strains with a single chromosomal copy of *gcoAB*, the PCR product using those two primers is 557 bp long. Bands corresponding to a shorter product represent DNA deletion in this region in ACN2148 populations A, B, and C. Control (gDNA); genomic DNA from ACN2148 population before starting liquid cultures used as template, control (plasmid); plasmid DNA of pBAC1618 (Table 3.2) used as template. The left most lane shows 1 kb plus standard DNA ladder (New England Biolabs).

Examination of mutations by natural transformation

As presented in chapter 2, analysis of mutations identified in an evolved Gua^+ mutant (ACN1850) revealed the role of a stationary phase gene (*rsfA*) during adaptive evolution. However, our analysis did not reveal the possible roles of the other mutations in ACN1850. Since a reconstructed strain that encodes both the *gcoAB* mutations and a deletion of *rsfA*, ACN2482, was unable to grow on guaiacol, this suggested that more than one mutation may be responsible for the observed phenotype. As discussed earlier, mutations in seven genes were identified (Table 2.1) but sequence information did not establish any potential role in the ability to grow on guaiacol. Reconstituting seven mutations in a wild-type background, in all possible combinations presents a challenging task of generating a large number of mutants. This process may require a long time to construct, characterize, and confirm the genotype of each mutant. This challenge highlights the need to develop genetic tools that enable the rapid investigation of the significance of mutations, especially in combinations. The high efficiency of natural transformation and recombination in ADP1 presents a platform to develop such a tool, building on assays previously developed in ADP1 during studies of gene amplification [5].

To identify which mutations may improve the ability to grow on guaiacol, a Gua^- strain with a single chromosomal copy of *gcoAB* (encoding G72D) is used as a recipient of transforming DNA. This transforming DNA includes fragments encoding the mutations under study. The recipient strain is plated on a guaiacol plate, on which it cannot grow unless it acquires beneficial mutations that enable growth. On different sections of the plate a single linear DNA fragment encoding one of the seven mutations from ACN1850 was dropped. In addition, combinations of the seven mutations were dropped on separate sections of the same plate. After incubation for up to two weeks, any Gua^+ colonies arising on the plate were analyzed genetically

for the uptake and incorporation of the mutation with which it was transformed. In initial trials, I used strains ACN1667 and ACN1886 as recipients, but no Gua^+ transformants were obtained after transformation with DNA. This might be due to either the requirement of more than one mutation or due to variable efficiencies in the uptake of multiple DNA fragments.

To increase the chance of obtaining Gua^+ transformants using this approach, I generated a new recipient strain, ACN2482 in which additional mutations were reconstructed. ACN2482 has the two point mutations in *gcoAB*, as well as a deletion of ACIAD_RS13910 (old locus tag ACIAD3076), which encodes a predicted ribosomal silencing factor. Using ACN2482 as a recipient, one Gua^+ transformant, ACN2571 was isolated (Figure 3.6). This transformant appeared on the section of the plat where a DNA fragment encoding a mutation in ACIAD_RS05060 (old locus tag ACIAD1101) was dropped. This gene encodes a penicillin-binding protein (PBP2) predicted to function in the crosslinking of peptidoglycan. The mutation encodes an early termination at position 365, which removes a large portion of the N-terminal catalytic domain. It is currently not clear whether this mutation is a loss of function or what role this gene may play in growth on guaiacol. Sequencing of the corresponding gene in ACN2571 confirmed that this mutated allele was indeed incorporated into the chromosome. In the ADP1 single knockout collection, the gene encoding PBP2 was found to be dispensable [6], although this result may be contingent on the growth conditions used for screening. In addition, recent studies in ADP1 show that a deletion of ACIAD_RS05060 results in the formation of abnormal giant cells [7]. However, we have not observed any morphological changes in ACN1850 or ACN2571. ACN2571 will be analyzed by whole-genome sequencing to determine if any other mutations were acquired. In addition, replicating the introduction of the mutated

ACIAD_RS05060 allele in ACN2482 is underway to confirm that combining the three mutations enables growth on guaiacol.

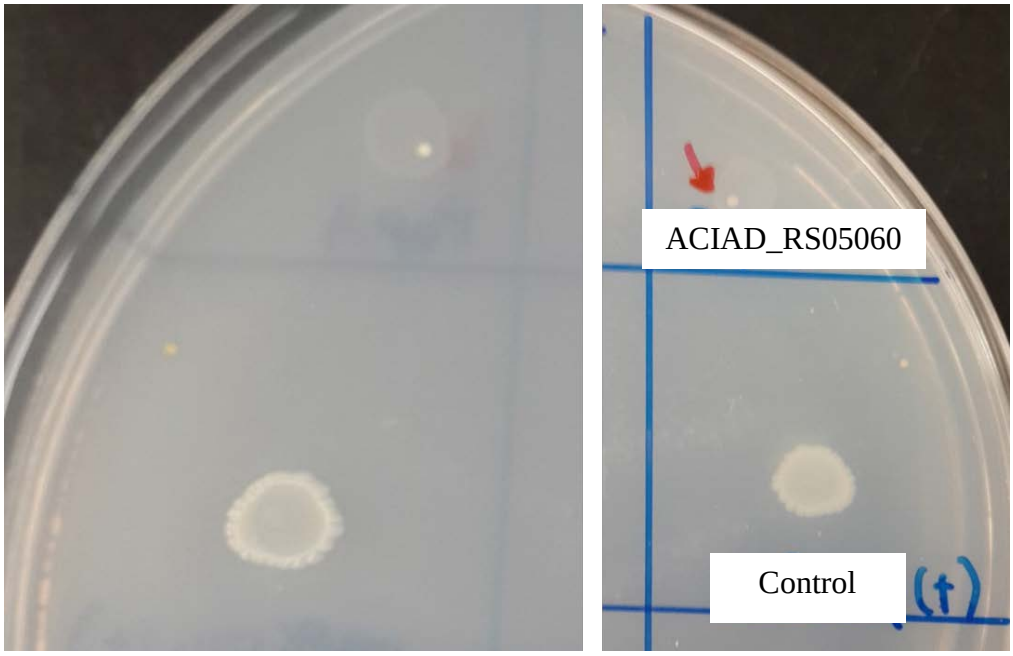


Figure 3.6. Isolation of a Gua$^+$ transformant using natural transformation. A lawn of ACN2482 (Gua$^-$) was spread on a guaiacol plate from overnight growth. On separate sections of the plate, DNA fragments encoding each of the seven mutations identified in ACN1850 were dropped. After incubation at 30 °C, one colony was isolated from the section where DNA encoding a mutated ACIAD_RS05060 (encoding a penicillin-binding protein) allele was added (top right section). This colony, ACN2571, was confirmed by PCR analysis and sequencing to have the mutated allele of ACIAD_RS05060. In the bottom right section, a linear DNA fragment encoding a *catA-gcoA* fusion was dropped as a control. This fusion was previously shown to yield Gua$^+$ phenotype in a single chromosomal copy [1]. The left and right panels are top and bottom views of the same plate.

Discussion

EASy modifications: an opportunity to expand experimental design

In this study, we tested different modifications to the experimental design of EASy cultures. These modifications included the addition of different types of mutated DNA to evolving cultures and modifying the sequence of *gcoAB* in the starting strains. We followed the amplicon copy number of populations evolving under each condition. While starting with a multi copy array where *gcoA* encodes the G72D variant (Figure 3.3, red line) seems to have a lower population average copy number compared to starting with WT *gcoA* sequence (Figure 3.3, black line), these results do not conclusively demonstrate a clear advantage of one condition over the other. Interestingly, all ACN2148 populations (WT *gcoAB*) eventually harbored a genetic fusion between *catA* and *gcoA*, which was frequently observed in our previous EASy experiment. However, ACN2150 populations (*gcoAB* encoding G72D) did not seem to have such DNA rearrangements. It appears that starting with a slightly beneficial allele of *gcoA* may bypass the need to generate this genetic fusion and enable the isolation of different mutations. Other possible variations in the *gcoAB* sequence may also be considered in future experiments. For example, *gcoAB* were obtained from an actinomycete with a much higher G+C content (73%) than ADP1 (42%). Therefore, using codon-optimized versions of *gcoAB* in the starting multi copy array may provide a more significant improvement over using the wild-type sequencing.

In the case of adding mutated DNA to evolving cultures, most types of mutated DNA did not demonstrate a clear advantage over cultures where no DNA was added. It is worth mentioning that adding randomly-mutated *gcoAB* DNA resulted in the lowest copy number among all populations (Figure 3.3, green line). This might be due to the higher genetic diversity in the added DNA, in addition to this diversity occurring in the genes encoding the selected

function. While deleterious mutations are expected to be more frequent than beneficial ones in randomly mutated DNA, the genetic redundancy provided by the multi-copy array of *gcoAB* buffers against loss-of-function and deleterious mutations being lethal.

In one case, a DNA pool encoding multiple mutations was added to evolving cultures (ACN1850 mutations). Presumably, some of those mutations may be neutral, while others may be beneficial and future efforts will investigate which mutations are fixed in the populations. A sample of each population was frozen at every time point, which may provide the ability to revive cells at any time point to identify genetic changes. Whole-genome sequencing may help to indicate which mutations are fixed after addition of the mutations mixture to evolving cultures. This genomic analysis will allow us to compare the results of this current EASy trial to our previously published data. Since serial transfers were conducted in a similar fashion to our previous EASy trial, a deletion of *rsfA* for example may still be beneficial and its fixation in the population may provide evidence further supporting the conclusions presented in chapter 2. It is possible that more time was required for some of the populations to eventually require a single copy of *gcoAB* for growth. The results presented here may help future design of EASy experiments. For example, it would be interesting to engineer starting strains with a deletion of *rsfA*, especially if serial transfers are to be conducted in a similar fashion.

Batch cultures employing serial transfers enable testing multiple conditions simultaneously. The conditions tested here are not comprehensive and do not cover every possible modification possible. Two or more EASy modifications may be combined in future experiments. For example, the addition of randomly mutated DNA to EASy cultures seems to result in a slightly lower average copy number than other conditions. Similarly, cultures that start from a strain with a multi copy array of altered *gcoA* sequence (encoding the G72D variant), had

an overall lower copy number than those cultures derived from a strain with a wild-type *gcoA* sequence. These two conditions, or any number of conditions, may be combined during future EASy experiments.

Analysis of adaptive mutations using natural transformation

In any laboratory evolution experiment, several mutations may be identified in a given evolved isolate. While sequence information may provide some clues on which mutations to further investigate, evolutionary dynamics are complex and may involve different types of interactions between mutations [4]. Furthermore, a systematic approach to investigating the significance of mutations in an evolutionary scenario requires reconstituting every possible combination of mutations. This task requires the generation of a large number of mutants and therefore, focus tends to be given to mutations in well characterized genes. However, those mutations that may not provide a clear role in adaptation may provide the most novel information.

In this chapter, I presented evidence on the feasibility of using natural transformation to identify mutations that contribute to a specific phenotype. In this case, it appears that three mutations may be required to enable growth on guaiacol: a point mutation in *gcoA* encoding a G72D change, a deletion of ACIAD_RS13910 which encodes a ribosomal silencing factor, and a point mutation in ACIAD_RS05060 that results in early termination of the encoded penicillin binding protein. A Gua^+ mutant that possesses all three mutations (ACN2571) is currently being prepared for whole-genome sequencing to identify whether it harbors other mutations. In addition, efforts are underway to generate an analogous reconstructed strain where all three mutations are introduced systematically.

This approach can be further expanded to use as the transforming DNA, fragments encoding multiple mutations. This transformation assay is also amenable to modifications, for example in the type of DNA to be used for transformation. In addition, automation may enable large scale analysis. Using microtiter plates, to each well a different transforming DNA fragment (or combination of fragments) may be added, while monitoring optical density as a proxy for growth. One obstacle to this approach may stem from the expected variability in transformation efficiency when more than one DNA fragment is used and may require further optimization. Overall, the results presented here provide a platform upon which a robust and large-scale assay may be developed.

Materials and methods

Culture conditions and evolution of Gua^+ strains

The ability to grow on guaiacol (Gua^+) was defined by the formation of colonies on plates and reaching an OD_{600} of at least 0.15 in liquid culture with 1 mM guaiacol as the sole source of carbon. Amplification mutants were grown at 30°C with shaking. The cells were allowed to reach stationary phase and serial transfers were done on a daily basis with a dilution of 1:50 (by volume) in fresh medium. Genomic DNA was isolated periodically from each culture and was analyzed by PCR for *gcoAB*, the synthetic bridging fragment (SBF), and the drug resistance marker. DNA was added to EASy cultures once a week immediately after serial transfers.

Copy Number Analysis in *A. baylyi* amplification mutants

Copy number was determined by a quantitative PCR (qPCR) method as described earlier [1]. Primers for qPCR were designed with PrimerExpress software (Applied Biosystems). Primer sequences are listed in (Table 3.3). Standard curves were obtained using genomic DNA from the parent single-copy strains ACN1667 or ACN2142 for *rpoA* and the drug resistance marker. The ratio of calculated concentration of Km^R relative to that of *rpoA* was presumed to indicate the copy number of that fragment.

Generation of mutated DNA by error-prone PCR

To generate randomly mutated DNA for addition to growing cultures, pBAC1619 was used as the template DNA. Primer pairs were designed to amplify different 400-800 bp long regions covering the entire length of *gcoA* (Table 3.3). For each primer pair, the standard epPCR protocol introduced by Cadwell and Joyce was used with a few modifications [3]. In a total volume of 100 µl, approximately 20 pg of template DNA was added to a PCR mixture containing 2.5 units of *Taq* polymerase, 1X Standard *Taq* reaction buffer (New England

Biolabs), Mg^{2+} (to a final concentration of 5 mM), 0.2 mM of dNTPs (each). $MnCl_2$ was added to the PCR mixture to a final concentration of 0.5 mM immediately before starting the thermal cycle to avoid precipitation. The thermal cycler program was setup as follows: initial denaturation at 95°C for 2 minutes, 30 cycles of 95°C for 30 seconds, 55-65°C (according to primer annealing) for 1 minute, and 68°C for 1.5 minutes. The final extension step was at 68°C for 5 minutes. To increase the diversity of the mutated DNA pool, after every 5 PCR cycles an aliquot of the amplified DNA was diluted into a fresh epPCR mixture (i.e. 6 transfers) [8]. To confirm the success of mutagenesis, a sample of the mutated DNA was cloned using a TOPO/TA cloning kit (Invitrogen) and transformed into XL-1 Blue competent cells using a standard heat-shock protocol [9]. Plasmid DNA was purified and sequenced from at least 10 colonies per reaction. Mutagenesis was considered successful if at least 70% of the cloned plasmids contained two point mutations per 400 bp.

Analysis of mutations by natural transformation

Liquid cultures of Gua^- recipient strains were started in minimal medium and pyruvate. At mid log phase, one ml of the liquid culture was centrifuged and the supernatent was discarded. The cell pellet was washed twice in minimal medium to remove any residual carbon source. The pellet was resuspended in 100 ml of minimal medium and the cells were added to guaiacol plates by either directly spotting the undiluted cell suspension on different sections of the plate, or by serially diluting the cell suspension and plating 100 µl of a dilution that yields approximately 10^8 cells on the plate surface. Next, purified DNA fragments generated by PCR and encoding each mutation to be tested were dropped individually or in combination on designated sections of the guaiacol plate. Approximately 200 ng total DNA was added per DNA fragment. Fragments were approximately 2 kb long, including 200 – 500 bp of homologous regions on each side of the gene

carrying the mutation. The plates were incubated at 30°C for up to 10 days. To account for background growth, cells were plated or spotted on designated sections of the plates without adding any DNA. As a positive control, pBAC1314 (Table 3.2) was linearized and used as the transofrming DNA. This plasmid carries a *catA-gcoA* fusion that confers the ability to grow on guaiacol. Positive growth was defined as visible colonies that grow within two days upon transferring to a fresh guaiacol plate.

Table 3.1 Strains used in this study

Strain	Relevant characteristics Donor DNA/Restriction enzyme, transforms (X) recipient strain	Source
ADP1	Wild-type (BD413)	[10, 11]
ACN1667	ACIAD1443::*gcoAB51661;* Km^R*51667* (ΩK^a cassette inserted downstream of *gcoB*), Gua^- parent strain single copy of *gcoAB* region pBAC1261/EcoRI X ACN1661 selected by Km^R	[1]
ACN1850	ACIAD1443::*gcoA51850* [encodes GcoA(G72D)], *gcoB51850* [encodes GcoB(A4T)]; Km^R*51667;* EASy-derived Gua^+ isolate from evolving population of ACN1686 Mutations identified in genome outside the *gcoAB* region	[1]
ACN1881	ACIAD1443::*gcoA51850* [encodes GcoA(G72D)], *gcoB51661*; Km^R*51667;* Reconstructed strain, Gua^- pBAC1456/NdeI X ADP1 selected by Km^R	[1]
ACN1886	ACIAD1443::*gcoA51850* [encodes GcoA(G72D)], *gcoB51850* [encodes GcoB(A4T)]; Km^R*51661;* Reconstructed strain, Gua^- pBAC1459/NdeI X ADP1 selected by Km^R	[1]
ACN2142	*gcoAB51667;* Km^R*51667;*(*gcoAB* – ΩK^a cassette replaces ACIAD1443), Gua^- pBAC1618/NdeI X ADP1 selected by Km^R	This study
ACN2143	*gcoA51850* [encodes GcoA(G72D)], *gcoB51661*; Km^R*51667;*(*gcoAB* - ΩK^a cassette replace ACIAD1443) pBAC1619/NdeI X ADP1 selected by Km^R	This study
ACN2692	*gcoAB51667;* Km^R*51667;*(*gcoAB* – ΩK^a cassette replaces ACIAD1443), Gua^-*;* 19 chromosomal copies of the amplicon including *gcoAB*; SBF*51676* pBAC1262/AatII X ACN2142 selected by high-level Km^R	This study
ACN2693	*gcoA51850* [encodes GcoA(G72D)], *gcoB51661*; Km^R*51667;*(*gcoAB* - ΩK^a cassette replace ACIAD1443); Gua^-*;* 39 chromosomal copies of the amplicon including *gcoAB*; SBF*51676* pBAC1262/AatII X ACN2143 selected by high-level Km^R	This study
ACN2148	*gcoAB51667;* Km^R*51667;*(*gcoAB* – ΩK^a cassette replaces ACIAD1443), Gua^- multiple chromosomal copies of the amplicon including *gcoAB*; SBF*51676*; Gua^+ Derived from ACN? By direct selection on guaiacol plates	This study
ACN2150	*gcoA51850* [encodes GcoA(G72D)], *gcoB51661*; Km^R*51667;*(*gcoAB* - ΩK^a cassette replace ACIAD1443); Gua^+*;* multiple chromosomal copies of the amplicon including *gcoAB*; SBF*51676* Derived from ACN?? By direct selection on guaiacol plates	This study

ACN2452	ACIAD3076::*sacB*-SmSp^R*52452;* counter-selectable marker inserted downstream of *hyi* to facilitate subsequent strain construction pBAC1769/NdeI X ADP1 selected by Sm^R	This study
ACN2453	ACIAD1443::*gcoA51850* [encodes GcoA(G72D)], *gcoB51850* [encodes GcoB(A4T)]; Km^R*51661,* ACIAD3076::*sacB*- SpSm^R *52452 ;* counter-selectable marker inserted downstream of *hyi* to facilitate subsequent strain construction*;* Gua$^-$ pBAC1769/NdeI X ACN1886 selected by Sm^R	This study
ACN2467	*catA-gcoA51738*; *gcoB51661*; Km^R*51667;* ACIAD3076::*sacB*-SpSm^R *52452 ;* counter-selectable marker inserted downstream of *hyi* to facilitate subsequent strain construction, Gua$^+$ pBAC1769/NdeI X ACN1739 selected by Sm^R	This study
ACN2482	ACIAD1443::*gcoA51850* [encodes GcoA(G72D)], *gcoB51850* [encodes GcoB(A4T)]; Km^R*51661,* unmarked deletion of ACIAD_RS13910 (ΔACIAD3076) (Δ*rsfA*)*;* Gua$^-$ pBAC1768/NdeI X ACN2453 selected by growth in the presence of 10% sucrose	This study
ACN2483	Unmarked deletion of ACIAD_RS13910 (ΔACIAD3076) (Δ*rsfA*) pBAC1768/NdeI X ACN2452 selected by growth in the presence of 10% sucrose	This study
ACN2484	*catA-gcoA51738*; *gcoB51661*; Km^R*51667;* unmarked deletion of ACIAD_RS13910 (ΔACIAD3076) (Δ*rsfA*); Gua$^+$ pBAC1768/NdeI X ACN2467 selected by growth in the presence of 10% sucrose	This study
ACN2571	ACIAD1443::*gcoA51850* [encodes GcoA(G72D)], *gcoB51850* [encodes GcoB(A4T)]; Km^R*51661*; unmarked deletion of ACIAD_RS13910 (ΔACIAD3076) (Δ*rsfA*)*;* Gua$^+$ PCR fragment encoding ACIAD_RS05060 (ACIAD1101) allele from ACN1850 X ACN2482	This study

[a]ΩK indicates the omega drug-resistance cassette for Km^R from pUI1637 [12]

[b]ΩS indicates the omega drug-resistance cassette for Sp^{R}Sm^R from pHP45 [13]

Table 3.2 Plasmids used in this study

Plasmid	Relevant characteristics	Source
pUC18, pUC19	Ap^R, cloning vectors	[14]
pMiniT 2.0	Ap^R, Cloning vector designed for cloning blunt-ended or single-base overhang PCR products	New England Biolabs
pUI1637	Source of omega Km^R cassette	[12]
pUI1638	Source of omega Sp^RSm^R cassette	[12]
pBAC1282	Ap^R, Km^R; This plasmid, linearized with XbaI and EcoRV, allows capture of the *A. baylyi* chromosomal region containing *gcoAB* using the gap-repair method.	[1]
pBAC1314	Ap^R, Km^R; *catA-gcoA51738*; *gcoB51661*; Km^R*51667;* DNA recovered from ACN1738 by the gap repair method using linearized pBAC1282	[1]
pBAC1446	Ap^R, Km^R; *gcoB51850*; [encodes GcoB(A4T)] made by replacing the SbfI-PspOMI fragment of pBAC1261 with comparable DNA carrying the *gcoB* mutation discovered in ACN1850	[1]
pBAC1456	Ap^R, Km^R; ACIAD1443::*gcoA51850* ; [encodes GcoB(A4T)] made by replacing the SbfI-BsiWI fragment of pBAC1261 with comparable DNA carrying the *gcoA* mutation discovered in ACN1850	[1]
pBAC1459	Ap^R, Km^R; ACIAD1443::*gcoA51850; gcoB51850* [encodes GcoA(G72D) and GcoB(A4T)] constructed by replacing the BbvCI-BsiWI fragment of pBAC1261 with comparable DNA carrying both mutations discovered in ACN1850	[1]
pBAC1580	Ap^R, Km^R; ACIAD_RS13910 (ACIAD3076) amplified from ADP1 gDNA by primers ALS157 and ALS158 and ligated to pCR4-TOPO vector (Thermo Fisher)	This study
pBAC1618	Ap^R, Km^R; *gcoAB51667;* Km^R*51667* (*gcoAB* – ΩK^a cassette replaces ACIAD1443) constructed by replacing the SwaI-SbfI fragment of pBAC1284 with a PCR generated *gcoAB* DNA fragment	This study
pBAC1619	Ap^R, Km^R; *gcoA51850* [encodes GcoA(G72D)], *gcoB51661;* Km^R*51667* (*gcoAB* – ΩK^a cassette replaces ACIAD1443) constructed by replacing the SwaI-SbfI fragment of pBAC1284 with a PCR generated *gcoAB* DNA fragment	This study
pBAC1633	Ap^R, Sp^R, Sm^R; ΩS cassette from pHP45 [13] inserted as a BamHI fragment in pBAC1580	This study
pBAC1670	Ap^R; ACIAD_RS04905 (ACIAD1064) allele amplified from ACN1850 genomic DNA with primers AA64 and AA65 cloned in pMiniT 2.0 vector (New England Biolabs)	This study
pBAC1671	Ap^R; ACIAD_RS05060 (ACIAD1101) allele amplified from ACN1850 genomic DNA with primers AA63 and AA66 cloned in pMiniT 2.0 vector (New England Biolabs)	This study

pBAC1672	Ap^R; ACIAD_RS08335 (ACIAD1807) allele amplified from ACN1850 genomic DNA with primers AA60 and AA61 cloned in pUC19.	This study
pBAC1764	Ap^R; ACIAD_RS05110 (ACIAD1115) allele amplified from ACN1850 genomic DNA with primers AA67 and AA68 cloned in pMiniT 2.0 (New England Biolabs)	This study
pBAC1768	Ap^R, unmarked deletion of ACIAD_RS13910 (ACIAD3076) (Δ*rsfA*), *in vivo* assembly of two ADP1 DNA fragments and a pUC18 fragment to generate an unmarked deletion of ACIAD3076	This study
pBAC1769	Ap^R, $SpSm^R$, *sacB*-$SpSm^R$, linearized fragment containing the counter selectable marker ligated to pBAC1768 linearized with PsiI, for insertion in ACIAD_RS13910 (ACIAD3076)	This study
pBAC1804	ApR; ACIAD_RS08335 (ACIAD3450) allele amplified from ACN1850 genomic DNA with primers ALS159 and ALS160 cloned in pMiniT 2.0 vector (New England Biolabs).	This study

Table 3.3 Primers used in this study

Primer	5′ → 3′ Sequence	Notes
AA26	GCAGTACGTCGAGCTGGTCG	Used for sequencing the 3′ end of *gcoB*
AA57	TGATGACGGCCTCGAAGT	Used with MTV112 to amplify a 557 bp fragment in the *catA* – *gcoA* region for detection of deletions
AA60	CTACGAATTCGCAGGCCCATGTAAAGGC	Used for amplifying ACIAD_RS08335 (ACIAD1807). Adds EcoRI cut site for cloning in pUC19, used with AA61
AA61	GCAATCTAGATGTGAAAACTGTCCAGACACTTCA	Used for amplifying ACIAD_RS08335 (ACIAD1807). Adds XbaI cut site for cloning in pUC19, used with AA60
AA63	GCAATCTAGACTTTAGCGAAGTCCTTTATCC	Used for amplifying ACIAD_RS05060 (ACIAD1101). Adds XbaI cut site for cloning in pUC19 or pMiniT 2.0, used with AA66
AA64	CTACGAATTCGTTTGTGTACTTCATCAAGC	Used for amplifying ACIAD_RS04905 (ACIAD1064). For cloning in pMiniT 2.0. Used with AA65
AA65	GCAATCTAGACCGTACTTCCTGTACCT	Used for amplifying ACIAD_RS04905 (ACIAD1064). For cloning in pMiniT 2.0. Used with AA64
AA66	CACTGCAGGTTTCACGAACACGATCT	Used for amplifying ACIAD_RS05060 (ACIAD1101). Adds PstI cut site for

		cloning in pUC19 or PMiniT 2.0, used with AA63
ALS151	CCGCTTTTGACTCTTCTACAG	Used for amplifying ACIAD_RS05110 (ACIAD1115). For cloning in pUC19 or pMiniT 2.0. Used with ALS152
ALS152	ATGCCCAGAGCTTGAGA	Used for amplifying ACIAD_RS05110 (ACIAD1115). For cloning in pUC19 or pMiniT 2.0. Used with ALS151
ALS157	AGCTTGCAGTCAATCTGTC	Used for amplifying ACIAD_RS13910 (ACIAD3076). For cloning in pBAC1580. Used with ALS158
ALS158	GCAACAGGTTATGCAGGTC	Used for amplifying ACIAD_RS13910 (ACIAD3076). For cloning in pBAC1580. Used with ALS157
ALS159	GCCACCTTGGTTATAGCCT	Used for amplifying ACIAD_RS08335 (ACIAD3450). For cloning in pMiniT 2.0. Used with ALS160
ALS160	TGGAAGTGCTAGATGCCTC	Used for amplifying ACIAD_RS08335 (ACIAD3450). For cloning in pMiniT 2.0. Used with ALS159
AA69	TGGTCGGCCTGTTCAGC	Used for epPCR to amplify a 472 bp fragment of *gcoA*, including 20 bp of the 5′ end of *gcoB*, used with MTV378
AA70	ACGAGCGGTTCAGCTTG	Used for epPCR to

		amplify a 539 bp fragment of *gcoA* including the 3′ end of *catA*, used with MTV707
AA90	GGGATCCTCTAGAGTCGACCTGCAATG AAAAAGTTTCGCGTCTTTAC	Used for constructing a clean deletion of ACIAD_RS05060, creates upstream fragment with overlapping sequence with pUC18
AA91	AGTCCATTTGTATGAAAAATCAACTGAGAATTATAG	Used for constructing a clean deletion of ACIAD_RS05060, creates upstream fragment, used with AA90
AA92	AGTTGATTTTTCATACAAATGGACTTAAATAG AGCGG	Used for constructing a clean deletion of ACIAD_RS05060, creates downstream fragment, used with AA93
AA93	GGCCAGTGCCAAGCTTGCATGCCTTACAGACAAAA CTAAGTGTATAGG	Used for constructing a clean deletion of ACIAD_RS05060, creates downstream fragment, overlapping sequence with pUC18
MTV112	TGCGCACATTCACTATTTTGTTT	Used with AA57 to amplify a 557 bp fragment in the *catA* – *gcoA* region for detection of deletions
MTV274	GCTCGACGCCTTCTATTTCAA	qPCR primer for detection of *rpoA* copy number, used

		with MTV275
MTV275	TTTACGTCGCATTCTATTGTCTTCTT	qPCR primer for detection of *rpoA* copy number, used with MTV274
MTV302	GCGTTGGCTACCCGTGATA	qPCR primer for detection of Km copy number, used with MTV303
MTV303	GGAAGCGGTCAGCCCATT	qPCR primer for detection of Km copy number, used with MTV302
MTV355	TACGAGGTGTACGAGCGGCT	Used with MTV356 to amplify a 482 fragment of *gcoA*, used in epPCR
MTV356	CCCTCGGGGTTGGCGAAC	Used with MTV355 to amplify a 482 fragment of *gcoA*, used in epPCR
MTV378	ACGCTGACCGCGAACG	Used for epPCR to amplify a 472 bp fragment of *gcoA*, including 20 bp of the 5' end of *gcoB*, used with AA69
MTV379	TTTAAGCGTATTGCCGAATCAG	qPCR primer for detection of fragment F62, used with MTV380
MTV380	CAAACTGCATCAGACGAACCA	qPCR primer for detection of fragment F62, used with MTV379
MTV707	CGATCTAAAATTGACGCGTTTG	Used for epPCR to amplify a 539 bp fragment of *gcoA* including the 3' end of *catA*, used with AA70

References

1. Tumen-Velasquez, M., et al., Accelerating pathway evolution by increasing the gene dosage of chromosomal segments. Proceedings of the National Academy of Sciences, 2018. **115**(27): p. 7105-7110.
2. Machovina, M.M., Enzymatic strategies for controlling and harnessing the oxidative power or O2 (Doctoral Dissertation). 2018.
3. Cadwell, R.C. and G.F. Joyce, Randomization of genes by PCR mutagenesis. PCR Methods Appl, 1992. **2**(1): p. 28-33.
4. Barrick, J.E. and R.E. Lenski, Genome dynamics during experimental evolution. Nature Reviews Genetics, 2013. **14**: p. 827.
5. Reams, A.B. and E.L. Neidle, Genome plasticity in *Acinetobacter:* new degradative capabilities acquired by the spontaneous amplification of large chromosomal segments. Mol Microbiol, 2003. **47**(5): p. 1291-304.
6. de Berardinis, V., et al., A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1. Molecular systems biology, 2008. **4**: p. 174-174.
7. Bailey, J., et al., Essential gene deletions producing gigantic bacteria. PLOS Genetics, 2019. **15**(6): p. e1008195.
8. Wilson, D.S. and A.D. Keefe, Random mutagenesis by PCR. Current Protocols in Molecular Biology, 2000. **51**(1): p. 8.3.1-8.3.9.
9. Chung, C.T., S.L. Niemela, and R.H. Miller, One-step preparation of competent *Escherichia coli*: transformation and storage of bacterial cells in the same solution. Proceedings of the National Academy of Sciences of the United States of America, 1989. **86**(7): p. 2172-2175.

10. Juni, E. and A. Janik, Transformation of *Acinetobacter calco-aceticus* (*Bacterium anitratum*). J Bacteriol, 1969. **98**(1): p. 281-8.

11. Vaneechoutte, M., et al., Naturally transformable *Acinetobacter* sp. strain ADP1 belongs to the newly described species *Acinetobacter baylyi*. Appl Environ Microbiol, 2006. **72**(1): p. 932-6.

12. Eraso, J.M. and S. Kaplan, *prrA*, a putative response regulator involved in oxygen regulation of photosynthesis gene expression in Rhodobacter sphaeroides. Journal of bacteriology, 1994. **176**(1): p. 32-43.

13. Prentki, P. and H.M. Krisch, In vitro insertional mutagenesis with a selectable DNA fragment. Gene, 1984. **29**(3): p. 303-313.

14. Yanisch-Perron, C., J. Vieira, and J. Messing, Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. Gene, 1985. **33**(1): p. 103-19.

CHAPTER 4

CONCLUSIONS AND FUTURE DIRECTIONS

Summary

My dissertation project focused on developing an accelerated system for laboratory evolution. The organism used in this study, *Acinetobacter baylyi* ADP1 is a naturally transformable and metabolically versatile soil bacterium [1, 2]. My initial studies involved the evolution of ADP1 mutants that acquired the ability to utilize a novel carbon source, guaiacol. This compound is of importance to the challenge of lignin valorization. The ability of ADP1 mutants to grow on guaiacol (Gua^+) was gained by inserting foreign genes, *gcoAB*, encoding a guaiacol-*O*-demethylase into the ADP1 chromosome, followed by engineered gene amplification [3]. One Gua^+ isolate, ACN1850, from an evolved population carried a point mutation in *gcoA* which required further investigation. When reconstituted in a single chromosomal copy in a wild-type background, this mutation was insufficient to confer growth on guaiacol. The focus of chapter 2 is the genetic analysis of ACN1850. Whole-genome sequencing data revealed mutations in seven genes in this isolate. Observations that are discussed in more detail in chapter 2 directed my focus to one gene. This gene, *rsfA* (ACIAD3076), encodes a putative ribosomal silencing factor (RsfA). This protein has not been previously characterized in ADP1 and all information available comes from different organisms. The main finding of chapter 2 is that under certain growth conditions, *rsfA* mutants display a competitive advantage relative to wild-type strains. This advantage was only observed during the transition from poor to rich medium when guaiacol was the sole carbon source. The significance of this result stems from the fact that

to date, all studies report that *rsfA* mutants suffer a fitness or growth disadvantage compared to wild-type strains [4]. In *E. coli*, *rsfA* mutants were outcompeted by wild-type strains at different rates, regardless of the growth conditions or the transition in nutrient availability. However, the same studies showed that the transition from poor to rich medium was better tolerated by the *rsfA* mutants, i.e. it took longer for the wild-type strains to outcompete the mutants.

In addition to *rsfA*, other genes harbored mutations in the same isolate. One gene encodes a penicillin-binding protein (PBP2) that is predicted to function in cell wall synthesis. A second gene encodes a putative transport protein that belongs to the MFS family. A third gene encodes a protease. In chapter 3, I presented some preliminary information on using natural transformation to assess the significance of adaptive mutations, especially when they occur in those genes that do not provide a clear role in the adaptive process. A mutant that harbored three of the mutations identified in ACN1850 was able to grow on guaiacol. These three mutations included a point mutation in *gcoA*, a deletion of the gene encoding RsfA, and a mutation encoding early termination of PBP2. In addition, chapter 3 provided preliminary results regarding modified conditions for improving EASy. These modifications included altering the sequence of the genes under study and the addition of mutated DNA to the evolving populations. Changes in copy number suggested that the addition of randomly mutated DNA might provide some benefit, possibly by increasing the genetic diversity available for evolutionary forces.

Adaptive mutations in the context of experimental conditions

My main finding that the lack of RsfA is beneficial under certain conditions highlights the importance of considering the role of adaptive mutations in the context of experimental setup. The gene encoding *rsfA* has been determined dispensable in both the ADP1 knockout collection study and a recent Tn-seq study in ADP1 [5, 6]. In addition, *rsfA* mutants have been

shown to be outcompeted by WT cells in *E. coli* in stationary phase and during the transition from rich to poor medium. In our experimental system, we used batch cultures that employ serial transfers. For logistic reasons, serial transfers were done every 24 hours, which allowed the cells to enter, and remain in, stationary phase for several hours. The model presented in Figure 2.7 discusses how the lack of RsfA under these conditions might be beneficial. Cells lacking a functional RsfA display derepressed protein synthesis upon transition to fresh medium, providing the mutants with a competitive growth advantage over wild-type cells (Figure 2.8A). When serial transfers were done during log phase, this advantage disappeared, and WT cells outcompeted *rsfA* mutants (Figure 2.8B). These results may inform the design of future laboratory evolution experiments. For example, starting strains may be constructed with a deletion of the gene encoding RsfA, if serial transfers are to be done during stationary phase. In addition, it would be interesting to identify the molecular signals for RsfA to bind to and dissociate from ribosomal subunits. Structural data from *Mycobacterim tuberculosis* showed that RsfA is a dimer in solution but a single monomer binds the large ribosomal subunit [7]. Elucidating the mechanism by which the RsfA dimer dissociates may allow further control of stationary phase regulation. An “omics” approach may identify interaction partners of RsfA through changes in gene expression in wild-type and *rsfA* mutants under different growth conditions. Additionally, other genetic targets may be identified for similar purposes. For example, additional ribosome hibernation factors warrant further investigation. In ADP1, ACIAD_RS03045 (ACIAD0658) encodes a protein that resembles a hibernation promotion factor in *E. coli*. Hibernation promotion factors are known to work in concert with ribosome modulation factors to form translationally-silent 100S ribosome dimers [8-10]. It would be interesting to investigate the effects of mutations in this gene in ADP1 under different contexts.

Revisiting gene essentiality data

Single knockout collections are available for several organisms such as *E. coli* [11], *Bacillus subtilis* [12], and *A. baylyi* ADP1 [13]. These studies employ a systematic approach to disrupting single genes for the identification of the minimal set of genes that allow an organism to grow under certain conditions. Results obtained from such studies can reveal the role of uncharacterized genes and help identify gene essentiality. However, discrepancies in gene essentiality data may arise from employing different growth conditions, even for the same organism. A recent study demonstrated improved genome engineering methods in ADP1 [6]. This study defined gene essentiality using a Tn-seq library to generate large genomic deletions in this organism. While this study used a rich medium, the ADP1 knockout collection utilized minimal medium. Interestingly, the essentiality data of several genes were inconsistent between the two studies. Particularly relevant to our results, the gene encoding PBP2 was annotated as dispensable in the knockout collection but determined to be essential in the Tn-seq study. In our evolved Gua^+ isolate discussed in chapter 2, a mutation was identified which introduces a stop codon at position 367 of PBP2. The early termination mostly affects the predicted transpeptidase domain but does not remove the conserved SXXK motif that includes the catalytic serine residue at position 328 in the ADP1 protein. While it has not been studied in the context of aromatic compounds metabolism, a recent study in ADP1 showed that a deletion of the gene encoding PBP2 results in the formation of giant cells [14]. We have not observed any morphological changes in ACN1850 colonies on plates or in cells under the microscope (data not shown), which suggests that the mutation in ACN1850 may not be a loss of function. While the role of PBP2 in the ability to grow on guaiacol is not clear, further investigation is underway. Interestingly, the two genes encoding PBP2 and RsfA are in the same operon in *E. coli*. While this is not the case

in ADP1, it presents an exciting avenue to further characterize RsfA and its interaction network. Overall, my results have important implications that extend to areas such as the design of laboratory evolution experiments, the significance of uncharacterized genes, and the currently available gene essentiality data.

References

1. Juni, E., Genetics and physiology of *Acinetobacter.* Annu Rev Microbiol, 1978. **32**: p. 349-71.
2. Juni, E. and A. Janik, Transformation of *Acinetobacter calco-aceticus* (*Bacterium anitratum*). J Bacteriol, 1969. **98**(1): p. 281-8.
3. Tumen-Velasquez, M., et al., Accelerating pathway evolution by increasing the gene dosage of chromosomal segments. Proceedings of the National Academy of Sciences, 2018. **115**(27): p. 7105-7110.
4. Hauser, R., et al., RsfA (YbeB) proteins are conserved ribosomal silencing factors. PLoS Genet, 2012. **8**(7): p. e1002815.
5. de Berardinis, V., et al., A complete collection of single-gene deletion mutants of *Acinetobacter baylyi* ADP1. Molecular systems biology, 2008. **4**: p. 174-174.
6. Suárez, G.A., et al., Rapid and assured genetic engineering methods applied to *Acinetobacter baylyi* ADP1 genome streamlining. bioRxiv, 2019: p. 754242.
7. Li, X., et al., Structure of Ribosomal Silencing Factor Bound to *Mycobacterium tuberculosis* Ribosome. Structure, 2015. **23**(12): p. 2387.
8. Maki, Y., H. Yoshida, and A. Wada, Two proteins, YfiA and YhbH, associated with resting ribosomes in stationary phase *Escherichia coli.* Genes Cells, 2000. **5**(12): p. 965-74.
9. Ueta, M., et al., Role of HPF (hibernation promoting factor) in translational activity in *Escherichia coli.* J Biochem, 2008. **143**(3): p. 425-33.

10. Wada, A., et al., Ribosome modulation factor: stationary growth phase-specific inhibitor of ribosome functions from *Escherichia coli.* Biochem Biophys Res Commun, 1995. **214**(2): p. 410-7.

11. Baba, T., et al., Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Molecular systems biology, 2006. **2**: p. 2006.0008-2006.0008.

12. Kobayashi, K., et al., Essential *Bacillus subtilis* genes. Proceedings of the National Academy of Sciences, 2003. **100**(8): p. 4678.

13. Barbe, V., et al., Unique features revealed by the genome sequence of *Acinetobacter* sp. ADP1, a versatile and naturally transformation competent bacterium. Nucleic Acids Res, 2004. **32**(19): p. 5766-79.

14. Bailey, J., et al., Essential gene deletions producing gigantic bacteria. PLOS Genetics, 2019. **15**(6): p. e1008195.